2008

Stochastic Differential Equations with Jumps

Jose L. Menaldi
Wayne State University, menaldi@wayne.edu

Stochastic Differential Equations with Jumps[1]

Jose-Luis Menaldi[2]

Current Version: 7 January 2014[3]

First Version: 15 May 2008 [4]

[2]Wayne State University, Department of Mathematics, Detroit, MI 48202, USA (e-mail: menaldi@wayne.edu).

[3]TEMPORARY: All sections in RED need some revision, specially Part III

[4]This book is being progressively updated and expanded. If you discover any errors or you have suggested improvements please e-mail the author.

Contents

Rationality for the book

In *Deterministic Control*, if time is regarded as either continuous or discrete then two models can be set, which combined yield the so called hybrid system. The state representation of the continuous model evolves following an ordinary differential equation (ODE) of the form

$$\dot{x}(t) = A(t)x(t) + B(t)v(t), \tag{1}$$

where $t \geq 0$ is the time, $x = x(t)$ is the state and $v = v(t)$ is the control. The state x (in $\mathbb{R}^n$) represents all variables needed to describe the physical system and the control v (in $\mathbb{R}^m$) contains all parameters that can be modified (as a controller's decision) as time passes. The matrices $A(t)$ and $B(t)$ are the coefficients of the system.

The first question one may ask is the validity of the model, which lead to the *identification* of the coefficients. Next, one may want to *control* the system, i.e., to start with an initial state $x(t_0) = x_0$ and to *drive* the system to a prescribed position $x(t_1) = x_0$. Variations of this question are well known and referred to as *controllability*.

Furthermore, another equation appear,

$$y(t) = C(t)x(t), \tag{2}$$

where $y = y(t)$ is the observation of the state and $C(t)$ is another coefficient. Clearly, y is in $\mathbb{R}^d$ with $d \leq n$. Thus, the problem is to reconstruct the state $\{x(t) : t_0 \leq t \leq t_1\}$ based on the observations $\{y(t) : t_0 \leq t \leq t_1\}$, which is called *observability*.

Another key question is the *stabilization* of the system, where one looks for a feedback, i.e., $v(t) = K(t)y(t)$ such that the closed system of ODE (1) and (2) is stable.

Variation of theses four basic questions: identification, controllability, observability and stabilization are solved in text books.

To each control (and state and observation) a cost (or profit) is associated with the intention of being minimized (or maximized), i.e., a performance index of the form

$$J = \int_0^T [y(t)]^* R(t)y(t)\mathrm{d}t + \int_0^T [v(t)]^* N(t)v(t)\mathrm{d}t \tag{3}$$

is to be optimized. This is called an *optimal control* problem.

Two methods are available to solve optimal control problems, namely, the Pontryagin maximum principle and the Bellman dynamic programming. The above (1), (2), (3) linear-quadratic model can be successfully solved by either method. The maximum principle transforms the given (infinite-dimensional optimization) problem into ODE with initial and terminal conditions and a finite-dimensional optimization problem, i.e., a Lagrange multiplier technique. The dynamic programming transforms the given problem into a non-linear partial differential equation (PDE). There is a vast bibliography under the subject *optimal control*, e.g. classic references such as the text book Bertsekas [23], and Fleming and Rishel [83] or more recently Bardi and Capuzzo-Dolcetta [9], among others.

The ODE defining the evolution equations (of the state and the observation) may be nonlinear and the performance index may have a more general form. Moreover, the state could be distribute, i.e., the evolution equation becomes a PDE. Again, there are many references on the subject.

Both, the maximum principle and the dynamic programming are innovations over the classic calculus of variations. The positive part of the maximum principle is the preservation of the equation type (i.e., if the evolution equation is an ODE then the maximum principle equation is an ODE), and the negative part is the open-loop solution (i.e., the optimal control is of the form $v = v(t)$). On the other hand, the positive part of the dynamic programming is the closed-loop or feedback control (i.e., the optimal control has the form $v = K(t, x(t))$), while the negative part is the new equation (i.e., if the evolution equation is an ODE then the dynamic programming equation is an PDE). It is clear that this material is built on the ODE theory.

In *Stochastic Control*, an *uncertainty* component is added to the previous model. The coefficients becomes *random* and the evolution equation includes a *noise*. Perhaps the most typical example is presented in *signal processing*, where the signal (say x) has some noise. The ODE becomes stochastic

$$\dot{x}(t) = g(t, x(t), v(t)) + (\text{noise}). \tag{4}$$

Since Gauss and Poisson distributions are the main examples of continuous and discrete distributions, the driving noise is usually a Wiener process or a Poisson measure. Again, the four basic questions are discussed. Observability becomes *filtering*, which is very importance. Perhaps the most practical situation is the case with a linear state space and linear observation, which produces the celebrated Kalman filter. Clearly, an average performance index is used for the optimal stochastic control. Again, there is a vast bibliography on stochastic control from variety of points of view, e.g., Fleming and Soner [84], Morimoto [184],Oksendal and Sulem [191], Yong and Zhou [261], Zabczyk [263], among others.

It is clear that stochastic control is mainly based on the theory of stochastic differential equations, which is the main subject of this book.

A Short History

This book-project was developed over a period of many years, first I wanted to collect a number of useful facts about diffusion with jumps, but in the doing so, other ideas appear and I decided to write the material in a lecture notes form, with more details, mainly for my own use. Part of this material was used to deliver some graduate course here and there, but, little by little the project becomes too big or too long for me.

With time, several improvement were made, and searching for possible collaborators, first in my mind and latter in the real world, I was disappointed with my first choice, but very please later on. In particular, with the help of Prof. Luciano Tubaro, a revision of several points were successfully implemented, but, instead of culminated with a regular book published (with Birkhäuser, where we were in the process of signing the book contract) several years ago, more attention was given to the research papers in progress and this book becomes inactive, again. Recently, I reshaped and shorted a little the content to submit it to the American Mathematical Society, and about a year later, a negative report (with no specific points discussed) was sent to me, staying in general lines, that the book should be rewritten and it is of no use as it is. Naturally, I was not happy with the decision, but the reviewer was certainly, an important mathematician specialized in probability. After recovering from this fact, I sent a revised version to another nonprofit publisher (Princeton Press) and after almost another year, while inquiring for a report on my book, I was told how tricky is to market books nowadays and I was asked for specific features that would make this book marketable.

At this point, perhaps I should mention my experience with my two previous books (with Prof. M.G. Garroni), where we dedicated a continuous effort of more than 7 years (although not exclusively!) to write about 600 pages on the Green function and having about 5% of the earning (to be shared with my co-author) of sell profits, essentially to libraries in the world. Thus, with all this short history, I intent to justify my decision of making this internet published book.

Thanks

I would like to express my sincere thanks to Luciano Tubaro for his help in several portions of this book, and to several colleagues (a long undisclosed list goes here) for their comments on certain part and versions of this “unfinished” book. Last, but not less, I owe a great deal of gratitude to my wife, Cristina, who gave me moral support during the many years of preparation and realization of this book.

Preface

The purpose of this book is to develop a solid foundation on Stochastic Differential Equations so that Stochastic Optimal Control can be widely treated. A solid course in measure theory and Lebesgue spaces is a prerequisite, while some basic knowledge in functional spaces and probability is desired. The beginner will find several entry levels to the text, while the expert may access any Chapter (or Section) with minimum reference to previous material. Each Chapter has been kept as independent as possible.

A quick look at the contents will convince the reader that actually, each chapter could be developed into a full lecture notes. However our choice is to comfortably place ourselves within probability (stochastic processes) and (functional) analysis, to be able to properly describe the mathematical aspect of the stochastic differential equations needed for the dynamic programming technique, specifically, for controlled diffusion processes with jumps in a d-dimensional region. In a way, this material may be called *Complements of Stochastic Analysis* which could be used as a neutral introduction to stochastic optimal control, without any direct application, only the state of the system is considered.

Starting at a good elementary level of measure theory, the first and second chapters are an introduction to Markov-Feller processes, in the form of an extended summary of what may be necessary to consult (to follow our main subject). Thus, Chapter 1 is essentially addressed to the analyst with little background in probability while Chapter 2 presents an overview of semigroups. Only precise references to proofs are given and therefore these two first chapters are hard to read for the first time. Moreover, they may be considered as service chapters, where notation and fundamental concepts are briefly discussed.

Our main material begins in Chapter 3, where we study continuous and right-continuous with left-hand limits stochastic processes, locally bounded variation processes, martingales, piecewise deterministic processes and Lévy processes. Chapter 4 treats Gaussian processes, random measures, stochastic integrals and differentials. Chapter 5 is dedicated to the construction of d-dimensional (controlled) diffusion processes with jumps, by means of stochastic ordinary differential equations and martingale problems. Basic proofs are kept to be a minimum while focussing on estimates and key concepts. Each of these first five chapters is rather independent, in the sense that the material covered in each one can be considered as a *step further* into the *neighborhood* of Markov processes.

The reader with a good knowledge in Markov processes should skip the

first three chapters, which should be taken as a short guided tour in stochastic analysis. Moreover, Chapter 3 can be viewed as a complement to the theory of processes with continuous sample paths, where more emphasis is given to Markov jump processes. Since some ideas of the main proofs are included, Chapter 3 is not so informal as the first two chapters. In Chapter 4, most of the details are given, the center is stochastic integration and differentials, but the focus is on estimates. This Chapter 4 can be skipped too, in a first reading, and coming back to it later if needed.

Someone familiar with Wiener processes and Itô stochastic integration will begin reading Chapter 4, and occasional go back to Chapter 3 to clear up some points in the discussion, and eventually to the end of Chapter 2, where second order integro-differential operators are briefly treated. Chapters 5 starts with stochastic differential equations, to end with the diffusion processes with jumps, as the main example of Markov-Feller processes. This is a formal chapter, the style theorems and proofs is used, and again, the focus is on estimates. Next, Chapters 6 and 7 discuss advanced topics in stochastic differential equations, e.g., oblique boundary conditions, backward equations and other more analytic tools.

In a way, the introduction to each chapter is essentially a delimiter (on the style) to test the reader's background necessary to understand what follows. To complement the theory, some exercises are found only in Chapter 1. Other chapters are more lecture style, where the main points are carefully treated, but the instructor (or reader) should complement the theory.

Depending on the background of the reader, this book begins in Chapter 1 (hard to start there, but possible), or in Chapter 3 (adequate level in stochastic processes), or in Chapter 4 (good level in Markov processes), or in Chapter 5 (good level in martingales), or in Chapter 6 (advanced level in stochastic differential equations). Eventually, the reader should go back to previous Chapters (and the references mentioned there to check the proofs of some basic results) and keep on the side the analysis view point of Chapter 2. Essentially, each of the five Chapters is an independent unit, which may be accessed directly. The last Chapter is mainly addressed to the expert in the field, where the theory is well-known.

Even if the heart of the book starts with Chapter 5, the beginner should realize that not only Chapters 5, 6, and 7 are relevant and worth to study in this area, but the whole path throughout the various topics briefly discussed in previous chapters are necessary to appreciate the full implication of the theory of stochastic differential equation with jumps.

Michigan (USA), *Jose-Luis Menaldi*, April 2008

Part I

Stochastic Processes with Jumps

Chapter 1

Probability Spaces

The purpose of this chapter is to supply the readers (who have taken a solid course in measure and integration theory but have little background in probability theory) with a *crash introduction* to stochastic processes. The focus is the *neighborhood* of Markov processes in continuous time. The first reading may be a little hard, since only references to proofs are given. The last section is only to complement the subject, some of the exercises are not so *simple exercises*, basically the reader should consult the references for the proofs if exercises are under consideration. Let us mention that a comprehensive introduction to probability, assuming *measure theory*, can be found in Stromberg [237] and in Stroock [238], among others. For instance, even without assuming measure theory, an introduction to probability can be found in Taylor [248], while an *analysis oriented* course on *diffusion processes* is given in Krylov [141]. An extensive classic study of the general theory of processes can be found in Dellacherie and Meyer [58], Gihman and Skorohod [97], Rao [208] and Sharpe [225]. For a complete discussion for foundation of probability, the reader may check the treatises De Finetti [57] and Loève[159], among many others.

1.1 Random Variables

Let $(\Omega, \mathcal{F})$ be a measurable space i.e., $\mathcal{F}$ is a σ-algebra of subsets in Ω. A *random variable* is a measurable mapping on $(\Omega, \mathcal{F})$, e.g. a real random variable x is a measurable function from $(\Omega, \mathcal{F})$ into $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B} = B(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$. Most of the information that we are interested in of a random variable x is contained in the σ-algebra generated by x i.e., $x^{-1}(\mathcal{B}) = \{F \in \mathcal{F} : x(F) \in \mathcal{B}\}$. Thus if x is a characteristic (or indicator) function

$$x(\omega) = \begin{cases} 1 & \text{if } \omega \in F, \\ 0 & \text{if } \omega \in \Omega \setminus F, \end{cases}$$

for some F in $\mathcal{F}$, then $x^{-1}(\mathcal{B}) = \{\Omega, \emptyset, F, \Omega \setminus F\}$. If $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are two measurable spaces, $\xi : \Omega \to \Omega'$ and $x : \Omega \to \mathbb{R}$ are two random variables,

then x is $\sigma(\xi)$-measurable, i.e., $x^{-1}(\mathcal{B}) \subset \xi^{-1}(\mathcal{F}')$, if and only if there exists a measurable map $\eta : \Omega' \to \mathbb{R}$ such that $x(\omega) = \eta(\xi(\omega))$ for any ω in Ω. This is proved by means of a monotone class argument, see Exercises 1.2 and 1.4. Moreover, this remains true if $\mathbb{R}$ is replaced by a Polish space, i.e., a complete separable metric space.

A *stochastic process* is a collection of random variables indexed by some set e.g., a real valued stochastic process $X = \{X_t : t \in T\}$ is a family of measurable functions $X_t : \Omega \to \mathbb{R}$, with $t \in T$. Sometimes, the same process is denoted by $X = \{X(t) : t \in T\}$. Certainly, we can replace $\mathbb{R}$ with $\mathbb{R}^d$ in the previous discussion with almost not conceptual changes. Usually, when the random variables are indexed by a discrete set (countable set of isolated and totally ordered points) i.e. $\{\dots, -1, 0, 1, \dots\}$ or $\{1, 2, \dots\}$, we speak of a random sequence or a time series. In this context, we can view a time series as a random variable with values in $\mathbb{R}^\infty$, the set of real valued sequences $\{(x_1, x_2, \dots) : x_i \in \mathbb{R},\ \forall i\}$. Here, we endowed $\mathbb{R}^\infty$ with the product topology and its associated Borel σ-algebra (e.g., Shiryayev [227]). A similar argument can be applied in general, but the discussion is more delicate. Thus, it is preferable to reserve the term *process* for uncountable index set T.

When the index set T is uncountable with a natural σ-algebra on it (for instance T is an interval), we restrict our attention to *measurable* stochastic process X i.e., we assume that the function $X : \Omega \times T \to \mathbb{R}$ is measurable. Moreover, if the index set T has a given topology and the stochastic process takes values in a topological space i.e., $\mathbb{R}^d$, then the following notions are necessary

Definition 1.1 (separable)**.** A d-dimensional *stochastic process* $\{X_t : t \in T\}$, $T \subset [0, +\infty)$ is *separable* if there exists a countable dense set of indexes $I \subset T$ (called separant) such that for any t in T and any ω in Ω there exists a sequence $\{t_n : n = 1, 2, \dots\}$ of elements in I which is convergent to t and such that $X(t_n, \omega)$ converges to $X(t, \omega)$. □

For instance, the reader may want to take a look at the book by Meyer [179, Chapter IV] to realize the complexity of this notion of separability.

Unless otherwise stated, when referring to a stochastic process $\{X_t : t \in T\}$ in a measurable space $(\Omega, \mathcal{F})$, when T is a topological space, we mean a measurable and separable stochastic process, as understood from the context. Thus we denote by $\mathcal{L}^0(\Omega \times T, \mathbb{R}^d)$ the set of measurable stochastic processes with values in $\mathbb{R}^d$. Naturally, we can identify $X_t(\omega)$ with a measurable function in t, for each fixed ω, so that $\mathcal{L}^0(\Omega \times T, \mathbb{R}^d) = \mathcal{L}^0(\Omega, \mathcal{L}^0(T, \mathbb{R}^d))$ with the corresponding product σ-algebra (see Exercise 1.5). Thus we may look at a d-dimensional stochastic process as a random variable with values in $\mathcal{L}^0(T, \mathbb{R}^d)$. On the other hand, we may need to consider processes continuous in probability (see Section 1.6 Versions of Processes) which are not expressible in terms of random variables.

Definition 1.2. A d-dimensional stochastic process $\{X_t : t \in T\}$, $T \subset [0, +\infty)$ is *continuous* if for any $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is continuous. □

Note that in the previous definition, the continuity is used as a global condition. Thus, if we denote by $\mathcal{C}^0(T, \mathbb{R}^d)$ the set of continuous functions, we may regard a d-dimensional stochastic process as a random variable with values in $\mathcal{C}^0(T, \mathbb{R}^d)$, provided a σ-algebra is defined on it. Similarly, we may define right (left) continuous and increasing (decreasing, locally bounded variation) processes.

When an order is given on the index set T, most of the information of a stochastic process X is contained in the history σ-algebra, i.e., the family $\mathcal{F}_t$ or $\mathcal{F}(t)$ defined as the minimal sub σ-algebra of $\mathcal{F}$ that makes the random variables $\{X_s : s \leq t\}$ measurable. This is an increasing family of σ-algebra i.e., $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$, which is called the natural *filtration* associated with the stochastic process. Most of the time, the index set $T = [0, +\infty)$. In this case, for a given measurable and separable process $\{X_t : t \geq 0\}$ we associate a natural filtration $\{\mathcal{F}_t : t \geq 0\}$ as before. Certainly, X is *adapted* to the natural filtration i.e., the random variable X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$. Also, X is *progressively measurable* with respect to the natural filtration i.e., the restriction of X to the set $\Omega \times [0, t]$ is measurable with respect to the product σ-algebra $\mathcal{F}_t \times \mathcal{B}([0, t])$, for any $t \geq 0$. Here, and in what follows, $\mathcal{B}(T)$ denotes the σ-algebra of Borel subsets of T, $T \subset \mathbb{R}^d$.

If the filtration is given a priori (independently of the stochastic process), then we will refer to as a stochastic process being *adapted* or *progressively measurable* with respect to the given filtration if the above conditions are satisfied. Moreover, we will see later that it is convenient to *normalize* the filtration to *standard* (or usual) conditions. As a caution, technical, we refers adapted as "adapted and measurable". However, note that sometimes it may be convenient to consider the notion of measurable independently of adapted, in this case, we may have a measurable process Y such that the mapping $\omega \mapsto Y(t, \omega)$ is $\mathcal{F}(t)$-measurable, but Y is not progressively measurable.

This is essentially how far the analysis can go on measurable spaces. As soon as a probability measure space $(\Omega, \mathcal{F}, P)$ is given, any random variable is identified with its equivalence class. The same applies to processes when considered as random variables on function spaces, e.g., $\mathbb{R}^T$ or $C(T, \mathbb{R})$. In general, we may say that a measurable function from the sample space $(\Omega, \mathcal{F})$ into another measurable space $(E, \mathcal{E})$ is a random variable, and it is called a stochastic process if the value spaces has the form $(E^T, \mathcal{E}^T)$, for some set of indexes T (usually a subset of $\mathbb{R}$). Moreover, when a probability measure P is given on the measurable space $(\Omega, \mathcal{F})$ then random variables and stochastic processes are identified with their corresponding P-equivalence classes. For a given E-valued random variable x, the probability measure defined by $P_x(B) = P\{x^{-1}(B)\}$, B in $\mathcal{E}$, is called the distribution of x. However, for a given E-valued stochastic process X the family of probability on $\mathcal{E}^n$, $n \geq 1$, defined by $P_X(B_1 \times \cdots \times B_n) = P\{X(t_1, \omega) \in B_1, \ldots, X(t_n, \omega) \in B_n\}$, $B_1, \ldots, B_n$ in $\mathcal{E}$, $t_1, \ldots, t_n$ in T, is called the finite-dimensional distributions of X.

As long as the index set T is countable, no more detail is needed, however, for an uncountable index set T, e.g., $T = [0, \infty)$, we need to use the notion of *version* and *realization* of a stochastic process. Indeed, for a given

(stochastic) process $\{X_t : t \in T\}$ on a probability space $(\Omega, \mathcal{F}, P)$ we say that $\{Y_t : t \in T\}$ is a *version* (or a modification) of the process $\{X_t : t \in T\}$ if $P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1$, for any t in T. However, given a set of a priori properties that a process should satisfy (e.g., its finite-dimensional distributions, an assumption of continuity or measurability on its paths, or some other condition) then *realization* is a process, a probability space and any other items (such as a filtration) used to verify all required properties. Sometimes, we will refer to processes (not necessary defined on the same probability space) having the same finite-dimensional distribution or what is essentially the same (provided some regularity on the paths is assumed) having the same law in E^T or $C(T, E)$, as discussed later.

Only the case where the value set E is a complete separable metric space (Polish space), e.g., $E \subset \mathbb{R}^d$, endowed with the Borel σ-algebra $\mathcal{B}(E)$, and the set of index T is a totally ordered complete separable metric space, e.g., $T = [0, \infty)$, will be discussed herein. When the set of index T is uncountable, we impose some property (e.g., separability or continuity) on processes so that the value space E^T is replaced by better a space, e.g., E^I, I countable and dense in T, or $C(T, E)$ as discussed later.

Sometimes when dealing with extended real-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$ we may need a definition of convergence in measure for random variables which may take values $\pm\infty$ with strictly positive probability. In this context we say that a sequence $\{x_n : n = 1, 2, \dots\}$ of random variables converges in measure to another random variable x if the sequence $\arctan(x_n)$ converges in measure to $\arctan(x)$ in the usual sense, equivalently, if

$$\lim_{n\to\infty} \mathbb{E}\{|\arctan(x_n) - \arctan(x)|\} = 0,$$

where $\mathbb{E}\{\cdot\}$ denote the *mathematical expectation*, i.e., the integral with respect to the probability measure P. The metric $d(x, y) = \mathbb{E}\{|\arctan(x) - \arctan(y)|\}$ on the space S of extended real-valued random variables make S a complete metric space, after the identification of two random variables whenever they are equal almost surely. Thus a measurable process $\{X_t : t \geq 0\}$ in the previous sense is (essentially) a Borel measurable mapping $t \mapsto X_t$ from $[0, +\infty)$ into S, we refer to Doob [60, pp. 407–410] for more details.

A typical generalization is to consider random variables with values in a *Polish space* (i.e, a complete and separable metric space), which is the analogous of stochastic processes if the Polish space is a function space. Stochastic processes are meant to model phenomenon which evolves in time in a random way. It is usually admitted that most often statistical experiments or physical considerations can only give information about the so-called finite-dimensional distributions of a process (see Section 1.10 and note that two processes may have the same finite-dimensional distributions but having not the same probability space of reference). Therefore the choice of the Polish space becomes relevant for mathematical considerations. For instance, consider the real-valued processes $X_t(\omega) = 1$ for every t in $[0, 1]$, and $Y_t(\omega) = 1$ only when $t \neq \omega$ and $Y_t(\omega) = 0$ otherwise. It is clear that X is a continuous process while Y is (Borel)

measurable but it is not separable. Moreover, if the probability measures (in which they are considered) have not atoms (each single event $\{\omega\}$) has zero probability) then these two processes have the same finite-dimensional distributions and from the phenomenological viewpoint they should be considered the same process. Mathematically, we prefers to take X.

Hence, in modeling a time-evolution random phenomenon, we are allowed to choose a *realization* of the process most suitable for our mathematical purpose. Questions like *is this process continuous?* really means *does there exist some process with the given finite-dimensional distributions whose paths are (almost sure) continuous?* or what is the same *is there a continuous realization of the process?*. This means that we can select the probability space $(\Omega, \mathcal{F}, P)$ and the map X among those satisfying the prescribed properties on the finite-dimensional distributions of the process. It will be clear by the end of this chapter, that there is a *canonical way* to performing this procedure of selecting a suitable realization such that the sample space Ω is a suitable Polish space and X is the *identity* as a random variable or the *coordinates mappings* if viewed as a stochastic process.

In what follows, we are going to denote *indistinctly* the notation $P(\{\cdot\})$, $P(\cdot)$ or $P\{\cdot\}$ for the probability measure, where the *dot* $\cdot$ represents a condition defining a set of events.

1.2 Distributions and Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability space i.e., P is a measure on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$, called a probability measure. A measurable set (or a set in $\mathcal{F}$) is called an event. When a probability measure is involved, the previous concept of random variables becomes *equivalence classes* of random variables. For instance we may use the Lebesgue Banach spaces $L^p = L^p(\Omega, \mathcal{F}, P)$, for any $1 \leq p \leq \infty$. However, the study of stochastic processes is more delicate, since the family of random variable may not be countable.

As mentioned early, the *distribution* (or law) of a given random variable x is the probability measure P_x induced by x on $\mathcal{B}$ i.e., if x is a real random variable then its distribution is given by

$$P_x(B) = P(\{\omega : x(\omega) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Perhaps the three most important one-dimensional laws on $\mathbb{R}$ are the Gaussian (or normal) distribution, with parameters m and $r > 0$ $[N(m, r)]$, which has support on $\mathbb{R}$ and is given by

$$P_g(B) = \int_B (2\pi r)^{-1/2} \exp(-\tfrac{|x-m|^2}{2r}) \mathrm{d}x,$$

the Poisson distribution, with parameter $\lambda > 0$, which has support on $\mathbb{N}$ and is given by

$$P_p(B) = \exp(-\lambda) \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \mathbb{1}_{(n \in B)},$$

and the exponential distribution, with parameter $\alpha > 0$, which has support on $\mathbb{R}_0^+ = [0, +\infty)$ and is given by

$$P_e(B) = \int_{B \cap \mathbb{R}_0^+} \alpha \exp(-\alpha x) \mathrm{d}x.$$

Thus the mean and the variance are as follows (see Exercise 1.7)

$$\int_{\mathbb{R}} x P_g(\mathrm{d}x) = m, \qquad \int_{\mathbb{R}} (x-m)^2 P_g(\mathrm{d}x) = r^2,$$
$$\int_{\mathbb{R}} x P_p(\mathrm{d}x) = \lambda, \qquad \int_{\mathbb{R}} (x-\lambda)^2 P_p(\mathrm{d}x) = \lambda,$$
$$\int_{\mathbb{R}} x P_e(\mathrm{d}x) = \alpha^{-1}, \qquad \int_{\mathbb{R}} [x-\alpha^{-1}]^2 P_e(\mathrm{d}x) = \alpha^{-2}.$$

The *characteristic* function (or Fourier transform) of a distribution (or probability law) P on $\mathbb{R}$ is the complex-value function

$$\widehat{P}(\xi) = \int_{\mathbb{R}} \mathrm{e}^{\mathtt{i}x\xi} P(\mathrm{d}x), \quad \forall \xi \in \mathbb{R},$$

with $\mathtt{i} = \sqrt{-1}$, and if the distribution P is on $\mathbb{R}_0^+$ then its *Laplace* transform is also defined

$$\widetilde{P}(\zeta) = \int_{\mathbb{R}_0^+} \mathrm{e}^{-x\zeta} P(\mathrm{d}x), \quad \forall \zeta \in \mathbb{R}_0^+.$$

For the previous distributions we have

$$\widehat{P_g}(\xi) = \exp\big(-\frac{1}{2} r\xi^2 + \mathtt{i}m\xi\big),$$
$$\widehat{P_p}(\xi) = \exp\big(\lambda(\mathrm{e}^{\mathtt{i}\xi} - 1)\big), \qquad \widetilde{P_p}(\zeta) = \exp\big(\lambda(\mathrm{e}^{-\zeta} - 1)\big),$$
$$\widehat{P_e}(\xi) = \frac{\lambda}{\lambda - \mathtt{i}\xi}, \qquad \widetilde{P_e}(\zeta) = \frac{\lambda}{\lambda + \zeta}.$$

There are others noted laws, such as the Cauchy distribution μ with parameters m and $c > 0$ and the Γ-distribution ν with parameters $c > 0$ and $\alpha > 0$ given by

$$\mu(B) = \pi^{-1} c \int_B \big[(x-m)^2 + c^2\big]^{-1} \mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$
$$\nu(B) = \frac{\alpha^c}{\Gamma(c)} \int_{B \cap \mathbb{R}_0^+} x^{c-1} \mathrm{e}^{-\alpha x} \mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}_0^+),$$

with

$$\widehat{\mu}(\xi) = \exp(-c|\xi| + \mathtt{i}m\xi),$$
$$\widehat{\nu}(\xi) = (1 - \mathtt{i}\alpha^{-1}\xi)^{-c}, \qquad \widetilde{\nu}(\zeta) = (1 + \alpha^{-1}\zeta)^{-c}.$$

The Cauchy distribution μ does not have a mean value (i.e, μ does not integrate the function $|x|$) and the Γ-distribution has mean value equal to c/α. The exponential distribution is a particular case of the Γ-distribution, $c = 1$, and the Γ-distribution with $c = n/2$ and $\alpha = 1/2$ is referred to as the χ^2-distribution with n degrees of freedom. All these distributions are infinitely divisible, e.g., see Sato [220, Section 1.2, pp. 7-14].

Definition 1.3 (independence)**.** A family $\mathcal{A}$ of measurable sets is (mutually) independent (relative to the probability P) if their elements are mutually independent, i.e., if for any finite number of sets $A_1, \dots, A_n$ in $\mathcal{A}$ we have

$$P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i). \tag{1.1}$$

Now, a family of σ-algebras is (mutually) independent if any finite number of σ-algebras $\mathcal{F}_1, \dots, \mathcal{F}_n$ in the family and any sets A_i in $\mathcal{F}_i$ we have (1.1). Similarly, a family of random variables is (mutually) independent if the family of their generated σ-algebras is (mutually) independent. □

Actually, for brevity we say *a family $\mathcal{A}$ of measurable sets is mutually independent relative to the probability P*, instead of saying a family $\mathcal{A}$ composed by measurable sets which are mutually independent relative to the probability P. However, in all what follows, we refer to *mutually independent* by saying only *independent*, i.e., we say a family of independent sets and a family of independent variables (or σ-algebras).

If $\mathcal{A}_i \subset \mathcal{F}$ is a family on a probability space $(\Omega, \mathcal{F}, P)$ indexed by $i \in I$, we define $\{\mathcal{A}_i : i \in I\}$ as independent if for any finite number of index $J \subset I$ and for any sets A_i in $\mathcal{A}_i$, $i \in J$, we have (1.1). It is clear that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras of $\mathcal{F}$, which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$ (i.e., $\sigma(\mathcal{H}_0) = \mathcal{H}$ and $\sigma(\mathcal{G}_0) = \mathcal{G}$ (recall that a π-system means a collection of subsets closed or stable under finite intersections) then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$, see Exercise 1.8.

Note that given a family $\mathcal{A}$ of three (or more) measurable sets, we may say that $\mathcal{A}$ is *pairwise* independent if any two subsets A_1 and A_2 of $\mathcal{A}$ are independent, i.e., $P(A_1 \cap A_2) = P(A_1)P(A_2)$. Clearly, this is distinct from the concept of *mutually* independent just defined. The same remark can be used for two or more families of either sub σ-algebras or random variables. On the other hand, two families $\mathcal{A}_1$ and $\mathcal{A}_2$ of measurable sets are (mutually or equivalently pairwise) independent $P(A_1 \cap A_2) = P(A_1)P(A_2)$ for any A_1 in $\mathcal{A}_1$ and A_2 in $\mathcal{A}_2$. Similarly, this definition can be extended to three or more families of measurable sets, where we need to distinguish between mutually and pairwise independent.

Note that if A and B are independent, i.e., $P(A \cap B) = P(A)P(B)$, then a simple calculation shows that $A' = \Omega \setminus A$ and B are also independent. As a consequence, if $\mathcal{F}_i$ denotes the σ-algebra generated by F_i, i.e., $\mathcal{F}_i = \{A_i, \Omega \setminus$

$A_i, \emptyset, \Omega\}$, then a family of measurable sets (events) $\{A_i : i \in I\}$ is independent if and only if the family of σ-algebras $\{\mathcal{F}_i : i \in I\}$ is independent.

Thus, a sequence of independent random variables $\{x_i : i \in I\}$ is independent if and only if

$$P\big(\bigcap_{j\in J}\{\omega : x_j(\omega) \in A_j\}\big) = \prod_{j\in J} P\big(\{\omega : x_j(\omega) \in A_j\}\big)$$

for any measurable sets A_j and any finite subset J of I. In term of the characteristic functions, this is equivalent to

$$\mathbb{E}\big\{\exp[\mathtt{i}\sum_{j\in J}\alpha_j x_j]\big\} = \prod_{j\in J}\mathbb{E}\big\{\exp[\mathtt{i}\alpha_j x_j]\big\},$$

for any constants α_j and any finite subset J of I, where $\mathtt{i} = \sqrt{-1}$. There is a very close connection between the concepts of *independence* and *Cartesian product*. If x and y are two real valued random variables, we may look at (x, y) as a two-dimensional real valued random variable, then a direct comparison with the definition of independence shows that the fact that x and y are independent may be very simply expressed by the equation

$$P_{(x,y)} = P_x \times P_y,$$

i.e., the joint distribution of x, y is equal to the Cartesian product of the single distributions of x and y.

• *Remark* 1.4. If x is a Normal distributed random variable with parameters m and $r > 0$ then it characteristic function is given by

$$P\{\exp(\mathtt{i}\xi x)\} = \exp(-\frac{1}{2}r^2\xi^2 + \mathtt{i}m\xi).$$

Hence, if x_i, $i = 1, \dots, k$ are independent Normal distributed random variables with parameters m_i and $r_i > 0$ then any linear combination $x = c_1x_1 + \cdots + c_kx_k$, with c_i real numbers, is indeed a Normal distributed random variables with parameters $m = m_1 + \cdots + m_k$ and $r = \sqrt{r_1^2 + \cdots + r_k^2}$. Similarly, if x is a Poisson distributed random variable with parameter $\lambda > 0$ then it characteristic function is given by

$$P\{\exp(\mathtt{i}\xi x)\} = \exp\big(\lambda(\mathrm{e}^{\mathtt{i}\xi} - 1)\big).$$

Thus, if x_i, $i = 1, \dots, k$ are independent Poisson distributed random variables with parameters λ_i then the sum $x = x_1 + \cdots + x_k$ is indeed a Poisson distributed random variables with parameter $\lambda = \lambda_1 + \cdots + \lambda_k$. However, if x_i, $i = 1, \dots, k$ are independent exponentially distributed random variables with the same parameter λ, i.e., with characteristic function

$$\mathbb{E}\{\exp(\mathtt{i}\xi x_1)\} = \frac{\lambda}{\lambda - \mathtt{i}\xi} = \big(1 - \mathtt{i}\lambda^{-1}\xi\big)^{-1},$$

then the sum $x = x_1 + \cdots + x_k$ has a Gamma distribution with parameters λ and k, i.e.,

$$\mathbb{E}\{\exp(\mathtt{i}\xi x)\} = \big(1 - \mathtt{i}\lambda^{-1}\xi\big)^{-k} \quad \text{or} \quad P\{x \in \mathrm{d}t\} = \frac{\lambda^k t^{k-1}\mathrm{e}^{-\lambda t}}{(k-1)!}\,\mathrm{d}t.$$

On the other hand, the process of counting an independent identically exponentially distributed sequence $\{x_i\}$ of random variables with parameter λ, i.e., $n(t) = \sum_i \mathbb{1}_{x_i \le t}$, produces a family of random variables, indexed by $t \ge 0$, identically Poisson distributed with parameter $t\lambda$. □

• *Remark* 1.5. Certainly, there many other useful distributions, e.g., (1) the deterministic or delta or Dirac distribution, which is concentrated at one point say $x = x_0$, i.e., $P\{x = x_0\} = 1$ and $P\{\exp(\mathtt{i}\xi{\cdot}x)\} = \exp(\mathtt{i}\xi{\cdot}x_0)$, (2) the uniform or Lebesgue distribution, which is uniformly distributed on a region with finite volume, e.g., a random variable uniformly distributed over an interval $[a, b]$ has distribution $P\{\alpha < x \le \beta\} = (\min\{b, \beta\} - \max\{a, \alpha\})/(b - a)$, for every real numbers $\alpha \le \beta$, and (3) the compound Poisson distribution, which is frequently used and can be described as follows: if n is a Poisson distributed random variable with parameter $\lambda > 0$ which is independent of a sequence of independent identically distributed random variables $\{x_i\}$ with F as its common distribution satisfying $F(0) = 0$, then the random sum $x(\omega) = x_1(\omega) + \cdots + x_{n(\omega)}(\omega)$, complemented with the condition $x(\omega) = 0$ if $n(\omega) = 0$, has a compound Poisson distribution with parameters λ and F. Note that the condition that $F(0) = 0$ ensures that $x = 0$ only when $n = 0$. If F is a distribution in $\mathbb{R}^m_* = \mathbb{R}^m \smallsetminus \{0\}$ then the k-convolution F^{*k} is the distribution of the independent sum $x_1 + \cdots + x_k$, and therefore, the *compound Poisson distribution* of a random variable x in $\mathbb{R}^m_*$ is given by

$$P\{x \in B\} = P_{cp}(B) = \exp(-\lambda)\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}F^{*k}(B),$$

with its characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot x)\} = \widehat{P_{cp}}(\xi) = \exp\big(\lambda(\mathrm{e}^{\widehat{F}(\xi)} - 1)\big),$$

where $\widehat{F}$ is the characteristic function of the distribution F. It is interesting to remark that the two parameters λ and F can be combined to produce finite measure π, with $\lambda = \pi(\mathbb{R}^m_*)$ and $F = \pi/\lambda$, i.e., a compound Poisson random variable x with parameter π has characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot x)\} = \exp\Big(\int_{\mathbb{R}^m_*}(\mathrm{e}^{\mathtt{i}x\cdot\xi} - 1)\pi(\mathrm{d}x)\Big),$$

which reduces to the (simple) Poisson distribution if F is a deterministic, e.g., if the finite-measure π is given by $\pi(B) = \lambda\delta_1(B) = \lambda\mathbb{1}_{1\in B}$, with B a Borel subset of $[0, \infty)$. It is also clear that a compound Poisson random variable x may or may not have first moment (or mean), but if it does, the expression $y = x - \mathbb{E}\{x\}$

produces what is called a *centered* or *compensated* Poisson distribution with parameter π and characteristic function

$$P\{\exp(\mathtt{i}\xi \cdot y)\} = \exp\Big(\int_{\mathbb{R}^m_*} \big(\mathrm{e}^{\mathtt{i}x\cdot\xi} - 1 - \mathtt{i}x\cdot\xi\big)\pi(\mathrm{d}x)\Big),$$

where the finite measure π must integrate the function $x \mapsto |x|$ and the random variable y has zero mean. At this point, let us mention that a distribution is infinitely divisible if and only if it is a limit of a sequence of compound Poisson distributions, e.g., see Prabhu [204, Chapter 4, pp.43–68]. □

The construction of examples of independent sequences of random variables involve some conditions (infinitely divisible) on the probability space $(\Omega, \mathcal{F}, P)$, for instance if the σ-algebra $\mathcal{F} = \{\emptyset, F, \Omega \smallsetminus F, \Omega\}$, with $P(F) > 0$, then any two independent sets A and B must be such that $A = \emptyset$ or $B = \emptyset$. There are many (classic) properties related to an independent sequence or series of random variables, commonly known as the (weak and strong) law of large numbers and the central limit theorem, e.g., the reader is referred to the classic probability books Doob [59], Feller [81] and Gnedenko [101], while an analytic view can be found in Dudley [62], Folland [85, Chapter 10], Halmos [104]), Stromberg [237] and Stroock [238].

• *Remark* 1.6 (conditional independence). In elementary probability theory we define the conditional probability of a event A (i.e., a measurable set) given another event C with $P(C) > 0$ as $P(A \,|\, C) = P(A \cap C)/P(C)$. A more sophisticate concept is the following: Given a measurable set C with $0 < P(C) < 1$, a family $\mathcal{A}$ of measurable sets is (mutually) conditional independent given C (relative to the probability P) if their elements are mutually conditional independent given C, i.e., if for any finite number of sets $A_1, \dots, A_n$ in $\mathcal{A}$ we have

$$P\big(\bigcap_{i=1}^{n} A_i \,|\, C\big) = \prod_{i=1}^{n} P\big(A_i \,|\, C\big). \tag{1.2}$$

Now, a family of σ-algebras is (mutually) conditional independent given C, if for any finite number of σ-algebras $\mathcal{F}_1, \dots, \mathcal{F}_n$ in the family and any sets A_i in $\mathcal{F}_i$ we have (1.2). Similarly, a family of random variables is (mutually) conditional independent given C, if the family of their generated σ-algebras is (mutually) conditional independent given C. Moreover, if (1.2) holds for every C with $0 < P(C) < 1$ in a σ-algebra $\mathcal{C}$ then the sets $A_1, \dots, A_n$ are mutually conditional independent given $\mathcal{C}$, and similarly, if this holds for every set A_i in $\mathcal{F}_i$ then the family of σ-algebras is (mutually) conditional independent given (the σ-algebra) $\mathcal{C}$. □

Note that if we allow $C = \Omega$ then we foldback to the case of complete independence Definition 1.3. It is also clear that condition (1.2) can be rewritten as

$$P\big(\bigcap_{j=1}^{n-1} A_{i_j} \,|\, A_{i_n},\, C\big) = \prod_{j=1}^{n-1} P\big(A_{i_n} \,|\, C\big),$$

for any permutation $\{i_1, \dots, i_n\}$ of $\{1, \dots, n\}$, provided $P(A_{i_n} \cap C) > 0$.

As in the case of complete independence, if $\mathcal{F}_i$ denotes the σ-algebra generated by F_i, i.e., $\mathcal{F}_i = \{A_i, \Omega \setminus A_i, \emptyset, \Omega\}$, then a family of events $\{A_i : i \in I\}$ is conditional independent given C if and only if the family of σ-algebras $\{\mathcal{F}_i : i \in I\}$ is conditional independent given C. However, if (1.2) condition holds for C then it does not necessarily hold for $\Omega \setminus C$, i.e., conditional independence with respect to an event C is not necessarily the same as conditional independence with respect to the σ-algebra generated by C, namely, $\mathcal{C} = \{\emptyset, \Omega, C, \Omega \setminus C\}$, see later Definition 1.13 on conditional independence.

Clearly, the conditional probability $P(\cdot \,|\, C)$ is itself a probability and thus, conditional independence given a set C is just independence under the conditional probability. Moreover, condition (1.2) can be rewritten in a less intuitive way as

$$P\big(\bigcap_{i=1}^{n} A_i \cap C\big)\,[P(C)]^{n-1} = \prod_{i=1}^{n} P\big(A_i \cap C\big),$$

which becomes $P(A_1 \cap A_2 \cap C)P(C) = P(A_1 \cap C)P(A_2 \cap C)$ when $n = 2$. In particular, the space $\Omega = A_1$ is conditional independent of any event $A = A_2$ given any event C with $P(C) > 0$. Also, if three events A, B and C are pairwise independent with $P(C) > 0$ then A and B are conditional independent given C, if and only if they are mutually independent according to Definition 1.3. Certainly, if $P(C) = 0$ the above equality could be used instead of condition (1.2), but it is trivially satisfied and the definition is meaningless. Some comments on this concept are given later, with the use of conditional expectation.

Going back to the three examples of distributions in $\mathbb{R}$, we can extend them to $\mathbb{R}^d$ as follows. Consider n independent identically distributed random variables $(\xi_1, \dots, \xi_n)$, a linear transformation q from $\mathbb{R}^n$ into $\mathbb{R}^d$, and then for a given $m \in \mathbb{R}^d$, we look at the distribution of the random variable $y = m + Q(\xi_1, \dots, \xi_n)$. Identifying the linear transformation Q with a canonical matrix, still denoted by Q, we deduce (see Exercise 1.9) that if the common distribution of $(x_1, \dots, x_n)$ is Gaussian, then

$$P_y(B) = \int_B p_n(x)\mathrm{d}x, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

where

$$p_n(x) = [2\pi \det(QQ^*)]^{-d/2} \exp\Big(-\frac{[(x-m)^*(QQ^*)^{-1}(x-m)]^2}{2}\Big), \quad (1.3)$$

and the * means the transpose (of a matrix), $\det(\cdot)$ is the determinant, and we have assumed that QQ^* is invertible. This is a Gaussian d-dimensional distribution. Similarly, d-dimensional Poisson (or exponential) distribution can be described.

Sums of independent random variables are studied with the purpose of generalizing and elaborating the *law of the large numbers*. Let $\{x_i : i \geq 1\}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$, and

let $\{s_n : n \geq 1\}$ be the sequence of partial sum, $s_n = x_1 + \cdots + x_n$. The first point is the Kolmogorov's *zero-one law*, namely, the series s_n converges almost surely or diverges almost surely (i.e., cannot converges for some ω and diverges for others ω). Next, if the two series of real numbers $\sum_i \mathbb{E}\{x_i\}$ (mean) and $\sum_i \mathbb{V}\{x_i\}$ (variance, $\mathbb{V}\{x\} = \mathbb{E}\{\big[x - \mathbb{E}\{x\}\big]^2\}$) converge then the series $\sum_i x_i$ converges almost surely. Another result, known as the *three series theorem*, affirms that $\sum_i x_i$ converges almost surely if and only if the following three series of real numbers $\sum_i \mathbb{E}\{x_i'\}$, $\sum_i \mathbb{V}\{x_i'\}$ and $\sum_i P\{x_i \neq x_i'\}$ are convergent, where $x_i' = x_i$ if $-1 \leq x_i \leq 1$ and $x_i' = 0$ otherwise. There are several variants of these theorems, e.g., the strong law of the large number, namely, if $\mathbb{V}\{x_i\}$ is bounded then $\big[s_n - \mathbb{E}\{s_n\}\big]/n$ converges to 0 almost surely, or if x_i are integrable identically distributed then s_n/n converges to $\mathbb{E}\{x_1\}$ almost surely. Further in this analysis is the *central limit theorems* and law of the iterated logarithm, where we define the sequence of random variables $t_n = \big[s_n - \mathbb{E}\{s_n\}\big]/\sqrt{\mathbb{V}\{s_n\}}$ and give conditions under which the probability distributions of t_n converges to the Gauss or normal distribution $N(0,1)$. For instance, if the limit

$$\frac{1}{\mathbb{V}\{s_n\}} \sum_{i=1}^{n} \mathbb{E}\big\{|x_i - \mathbb{E}\{x_i\}|^2,\ |x_i - \mathbb{E}\{x_i\}| \geq \varepsilon\sqrt{\mathbb{V}\{s_n\}}\big\} \to 0$$

holds true for every $\varepsilon > 0$, then the probability distributions of t_n converges to the Gauss or normal $N(0,1)$, i.e.,

$$\lim_n P(a < t_n < b) = (2\pi)^{-1/2} \int_a^b e^{-x^2/2} \mathrm{d}x, \quad \forall b > a,$$

however we have

$$\limsup_n t_n = +\infty \quad \text{and} \quad \liminf_n t_n = -\infty$$

almost surely. This is used in the Gauss' theory of errors, namely, for every $n \geq 1$ let $\xi_{1,n}, \ldots, \xi_{n,k(n)}$ be independent random variables and define $\sigma_n = \sum_{i=1}^{k(n)} \xi_{i,k(n)}$. If $\varepsilon_n := \sup_{i,\omega} |\xi_{i,k}(\omega)| \to 0$, $\mathbb{E}\{\sigma_n\} \to m$ and $\mathbb{V}\{s_n\} \to v$ then the probability distribution of σ_n converges to the Gauss or normal distribution $N(m,v)$. On the other hand, if the variables $x_{i,k}$ take only two values, i.e., assuming $P\{\xi_{i,k} = 1\} = p_{i,k}$ and $P\{\xi_{i,k} = 0\} = 1 - p_{i,k}$, and if $\bar{p}_n := \max_k p_{i,k}$ and $\sum_{i=1}^{k(n)} p_{i,k(n)} \to \lambda$ then the probability distribution of σ_n converges to the Poisson distribution with parameter λ, this last result is know as *Poisson's law of rare events*. Proofs of the above theorems can be found in several text books in probability, e.g, Breiman [31, Chapter 9, pp. 185–190] or Itô [111, Chapter 4, pp. 165–211].

It should be clear that given a probability space $(\Omega, \mathcal{F}, P)$, it is not possible to ensure the existence of (independent) random variables (or stochastic processes) with a prescribed distribution. However, the typical (universal) probability space where realization are shown is the Lebesgue space on the interval $[0,1]$. A well known example is to write any ω in $\Omega = [0,1]$ in binary, i.e., $\omega = \sum_k 2^{-k}\omega_k$. Then the sequence of variables $\pi_n(\omega) = \omega_n$ for $n = 1, 2, \ldots$ are independent

coin-tossing variables each taking the values 0 or 1 with probability 1/2. Thus, given a mapping $i, j \mapsto k(i,j)$ which is injective from $\{1,2,\ldots\} \times \{1,2,\ldots\}$ into $\{1,2,\ldots\}$, the expression $X_i = \sum_j 2^{-k(i,j)}\omega_{k(i,j)}$ for $i = 1,2,\ldots$ defines an independent sequence of random variables, each with the same distribution as X, $X(\omega) = \omega$, i.e., each with the uniform distribution on $[0,1]$. In general, if S_i is a Borel space (i.e., a measurable space isomorphic to a Borel subset of $[0,1]$, for instance any complete separable metric space), P_i is a probability measure on the Borel σ-algebra $\mathcal{B}_i(S_i)$, for $i = 1,2,\ldots$ then there exists a sequence $\{\xi_1, \xi_2, \ldots\}$ of independent random variables defined on the universal Lebesgue probability space $[0,1]$ such that $P_i(B) = P(\{\omega : \xi_i(\omega) \in B\})$, for any B in $\mathcal{B}_i(S_i)$, $i = 1,2,\ldots$, i.e., the distribution of ξ_i is exactly P_i, e.g., see Kallenberg [121, Theorem 3.19, pp. 55–57].

Let ξ be a random (vector) variable having a given (joint) density distribution p_ξ. Sometimes we are interested in computing

$$\mathbb{E}\{g(\xi)\} = \int g(x)p_\xi(x)\mathrm{d}x,$$

for some real-valued function g. In many situation, it is not analytically possible either to compute the above (multiple) integral exactly or even to numerically approximate it within a given accuracy. Another way to approximating $\mathbb{E}\{g(\xi)\}$ is by means of the co-called *Monte Carlo simulation* method. This goes as follows: start by generating a random (vector) variable ξ^1 having the (joint) density g, and then compute $\eta^1 := g(\xi^1)$. Now generate a second random (vector) variable ξ^2, independent of the first, and compute $\eta^2 := g(\xi^2)$. Keep on doing this, for a fix number of times n, to generate the independent and identically distributed random (vector) variables $\eta^i := g(\xi^i)$, $i = 1,\ldots,n$. As mentioned later, the strong law of large number applies and we find

$$\lim_n \frac{\eta^1 + \cdots + \eta^n}{n} = \mathbb{E}\{\eta^i\} = \mathbb{E}\{g(\xi)\}.$$

Clearly, the remaining problem is how to generate, or *simulate* random (vector) variables having a specific (joint) distribution. The first step in doing this is to be able to generate random variables from a uniform distribution on $(0,1)$, i.e., a random variable u with valued in the interval $(0,1)$ such that $P\{u \le \lambda\} = \lambda$ for every λ in $[0,1]$. One way of doing this would be to take 10 identical slips of paper, numbered $0,1,\ldots,9$, place them in a hat and then successively select n slips, with replacement, from the hat. The sequence of digits obtained (with a decimal point in from) can be regarded as the value of a uniform $(0,1)$ random variable rounded off to the nearest 10^{-n}. This constitutes the so-called *random number tables*. Nowadays, digital computers simulate *pseudo random numbers* instead of the truly random numbers. Most of the random number generators start with an initial value ξ_0, called the seed, and then recursively compute values by specifying positive integers a and b and m, and then letting x_{n+1} be the remainder of $ax_n + b$ divided by m, i.e., $x_{n+1} := (ax_n + b) \mod (m)$. The quantity $u_n := x_n/m$ is taken as an approximation to a uniform $(0,1)$ random

variable. All other distributions are generated from uniform $(0,1)$ random variables u. Indeed, the inverse transformation method is based on the fact that for any right-continuous distribution F, the random variable $\xi := F^{-1}(u)$ has distribution F. Note the definition of $F^{-1}(\lambda) := \inf\{s \in (0,1) : F(s) = \lambda\}$, so that $t < F(\lambda)$ if and only if $F^{-1}(\lambda) < F(t)$.

The rejection method simulate a random variable η having density f on the basis of a random variable ξ having density g, it uses a two-step iteration as follows: for two distributions f and g satisfying $f(x) \le cg(x)$ for every x and some constant c :

Step 1: simulate ξ having density g and simulate a random number u.

Step 2: if $u \le f(\xi)/[cg(\xi)]$ set $\eta = \xi$, otherwise return to Step 1.

This generates a random variable ξ having density f.

For a continuously differentiable distribution F, the *hazard rate function* of F is defined by

$$\lambda(t) := \frac{F'(t)}{F(t)}, \quad \forall t \ge 0.$$

The hazard rate method simulates a random variable ξ having $\lambda(t)$ as its hazard rate function, provided λ is a given nonnegative bounded function satisfying

$$\int_0^\infty \lambda(t)\mathrm{d}t = \infty.$$

This is defined as follows:

Step 1: select $r \ge \lambda(t)$ for every $t \ge 0$ and simulate independent random variables u^i, η^i, $i = 1, \ldots,$ where u^i are $(0,1)$ uniform and η^i are exponential with rate r.

Step 2: stopping at $\tau := \min\left\{n : u_n r \le \lambda(\sum_{i\le n} \eta_i)\right\}$ define $\xi := \sum_{i\le\tau} \eta_i$.

It is proven that ξ has hazard rate function $\lambda(t)$. For instance, we refer to Ross [217], among others.

1.3 Filtrations and Optional Times

Let $(\Omega, \mathcal{F})$ be a measurable space and T be an partially order index set, with a first element called 0. Generally, the index set is the positive integers or a real interval, i.e. $[0,T]$, $0 < T \le +\infty$. Suppose we are given an increasing family of sub σ-algebras $\{\mathcal{F}(t) : t \in T\}$, i.e. $s \le t$ implies $\mathcal{F}(s) \subset \mathcal{F}(t)$. Sometimes the notation $\mathcal{F}_t = \mathcal{F}(t)$ is used. Define $\mathcal{F}^+(t) = \cap_{s>t}\mathcal{F}(s)$ for t in T to get another filtration with $\mathcal{F}(t) \subset \mathcal{F}^+(t)$ for any t in T. The filtration is said to be *right continuous* if $\mathcal{F}(\cdot) = \mathcal{F}^+(\cdot)$ and it is also denoted by $\mathcal{F}(t+)$. In particular, $\{\mathcal{F}^+(t) : t \in T\}$ is right continuous.

When a probability measure P is given on $(\Omega, \mathcal{F})$, the hypothesis that $\mathcal{F}(0)$ contains the null sets implies that the restriction of a given measure on each σ-algebra $\mathcal{F}(t)$ is *complete*, but the converse may be false. In most of the cases, we may assume that $\mathcal{F}(0)$ contains the null sets, at a cost of enlarging each

σ-algebra to the σ-algebra generated by $\mathcal{F}(t)$ and the null sets. If the index set is a real interval $[0,T]$, or $[0,\infty)$, then it is possible without loss of generality to replace $\mathcal{F}(\cdot)$ by the right continuous filtration $\mathcal{F}^+(\cdot)$. If the index set does not have a last element, i.e. $[0,+\infty)$, then we add a last element denoted by $+\infty$ (or ∞) with $\mathcal{F}(+\infty)$ equal to the sub σ-algebra generated by all the $\mathcal{F}(t)$, $t \geq 0$. Thus we will refer to a filtration satisfying the *usual conditions* or a standard filtration when the filtration is completed and right continuous.

Definition 1.7 (filtration)**.** In a (complete) probability space $(\Omega, \mathcal{F}, P)$, a family of sub σ-algebras $\{\mathcal{F}(t) : t \in T\}$ is called a *filtration* if $s \leq t$ implies $\mathcal{F}(s) \subset \mathcal{F}(t)$, $\mathcal{F}(0)$ contains all sets of probability zero, and $\mathcal{F}(t) = \cap_{s>t}\mathcal{F}(s)$ i.e., unless explicitly stated we assume the usual conditions are satisfied. A family of random variables $\{X(t) : t \in T\}$ is said *adapted* to the filtration if $X(t)$ is $\mathcal{F}(t)$-measurable for any t in T. □

Given a stochastic process and a filtration we can talk about a stochastic process being adapted to a filtration, being progressively measurable, and so on. Several operations can be performed with processes and filtrations. For a family $\{X_\gamma(\cdot) : \gamma \in \Gamma\}$ of processes adapted to a common filtration $\mathcal{F}(\cdot)$ we may define the process *essential infimum* and *essential supremum*. For instance

$$X(t) = \operatorname*{ess\,sup}_{\gamma\in\Gamma} X_\gamma(t),$$

which can be taken adapted to the same common filtration $\mathcal{F}(\cdot)$. Similarly, the sample integral can be defined for a progressively measurable (see definition later on) integrable process $\{X(t), \mathcal{F}(t) : t \geq 0\}$. The resulting process

$$Y(t) = \int_0^t X(s,\omega)\mathrm{d}s$$

can be taken progressively measurable with respect to the same filtration $\mathcal{F}(\cdot)$.

The problems of defining what is meant by a random time τ corresponding to the arrival time of an event whose arrival is determined by the preceding events and of defining the class $\mathcal{F}(\tau)$ of preceding events are solved by the following definition.

Definition 1.8. An *optional time* (*stopping* or *Markov time*) τ with respect to a filtration $\mathcal{F}(\cdot)$ is a function from Ω into $[0,+\infty]$ satisfying

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}(t) \quad \forall t \geq 0.$$

If an optional time τ is given, then $\mathcal{F}(\tau)$, respectively $\mathcal{F}(\tau-)$, is the σ-algebra of subsets A in $\mathcal{F}(+\infty)$ (or in $\mathcal{F}$) for which

$$A \cap \{\tau \leq t\} \in \mathcal{F}(t), \text{ respectively } A \cap \{\tau < t\} \in \mathcal{F}(t),$$

for every $t \geq 0$. □

Sometime, optional times are defined as nonnegative random variables satisfying $\{\omega : \tau(\omega) < t\} \in \mathcal{F}(t)$ for every $t > 0$, e.g., see Karatzas and Shreve [124, Section 1.2, pp. 6-11]. Since $\{\tau \le t\} = \cup_{n\ge1}\{\tau < t - 1/n\}$ and $\mathcal{F}(t - 1/n) \subset \mathcal{F}(t)$, we see that stopping time is stronger than optional time. Conversely, under the right-continuity condition, i.e., $\mathcal{F}(t) = \mathcal{F}(t+)$, for every $t \ge 0$, the equality $\{\tau \le t\} = \cap_{n\ge1}\{\tau < t + 1/n\}$ shows that any optional time is also a stopping time. Thus, unless specially mentioned, we do not differentiate between optional and stopping times.

Most of the time we use the σ-algebra $\mathcal{F}(\tau)$, however, when dealing with jump processes we may need $\mathcal{F}(\tau-)$. Note that we have $\cap_{\varepsilon>0}\mathcal{F}(\tau + \varepsilon) := \mathcal{F}(\tau+) = \mathcal{F}^+(\tau)$ for any optimal time τ. If τ_1 and τ_2 are two optional times with $\tau_1 \le \tau_2$, the *stochastic interval* $[\![\tau_1, \tau_2]\!]$, is defined by

$$[\![\tau_1, \tau_2]\!] = \{(t, \omega) \in \mathbb{R}^+ \times \Omega : \tau_1 \le t \le \tau_2\}.$$

Similarly, we define the open stochastic interval $]\!]\tau_1, \tau_2[\![$ and the half-open ones $[\![\tau_1, \tau_2[\![$, and $]\!]\tau_1, \tau_2]\!]$. Several properties are satisfied by optional times, we will list some of them (see Exercise 1.6).

(a) If τ is optional, then τ is $\mathcal{F}(\tau)$-measurable.

(b) If τ is optional and if τ_1 is a random variable for which $\tau_1 \ge \tau$ and τ_1 is $\mathcal{F}(\tau)$ measurable, then τ_1 is optional.

(c) If τ_1 and τ_2 are optional, then $\tau_1 \vee \tau_2$ (max) and $\tau_1 \wedge \tau_2$ (min) are optional.

(d) If τ_1 and τ_2 are optional and $\tau_1 \le \tau_2$, then $\mathcal{F}(\tau_1) \subset \mathcal{F}(\tau_2)$; if $\tau_1 < \tau_2$, then $\mathcal{F}(\tau_1+) \subset \mathcal{F}(\tau_2)$.

(e) If τ_1 and τ_2 are optional, then $\mathcal{F}(\tau_1) \cap \mathcal{F}(\tau_2) = \mathcal{F}(\tau_1 \wedge \tau_2)$. In particular, $\{\tau_1 \le t\} \in \mathcal{F}(\tau_1 \wedge t)$.

(f) If τ_1 and τ_2 are optional, then the sets $\{\tau_1 < \tau_2\}$, $\{\tau_1 \le \tau_2\}$ and $\{\tau_1 = \tau_2\}$ are in $\mathcal{F}(\tau_1 \wedge \tau_2)$.

(g) If τ_1 and τ_2 are optional and if $A \in \mathcal{F}(\tau_1)$, then $A \cap \{\tau_1 \le \tau_2\} \in \mathcal{F}(\tau_1 \wedge \tau_2)$.

(h) Let τ_1 be optional and finite valued, and let τ_2 be random variable with values in $[0, +\infty]$. The optionality of $\tau_1 + \tau_2$ implies optionality of τ_2 relative to $\mathcal{F}(\tau_1 + \cdot)$. Moreover, the converse is true if $\mathcal{F}(\cdot)$ is right continuous i.e., if τ_2 is optional for $\mathcal{F}_{\tau_1}(\cdot) := \mathcal{F}(\tau_1 + \cdot)$, then $\tau_1 + \tau_2$ is optional for $\mathcal{F}(\cdot)$ and $\mathcal{F}(\tau_1 + \tau_2) = \mathcal{F}_{\tau_1}(\tau_2)$.

(i) Let $\{\tau_n : n = 1, 2, \dots\}$ be a sequence of optional times. Then $\sup_n \tau_n$ is optional, and $\inf \tau_n$, $\liminf_n \tau_n$, $\limsup_n \tau_n$ are optional for $\mathcal{F}^+(\cdot)$. If $\lim_n \tau_n = \tau = \inf_n \tau_n$, then $\mathcal{F}^+(\tau) = \cap_n \mathcal{F}^+(\tau_n)$. If the sequence is decreasing [resp., increasing] and $\tau_n(\omega) = \tau(\omega)$ for $n \ge n(\omega)$, then τ is optional and $\mathcal{F}(\tau) = \cap_n \mathcal{F}(\tau_n)$ [resp., $\mathcal{F}(\tau)$ is equal to the smaller σ-algebra containing $\cup_n \mathcal{F}(\tau_n)$].

There are many relations between optional times, progressively measurable stochastic processes and filtration, we only mention the following result (see Doob [60, pp. 419–423])

Theorem 1.9 (exit times). *Let B be a Borel subset of $[0,T]\times\mathbb{R}^d$ and $\{X(t) : t\in[0,T]\}$ be a d-dimensional progressively measurable stochastic process with respect to a filtration $\mathcal{F}(\cdot)$ satisfying the usual conditions on a probability space $(\Omega,\mathcal{F})$, Then the hitting, entry and exit times are optional times with respect to $\mathcal{F}(\cdot)$, i.e., for the hitting time*

$$\tau(\omega)=\inf\{t>0:(t,X(t,\omega))\in B\},$$

where we take $\tau(\omega)=+\infty$ if the set in question is empty. Similarly, the entry time is define with $t>0$ replaced by $t\geq 0$ and the exit time is the entry time of complement of B, with the convention of being equal to T if the set in question is empty. □

Note that the *last hitting time* of a Borel set B, which is defined by

$$\hat{\tau}(\omega)=\sup\{t>0:(t,X(t,\omega))\in B\},$$

is not in general an optional time. However, if τ_c denotes the hitting time of B by the process $(t+c,X(t+c,\omega))$ then $\{\hat{\tau}>c\}=\{\tau_c<+\infty\}$ so that measurability properties for the last hitting time can be considered.

1.4 Conditional Expectation

The concept of *independence* is fundamental for probability theory and in fact distinguishes it from classical measure theory. In a sense we may say that *conditioning* is basic and fundamental to probability. Conditional measures is related to disintegration of probability measure, and it is a key concept to study Markov processes, we refer to Rao [207] for a comprehensive discussion on this subject.

If x is a real integrable random variable on a probability space $(\Omega,\mathcal{F},P)$ and if $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$, then the *conditional expectation* $\mathbb{E}\{x\,|\,\mathcal{G}\}$ is (uniquely determined up to null sets) a $\mathcal{G}$-measurable random variable satisfying

$$\int_A \mathbb{E}\{x\,|\,\mathcal{G}\}(\omega)P(\mathrm{d}\omega)=\int_A x(\omega)P(\mathrm{d}\omega),\quad \forall A\in\mathcal{G}.$$

We can also consider random variables that are only one sided integrable. The notation $\mathbb{E}\{x\,|\,\mathcal{G}\}$ always refers to any of the possible choices of the conditional expectation rather than to the equivalence class. The Radon-Nikodym theorem (see Exercise 1.10) ensures the existence and uniqueness up to null sets of conditional expectations, i.e., given x and $\mathcal{G}$ there exists a null set N (which may depends on both x and $\mathcal{G}$) such that $\omega\to\mathbb{E}\{x\,|\,\mathcal{G}\}(\omega)$ is uniquely defined for ω in $\Omega\setminus N$. If x,y and z are d-dimensional random variables then the relation $x=\mathbb{E}\{y\mid z\}$ means that except a set of probability zero, the random variable x is equal to the conditional expectation of y with respect to the σ-algebra $\mathcal{F}_z$ generated by z. This can be characterized by the condition

$$\mathbb{E}\{y\varphi(z)\}=\mathbb{E}\{x\varphi(z)\},\quad \forall\varphi \text{ bounded continuous.}$$

Actually, we may take only φ continuous with compact support, see Exercise 1.11. Some properties follows easily from the definition (see Exercise 1.12, 1.14).

(a) $\mathbb{E}\{y|\mathcal{G}\} = y$ a.s. if y is $\mathcal{G}$-measurable, in particular if y is a constant function.

(b) If y is bounded and $\mathcal{G}$-measurable, then $\mathbb{E}\{xy \mid \mathcal{G}\} = y\mathbb{E}\{x \mid \mathcal{G}\}$ a.s.

(c) $\mathbb{E}\{x + y \mid \mathcal{G}\} = \mathbb{E}\{x \mid \mathcal{G}\} + \mathbb{E}\{y \mid \mathcal{G}\}$ a.s.

(d) $x \leq y$ a.s. implies $\mathbb{E}\{x \mid \mathcal{G}\} \leq \mathbb{E}\{y \mid \mathcal{G}\}$ a.s.

(e) If $A \in \mathcal{G}$ and if $x = y$ a.s. on A, then $\mathbb{E}\{x \mid \mathcal{G}\} = \mathbb{E}\{y \mid \mathcal{G}\}$ a.s. on A.

(f) If $A \in \mathcal{G}_1 \cap \mathcal{G}_2$ and $A \cap \mathcal{G}_1 = A \cap \mathcal{G}_2$ (i.e., if any subset of A is in $\mathcal{G}_1$ if and only if the subset is in $\mathcal{G}_2$), then $\mathbb{E}\{x \mid \mathcal{G}_1\} = \mathbb{E}\{x \mid \mathcal{G}_2\}$ a.s. on A.

(g) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then

$$\mathbb{E}\{\mathbb{E}\{x \mid \mathcal{G}_1\} \mid \mathcal{G}_2\} = \mathbb{E}\{\mathbb{E}\{x \mid \mathcal{G}_2\} \mid \mathcal{G}_1\} = \mathbb{E}\{x \mid \mathcal{G}_1\} \quad \text{a.s.} \tag{1.4}$$

The next properties are less immediate (see Exercise 1.13)

(h) If x is independent of $\mathcal{G}$, then $\mathbb{E}\{x \mid \mathcal{G}\} = \mathbb{E}\{x\}$ a.s., in particular, if z is a $\mathcal{G}$-measurable random variable which is independent of $\mathcal{G}$ then $z = \mathbb{E}\{z\}$ a.s.

(i) If f is a convex function and if x and $f(x)$ are integrable, then Jensen's inequality holds i.e.,

$$f\big(\mathbb{E}\{x \mid \mathcal{G}\}\big) \leq \mathbb{E}\{f(x) \mid \mathcal{G}\} \quad \text{a.s..} \tag{1.5}$$

In particular we may take $f(\cdot) = |\cdot|^p$ to deduce that $\mathbb{E}\{x \mid \mathcal{G}\} \in L^p$ if $x \in L^p$. Certainly, if $0 < q \leq p$ then $(\mathbb{E}\{|x|^q \mid \mathcal{G}\})^{1/q} \leq (\mathbb{E}\{|x|^p \mid \mathcal{G}\})^{1/p}$, which is some referred to as Lyapunov's inequality.

(j) If $x_{\mathcal{G}} = \mathbb{E}\{x \mid \mathcal{G}\}$, the family of all conditional expectations $\{x_{\mathcal{G}} : \mathcal{G} \text{ sub } \sigma\text{-algebra of } \mathcal{F}\}$ is uniformly integrable.

(k) If $x \in L^p$ and $y \in L^q$, with $1/p + 1/q = 1$, then Hölder's inequality holds i.e., xy is integrable and

$$|\mathbb{E}\{xy \mid \mathcal{G}\}| \leq \big(\mathbb{E}\{|x|^p \mid \mathcal{G}\}\big)^{1/p}\big(\mathbb{E}\{|y|^q \mid \mathcal{G}\}\big)^{1/q} \quad \text{a.s.} \tag{1.6}$$

(l) If $x, y \in L^p$, then Minkowski inequality holds i.e., $x + y \in L^p$ and

$$\big(\mathbb{E}\{|x + y|^p \mid \mathcal{G}\}\big)^{1/p} \leq \big(\mathbb{E}\{|x|^p \mid \mathcal{G}\}\big)^{1/p} + \big(\mathbb{E}\{|y|^p \mid \mathcal{G}\}\big)^{1/p} \quad \text{a.s.} \tag{1.7}$$

(m) If $x \in L^2$ then $\mathbb{E}\{x \mid \mathcal{G}\} \in L^2$, and $x - \mathbb{E}\{x \mid \mathcal{G}\}$ is orthogonal to every $\mathcal{G}$-measurable random variable in L^2. Moreover, if z is a $\mathcal{G}$-measurable random variable in L^2 such that $x - z$ is orthogonal to any $\mathcal{G}$-measurable random variable in L^2 then $z = \mathbb{E}\{x \mid \mathcal{G}\}$.

Recall that when working with the spaces $L^p(\Omega, \mathcal{F}, P)$ the elements are actually classes of equivalence, and that we are implicitly assuming that the σ-algebra $\mathcal{F}$ contains all sets of null probability (measure), i.e., $(\Omega, \mathcal{F}, P)$ is a complete probability space. However, the separability of the space $L^p(\Omega, \mathcal{F}, P)$,

$1 \leq p < \infty$, is always an issue. Moreover, the concept of separability of a complete σ-algebra has to be reviewed. Thus, we say that the σ-algebra $\mathcal{F}$ is *(P-)separable* or that the probability space $(\Omega, \mathcal{F}, P)$ is *separable* if the space $L^1(\Omega, \mathcal{F}, P)$ is separable, i.e., if there exists a sequence of finite-dimensional subspaces $\{L^1_n : n \geq 1\}$ such that the closure of its union $\bigcup_n L^1_n$ is the whole space $L^1(\Omega, \mathcal{F}, P)$.

• *Remark* 1.10. It is proved (e.g., see Malliavin[162, Section IV.6.0, pp. 219-220]) that a probability space $(\Omega, \mathcal{F}, P)$ is separable if and only if there exists an increasing sequence of finite σ-algebras $\mathcal{F}(1) \subset \cdots \subset \mathcal{F}(n) \subset \cdots$ such that $\mathbb{E}\{x \mid \mathcal{F}(n)\} \to x$ for evert x in $L^1(\Omega, \mathcal{F}, P)$. In this case, we say that $\mathcal{F}$ is P-generated by the sequence $\{\mathcal{F}(n) : n \geq 1\}$. □

Some continuity properties for extended real valued random variable are stated as follows (see Doob [60, pp. 393–397]). If $\{\mathcal{F}(n) : n = 0, 1, \dots\}$ is a monotone sequence of sub σ-algebras of $\mathcal{F}$, then we denote by $\mathcal{F}(\infty)$ the σ-algebras either $\mathcal{F}(+\infty)$ or $\mathcal{F}(-\infty)$ according as $\mathcal{F}(\cdot)$ is monotone increasing or decreasing, where $\mathcal{F}(+\infty)$ is defined as the minimal σ-algebras containing all $\mathcal{F}(n)$, and $\mathcal{F}(-\infty)$ as the intersection of all $\mathcal{F}(n)$.

Theorem 1.11 (Fatou)**.** *Let $\{\mathcal{F}(n) : n = 0, 1, \dots\}$ be a monotone sequence of sub σ-algebras of $\mathcal{F}$ and let $\{x(n) : n = 0, 1, \dots\}$ be a sequence of positive extended real valued random variables. Then*

$$\liminf_{n\to\infty} \mathbb{E}\{x(n) \mid \mathcal{F}(n)\} \leq \mathbb{E}\{\liminf_{n\to\infty} x(n) \mid \mathcal{F}(\infty)\}$$

almost surely. □

Theorem 1.12 (Lebesgue)**.** *Let $\{\mathcal{F}(n) : n = 0, 1, \dots\}$ be a monotone sequence of sub σ-algebras of $\mathcal{F}$ and let $\{x(n) : n = 0, 1, \dots\}$ be a sequence of positive extended real valued random variables, with almost sure limit $x(\infty)$ and $\mathbb{E}\{\sup_n |x(n)|\} < \infty$. Then*

$$\lim_{n\to\infty} \mathbb{E}\{x(n) \mid \mathcal{F}(n)\} = \mathbb{E}\{x(\infty) \mid \mathcal{F}(\infty)\}$$

almost surely. □

In the above dominated convergence theorem we can replace the integrability of the function $\sup_n |x(n, \omega)|$ by the assumption that the sequence $\{x(n) : n = 0, 1, \dots\}$ is uniformly integrable. For instance, a typical application of this result is as follows. Let x be a real random variable independent of a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and f be a bounded Borel measurable function in $\mathbb{R}^2$. If we define $f_1(\eta) = \mathbb{E}\{f(x, \eta)\}$, then f_1 is Borel measurable and $f_1(y) = \mathbb{E}\{f(x, y) \mid \mathcal{G}\}$ a.s. (see Exercise 1.16).

Sometimes, if x is a integrable random variable and $\{x(i) : i \in I\}$ is a family of random variables, then we denote by $\mathbb{E}\{x \mid x(\cdot)\}$ or $\mathbb{E}\{x \mid x(i),\ i \in I\}$ the conditional expectation with respect to the σ-algebra $\mathcal{G} = \sigma\{x(i) : i \in I\}$ (or simply $\sigma\{x(\cdot)\}$), generated by the family $\{x(i) : i \in I\}$.

It is rather relevant to observe that the *conditional expectation* is an operator defined and valued on classes of equivalence of random variables, i.e., an operator

on Lebesgue spaces, from $L^p(\Omega, \mathcal{F}, P)$ into $L^p(\Omega, \mathcal{G}, P)$, for any $1 \leq p \leq \infty$. It can be extended to *functions* such that the positive (or negative) part belongs to the above Lebesgue spaces.

Another key property of the conditional expectation is the fact that for any random variable x in $L^1(\Omega, \mathcal{F}, P)$, the family of random variables $\{y = \mathbb{E}\{x|\mathcal{G}\} : \mathcal{G} \text{ is a sub } \sigma\text{-algebra of } \mathcal{F}\}$ is uniformly integrable (see Exercise 1.15).

Now, we can discuss the concept of conditional independence (for two events or σ-algebras or random variables) given another σ-algebra or random variable).

Definition 1.13 (conditional independence)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{C}$ be sub σ-algebras of $\mathcal{F}$. We say that two measurable sets A and B are (conditional) independent given $\mathcal{C}$ if

$$\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, \mathcal{C}\} = \mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{C}\}\, \mathbb{E}\{\mathbb{1}_B \,|\, \mathcal{C}\}, \quad \text{a.s.} \tag{1.8}$$

holds. Moreover, two sub σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are (conditional) independent given $\mathcal{C}$ (relative to the probability P) if (1.8) is satisfied for any sets $A \in \mathcal{H}$, $B \in \mathcal{G}$. Particularly, if the sub σ-algebras are generated by a family of random variables, i.e., $\mathcal{H} = \sigma(x(t) : t \in T)$, $\mathcal{G} = \sigma(y(s) : s \in S)$ and $\mathcal{C} = \sigma(z(r) : r \in R)$, then (1.8) is equivalent to

$$\begin{aligned}\mathbb{E}\Big\{ \prod_i h_i(x(t_i)) \prod_j g_j(y(s_j)) \prod_k c_k(z(r_k))\Big\} = \\ = \mathbb{E}\Big\{\mathbb{E}\{\prod_i h_i(x(t_i)) \,|\, \mathcal{C}\}\, \mathbb{E}\{\prod_j g_j(y(s_j)) \,|\, \mathcal{C}\} \prod_k c_k(z(r_k))\Big\},\end{aligned}$$

where all products are extended to any finite family of subindexes and any real-valued bounded measurable functions h_i, g_j and c_k. □

Certainly this concept extends to a family of measurable sets, a family of either sub σ-algebras or random variables, where mutually or pairwise (conditional independent given $\mathcal{C}$) are not the same.

Recall that $\mathbb{E}\{\prod_i h_i(x(t_i)) \,|\, \mathcal{C}\}$ and $\mathbb{E}\{\prod_j g_j(y(s_j)) \,|\, \mathcal{C}\}$ are defined (almost surely) as $\mathcal{C}$ measurable integrable (also, bounded because h_i and g_j are so) functions satisfying

$$\mathbb{E}\Big\{ \prod_k c_k(z(r_k)) \prod_i h_i(x(t_i))\Big\} = \mathbb{E}\Big\{ \prod_k c_k(z(r_k))\, \mathbb{E}\{\prod_i h_i(x(t_i)) \,|\, \mathcal{C}\}\Big\},$$

and

$$\mathbb{E}\Big\{ \prod_k c_k(z(r_k)) \prod_j g_j(y(s_j))\Big\} = \mathbb{E}\Big\{ \prod_k c_k(z(r_k))\, \mathbb{E}\{\prod_j g_j(y(s_j)) \,|\, \mathcal{C}\}\Big\}.$$

for any functions h_i, g_i and c_k as above.

The definition of conditional independence applies to two random variables and a σ-algebra, i.e., a random variable x is (conditional) independent of another random variable y given a sub σ-algebra $\mathcal{C}$ in the probability space $(\Omega, \mathcal{F}, P)$ if for any bounded and measurable functions f and g we have $\mathbb{E}\{f(x)\, g(y) \,|\, \mathcal{C}\} = \mathbb{E}\{f(x) \,|\, \mathcal{C}\}\, \mathbb{E}\{g(y) \,|\, \mathcal{C}\}$ almost surely.

It is also clear that if $\mathcal{C} = \{\Omega, C, \Omega \smallsetminus C, \emptyset\}$, then the random variable $\mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{C}\}$ is everywhere defined as

$$\mathbb{E}\{\mathbb{1}_A \,|\, \mathcal{C}\} = \begin{cases} P(A \cap C)/P(C) & \text{if } \omega \in C, \\ P\big(A \cap (\Omega \smallsetminus C)\big)/P(\Omega \smallsetminus C) & \text{if } \omega \in \Omega \smallsetminus C, \end{cases}$$

when $0 < P(C) < 1$ and otherwise, only almost surely defined (using only one part of the above formula).

This means that the concept of two measurable sets A and B being conditional independent given (another measurable set) C (relative to the probability P) is properly defined by means of $\mathbb{1}_A$ and $\mathbb{1}_B$ as random variables and $\mathcal{C} = \{\Omega, C, \Omega \smallsetminus C, \emptyset\}$, the σ-algebra generated by C (or equivalently $\mathbb{1}_C$). Indeed, use the equality $P(A \cap B \cap C)\, P(C) = P(A \cap C)\, P(B \cap C)$ and add the complement equality $P(A \cap B \smallsetminus C)\, (1 - P(C)) = P(A \smallsetminus C)\, P(B \smallsetminus C)$, to deduce the validity of condition (1.8). In particular, when $C = \Omega$ the conditional independence coincides with the independence concept of Definition 1.3 and each of the previous equalities is trivially satisfied.

Nevertheless, in analogy with (elementary) conditional probability, where $P(A \,|\, C) = P(A \cap C)/P(C)$, we define the *conditional expectation* of a random variable x relative to a set C (with positive probability), instead of a σ-algebra $\mathcal{C}$, by means of

$$\mathbb{E}\{x \,|\, C\} = \frac{\mathbb{E}\{x\, \mathbb{1}_C\}}{P(C)},$$

i.e., expectation with respect to the conditional probability $P(\cdot \,|\, C)$. With this notation, two measurable sets A and B are conditional independent given another measurable set C if $\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, C\} = \mathbb{E}\{\mathbb{1}_A \,|\, C\}\, \mathbb{E}\{\mathbb{1}_B \,|\, C\}$, or equivalently, either $P(A \cap B \,|\, C) = P(A \,|\, C)\, P(B \,|\, C)$ or $P(A \cap B \cap C)\, P(C) = P(A \cap C)\, P(B \cap C)$. Similarly, two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ (or two random variables x and y, where $\mathcal{A}$ and $\mathcal{B}$ are the generated by x and y) are conditional independent given a set C if the previous condition holds for any A in $\mathcal{A}$ and B in $\mathcal{B}$. However, we cannot use the condition $\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, C\} = \mathbb{E}\{\mathbb{1}_A \,|\, C\}\, \mathbb{E}\{\mathbb{1}_B \,|\, C\}$, for any C in $\mathcal{C}$, as definition of conditional independent given a σ-algebra $\mathcal{C}$, since this would include $C = \Omega$ and then A and B would be independent, not just conditional independent, see Remark 1.6. Thus, we need to recall that conditioning with respect to a measurable set (event) C yields a number, an evaluation operator. While, conditioning with respect to a σ-algebra (or a random variable) is an operator (with values into the sets of random variables) defined almost surely.

Exercise 1.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{C}$ be sub σ-algebras of $\mathcal{F}$. Prove that two sub σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are (conditional) independent given $\mathcal{C}$ (relative to the probability P) if and only if $\mathbb{E}\{\mathbb{1}_G \,|\, \mathcal{H} \vee \mathcal{C}\}$ is $\mathcal{C}$-measurable, for every G in $\mathcal{G}$, where $\mathcal{H} \vee \mathcal{C}$ is the σ-algebra generated by $\mathcal{H}$ and $\mathcal{C}$. See Çıinlar [46, Proposition 3.2, Section 4.3]. □

As mentioned early, conditional expectation can be derived from the orthogonal projection, i.e., if $\mathcal{G}$ is a sub σ-algebra in $(\Omega, \mathcal{F}, P)$, x is an element in the

Lebesgue space $L^2(\Omega, \mathcal{F}, P)$, and $L^2(\mathcal{G})$ denotes the subspace of $L^2(\Omega, \mathcal{F}, P)$ composed by all $\mathcal{G}$-measurable functions (actually, equivalent classes) then $y = \mathbb{E}\{x \,|\, \mathcal{G}\}$ if and only if $x = y + z$ where y belongs to $L^2(\mathcal{G})$ and z is orthogonal to $L^2(\mathcal{G})$, namely

$$y \in L^2(\mathcal{G}) \qquad \text{and} \qquad \mathbb{E}\{(y - x)g\} = 0, \quad \forall g \in L^2(\mathcal{G}).$$

Clearly, if x belongs only to $L^1(\mathcal{G})$ then the above condition becomes

$$y \in L^1(\mathcal{G}) \qquad \text{and} \qquad \mathbb{E}\{(y - x)g\} = 0, \quad \forall g \in L^\infty(\mathcal{G}),$$

where $L^\infty(\mathcal{G})$ is the space of bounded $\mathcal{G}$-measurable functions (actually, equivalent classes). Note that a simple argument of monotone class shows that if $\mathcal{G}$ and $\mathcal{H}$ are two sub σ-algebras and x is an element in $L^1(\Omega, \mathcal{F}, P)$, then $y = \mathbb{E}\{x \,|\, \mathcal{G} \vee \mathcal{H}\}$ if and only if

$$y \in L^1(\mathcal{G} \vee \mathcal{H}) \qquad \text{and} \qquad \mathbb{E}\{(y - x)\mathbb{1}_G \mathbb{1}_H\} = 0, \quad \forall G \in \mathcal{G},\ H \in \mathcal{H},$$

where $\mathcal{G} \vee \mathcal{H}$ denotes the σ-algebra generated by $\mathcal{G}$ and $\mathcal{H}$.

Sometimes, we need to extend notion of conditional expectation to random variables x which are only σ-integrable with respect to a given sub σ-algebra $\mathcal{G}$, i.e., x is a measurable functions on Ω such x is integrable on G_n, for any n, where $\{G_n\}$ is some increasing sequence of $\mathcal{G}$-measurable. In this case, and assuming $x \geq 0$, we define $\xi = \mathbb{E}\{x \,|\, \mathcal{G}\}$ as the monotone limit of $\mathbb{E}\{x\mathbb{1}_{G_n} \,|\, \mathcal{G}\}$. Certainly, the random variable ξ is the unique $\mathcal{G}$-measurable function satisfying $\mathbb{E}\{x\mathbb{1}_G\} = \mathbb{E}\{\xi\mathbb{1}_G\}$, for any G in $\mathcal{G}$ with $\mathbb{E}\{|x|\mathbb{1}_G\} < \infty$, e.g., see He et al. [105, Section I.4, pp. 10–13].

1.5 Regular Conditional Probability

As before, let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$ and consider the conditional expectation of $\mathbb{E}\{f \mid \mathcal{G}\}$ for the special case where the random variable f is the indicator function $\mathbb{1}_A(\cdot)$ of a set A in $\mathcal{F}$. We will refer to the conditional expectations as the *conditional probability* and denote it by $P\{A \mid \mathcal{G}\}$. On the other hand, we may begin with the conditional probability, i.e., $A \mapsto \mathbb{E}\{\mathbb{1}_A \mid \mathcal{G}\} = P\{A \mid \mathcal{G}\}$ a random linear operator with values in $[0, 1]$ (defined almost surely) such that

$$P\{A \cap B\} = \mathbb{E}\big\{P\{A \mid \mathcal{G}\}\mathbb{1}_B\big\}, \quad \forall A \in \mathcal{F},\ B \in \mathcal{G},$$

then we define $\mathbb{E}\{f \mid \mathcal{G}\}$ for simple functions f and we pass to the limit for any integrable f, by using the fact that the operator $f \mapsto \mathbb{E}\{f \mid \mathcal{G}\}$ is a contraction in $L^1(\Omega, \mathcal{F}, P)$.

In any way, the conditional probability has some elementary properties inherited from the properties of the conditional expectation. For instance, if A and B are two disjoint sets in $\mathcal{F}$ then

$$P\{A \cup B \mid \mathcal{G}\} = P\{A \mid \mathcal{G}\} + P\{B \mid \mathcal{G}\} \quad \text{a.s.}$$

However, $P\{A \mid \mathcal{G}\}$ can be altered on a set of measure zero for each A in $\mathcal{F}$, we cannot conclude that $P\{A \mid \mathcal{G}\}$ (which is a random variable for each fixed A) is a countably (or finitely) additive probability measure on A in $\mathcal{F}$ for each ω outside of a null set. Technically, we have a function $P\{A \mid \mathcal{G}\}(\omega)$ of two variables A and ω, which is defined $A \mapsto P\{A \mid \mathcal{G}\}(\cdot)$ as a function of A taking values in a "class-of-equivalence" space in ω and now we want to consider this function as $\omega \mapsto P\{\cdot \mid \mathcal{G}\}(\omega)$ taking values in the space of probability measures, for each ω or even for almost every ω. For this to work, we need first to define the function $P\{A \mid \mathcal{G}\}(\omega)$ in a "dense" countable set of (A, ω) and then to extend its definition in a suitable way. A countably generated sub σ-algebra $\mathcal{G}$ is a suitable choice to handle the variable A, but some topology is required in the base space Ω to deal with ω. In short, this means that we look for a member from the above equivalence class of functions in such a way that *additivity* property (in particular order preserving and positivity) is preserved, see Taylor [248, pp. 210–226].

Theorem 1.14 (regular)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is a complete separable metric (Polish) space and $\mathcal{F} = \mathcal{B}(\Omega)$ is its Borel σ-algebra. Then, for any countably generated sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$ we can choose a* regular *conditional probability $P\{A \mid \mathcal{G}\}$ i.e.,* (a) *for each A in $\mathcal{F}$ the function $\omega \mapsto P\{A \mid \mathcal{G}\}(\omega)$ is $\mathcal{G}$-measurable,* (b) *for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$ we have*

$$P(A \cap B) = \int_B P\{A \mid \mathcal{G}\}(\omega) P(\mathrm{d}\omega),$$

and (c) *for each ω in Ω the function $A \mapsto P\{A \mid \mathcal{G}\}(\omega)$ is a probability measure on Ω and $P\{B \mid \mathcal{G}\}(\omega) = \mathbb{1}_B(\omega)$, for any ω in Ω and B in $\mathcal{G}$.* □

Note that if we can define $P\{A \mid \mathcal{G}\}(\omega)$ for ω in $\Omega \setminus N$, with $P(N) = 0$, satisfying (a), (b) and (c), then we can extend its definition to the whole Ω by simply setting it equal to P on the null set N. The hard point in the proof of the above theorem in precisely to construct such a null set (see Stroock and Varadhan [241]). Essentially we argue as follows: first for any A in $\mathcal{F}$ there exists a null set N_A such that $P\{A \mid \mathcal{G}\}(\omega)$ is uniquely determinate for ω outside of N_A, and if the σ-algebra $\mathcal{F}$ is generated by a countably π-class (i.e., closed under finite intersections) $\mathcal{F}_g$, then the same holds true for any A in $\mathcal{F}_g$ and ω outside of the null set $N = \cup\{N_A : A \in \mathcal{F}_g\}$. Finally, because $\mathcal{F}_g$ generates $\mathcal{F}$, the probability measure $A \mapsto P\{A \mid \mathcal{G}\}(\omega)$ is also uniquely determinate for any ω in $\Omega \setminus N$.

The first two conditions (a) and (b) refer to the *conditional probability distributions*, while the last condition (c) means *regular* property. A similar point of view can be found in Malliavin [162, Chapter IV, Theorem 6.5.1], where an analogous result is proved under the assumption that $L^1(\Omega, \mathcal{F}, P)$ is separable (as a Banach space). Usually, a probability space $(\Omega, \mathcal{F}, P)$ is called *separable* if the Banach space $L^1(\Omega, \mathcal{F}, P)$ is separable (i.e., there exists a countable dense set), which is equivalent to existence of an increasing sequence of finite sub σ-algebras $\{\mathcal{F}_n : n = 1, 2 \ldots\}$ of $\mathcal{F}$ such that $\mathbb{E}\{|\mathbb{E}\{f \mid \mathcal{F}_n\} - f|\} \to 0$ for every f in $L^1(\Omega, \mathcal{F}, P)$.

Given a filtration $\{\mathcal{G}(t) : t \geq 0\}$ (of countably generated σ-algebras) satisfying the usual conditions, with the same technique of the above theorem we can prove that a regular conditional probability $P\{A \,|\, \mathcal{G}(t)\}$ exists i.e., (a) for each A in $\mathcal{F}$ the function $\omega \mapsto P\{A \,|\, \mathcal{G}(t)\}(\omega)$ is $\mathcal{B} \times \mathcal{G}(t)$-measurable, (b) for every $t \geq 0$, $A \in \mathcal{F}$ and $B \in \mathcal{G}(t)$ we have

$$P(A \cap B) = \int_B P\{A \,|\, \mathcal{G}(t)\}(\omega) P(\mathrm{d}\omega),$$

and (c) for each $t \geq 0$, ω in Ω the function $A \mapsto P\{A \,|\, \mathcal{G}(t)\}(\omega)$ is a probability measure on Ω. Indeed, since $\bigcap_{s>t} \mathcal{G}(s) = \mathcal{G}(t)$ for any $t \geq 0$, it is enough to define $P\{A \,|\, \mathcal{G}(t)\}$ for t rational and use the monotone convergence.

Remark that if the Borel σ-algebra $\mathcal{B}(S)$ of a compact metrizable space S is countably generated, and because a Polish (Lusin) space is homeomorphic to a countable intersection of open sets (Borel set) of a compact metric space, the same is true for Polish or Lusin spaces S.

Let G be a set in $\mathcal{F}$ such that both G and $\Omega \setminus G$ have positive probability. In elementary probability, we define the conditional probability of a set A (in $\mathcal{F}$) *given* G by the formula $P(A \,|\, G) := P\{A \cap G\}/P\{G\}$. On the other hand, if $\mathbb{1}_A(\cdot)$ denotes the indicator (or characteristic) function of the set A, and $\sigma(G) = \{\Omega, \emptyset, G, \Omega \setminus G\}$ the σ-algebra generated by the set G, then $P\{A \,|\, \sigma(G)\} = \mathbb{E}\{\mathbb{1}_A \,|\, \sigma(G)\}$ and

$$P\{A \,|\, \sigma(G)\}(\omega) = \begin{cases} P\{A \cap G\}/P\{G\} & \text{if } \omega \in G, \\ P\{A \setminus G\}/P\{\Omega \setminus G\} & \text{if } \omega \in \Omega \setminus G, \end{cases}$$

so that both concepts are reconcilable. However, we should recall that the conditional probability given a set C is an evaluation, while given a σ-algebra is an operator (with values into the set of probability measures) defined almost surely. Simple considerations on the random variable $g(\omega) = P(A \,|\, G)\mathbb{1}_G(\omega) + P(A \,|\, \Omega \setminus G)\mathbb{1}_{\Omega \setminus G}(\omega)$ establishes that g is $\sigma[G]$-measurable and uniquely determined (almost surely) by the condition

$$\int_A g(\omega) P(\mathrm{d}\omega) = P(A \cap G), \quad \forall A \in \sigma[G].$$

It is remarkable to note that the above expression makes perfectly sense when G is negligible and gives the precise generalization quoted in the previous section. Moreover, this is better seen if the σ-algebra $\mathcal{G}$ is finitely-generated, i.e., $\mathcal{G} = \sigma\{G_1, \ldots, G_n\}$, where $P\{A \,|\, \mathcal{G})\}$ can be explicitly defined and the σ-additive condition is easily checked, see Exercise 1.17.

Given any event A, when $P(B) > 0$ and $B \in \mathcal{G}$, we have

$$P\{A \,|\, B\} = \frac{1}{P(B)} \int_B P\{A \,|\, \mathcal{G}\}(\omega) P(\mathrm{d}\omega).$$

Moreover we recall that two events A and B are said conditionally independent with respect to (or given) the *sub σ-algebra* $\mathcal{L}$ if

$$P(A \cap B \,|\, \mathcal{L}) = P(A \,|\, \mathcal{L})\, P(B \,|\, \mathcal{L}). \tag{1.9}$$

Analogously, $\mathcal{H}$ and $\mathcal{G}$ are called independent given (or with respect to) $\mathcal{L}$, a sub σ-algebra of $\mathcal{F}$, (relative to the probability P) if (1.9) is true for any sets $A \in \mathcal{H}$, $B \in \mathcal{G}$, see Definition 1.13.

It is interesting to note that given two random variables X and Y with a joint probability density function $f_{X,Y}(x,y)$, the functions

$$f_X(y) = \int f_{X,Y}(x,y)\mathrm{d}y, \qquad f_Y(y) = \int f_{X,Y}(x,y)\mathrm{d}x,$$

are the (marginal) probability density for X and Y, and the *elementary conditional* probability density function $f_{X|Y}$ of X given Y is defined by

$$f_{X|Y}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then for any Borel measurable function h such that

$$\mathbb{E}\{|h(X)|\} = \int |h(x)|\, f_X(x)\mathrm{d}x < \infty,$$

we can define the function

$$g(y) = \int h(x) f_{X|Y}(x,y)\mathrm{d}x$$

which provides a version of the conditional expectation of $h(X)$ given $\sigma(Y)$, i.e., $g(Y) = \mathbb{E}\{h(X) \mid Y\}$. Moreover, the function

$$\omega \mapsto \int_A f_{X|Y}(x, Y(\omega))\mathrm{d}x$$

is a regular version of the conditional probability of X given Y or given $\sigma(Y)$, usually denoted by $P\{X \in A \mid Y\}$.

• *Remark* 1.15. For instance, if $\{T_i : i = 1, \dots, n\}$ is a finite sequence of independent random variables, each with exponential distribution with mean $\mathbb{E}\{T_i\} = 1/\alpha_i$, then the above calculation can be used to show that $P\{S = T_j\} = \alpha_j/\alpha$, where $S = \min_i\{T_i\}$ and $\alpha = \sum_i \alpha_i$. Indeed, the assumptions on the random variables T_i yield $P\{T_i > T_j,\, \forall i \neq j \mid T_j\} = \mathrm{e}^{-(\alpha-\alpha_i)T_j}$. Hence,

$$P\{S = T_j\} = \mathbb{E}\big\{\mathbb{E}\{\mathbb{1}_{\{S=T_j\}} \mid T_j\}\big\} = \mathbb{E}\big\{\mathrm{e}^{-(\alpha-\alpha_i)T_j}\big\} = \frac{\alpha_j}{\alpha},$$

as desired. Since the event $\{T_j = t\}$ has zero probability, the natural notation $P\{T_i > T_j,\, \forall i \neq j \mid T_j = t\} = \mathrm{e}^{-(\alpha-\alpha_i)t}$ is not clearly justified. □

• *Remark* 1.16. It is clear that the concept of *conditional expectation* or *regular conditional probability* can be applied to σ-finite measures, instead of just probability measures. □

• *Remark* 1.17 (regular conditional distribution). Let $(\Omega, \mathcal{F}, P)$ be a probability space as in Theorem 1.14, and let X be a given a random variable with values in some Polish space $(E, \mathcal{E})$. Then the regular conditional probability (in this case, also called regular conditional distribution given X) exists for $\mathcal{G} = X^{-1}(\mathcal{F})$, the σ-algebra generated by X. Thus, for any A in $\mathcal{F}$, the function $\omega \mapsto P\{A \,|\, \mathcal{G}\} = P\{A \,|\, X\}$ is a real-valued measurable with respect to $X^{-1}(\mathcal{F})$ and therefore (see Exercises 1.4) there exists a (real valued) Borel measurable function $x \mapsto P(x, A)$ on $(E, \mathcal{E})$ (which depends on X) such that $P\{A \,|\, X\}(\omega) = P(X(\omega), A)$ almost surely. This is called the transition probability function $P(x, A)$ given X, and usually denoted by $P\{A \,|\, X = x\}$. Note the two defining properties: (a) for each A in $\mathcal{F}$, the function $x \mapsto P(x, A)$ is measurable from $(E, \mathcal{E})$ into $[0, 1]$, and (b) for any x in E, the function $A \mapsto P(x, A)$ is a probability measure on $(\Omega, \mathcal{F})$. Clearly, if P_X denotes the probability distribution of X then for any A in $\mathcal{F}$ the function $x \mapsto P\{A \,|\, X = x\}$ is uniquely determinate outside of a P_X-negligible set. Moreover, condition (b) in Theorem 1.14 can be rewritten as

$$P\big\{A \cap \{X \in C\}\big\} = \int_{\{X \in C\}} P\{A \,|\, X\} \, \mathrm{d}P, \quad \forall A \in \mathcal{F}, \, C \in \mathcal{E},$$

and in particular, for $C = \{x\}$, we have

$$P\{A \,|\, X = x\} = \frac{P\big\{A \cap \{X = x\}\big\}}{P\{X = x\}},$$

for any x in E with $P\{X = x\} > 0$ and any A in $\mathcal{F}$. The reader may check the book Taira [245, Chapter 1] for a complete discussion on this point. □

• *Remark* 1.18. Related to the previous discussion on regular conditional distribution, of particular interest is the case of two E-valued random variables X and Y, for which we consider the joint distribution P_{XY} on the product Polish space E^2. Independently of the initial probability space, the image $(E^2, \mathcal{E}^2, P_{XY})$ is a probability space satisfying the assumptions of Theorem 1.14. Thus, we may consider the conditional probability distribution given the sub σ-algebra $\mathcal{E}_x$ generated by the projection $(x, y) \mapsto x$ of for the first coordinate, i.e., $P_{XY}\{A \,|\, \mathcal{E}_x\} = P\{(X, Y) \in A \,|\, X\}$, with A in $\mathcal{E}^2$. Thus, we obtain the conditional probability distribution of Y given X, namely $P_{Y|X}(B) = P\{Y \in B \,|\, X\}$, after choosing $A = E \times B$. Hence, there exists a transition function $P(x, B)$ of Y given X, which depends on X and Y and is denoted by $P\{Y \in B \,|\, X = x\}$, i.e., $P\{Y \in B \,|\, X\}(\omega) = P(X(\omega), B)$, for any B in $\mathcal{E}$, see Exercise 1.17. □

It is now clear that an essential point is the use of the following two properties: (1) the σ-algebra $\mathcal{E}$ is generated by a countable algebra and (2) any $(E, \mathcal{E})$-valued random variable x on any probability space $(\Omega, \mathcal{F}, P)$ admits a regular conditional distribution relative to a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$. This *disintegration property* (2) can be re-stated as: for any positive and finite measure m on the product space $(E \times B, \mathcal{E} \times \mathcal{B})$ there exist a measurable kernel $k(\mathrm{d}x, b)$ such that $m(\mathrm{d}x, \mathrm{d}b) = k(\mathrm{d}x, b) \, m_B(\mathrm{d}b)$, where $m_B(\mathrm{d}b) := m(E, \mathrm{d}b)$ is the B-marginal distribution of m. Any Polish space possess these properties, and in general, based on these properties we define the so-called *Blackwell spaces*, see

Dellacherie and Meyer [58]. Let us mention that only the case of locally compact Polish spaces will be mainly used here. Clearly, in most of our considerations, it will be implicitly assumed that we are working on probability spaces having these two properties.

1.6 Versions of Processes

To fully understand the previous sections in a more specific context, the reader should acquire some basic background on the very essential about probability, perhaps the beginning of books such as Jacod and Protter [116] or Williams [253], among many others, is a good example. This is not really necessary for what follows, but it is highly recommended.

On a probability space $(\Omega, \mathcal{F}, P)$, sometimes we may denote by $X(t,\omega)$ a stochastic process $X_t(\omega)$. Usually, equivalent classes are not used for stochastic process, but the definition of *separability* and *continuity* of a stochastic process have a natural extension in the presence of a probability measure, as *almost sure* (a.s.) properties, i.e., if the conditions are satisfied only for $\omega \in \Omega \setminus N$, where N is a null set, $P(N) = 0$. This is extremely important since we are actually working with a particular element of the equivalence class. Moreover, the concept of version is used, which is not exactly the same as equivalence class, unless some extra property (on the path) is imposed, e.g., separability or continuity. Actually, the member of the equivalence class used is ignored, but a *good* version is always needed. We are going to work mainly with d-dimensional valued stochastic process with index sets equal to continuous times intervals e.g., a measurable and separable function $X : \Omega \times [0, +\infty] \to \mathbb{R}^d$.

It is then clear when two processes X and Y should be considered equivalent (or simply equal, $X = Y$), if

$$P(\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}) = 1.$$

This is often referred as X being *indistinguishable* from Y, or that $X = Y$ up to an *evanescent* set. So that any property valid for X is also valid for Y. When the index set is uncountable, this notion differs from the assertion X or Y is *a version* (or a modification) of the given process, where it is only required that

$$P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1, \quad \forall t \in T, \tag{1.10}$$

which implies that both processes X and Y have the same family of finite-dimensional distributions. For instance, *sample path* properties such as (progressive) measurability and continuity depend on the version of the process in question.

Furthermore, the integrand of a stochastic integral is thought as an equivalence class with respect to a product measure in $(0,\infty) \times \Omega$ of the form $\mu = \mathrm{d}\alpha(t,\omega)P(\mathrm{d}\omega)$, where $\alpha(t,\omega)$ is an integrable nondecreasing process. In this case, two processes may belong to the same μ-equivalence class without being a version of each other. Conversely, two processes, which are versions of each other, may not belong to the same μ-equivalence class. However, any two

indistinguishable processes must belong to the same μ-equivalence class. Moreover, a measure μ in the product space $(0,\infty)\times\Omega$ vanishes on every evanescent set if and only if it has the product form $\mu = \mathrm{d}\alpha(t,\omega)P(\mathrm{d}\omega)$ for some integrable nondecreasing process α. This is discussed in some detail later, in Chapter 4.

The finite-dimensional distributions are not sufficient to determine the sample paths of a process, and so, the idea of separability is to use a countable set of time to determine the properties of a process. The following result (see Doob [59, Theorem 2.4, pp. 60], Billingsley [27, Section 7.38, pp. 551-563] or Neveu [188, Proposition III.4.3, pp. 84-85]) is necessary to be able to assume that we are always working with a separable version of a process.

Theorem 1.19 (separability)**.** *Any d-dimensional stochastic process has a version which is separable i.e., if X is the given stochastic process indexed by some real interval T, then there exists a $\bar{\mathbb{R}}^d$-valued stochastic process Y satisfying (1.10) and the condition of separability in Definition 1.1, which may be rephrased as follows: there exist a countable dense subset I of T and a null measurable set N, $P(N)=0$, such that for every open subset O of T and any closed subset C of $\mathbb{R}^d$ the set $\{\omega\in\Omega : Y(t,\omega)\in C,\ \forall t\in O\setminus I\}$ is a subset of N.* □

By means of the above theorem, we will always assume that we have taken a (the qualifier a.s. is generally omitted) separable version of a (measurable) stochastic process provided we accept processes with values in $\bar{\mathbb{R}}^d = [-\infty,+\infty]^d$. Moreover, if we insist in calling stochastic process X a family of random variables $\{X_t\}$ indexed by t in T then we have to deal with the separability concept. Actually, the set $\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t\in T\}$ used to define equivalent or indistinguishable processes may not be measurable when X or Y is not a measurable process. Even working only with measurable processes does not solve completely our analysis, e.g., a simple operation as $\sup_{t\in T} X_t$ for a family of uniformly bounded random variables $\{X_t\}$ may not yields a measurable random variable. The separability notion solves all these problems.

Furthermore, this generalizes to processes with values in a separable locally compact metric space (see Gikhman and Skorokhod [98, Section IV.2]), in particular, the above separable version Y may be chosen with values in $\mathbb{R}^d\cup\{\infty\}$, the one-point compactification of $\mathbb{R}^d$, and with $P\{Y(t)=\infty\}=0$ for every t, but not necessarily $P\{Y(t)=\infty\ \forall t\in T\}=0$. Thus in most cases, when we refer to a stochastic process X in a given probability space $(\Omega,\mathcal{F},P)$, actually we are referring to a measurable and separable version Y of X. Note that in general, the initial process X is not necessarily separable or even measurable. By using the separable version of a process, we see that most of the *measurable* operations usually done with a function will make a proper sense. The construction of the separant set used (in the proof of the above theorem) may be quite complicate, e.g., see Neveu [188, Section III.4, pp. 81–88].

A process X which is continuous in probability i.e., for all $t\in T$ and $\varepsilon>0$ we have

$$\lim_{s\to t} P(\{\omega\in\Omega : |X(s,\omega)-X(t,\omega)|\geq\varepsilon\}) = 0.$$

is called *stochastically continuous*. Similarly, we define left or right stochastically continuous. Actually, if the interval T is compact, then the process is uniformly stochastically continuous. In most of the cases, a stochastic process X will be (right or left) continuous in probability (see below) and then (see Exercise 1.18) any dense set in T will be separant. Note that the concept of stochastic continuity (or continuity in probability) is not a sample path property, it does not depend on the particular version of the process involved. On the contrary, most of the smoothness properties such as separability, measurability or continuity are conditions on the sample paths and depend on the version of the process used to test the property.

It is known (e.g., see Da Prato and Zabczyk [51, p. 72–75], Gikhman and Skorokhod [98, Section IV.3]) that

Theorem 1.20 (measurability)**.** *Any (right or left) stochastically continuous d-dimensional stochastic process has a version which is measurable. Moreover, if the stochastic process is adapted then there is a version which is progressively measurable.* □

Sometimes we can take (a.s.) continuous modification of a given process on a bounded interval $[0,T]$

Theorem 1.21 (continuity)**.** *Let $\{X_t : t \in [0,T]\}$ be a d-dimensional stochastic process in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$E|X_t - X_s|^\alpha \le C|t-s|^{1+\beta}, \quad \forall s,t \in [0,T], \tag{1.11}$$

for some positive constants α, β and C. Then there exists a continuous version $Y = \{Y_t : t \in [0,T]\}$ of X, which is locally Hölder continuous with exponent γ, for any $\gamma \in (0, \beta/\alpha)$ i.e., there exist a null set N, with $P(N) = 0$, an (a.s.) positive random variable $h(\omega)$ and a constant $K > 0$ such that for all $\omega \in \Omega \setminus N$, $s,t \in [0,T]$ we have

$$|Y_t(\omega) - Y_s(\omega)| \le K|t-s|^\gamma \quad \text{if } 0 < |t-s| < h(\omega). \ \square$$

The previous result is essentially based on the following arguments, e.g., Karatzas and Shreve [124, pp. 53–55]). Estimate (1.11) and the dyadic construction $\{X(k2^{-n}) : k = 0, 1, \dots, 2^n,\ n = 1, 2, \dots\}$ yields

$$P\{\max_{1\le k\le 2^n} |X(k2^{-n}) - X((k-1)2^{-n})| \ge 2^{-\gamma}\} \le$$
$$\le \sum_{k=1}^{2^n} P\{|X(k2^{-n}) - X((k-1)2^{-n})| \ge 2^{-\gamma}\} \le C2^{-n(\beta-\alpha\gamma)},$$

for any $\gamma > 0$ such that $\beta > \alpha\gamma$. Hence, the Borel-Cantelli lemma shows that there exists a measurable set Ω^* of probability 1 such that for any ω in Ω^* there is an index $n^*(\omega)$ with the property

$$\max_{1\le k\le 2^n} |X(k2^{-n},\omega) - X((k-1)2^{-n},\omega)| < 2^{-\gamma}, \quad \forall n \ge n^*(\omega).$$

This proves that for t of the form $k2^{-n}$ we have a uniformly continuous process which gives the desired modification. Certainly, if the process X itself is separable, then we get do not need a modification, we obtain an equivalent continuous process.

An interesting point in this result, is the fact that the condition (1.11) on the given process X can be verified by means of the so-called two-dimensional distribution of the process (see below). Moreover, the integrability of the process is irrelevant, i.e., (1.11) can be replaced by

$$\lim_{\delta\to 0} P\Big\{ \sup_{|t-s|<\delta} |X(t)-X(s)| > \varepsilon \Big\} = 0, \quad \forall \varepsilon > 0.$$

This condition is stronger that

$$\lim_{\delta\to 0} \sup_t P\Big\{ \sup_{|s|<\delta} |X(t)-X(t+s)| > \varepsilon \Big\} = 0, \quad \forall \varepsilon > 0,$$

which only yields almost surely continuity at every time t. In any case, if the process X is separable then the same X is continuous, otherwise, we construct a version Y which is continuous.

Recall that a real function on an interval $[0,T)$ (respectively $[0,\infty)$ or $[0,T]$) has only *discontinuities of the first kind* if (a) it is bounded on any compact subinterval of $[0,T)$ (respectively $[0,\infty)$ or $[0,T]$), (b) left-hand limits exist on $(0,T)$ (respectively $(0,\infty)$ or $(0,T]$) and (c) right-hand limits exist on $[0,T)$ (respectively $[0,\infty)$ or $[0,T)$). After a normalization of the function, this is actually equivalent to a right continuous functions having left-hand limits, these functions are called cad-lag.

It is interesting to note that continuity of a (separable) process X can be localized, X is called continuous (or a.s. continuous) at a time t if the set N_t of ω such that $s \mapsto X(s,\omega)$ is not continuous at $s = t$ has probability zero (i.e., N_t is measurable, which is always true if X is separable, and $P(N_t) = 0$). Thus, a (separable) process X may be continuous at any time (i.e., $P(N_t) = 0$ for every t in T) but not necessarily continuous (i.e., with continuous paths, namely $P(\cup_t N_t) = 0$). Remark that a cad-lag process X may be continuous at any (deterministic) time (i.e., $P(N_t) = 0$ for every t in T) without having continuous paths, as we will se later, a typical example is a Poisson process.

Analogously to the previous theorem, a condition for the case of a modification with only discontinuities of the first kind can be given (e.g., see Gikhman and Skorokhod [98, Section IV.4], Wong [255, Proposition 4.3, p. 59] and its references)

Theorem 1.22 (cad-lag)**.** *Let $\{X_t : t \in [0,T]\}$ be a d-dimensional stochastic process in a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$\mathbb{E}\{|X_{t+h} - X_s|^\alpha |X_s - X_t|^\alpha\} \le Ch^{1+\beta}, \quad \forall\, 0 \le t \le s \le t+h \le T, \quad (1.12)$$

for some positive constants α, β and C. Then there exists a cad-lag version $Y = \{Y_t : t \in [0,T]\}$ of X. □

Similarly, for processes of locally bounded variation we may replace the expression $|\cdot|$ in (1.11) by the *variation* to get a corresponding condition. In general, by looking at a process as a random variable in $\mathbb{R}^T$ we can use a complete separable metric space $D \subset \mathbb{R}^T$ to obtain results analogous to the above, i.e., if (1.11) holds for the metric $\mathrm{d}(X_t, X_s)$ instead of the Euclidean distance $|X_t - X_s|$, then the conclusions of Theorem 1.21 are valid with $\mathrm{d}(Y_t, Y_s)$ in lieu of $|Y_t - Y_s|$, e.g., see Durrett [67, p. 5, Theorem 1.6].

The statistics of a stochastic process are characterized by its *finite-dimensional* distributions, i.e., by the family of probability measures

$$P_s(B) = P(\{(X(s_1,\omega),\dots,X(s_n,\omega)) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

with $s = (s_1,\dots,s_n)$, $n = 1, 2, \dots$, for a real valued process $\{X(t,\omega) : t \in T\}$. This family of finite-dimensional distributions essentially determines a stochastic processes (i.e., modulo all possible version of a process), but not the process itself. The above results allow the verification of the (path) continuity properties of a given stochastic process in term of its two-dimensional distribution.

A typical (sample) *path* of a stochastic process is the function $X(\cdot,\omega)$ for each fixed ω, and so, a stochastic process (with prescribed finite-dimensional distributions) can always be constructed in the product space $\mathbb{R}^T$, endowed with the σ-algebra $\mathcal{B}^T(\mathbb{R})$ generated by the algebra of cylindrical sets, which may be smaller than the Borel σ-algebra $\mathcal{B}(\mathbb{R}^T)$. Thus we can view a stochastic process X as probability measure P_X on $(\mathbb{R}^T, \mathcal{B}^T(\mathbb{R}))$, but in general the σ-algebra $\mathcal{B}^T(\mathbb{R})$ is not appropriated, it is *too small* comparatively with the *big* space $\mathbb{R}^T$ of all functions.

If we can find a proper subset $\Omega \subset \mathbb{R}^T$ containing almost every paths of X, i.e., such that $P_X^*(\Omega) = 1$ (where P_X^* is the exterior probability measure defined for any subset of $\mathbb{R}^T$), then the stochastic process X becomes a probability measure P on $(\Omega, \mathcal{B})$, where $\Omega \subset \mathbb{R}^T$ and $\mathcal{B} = \Omega \bigcap \mathcal{B}^T(\mathbb{R})$ is the restriction of $\mathcal{B}^T(\mathbb{R})$ to Ω with $P = P_X^*$, i.e., $P(\Omega \cap B) = P_X(B)$. It turn out that $\mathcal{B}$ contains only sets that can be described by a countable number of restrictions on $\mathbb{R}$, in particular a singleton (a one point set, which is closed for the product topology) may not be measurable. Usually, $\mathcal{B}$ is enlarged with all subsets of negligible (or null) sets with respect to P, and we can use the completion $\mathcal{B}^*$ of $\mathcal{B}$ as the measurable sets. Moreover, if Ω is an appropriate separable topological space by itself (e.g., continuous functions) so that the process have some regularity (e.g., continuous paths), then the Borel σ-algebra $\mathcal{B}(\Omega)$, generated by the open sets in Ω coincides with the previous $\mathcal{B}$. Note that another way to describe $\mathcal{B}$ is to see that $\mathcal{B}$ is the σ-algebra generated by sets (so-called cylinders in Ω) of the form $\{\omega \in \Omega : (\omega(s_1),\dots,\omega(s_n)) \in B\}$ for any $B \in \mathcal{B}(\mathbb{R}^n)$, with $s = (s_1,\dots,s_n)$, $n = 1, 2, \dots$.

Note that the arguments in Theorems 1.20, 1.21 or 1.22 are such that if we begin with a separable process, then we find that the measurable, continuous or cad-lag version Y is actually indistinguishable from the initial process X, i.e., $P(\{\omega : X_t(\omega) = Y_t(\omega),\ \forall t \in T\}) = 1$.

Definition 1.23 (process)**.** Given an index set T (usually $T \subset \mathbb{R}$), a measurable space $(E, \mathcal{E})$ (usually $E \subset \mathbb{R}^d$) and a probability space $(\Omega, \mathcal{F}, P)$, an E-valued *general stochastic process* is a measurable function X from $(\Omega, \mathcal{F})$ into $(E^T, \mathcal{E}^T)$, i.e. a family of E-valued random variables $\{X_t : t \in T\}$. Moreover, if E is a Hausdorff topological space, $\mathcal{E}$ is its Borel σ-algebra and there exits a topological sub-space B of the product space E^T (which is called *sample space* and endowed with its Borel σ-algebra $\mathcal{B}$) such that the restriction to B of the function $\omega \mapsto X(\cdot, \omega)$ (to emphasized, now denoted by $\bar{X}$) is a B-valued random variable, then $\bar{X}$ (or X) is called an E-valued *stochastic process with paths in* B. Usually B does not belong to the product σ-algebra $\mathcal{B}^T(E)$ (generated by all Borel cylindrical sets), and $\bar{X}$ (considered with values in $E^T \supset B$) is a version of the general process X. Actually $\bar{X}$ is identified with its P-equivalence class, and for each t in T, the *canonical* (coordinate, evaluation or projection) mapping $\bar{X} \mapsto \bar{X}_t$ from B into E is defined. The probability measure on B induced by $\bar{X}$ (denoted by P_X) is called the *law* of the process. Furthermore, if the index set $T = [0, \infty)$ then the minimal filtration satisfying the usual conditions (complete and right-continuous) $(\mathcal{F}_X(t) : t \geq 0)$ such that the E-valued random variables $\{\bar{X}_s : 0 \leq s \leq t\}$ are measurable is called the *canonical* filtration associated with the given process. On the other hand, given a family of finite-dimensional distributions on E^T of some (general) stochastic process X, a *realization* of a stochastic process X with paths in B and the prescribed finite-dimensional distributions is the probability space $(\Omega, \mathcal{F}, P)$ and the stochastic process $\bar{X}$ as above. □

Clearly, the passage from general stochastic processes (i.e., a family of random variables X) to stochastic processes is very subtle (however very significant, since X becomes a random variable in some topological space). Technically, if we take the trivial choice $B = E^T$ then to label X a *stochastic process*, we need to know that X is also measurable for Borel σ-algebra $\mathcal{B}(E^T)$, which may be larger than $\mathcal{B}^T(E)$, and so not a priori satisfied. Note that we abandon the trivial choice $B = E^T$ because in several arguments, we need that the Borel σ-algebra $\mathcal{B}$ of the topological space B (the sample space, where all paths are) coincides with the cylindrical Borel σ-algebra in B, i.e., a B-valued function Y is $\mathcal{B}(B)$-measurable if and only if for each t, the E-valued function $Y(t)$ is $\mathcal{B}(E)$-measurable. Usually, the definition of the (topological) sample space B involves some topology on the index set T. Also, if the index set $T \subset \mathbb{R}^n$, $n \geq 2$ then the name *random field* is preferable. The notion of general stochastic processes is as general as possible, however the concept of stochastic processes imposes some *path regularity*. Depending on the interest and the objective of the study undertaken, there are other possible approaches, for instance, when $E = \mathbb{R}^d$ we may consider (right- or left-)continuous processes, in probability or in mean-square, where processes are treated as continuous functions from T into $L^0(\Omega, \mathcal{F}, P)$, the complete metric space of $\mathbb{R}^d$-valued random variables endowed with the convergence in probability (called stochastically continuous processes), or into the Hilbert space $L^2(\Omega, \mathcal{F}, P)$ of square-integrable random variables (usually called second-order processes). This type of analysis is rarely used here, the interested

reader may check the book Gikhman and Skorokhod [98], among others.

Starting from a general stochastic process X, usually, a two-step procedure is applied: first a *good* version $\bar{X}$ is chosen and then a *good* sample space B is found. For instance, given a $\mathbb{R}$-valued general stochastic process X with index set $T = [0, \infty)$, we can apply Theorem 1.19 to get a separable version of $\bar{X}$ (with extended values, i.e., in $[-\infty, +\infty]^T$). Thus if I is a denumerable separant subset of indexes then we may consider $\bar{X}$ as taking valued in $[-\infty, +\infty]^I$ (or the one-point compactification $\mathbb{R}^I \cup \{\infty\}$), which is not exactly a subset of $\mathbb{R}^T$, but it is essentially the space of all (extended) real-valued sequences. This point of view is not considered in this book. Alternatively, if we know that the initial general stochastic process X is right-(or left-)continuous in probability (so when X is also separable any dense set in T is a separant set) then we may apply Theorem 1.20 to get a measurable version. Still, this point of view is rarely used in this book. Finally, the two-step procedure is as follows: first, we assume that Theorems 1.21 or 1.22 is applicable to X, so that a continuous or at least cad-lag version $\bar{X}$ can be found (i.e., first $\bar{X}(t, \omega) = X(t, \omega)$ is defined for every ω but only for t in a countable dense subset of $[0, \infty)$ and then $\bar{X}(t, \omega)$ is extended for every t in $[0, \infty)$ but only for ω outside of a negligible set). Second, we take as B the space of continuous or cad-lag functions, with a suitable topology and we forget about the initial general stochastic process X, because X (technically its version $\bar{X}$) is considered now as a random variable with values in a complete separable metric space B endowed with its Borel σ-algebra $\mathcal{B}$. So, essentially, a version (or modification) of a process is allowed only once, and then all stochastic processes (with the same path regularity) indistinguishable of this good version are considered equals. In general we use the following result, e.g., Kallenberg [121, Lemma 3.24, pp. 58]): Let T be a set of index, and $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be family of random variables (perhaps defined on distinct probability spaces) taking values in some separable metric space E such that X_t and Y_t have the same finite-dimensional distribution. If the paths $t \mapsto Y_t$ lie in some Borel subset B of $\mathcal{B}^T(E)$ (σ-algebra generated by cylindrical Borel sets) then there exists a family of random variables $\{\tilde{X}_t : t \in T\}$ with paths in B such that $P(X_t = \tilde{X}_t) = 0$ for every t in T, i.e., $\tilde{X}$ is a version of X.

Note that the initial probability space is irrelevant in the above context, we can always reduce to a canonical space of functions. Moreover, by taking the image measure through the map X if necessary, we can always reduce to the canonical process, i.e., the probability space becomes $(B, \mathcal{B}, P_X)$ and the process is the random variable $\omega \mapsto X_t(\omega) = \omega(t)$ from B into itself. So that the *law* of a process carried all necessary information. If two or more processes are involved, then we have to deal with more that one probability measure on the sample space $(B, \mathcal{B})$. Thus only one process can be reduced to the canonical process and the others can be viewed as probability measures or as measurable functions from B into itself. In most of the cases, the sample space B is a Polish (complete separable metric) space (better than E^T) where the motions of versions and equivalence classes coincide.

Clearly, most of the properties required for a stochastic process are stated relative to the probability P_X on the (Polish) sample space. However, for in-

stance, when we state some property relative to the sample path of a process (such as integrability in one of the variables), the difficulty is the fact we refer either to the map $t \mapsto X(t,\omega)$ for a fixed ω or to the map $\omega \mapsto X(t,\omega)$ for any fixed t. Thus, sometime we are forced to comeback to the initial setting of general stochastic process as a family of E-valued random variables. Alternatively, we may define processes as functions from the *base space* $T \times \Omega$ into E, this is, we consider X as a function of two-variables $X = X(t,\omega)$, and immediately we restrict the attention to (joint) measurable functions, i.e., the so-called measurable processes. This approach yields delicate measurability problems when dealing with stochastic integration, as carefully discussed in Bichteler [25].

The following type of processes may be useful

Definition 1.24 (Gaussian)**.** A real valued process $\{X(t) : t \in T\}$ is a *Gaussian* process if for any finite sub-family $(t_1, \ldots, t_n)$ of indexes in T, the random variable $(X(t_1), \ldots, X(t_n))$ has a Gaussian n-dimensional distribution. Its *mean* is $m(t) := \mathbb{E}\{X(t)\}$ and its *covariance* is defined by $\Gamma(s,t) := \mathbb{E}\{[X(s) - m(s)][X(t) - m(t)]\}$, for any s, t in T. The process is called *centered* if $\mathbb{E}\{X(t)\} = 0$ for any t in T. □

An important property of the Gaussian processes is the fact that its covariance function is always semi-definite positive, i.e., for any $(t_1, \ldots, t_n)$, any n, the matrix $\{\Gamma(t_i, t_j)\}$ is semi-definite positive. Moreover, any symmetric semi-definite positive function is the covariance of a centered Gaussian process, see Revuz and Yor [212, p. 36, Chapter 1].

Another important class of processes is the following

Definition 1.25 (stationary)**.** A E-valued process $\{X(t) : t \geq 0\}$ is called *stationary* if for every $t_1, \ldots, t_n$ and t we have

$$P(\{X(t_1 + t) \in A_1, \ldots, X(t_n + t) \in A_n\}) = \\ = P(\{X(t_1) \in A_1, \ldots, X(t_n) \in A_n\}),$$

for any Borel sets $A_1, \ldots, A_n$ in E, i.e., its finite-dimensional distribution (see Section 1.10) is invariant by a time translation. □

These processes play a central role in the study of *ergodicity* or *stability*, e.g., see the books Khasminskii [130] and Skorokhod [230].

1.7 Markov Chains

Let $\{X(t) : t \in T\}$, $T \subset \mathbb{R}$ be an E-valued stochastic process, i.e. a (complete) probability measure P on $(E^T, \mathcal{B}^T(E))$. If the cardinality of the state space E is finite, we say that the stochastic process takes finitely many values, labeled $1, \ldots, n$. This means that the probability law P on $(E^T, \mathcal{B}^T(E))$ is concentrated in n points. Even in this situation, when the index set T is uncountable, the σ-algebra $\mathcal{B}^T(E)$ is rather small (see Exercise 1.33, a set of a single point is not measurable). A typical path takes the form of a function $t \mapsto X(t,\omega)$ and

cannot be a continuous function in t. As discussed later, it turn out that cad-lag functions are a good choice. The characteristics of the stochastic processes $\{X(t) : t \in T\}$ are the functions $t \mapsto x_i(t) := P\{X(t) = i\}$, for any $i = 1, \dots, n$, with the property $\sum_{i=1}^{n} x_i = 1$. We are interested in the case where the index set T is usually an interval of $\mathbb{R}$.

Now, we turn our attention where the stochastic process describes some evolution process, e.g., a dynamical system. If we assume that the dimension of X is sufficiently large to include all relevant information and that the index t represents the time, then the knowledge of $X(t)$, referred to as the state of the system at time t, should summarize all information up to the present time t. This translated mathematically to

$$P\{X(t) = j \mid X(r),\, r \le s\} = P\{X(t) = j \mid X(s)\}, \tag{1.13}$$

almost surely, for every $t > s$, $j = 1, \dots, n$. At this point, the reader may consult the classic book Doob [59, Section VI.1, pp. 235–255] for more details. Thus, the evolution of the system is characterized by the transition function $p_{ij}(s,t) = P\{X(t) = j \mid X(s) = i\}$, i.e., a transition from the state j at time s to the state i at a later time t. Since the stochastic process is assumed to be cad-lag, it seems natural to suppose that the functions $p_{ij}(s,t)$ satisfies for every $i, j = 1, \dots, n$ conditions

$$\begin{cases} \displaystyle\sum_{j=1}^{n} p_{ij}(s,t) = 1, \ \forall t > s, \\ \displaystyle\lim_{(t-s)\to 0} p_{ij}(s,t) = \delta_{ij}, \ \forall t > s, \\ \displaystyle p_{ij}(s,t) = \sum_{k=1}^{n} p_{ik}(s,r) p_{kj}(r,t), \ \forall t > r > s. \end{cases} \tag{1.14}$$

The first condition expresses the fact that $X(t)$ takes values in $\{1, \dots, n\}$, the second condition is a *natural regularity* requirement, and the last conditions are known as the Chapman-Kolmogorov identities. Moreover, if $p_{ij}(s,t)$ is smooth in s, t so that we can differentiate either in s or in t the last condition, and then let $r - s$ or $t - r$ approaches 0 we deduce a system of ordinary differential equations, either the *backward* equation

$$\begin{cases} \partial_s p_{ij}(s,t) &= \displaystyle\sum_{k=1}^{n} \rho_{ik}^{+}(s)\, p_{kj}(s,t), \quad \forall t > s, \ i, j, \\ \rho_{ij}^{+}(s) &= \displaystyle\lim_{r \to s} \partial_s p_{ij}(s,r) \quad \forall s, \ i, j, \end{cases} \tag{1.15}$$

or the *forward* equation

$$\begin{cases} \partial_t p_{ij}(s,t) &= \displaystyle\sum_{k=1}^{n} p_{ik}(s,t)\, \rho_{kj}^{-}(t), \quad \forall t > s, \ i, j, \\ \rho_{ij}^{-}(t) &= \displaystyle\lim_{r \to t} \partial_t p_{ij}(r,t) \quad \forall t, \ i, j, \end{cases} \tag{1.16}$$

The quantities $\rho_{ij}^+(s)$ and $\rho_{ij}^-(s)$ are the characteristic of the process, referred to as *infinitesimal rate*. The initial condition of (1.14) suggests that $\rho_{ij}^-(s) = -\rho_{ij}^+(t) := \rho_{ij}(t)$, if $s = t$. Since $\sum_{j=1}^n p_{ij}(s,t) = 1$ we deduce

$$\rho(t,i,j) \geq 0, \quad \forall i \neq j, \qquad \rho(t,i,i) = -\sum_{j \neq i} \rho(t,i,j). \tag{1.17}$$

Using matrix notation, $R(\cdot) = \{\rho_{ij}\}$, $P(s,t) = \{p_{ij}(s,t)\}$ we have

$$\begin{cases} \partial_s P(s,t) &= -R(s)P(s,t), \quad \forall s < t, \\ \partial_t P(s,t) &= P(s,t)R(t), \quad \forall t > s, \\ \lim\limits_{t-s\to 0} P(s,t) &= I, \quad \forall t > s. \end{cases} \tag{1.18}$$

Conversely, given the integrable functions $\rho_{ij}(t)$, $i,j = 1,\dots,n$, $t \geq 0$ satisfying (1.17), we may solve the system of (non-homogeneous and linear) ordinary differential equations (1.15), (1.16) or (1.18) to obtain the transition (matrix) function $P(s,t) = \{p_{ij}(s,t)\}$ as the fundamental solution (or Green function). For instance, the reader may consult the books by Chung [43], Yin and Zhang [259, Chapters 2 and 3, pp. 15–50].

Since $P(s,t)$ is continuous in $t > s \geq 0$ and satisfies the conditions in (1.14), if we give an initial distribution, we can find a cad-lag realization of the corresponding Markov chain, i.e., a stochastic process $\{X(t) : t \geq 0\}$ with cad-lag paths such that $P\{X(t) = j \mid X(s) = i\} = p_{ij}(s,t)$, for any $i,j = 1,\dots,n$ and $t \geq 0$. In particular, if the rates $\rho_{ij}(t)$ are independent of t, i.e., $R = \{\rho_{ij}\}$, then the transition matrix $P(s,t) = \exp[(t-s)R]$. In this case, a realization of the Markov chain can be obtained directly from the rate matrix $R = \{\rho_{ij}\}$ as follows. First, let Y_n, $n = 0,1,\dots$ be a sequence of E-valued random variables with $E = \{1,\dots,n\}$ and satisfying $P(Y_n = j \mid Y_{n-1} = i) = \rho_{ij}/\lambda$, if $i \neq j$ with $\lambda = -\inf_i \rho_{ii}$, $i > 0$, and Y_0 initially given. Next, let $\tau_1, \tau_2, \dots$ be a sequence of independent identically distributed exponentially random variables with parameter λ i.e., $P(\tau_i > t) = \exp(-\lambda t)$, which is independent of $(Y_0, Y_1, \dots)$. If we define $X(t) = Y_n$ for t in the stochastic interval $[\![T_n, T_{n+1}[\![$, where $T_0 = 0$ and $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$, then $X(t)$ gives a realization of the Markov chain with the above infinitesimal rates.

A typical setting includes $T = [0,\infty)$ and a denumerable state space E (with the discrete topology). This type of processes are very useful in modeling dynamical systems, referred to either as *queueing systems* or as *point processes* very well known in the literature, e.g., Bremaud [33], Kemeny and Snell [129], Kleinrock [132], Nelson [187].

This study is simplified if the time is discrete, i.e., the Markov chain X_n, $n = 0,1,\dots$, with values in subset E of $\mathbb{R}^d$ is defined by recurrence by

$$P\{X_{n+1} \in \cdot \mid X_n\} = P(X_n, \cdot), \quad n \geq 1,$$

where the *stochastic kernel* $P(x,A)$ satisfies:

(a) $P(x,\cdot)$ is a probability measure on $\mathcal{B}(E)$ for any x in E

(b) $P(\cdot, A)$ is measurable in E for any A in $\mathcal{B}(E)$.

The finite-dimensional distributions (as discussed later in Section 1.10) of a Markov chain is given by

$$\begin{cases} P\{X_0 \in A_0,\, X_1 \in A_1, \dots\, X_n \in A_n\} = \\ \qquad = \displaystyle\int_{A_0} \nu(\mathrm{d}x_0) \int_{A_1} P(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P(x_{n-1}, \mathrm{d}x_n), \end{cases} \tag{1.19}$$

for any $A_0, A_1, \dots, A_n$ in $\mathcal{B}(E)$, and where $\nu(\cdot)$ is the initial distribution. Thus, given the stochastic kernel we can use Kolmogorov's construction theorem (see Theorem 1.30 below) to obtain a Markov chain X_n for $n = 0, 1, \dots$ satisfying the above equation (1.19). Moreover, we have

$$P\{X_n \mid X_0 = x\} = P^n(x, \cdot)$$

where $P^n(x, A)$ denote the n kernel convolutions, defined by induction by

$$P^n(x, A) := \int_E P^{n-1}(x, \mathrm{d}y) P(y, A). \tag{1.20}$$

The reader may consult the book by Chung [43] and Shields [226], among others, for a more precise discussion.

Before going further, let us mention a couple of classic simple processes which can be viewed as Markov chains with denumerable states, e.g., see Feller [81, Vol I, Sections XVII.2–5, pp. 400–411]. All processes below $\{X(t) : t \geq 0\}$ take values in $\mathbb{N} = \{0, 1, \dots\}$, with an homogeneous transition given by $p(j, t-s, n) = P\{X(t) = j \mid X(r),\, 0 \leq r < s,\, X(s) = n\}$, for every $t > s \geq 0$ and j, n in $\mathbb{N}$. Thus, these processes are completely determined by the knowledge of the characteristics $p(t, n) := P\{X(t) = n\}$, for every $t \geq 0$ and n in $\mathbb{N}$, and a description on the change of values.

The first example is the Poisson process where there are only changes from n to $n+1$ (at a random time) with a fix rate $\lambda > 0$, i.e.,

$$\begin{cases} \partial_t p(t, n) = -\lambda \big[p(t, n) - p(t, n-1) \big], \\ \partial_t p(t, 0) = -\lambda p(t, 0), \end{cases} \tag{1.21}$$

for every $t \geq 0$ and n in $\mathbb{N}$. Solving this system we obtain

$$p(t, n) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad t \geq 0,\ n \in \mathbb{N},$$

which is the Poisson distribution.

The second example is a *pure birth* process where the only variation relative to the Poisson process is the fact that the rate is variable, i.e., the rate is $\lambda_n > 0$ when $X(t) = n$. This means that (1.21) becomes

$$\begin{cases} \partial_t p(t, n) = -\lambda_n p(t, n) + \lambda_{n-1} p(t, n-1), \\ \partial_t p(t, 0) = -\lambda p(t, 0), \end{cases} \tag{1.22}$$

for every $t \geq 0$ and n in $\mathbb{N}$. Certainly, this system can be solved explicitly, but the expression is rather complicate in general. If X represents the size of a population then the quantity λ_n is called the average rate of growth. An interesting point is the fact that $\{p(t,n) : n \in \mathbb{N}\}$ is indeed a probability distribution, i.e.,

$$\sum_{n=1}^{\infty} p(t,n) = 1$$

if and only if the coefficients λ_n increase sufficiently fast, i.e., if and only if the series $\sum_n \lambda_n^{-1}$ diverges.

The last example is the *birth-and-death* process, where the variation is the fact that either a change from n to $n+1$ (birth) with a rate λ_n or from n to $n-1$, if $n \geq 1$ (death) with a rate μ_n may occur. Again, the system (1.21) is modifies as follows

$$\begin{cases} \partial_t p(t,n) = -(\lambda_n + \mu_n)p(t,n) + \lambda_{n-1}p(t,n-1) + \mu_{n+1}p(t,n+1), \\ \partial_t p(t,0) = -\lambda p(t,0) + \mu_1 p(t,1), \end{cases} \tag{1.23}$$

for every $t \geq 0$ and n in $\mathbb{N}$. As in the case of a pure birth process, some conditions are needed on the rates $\{\lambda_0, \lambda_1, \ldots\}$ and $\{\mu_1, \mu_2, \ldots\}$ to ensure that the birth-and-death process does not get trap in 0 or ∞ and the characteristics $\{p(t,n) : n \in \mathbb{N}\}$ is a probability distribution.

1.8 Processes without after-effect

Markov processes, or stochastic processes without aftereffect, or non-hereditary (or memory less) stochastic processes refer always to the same property: any additional knowledge concerning the states of a process $\{X(\cdot)\}$ at previous times $s < t_0$ relative to the present time t_0 does not affect the statistics of the process at future times $s > t_0$. As will be noticed later, this means that a stochastic processes without aftereffect is completed determined by its transition function, i.e., a function $P(s,x,t,\mathrm{d}y)$, which is the condition (transition) probability of $X(t)$ knowing that $X(s) = x$. For real valued process, we may use the transition distribution function $F(s,x,t,y)$, for $s < t$ and x, y in $\mathbb{R}$, associated with $\{X(\cdot)\}$, which is the probability that $X(t) < y$ assuming that $X(s) = x$. Clearly the function F should satisfies

$$\lim_{y \to -\infty} F(s,x,t,y) = 0, \qquad \lim_{y \to \infty} F(s,x,t,y) = 1,$$

and continuity from the left in y

$$\lim_{z \uparrow y} F(s,x,t,z) = F(s,x,t,y)$$

as any distribution function. To be able to use Bayes' formula for conditional distribution it is convenient to assume that F is continuous with respect to the variables s, x and t. This yields the Chapman-Kolmogorov equation (or identity)

$$F(s,x,t,y) = \int_{\mathbb{R}} F(s,x,\tau,\mathrm{d}\zeta)\, F(\tau,\zeta,t,y),$$

valid for any $s < \tau < t$ and x, y in $\mathbb{R}$. So, we may complete the definition of $F(s,x,t,y)$ for $s = t$ by a limit continuation as follows

$$\lim_{t\to s+} F(s,x,t,y) = \lim_{s\to t-} F(s,x,t,y) = F_0(x,y), \tag{1.24}$$

where $F_0(x,y) = 1$ if $y > x$ and $F_0(x,y) = 0$ otherwise. In what follows, this condition (1.24) is always assumed.

If a density function $f(s,x,t,y) = \partial_y F(s,x,t,y)$ exists then

$$F(s,x,t,y) = \int_{-\infty}^{y} f(s,x,t,\zeta)\mathrm{d}\zeta$$

and clearly

$$f(s,x,t,y) = \int_{\mathbb{R}} f(s,x,\tau,\zeta) f(\tau,\zeta,t,y)\mathrm{d}\zeta,$$

for any $s < \tau < t$ and x, y in $\mathbb{R}$.

For instance, the interested reader should consult the classic books Feller [81, Vol II, Sections X.3–5, pp. 316–331] and Gnedenko [101, Sections 53–55, pp. 358–376], for a more detailed discussion and proofs of most the results in this section.

For a *continuous* stochastic process we assume

$$\begin{cases} \lim_{t\to s+} \dfrac{1}{t-s} \displaystyle\int_{|y-x|\geq\delta} F(s,x,t,\mathrm{d}y) = 0, \\ \lim_{s\to t-} \dfrac{1}{t-s} \displaystyle\int_{|y-x|\geq\delta} F(s,x,t,\mathrm{d}y) = 0, \end{cases} \tag{1.25}$$

for every s, x and t, plus the following two conditions: (a) the first and second partial derivatives $\partial_x F(s,x,t,y)$ and $\partial_x^2 F(s,x,t,y)$ exit and are continuous for any $s < t$, x and y, and (b) for every $\delta > 0$ the limits

$$\begin{cases} \lim_{t\to s+} \dfrac{1}{t-s} \displaystyle\int_{|y-x|<\delta} (y-x)^2 F(s,x,t,\mathrm{d}y) = 2a(s,x), \\ \lim_{s\to t-} \dfrac{1}{t-s} \displaystyle\int_{|y-x|<\delta} (y-x)^2 F(s,x,t,\mathrm{d}y) = 2a(t,x), \end{cases} \tag{1.26}$$

and

$$\begin{cases} \lim_{t\to s+} \dfrac{1}{t-s} \displaystyle\int_{|y-x|<\delta} (y-x) F(s,x,t,\mathrm{d}y) = b(s,x), \\ \lim_{s\to t-} \dfrac{1}{t-s} \displaystyle\int_{|y-x|<\delta} (y-x) F(s,x,t,\mathrm{d}y) = b(t,x), \end{cases} \tag{1.27}$$

exit uniformly in x for every $s < t$ fixed. Note that the limiting coefficients a and b in (1.26) and (1.27) seem to depend on the value δ, but in view of the continuity condition (1.25) is merely apparent.

Under the above assumptions (1.25), (1.26) and (1.27) we can easily prove the validity of the *backward* Kolmogorov equation, namely

$$\partial_s F(s,x,t,y) + a(s,x)\partial_x^2 F(s,x,t,y) + b(s,x)\partial_x F(s,x,t,y) = 0, \qquad (1.28)$$

for any $s < t$, x and y. Actually, only the first part of conditions (1.26) and (1.27) play any role here.

However, to deduce the *forward* Kolmogorov (also called Fokker-Planck) equation satisfied by the (probability) density function, namely

$$\partial_t f(s,x,t,y) + \partial_y\big[b(t,y)f(s,x,t,y)\big] = \partial_y^2\big[a(t,y)f(s,x,t,y)\big], \qquad (1.29)$$

for any $s < t$, x and y, we do need the second part of conditions (1.26) and (1.27), as well as the existence and continuity of the derivatives $\partial_t f(s,x,t,y)$, $\partial_y\big[b(t,y)f(s,x,t,y)\big]$ and $\partial_y^2\big[a(t,y)f(s,x,t,y)\big]$.

If the continuity condition (1.25) is slightly strengthen into

$$\lim_{t\to s+}\frac{1}{t-s}\int_{|y-x|\geq\delta}(y-x)^2F(s,x,t,\mathrm{d}y) = 0,$$
$$\lim_{s\to t-}\frac{1}{t-s}\int_{|y-x|\geq\delta}(y-x)^2F(s,x,t,\mathrm{d}y) = 0,$$

then conditions (1.26) and (1.27) are valid also for $\delta = 0$, which mean

$$\mathbb{E}\{[X(t)-X(s)] - (t-s)b(s,X(s))\} = (t-s)\varepsilon_1,$$
$$\mathbb{E}\{[X(t)-X(s)]^2 - 2(t-s)a(s,X(s))\} = (t-s)\varepsilon_2,$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $t-s \to 0+$, in term of the first and second moment of the Markov process $x(\cdot)$. As seem later, the actual construction of the stochastic process $\{X(t) : t \geq 0\}$ is quite delicate, in particular if $a > 0$ then almost surely the path $t \mapsto X(t,\omega)$ are continuous, but with infinite variation.

The transition distribution F of a *purely jump* Markov process $\{X(\cdot)\}$ can be expressed as follows:

$$\begin{cases} F(s,x,t,y) = [1-(t-s)\lambda(s,x)]F_0(x,y) + \\ \qquad\qquad +(t-s)\lambda(s,x)\rho(s,x,y) + (t-s)\varepsilon, \end{cases} \qquad (1.30)$$

where $\varepsilon \to 0$ as $(t-s) \to 0^+$, uniformly in x and y. Thus, on any time interval (s,t) the value $X(\cdot)$ remains constant and is equal to $X(s) = x$ with probability $1-(t-s)\lambda(s,x)+(t-s)\varepsilon$, with $\varepsilon \to 0$ as $t-s \to 0^+$ (so that it may only have a jump with a probability $(t-s)\lambda(s,x)+(t-s)\varepsilon$). The coefficient $\rho(s,x,y)$ denotes the conditional distribution function of $X(s)$ under the condition that a jump has indeed occurred at time s and that immediately before the jump $X(\cdot)$ was equal to $X(s-) = x$.

Certainly, $\lambda(s,x)$ and $\rho(s,x,y)$ are nonnegative, $y \mapsto \rho(s,x,y)$ is a nondecreasing function continuous from the left, $\rho(s,x,-\infty) = 0$, $\rho(s,x,\infty) = 1$, and we assume that $\rho(s,x)$ is bounded, and that both $\rho(s,x)$ and $\rho(s,x,y)$ are continuous in s and Borel measurable in x. Under these conditions, (1.30) and

the continuity (1.24), the following *backward* and *forward* Kolmogorov *integro-differential* equations

$$\begin{cases} \partial_s F(s,x,t,y) = \\ \qquad = \lambda(s,x) \displaystyle\int_{\mathbb{R}} \big[F(s,x,t,y) - F(s,\zeta,t,y)\big]\rho(s,x,\mathrm{d}\zeta), \\ \partial_t F(s,x,t,y) = \displaystyle\int_{\mathbb{R}} \big[\lambda(s,\zeta) - \lambda(t,\zeta)\rho(t,\zeta,y)\big]F(s,x,t,\mathrm{d}\zeta), \end{cases} \tag{1.31}$$

for any $s < t$, x and y. In the case of purely jump, the construction of the process $\{X(t) :\geq 0\}$ is relatively simple and the path $t \mapsto X(t,\omega)$ are piecewise constant (and normalized to be left-hand continuous) almost surely, however, the average or mean $t \mapsto \mathbb{E}\{X(t)\}$ is a continuous function.

Since Gaussian and Poisson processes are the prototypes of continuous and jump processes, we would like to quote some results related to the convergence of a sum of independent random variables to the Gaussian and Poisson distributions, e.g., see Gnedenko [101, Sections 49, pp. 336–339].

Let $\{\xi_{n,1}, \xi_{n,2}, \dots, \xi_{n,m_n}\}$ be a set of m_n independent random variables for $n = 1, 2, \dots$, and set $\zeta_n = \xi_{n,1} + \xi_{n,2} + \cdots \xi_{n,m_n}$ and denote by $F_{n,m}(x)$ the distribution function of the random variable $\xi_{n,m}$.

Suppose that we normalize the sequence so that

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\{\xi_{n,m}\} = 0,$$
$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\big\{\big[\xi_{n,m} - \mathbb{E}\{\xi_{n,m}\}\big]^2\big\} = a > 0.$$

Then the sequence of distribution functions $G_n(x)$ of the sum of random variables ζ_n converge to the Gaussian (or normal) distribution

$$N(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\big(-\frac{y^2}{2a}\big)\mathrm{d}y$$

if and only if

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \int_{|x|>\delta} x^2 F_{n,m}(\mathrm{d}x) = 0,$$

for every $\delta > 0$.

Alternately, suppose that we normalize the sequence so that

$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\{\xi_{n,m}\} = \lambda > 0,$$
$$\lim_{n\to\infty} \sum_{m=1}^{m_n} \mathbb{E}\big\{\big[\xi_{n,m} - \mathbb{E}\{\xi_{n,m}\}\big]^2\big\} = \lambda.$$

Then the sequence of distribution functions $G_n(x)$ of the sum of random variables ζ_n converge to the Poisson distribution

$$P(x) := \begin{cases} 0 & \text{for} \quad x \leq 0, \\ \mathrm{e}^{-\lambda} \sum\limits_{0 \leq k < x} \dfrac{\lambda^k}{k!} & \text{for} \quad x > 0 \end{cases}$$

if and only if

$$\lim_{n \to \infty} \sum_{m=1}^{m_n} \int_{|x-1|>\delta} x^2 F_{n,m}(\mathrm{d}x + \lambda_{n,m}) = 0,$$

for every $\delta > 0$, where $\lambda_{n,m} := \mathbb{E}\{\xi_{n,m}\}$.

Clearly, plenty of general results existent in the current literature regarding limits of sum of random variables, the point we remark is the fact that a normalization, the sum converges to a Gaussian distribution if the mean and the variance have limits of different order, one zero and the other non-zero. However, if the mean and the variance have limits of the same order (non-zero) the sum converges to a Poisson distribution. This gives the two flavors, continuous and discontinuous (jumps) processes.

Now, let us discuss (Markov) jump processes from sample space point of view as in the classic book Doob [59, Section VI.2, pp. 255–273]. Similarly to the transition distribution function, an homogeneous transition probability function $P(x, t, A)$, x in a E (usually an open or closed subset of $\mathbb{R}^d$), $t > 0$ and A in $\mathcal{B}(E)$, the Borel σ-algebra in E, satisfies: (a) for each $t > 0$ and x in E the function $A \mapsto P(x, t, A)$ is a probability measure on $(E, \mathcal{B}(E))$, (b) for each $t > 0$ and A in $\mathcal{B}(E)$ the function $x \mapsto P(x, t, A)$ is a (Borel) measurable, (c) for any x in E and A in $\mathcal{B}(E)$ we have

$$\lim_{t \to 0} P(x, t, A) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise, (d) for each $t, s > 0$, x in E and A in $\mathcal{B}(E)$ we have

$$P(t + s, x, A) = \int_E P(t, x, \mathrm{d}y)\, P(s, y, A)$$

which is referred to as the Chapman-Kolmogorov identity.

If the continuity condition (c) above is strengthen into (or replace by)

$$\lim_{t \to 0} \inf_{x \in E} P(x, t, \{x\}) = 1, \quad \forall x \in E, \tag{1.32}$$

then Doeblin's result proves that there is a homogeneous piecewise constant Markov process $\{X(t) : t \geq 0\}$ with a cad-lag path and transition probability function $P(x, t, A)$. By piecewise constant (cad-lag) paths we means that if $X(t, \omega) = x$ then there exists a positive constant $\delta = \delta(t, \omega)$ such that $X(s) = x$ for every s in the stochastic interval $[t, t + \delta)$.

By means of the Chapman-Kolmogorov identity and under the continuity assumption (1.32), where the uniform limit in x is essential, we may use the transition probability function $P(x,t,A)$ to define the pair of functions $\lambda(x)$ and $\Lambda(x,A)$, for x in E and A in $\mathcal{B}(E)$ as follows:

$$\begin{cases} \lim\limits_{t\to 0^+} \dfrac{1-P(x,t,\{x\})}{t} = \lambda(x), \\ \lim\limits_{t\to 0^+} \dfrac{P(x,t,A)}{t} = \Lambda(x,A), \end{cases} \tag{1.33}$$

for any x in E, A in $\mathcal{B}(E\smallsetminus\{x\})$, where the definition is completed by setting $\Lambda(x,\{x\})=0$. Thus the function $x\mapsto\lambda(x)$ results bounded in E, the function $x\mapsto\Lambda(x,A)$ is Borel measurable for every fixed A in $\mathcal{B}(E)$, the function $A\mapsto\Lambda(x,A)$ is finite Borel measure on E (or in $E\smallsetminus\{x\}$) for every fixed x in E, and $\lambda(x)=\Lambda(x,E)=\Lambda(x,E\smallsetminus\{x\})$, so that $0\le\Lambda(x,A)\le\lambda(x)$, for every x in E and $\mathcal{B}(E)$. Moreover, both convergence in (1.33) is uniform in x and A in $\mathcal{B}(E\smallsetminus\{x\})$, x in E. Usually, we normalize the coefficients so that $\bar\Lambda(x,A):=\Lambda(x,A)/\Lambda(x,E)$ is a probability measure (with a distribution $\rho(x,A)$ if $E=\mathbb{R}$ as in the previous discussion) and $\bar\lambda(x)=\lambda(x)\Lambda(x,E)$.

Note that as in the initial discussion with the transition distribution function, for every x in E and A in $\mathcal{B}(E)$ we do have

$$P(x,t,A)=\big[1-t\lambda(x)\big]\delta(x,A)+t\Lambda(x,A)+t\varepsilon,$$

with $\varepsilon\to 0$ as $t\to 0^+$, uniformly in x and A in $\mathcal{B}(E)$. The *Backward* and *Forward* Kolmogorov *integro-differential* equations have the form

$$\begin{cases} \partial_t P(x,t,A)=\lambda(x)\displaystyle\int_E \big[P(x,t,A)-P(\zeta,t,A)\big]\Lambda(x,\mathrm{d}\zeta), \\ \partial_t P(x,t,A)=\displaystyle\int_E \lambda(\zeta)\big[\Lambda(\zeta,E)-\Lambda(\zeta,A)\big]P(x,t,\mathrm{d}\zeta), \end{cases} \tag{1.34}$$

for any $s<t$, x in E and A in $\mathcal{B}(E)$. The last equation takes also the form

$$\partial_t P(x,t,A)=\int_{E\smallsetminus A}\Lambda(\zeta,A)P(x,t,\mathrm{d}\zeta)-\int_A\Lambda(\zeta,E\smallsetminus A)P(x,t,\mathrm{d}\zeta).$$

Moreover, with suitable assumptions all this extends to non-homogeneous transition probability functions on Polish (separable, complete and metric) spaces.

1.9 Markov Processes

There is an important class of processes known as Markov process which are used to model dynamical systems under disturbances. They are based on the principle that the *future* is independent of the *past* when we know the *present*. Similar to the *state variable* model for deterministic dynamical systems. Essentially, it is a matter of what is called *state* so that any dynamical process can be view a Markov process with a *larger* state. However, the price of the Markov

character is the lack of differentiability in time of the process as we will see later. It is convenient to assume that state-space is a complete metric space (i.e. a Polish space) and that the index set T has a natural order e.g., T is a subset of $\mathbb{R}$. In most of our cases $T = [0, \infty)$ and E is a either closed or open subset of $\mathbb{R}^d$.

A stochastic process X on a (complete) probability space $(\Omega, \mathcal{F}, P)$ and values in a Polish space E satisfies the *Markov property* if for any $n = 1, 2 \dots$, any bounded measurable (actually continuous suffices, because E is a complete metric space) functions $f_1, \dots, f_n$, $g_1, \dots, g_n$, h, and times $s_1 \le \cdots \le s_n \le t \le t_1 \le \cdots \le t_n$ we have

$$\left\{ \begin{aligned} &\mathbb{E}\big\{h(X_t)\big(\prod_{i=1}^{n} f(X_{s_i})\big)\big(\prod_{i=1}^{n} g(X_{t_i})\big)\big\} = \\ &\qquad = \mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t)\}\prod_{i=1}^{n} g(X_{t_i})\big\}, \end{aligned} \right. \tag{1.35}$$

where $\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t\}$ is X_t-measurable functions satisfying

$$\mathbb{E}\big\{h(X_t)\prod_{i=1}^{n} f(X_{s_i})\big\} = \mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n} f(X_{s_i}) \,|\, X_t)\}\big\},$$

i.e., it is the conditional expectations with respect to the σ-algebra generated by the random variable X_t. This is briefly expressed by saying that the future is independent of the past given the present. Clearly, this condition involves only the finite-dimensional distributions of the process, and (1.35) is equivalent to (e.g., see Blumental and Getoor [28, Thm 1.3, pp. 12-14]) either

$$P(X_t \in B \,|\, X_{s_1}, \dots, X_{s_n}, X_s) = P(X_t \in B \,|\, X_s), \quad \text{a.s.}$$

for every $t > s \ge s_n > \cdots > s_1$, B in $\mathcal{B}(E)$, or

$$\mathbb{E}\{f(X_t) \,|\, X_{s_1}, \dots, X_{s_n}, X_s\} = \mathbb{E}\{f(X_t) \,|\, X_s\}, \quad \text{a.s.},$$

for every $t > s \ge s_n > \cdots > s_1$, and for any arbitrary bounded and continuous (actually, with compact support when E is locally compact) function f from E into $\mathbb{R}$.

Definition 1.26 (history)**.** Given a stochastic process X on a (complete) probability space $(\Omega, \mathcal{F}, P)$ we can define the *history* (or internal history or strict history) of the process as the increasing family of σ-algebras $\{\mathcal{H}(t) : t \in T\}$, where each $\mathcal{H}(t)$ is generated by the random variables $\{X(s) : s \le t\}$ and the null sets. Similarly, the *innovation* $\{\mathcal{H}^{\perp}(t) : t \in T\}$ is the decreasing family of σ-algebras, where each $\mathcal{H}^{\perp}(t)$ is generated by all sets in some $\mathcal{H}(s)$ for $s > t$ which are independent of $\mathcal{H}(t)$ (see Exercises 1.19 and 1.20). □

The internal history $\{\mathcal{H}(t) : t \in T\}$ of a process X is also denoted by $\{\mathcal{F}^X(t) : t \in T\}$ (or $\{\mathcal{F}_X(t) : t \in T\}$ or with $\mathcal{H}$ replacing $\mathcal{F}$) and contains (or records) all events linked to the process X, up to (and including) the time

t, i.e., past and present. From the system-science point of view, the history $\{\mathcal{F}^X(t) : t \in T\}$ is best thought as an *increasing information pattern*. On the other hand the innovation $\{\mathcal{H}^\perp(t) : t \in T\}$ records all events linked to the process X, after time t and is unrelated to (independent of) the past.

Based on the observation of a stochastic process up to the present time we can know whether a *causal* phenomenon has (or has not) already taken place. If *causally* is understood in this way, a random variable τ with values in $[0, \infty]$ can be interpreted as a random time of occurrence of some phenomenon depending causally upon the process X when the event $\{\omega : \tau(\omega) \leq t\}$ is $\mathcal{F}^X(t)$-measurable, which correspond to the notion of optional or stopping times previously mentioned.

Most of the processes that we are going to discuss will be cad-lag (see Exercises 1.21), and the history $\{\mathcal{H}(t) : t \in T\}$ will be right-continuous and therefore be equal to the canonical filtration (associated with the given process), after being augmented with all zero-measure sets. By construction $\mathcal{H}(t)$ is independent of $\mathcal{H}^\perp(t)$ for any $t \in T$, $\mathcal{H}(t)$ represents the *past* and *present* information at time t and $\mathcal{H}^\perp(t)$ is the *future* new information to come.

Thus, another process Y is said to be *adapted* to X if $Y(t)$ is measurable with respect to $\mathcal{H}(t)$ for any $t \in T$. Similarly, the process Y is *non-anticipative* with respect to X if the random variable $Y(t)$ is independent of $\mathcal{H}^\perp(t)$ for any $t \in T$. It is clear that if Y is adapted to X then Y is non-anticipative with respect to X, but the converse does not hold in general.

Actually, we do not need a process X to define the *innovation*, if we start from a filtration $\{\mathcal{F}(t) : t \in T\}$ we can define its innovation or *independent complement* $\{\mathcal{F}^\perp(t) : t \in T\}$, and then we can say that a process X is either adapted or non-anticipative with respect to the filtration $\{\mathcal{F}(t) : t \in T\}$.

At this point, the Markov property (1.35) can be re-phrased as

$$P(X_t \in B \mid \mathcal{H}_s) = P(X_t \in B \mid X_s), \quad \text{a.s.} \quad \forall t > s,\ B \in \mathcal{B}(E),$$

where $\mathcal{H}_t = \mathcal{H}(t) = \mathcal{H}_X(t)$. The Markov property identifies only how the finite-dimensional distributions of the process interact themselves or evolve in time.

Definition 1.27 (Markov)**.** A Markov process with states in $E \subset \mathbb{R}^d$ is a (complete) probability measure P on $(\Omega, \mathcal{F})$, together with a measurable mapping X (P-equivalence class) from $(\Omega, \mathcal{F})$ into $(E^T, \mathcal{B}^T(E))$ and an increasing family of completed σ-algebras $(\mathcal{F}_t : t \in T)$ on $(\Omega, \mathcal{F})$ satisfying the *Markov property*

$$P(X_t \in B \mid \mathcal{F}_s) = P(X_t \in B \mid X_s), \quad \text{a.s.} \quad \forall t > s,\ B \in \mathcal{B}(E).$$

If the family of σ-algebras $(\mathcal{F}_t : t \in T)$ is not mentioned, then it is assumed that $(\mathcal{F}_t : t \in T)$ is the history $(\mathcal{H}(t) : t \in T)$ of the process X. Moreover, if $(\mathcal{F}_t : t \in T)$ is a filtration satisfying the usual conditions and the paths of $(X_t : t \in T)$ are cad-lag, except in a set of P-probability zero, then $(P, X_t : t \in T)$ is called a *cad-lag* Markov process. □

As mentioned early, we are concerned with E-valued Markov processes where $E \subset \mathbb{R}^d$, and because cad-lag is usually assumed, the sample space Ω will be a

Polish (separable, complete and metric) space as discussed later. However, the above definition is meaningful when E is a Polish, and even when E is only a Lusin space (homeomorphic to a Borel subset of a compact metric space).

In the above Markov property, assuming we have taken a regular conditional probability, the equality is true except on a set of probability zero which may depend on t, s. Thus some regularity is necessary on path of the process to completely identify the process in term of its finite-dimensional distributions. In order to avoid extra difficulties, we consider only cad-lag Markov processes, where the Markov property is satisfied with a complete and right-continuous increasing family of σ-algebras and the path of the process may have only discontinuities of first kind, which are normalized to be cad-lag. The larger the σ-algebras of the filtration $(\mathcal{F}_t : t \in T)$ are, the more significant is the assertion that $(P, X_t, \mathcal{F}_t : t \in T)$ has the Markov property. Thus, the process $(X_t : t \in T)$ is adapted to $(\mathcal{F}_t : t \in T)$ and the filtration $(\mathcal{F}_t : t \in T)$ is non-anticipative i.e., $\mathcal{F}_t$ is independent of $\mathcal{H}^{\perp}(t)$ for any t in T. Note that the Markov property can be re-phased as follows: for every time t the σ-algebra $\mathcal{F}_t$ is independent of $\sigma(X_s : s \geq t)$ given X_t.

In most of the literature, the word *standard* Markov processes refer to cad-lag Markov processes which are also *quasi-left-continuous* and satisfy the strong Markov property (as discussed later). It will become clear that the strong Markov property is highly desired, however, some applications involving *deterministic impulses* yield cad-lag Markov processes which are not quasi-left-continuous.

Usually, when talking about a Markov process we do not refer to a single process, we really mean a family of processes satisfying the Markov property and some given initial distribution. The following concept of transition function is then relevant if we can explicitly write

$$P\{X_t \in A \,|\, X_s = x\} = P(s, x, t, A), \quad \forall s < t,\ x \in E,\ A \in \mathcal{B}(E),$$

for some function $P(s, x, t, A)$. Note that

$$P\{X_t \in A \,|\, X_s = x\} = \frac{P(\{X_t \in A,\ X_s = x\})}{P(\{X_s = x\})} = \\ = \frac{1}{P(\{X_s = x\})} \int_{\{X_s = x\}} P\{X_t \in A \,|\, X_s\}(\omega) P(\mathrm{d}\omega),$$

whenever $P(\{X_s = x\}) > 0$ and $P\{X_t \in A \,|\, X_s = x\} = 0$ when $P(\{X_s = x\}) = 0$, under the condition that a regular conditional probability exists.

Definition 1.28 (transition)**.** A *transition probability function* on a given measurable space $(E, \mathcal{F})$, is a function $P(s, x, t, A)$ defined for $s < t$ in T (T is equal to $[0, +\infty)$ or $(-\infty, +\infty)$ in most of our cases), x in E and A in $\mathcal{F}$ such that

(a) for each $s < t$ in T and x in E the function $A \mapsto P(s, x, t, A)$ is a probability measure on $(E, \mathcal{F})$,

(b) for each $s < t$ in T and A in $\mathcal{F}$ the function $x \mapsto P(s, x, t, A)$ is a measurable,

(c) for any s in T, x in E and A in $\mathcal{F}$ we have

$$\lim_{t\to s} P(s,x,t,A) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise,

(d) for each $s < r < t$ in T, x in E and A in $\mathcal{F}$ we have

$$P(s,x,t,A) = \int_E P(s,x,r,\mathrm{d}y)\, P(r,y,t,A),$$

which is referred to as the Chapman-Kolmogorov identity. It is called *homogeneous* if $P(s,x,t,A) = P(0,x,t-s,A)$ for any $t > s$ in $T = [0,+\infty)$ (or $T = \{0,1,2,\dots\}$), x in E and any Borel measurable subset A of E, in this case we will denote $P(0,x,r,A)$ by $P(r,x,A)$. In most of the cases, the space E is a Polish space and $\mathcal{F} = \mathcal{B}(E)$, its Borel σ-algebra. We say that $P(s,x,t,A)$ is a *Feller* transition probability function

(e) if the function $(s,x) \mapsto P(s,x,t,f)$, with

$$P(s,x,t,f) := \int_E f(y) P(s,x,t,\mathrm{d}y),$$

is continuous from $[0,t] \times E$ into $\mathbb{R}$, for any fixed t in $(0,\infty)$ and any bounded continuous function f from E into $\mathbb{R}$. □

Note that conditions (a) and (b) are natural weak regularity assumptions, the limit in (c) is a more restrictive (but necessary) initial condition, and the Chapman-Kolmogorov identity (see Exercise 1.22) follows from the Markov property in Definition 1.27. Usually, when the space E is locally compact Polish space and $T = [0,\infty)$, we replace the condition (c) by a stronger assumption, namely, for any compact subset K of E, any s in $[0,\infty)$ and any $\varepsilon > 0$ we have

$$\begin{cases} \text{(a)} \ \lim\limits_{t\to s} \sup\limits_{x\in K} [1 - P(s,x,t,B(x,\varepsilon))] = 0, \\ \text{(b)} \ \lim\limits_{x\to\infty} \sup\limits_{0\le s<t\le 1/\varepsilon} P(s,x,t,K) = 0, \end{cases} \tag{1.36}$$

where $B(x,\varepsilon)$ is the ball of radius ε and center x, and neighborhood of ∞ are of the form $E \setminus K$ for some compact K of E. In (1.36), the first condition (a) is referred to as local *uniformly stochastic continuity* property, while condition (b) is only necessary when E is not compact. Note that by adding one dimension to the space E, e.g., replacing E by $E \times T$, we can always assume that the transition is homogeneous (see Exercise 1.23).

Theorem 1.29 (strong Markov)**.** *Let $(P, X_t, \mathcal{F}_t : t \in T)$ be a Markov process on a Polish space E with cad-lag paths and homogeneous transition function $P(t,x,A)$. If either $P(t,x,A)$ is a Feller transition, i.e., condition (e) holds, or at least the process $s \mapsto P(t,X_s,f)$ is cad-lag for every $t > 0$ and any bounded continuous function f, then (1) $(P, X_t, \mathcal{F}_t^+ : t \in T)$ is a Markov process with transition $P(t,x,A)$, where $\mathcal{F}_t^+ = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$, and (2) $\mathcal{F}_t^+ = \mathcal{F}_t$, for every $t \ge 0$,*

when ever $\mathcal{F}_t$ is the σ-algebra generated by the null sets and $\{X_s : s \le t\}$. Moreover, if the cad-lag Markov process $(P, X_t, \mathcal{F}_t : t \in T)$ exists for every initial condition $X_0 = x$, any x in E, and the transition function is Feller or at least the process $s \mapsto P(t, X_s, f)$ is cad-lag for every $t > 0$ and any initial condition $X_0 = x$, then $(P, X_t, \mathcal{F}_t^+ : t \in T)$ is a strong Markov process, i.e., for any τ optional (or stopping) time, any $t > 0$, and every Borel measurable set A in E,

$$P(X(\tau + t) \in A \mid \mathcal{F}_\tau^+) = P(t, X(\tau), A), \quad a.s. \text{ on } \{\tau < \infty\},$$

where $\mathcal{F}_\tau^+$ is the σ-algebra generated by the optional time τ, relative to $\{\mathcal{F}_t^+ : t \ge 0\}$. □

A proof of the above *strong Markov* property can be founded in Doob [60, Theorems 8 and 9, pp. 556-560] or in Blumental and Getoor [28, Chapter 1, Theorem 8.1, pp. 41-42], where almost surely right continuous (instead of cad-lag) processes is only assumed. Moreover, adding another coordinate to include time as a new state variable, this result is extended to non-homogenous Markov processes with almost no changes. Indeed, if $P(s, x, t, \mathrm{d}\xi)$ is a non-homogeneous transition probability function then $\dot{P}(\dot{x}, t, \mathrm{d}\dot{\xi}) = P(s, x, \tau, \mathrm{d}\xi)\delta(t - \tau)\mathrm{d}\tau$ is a homogeneous transition probability function associated to the Markov process $\dot{X}(t) = (t - s, X(t - s))$ with initial condition $\dot{X}(s) = (s, X(0))$, where $\delta(t-\tau)\mathrm{d}\tau$ is the Dirac measure at $\{t\}$, $\dot{x} = (s, x)$, $\dot{\xi} = (\tau, \xi)$ and $\dot{E} = [0, \infty) \times E$, and the associated Markov process $\dot{X}(t) = (t - s, X(t - s))$ with initial condition $\dot{X}(s) = (s, X(0))$.

In most of the cases, the Markov process takes values in a locally compact metric space E endowed with its Borel σ-algebra. Using the fact that Radon measures can be regarded as linear continuous functions on the space of continuous functions with compact support, the properties in the Definition 1.28 of transition function including condition (e) and (1.36) can be rephrased as a family of linear operators $P(t, s) : C_0(E) \longrightarrow C_0(E)$, where $C_0(E)$ is the space of continuous functions vanishing at infinity (i.e., for any $\varepsilon > 0$ there exists a compact subset K of E such that $|\varphi(x)| \le \varepsilon$ for every x in $E \setminus K$), such that

(a) $0 \le P(t, s)\varphi \le 1$, for every φ in $C_0(E)$ with $0 \le \varphi \le 1$

(b) $\lim_{t \to s} P(t, s)\varphi(x) = \varphi(x)$, for any x in E and φ in $C_0(E)$

(c) $P(t, s) = P(t, r)P(r, s)$, for any $s < r < t$.

Thus, if the transition function is homogeneous, i.e., $P(t, s) = P(t - s)$, we have a one-parameter semigroup in $C_0(E)$.

Sometimes, it is convenient to consider processes with values in a enlarged space $\bar{E} = E \cup \{\infty\}$, with an isolated point ∞ (usually, the one-point compactification), and even defined in the whole $[0, \infty]$. In this case, the *lifetime formalism* is used, i.e., define the *lifetime* of a process $X(\cdot)$ as $\varsigma(\omega) := \{t \ge 0 : X(t) = \infty\}$, and assume that $X(t) = \infty$ for every $t \ge \varsigma$. This allow to relax the condition (a) of the definition of transition function, only the fact that $P(s, x, t, \cdot)$ is a measure with total mass not larger than 1 (instead of a probability measure) is actually necessary.

Usually, the transition function $P(s,x,t,A)$ associated with a Markov process X_t is continuous in time (i.e., in s and t) and a standard realization makes X_t a cad-lag process. In this case, an extra property is desirable, the process $X_t = X(t)$ is *quasi-left* continuous, i.e., $X(T_n)$ converges to $X(T)$ on the set where $T < \infty$, for any increasing sequence of stopping time T_n converging to T, with $T_n < T$. It is clear that here the key is fact that T_n are stopping times. In this sense, the process X_t do not have any *deterministic* jumps.

If a stochastic process represents some kind of phenomenological process then it should characterized by its finite-dimensional distributions. Then, a mathematical model is a *realization* of such a process in a convenient sample space, see Section 1.12. However, a Markov process is also characterized by either its transition function or it infinitesimal generator (see next chapter). It is important to recognize that when only one stochastic process (or variable) is involved, it finite-dimensional distributions determine the process in an appropriate sample space (usually refer to as a *realization of the process*), however, when two or more stochastic processes (or variables) are discussed, it is important to know its *joint distribution*. Thus the concept of *version* or *modification* of a process is relevant, i.e., at the end we are always working with stochastic processes as random variables which are *almost surely* equals. Recall that two stochastic processes need not to be defined in the same probability space to have the same finite-dimensional distributions, but they do have the same law, once the sample space has been chosen. However, to be a version (or a modification) one of each other, they do need to be defined in the same probability space. After a sample space has been chosen, the stochastic process are treated as random variable, with values in the sample space. The procedure of selecting a sample space on which a probability is constructed satisfying its characteristic properties (e.g., the finite-dimensional distributions are given, or in the case of a Markov process, its transition function or its infinitesimal generator is given) is called a *realization* of the stochastic process.

The reader may consult the classic books by Blumental and Getoor [28], Dynkin [69, 70] or more recent books, such as, Çınlar [46], Chung [44], Dellacherie and Meyer [58], Ethier and Kurtz [76], Marcus and Rosen [166], Rogers and Williams [214], Taira [245], among many others.

1.10 Construction of Processes

It is clear that a family of finite-dimensional distributions $\{P_s : s \in T^n,\ n = 1,2,\dots\}$, on a Borel (usually open or closed) subset of $\mathbb{R}^d$, derived from a stochastic process form a *projective family*, i.e., they must satisfy a set of (natural) *consistency conditions*, namely

(a) if $s = (s_{i_1},\dots,s_{i_n})$ is a permutation of $t = (t_1,\dots,t_n)$ then for any B_i in $\mathcal{B}(E)$, $i = 1,\dots,n$, we have $P_t(B_1 \times \cdots \times B_n) = P_s(B_{i_1} \times \cdots \times B_{i_n})$,

(b) if $t = (t_1,\dots,t_n,t_{n+1})$ and $s = (t_1,\dots,t_n)$ and B in $\mathcal{B}(E^n)$ then $P_t(B \times E) = P_s(B)$.

If a total order is given on the index set T, it is enough to have the finite-dimensional distributions defined only for $(s_1, s_2, \dots, s_n)$ such that $s_1 < s_2 < \cdots < s_n$ and to satisfy only a consistency conditions of the form

(b') if $t = (t_1, \dots, t_n)$ and $s = (s_1, \dots, s_m)$ with $t_1 < \cdots < t_n < r < s_1 < \cdots < s_m$ and $A \times B$ in $\mathcal{B}(E^n) \times \mathcal{B}(E^m)$ then $P_{(t,r,s)}(A \times E \times B) = P_{(t,s)}(A \times B)$, for any $n, m = 0, 1, \dots$.

The converse of this assertion is given by the following classic Kolmogorov (sometime called Daniell-Kolmogorov or Čentsov-Kolmogorov) construction or the *coordinate method* of constructing a process (see Kallenberg [121], Karatzas and Shreve [124], Malliavin [163], Revuz and Yor [212], among others, for a comprehensive treatment).

Theorem 1.30 (Kolmogorov)**.** *Let $\{P_s : s \in T^n,\ n = 1, 2, \dots\}$ be a consistent family of finite-dimensional distributions on a Borel subset E of $\mathbb{R}^d$. Then there exists a probability measure P on $(E^T, \mathcal{B}^T(E))$ such that the canonical process $X_t(\omega) = \omega(t)$ has $\{P_s\}$ as its finite-dimensional distributions.* □

Under the consistency conditions, an additive function can be easily defined on product space $(E^T, \mathcal{B}^T(E))$, the question is to prove its σ-additive property. In this respect, we point out that one of the key conditions is the fact that the (Lebesgue) measure on the *state* space $(E, \mathcal{B}(E))$ is *inner regular* (see Doob [60, pp. 403, 777]). Actually, the above result remains true if E is a Lusin space, i.e., E is homeomorphic to a Borel subset of a compact metrizable space. Note that a Polish space is homeomorphic to a countable intersection of open sets of a compact metric space and that every probability measure in a Lusin space is inner regular, see Rogers and Williams [214, Chapter 2, Sections 3 and 6].

We recall that a cylinder set is a subset C of E^T such that ω belongs to C if and only if there exist an integer n, an n-uple $(t_1, t_2, \dots, t_n)$ and $B \in \mathcal{B}(E^n)$ such that $(\omega(t_1), \omega(t_2), \dots, \omega(t_n))$ belongs to B for any $i = 1, \dots, n$. The class of cylinder sets with $t_1, \dots, t_n$ fixed is equivalent to product σ-algebra in $E^{\{t_1,\dots,t_n\}} \simeq E^n$ and referred to as a finite-dimensional projection. However, unless T is a finite set, the class of all cylinder sets is only an algebra. Based on cylinder sets, another way of re-phrasing the Kolmogorov's construction theorem is saying that any (additive) set function defined on the algebra of cylinder sets such that any finite-dimensional projection is a probability measure, has a unique extension to a probability measure on E^T. In particular, if $T = \{1, 2, \dots\}$ then the above Kolmogorov's theorem shows the construction of an independent sequence of random variables with a prescribed distribution. In general, this is a realization of processes where the distribution at each time is given.

Note that a set of only one element $\{a\}$ is closed for the product topology of E^T and so it belongs to the Borel σ-algebra $\mathcal{B}(E^T)$ (generated by the product topology in E^T). However, the product σ-algebra $\mathcal{B}^T(E)$ (generated by cylinder sets) contains only sets that can be described by a countable number of restrictions on E, so that $\{a\}$ is not in $\mathcal{B}^T(E)$ if T is uncountable. Thus we see the importance of finding a subset Ω of E^T having the full measure under the outer measure P^* derived from P, which is itself a topological space. For instance,

combining the last two theorems, if a consistent family of distributions satisfies the a priori estimate (1.11), then we can redefine the probability measure P of the above theorem on the space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, of continuous functions.

There is a very important class of stochastic processes so-called *Lévy processes* which provides prototype examples, we refer to Bertoin [21] and Sato [220] for a comprehensive study.

Definition 1.31 (Lévy process)**.** An E-valued ($E \subset \mathbb{R}^d$) *Lévy (stochastic) process* is a couple (P_X, X) as in Definition 1.23 satisfying:

(a) its *increments are independent of the past* i.e., for any $t > s \geq 0$ the random variable $X_t - X_s$ is independent of X_s under P_X,

(b) it has *stationary increments* i.e., for any $t > s \geq 0$ the random variable $X_t - X_s$ has the same distribution as X_{t-s} under P_X,

(c) its paths are *continuous in probability* (referred to as *stochastically continuous*) i.e., for any $\varepsilon > 0$ and $s \geq 0$ we have

$$\lim_{t \to s} P(|X_t - X_s| \geq \varepsilon) = 0.$$

Usually the condition $P_X(X_0 = 0) = 0$ is added to normalize the process. □

Note that a process with independent increments, property (a), has its finite-dimensional distributions completely determined by the distributions of $(X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_k} - X_{t_{k-1}})$, for any $t_0 \leq t_1 \leq \cdots \leq t_k$ in $[0, +\infty)$. Moreover, if the process has stationary increments, property (b), then the distribution of the random variables $\{X_t : t \geq 0\}$ in E completely characterize its finite-dimensional distributions. Adding the continuity condition (c), the Lévy process is identified. Now, instead of looking at the distributions of X_t under P_X we may consider its *characteristic functions*, which is its Fourier transform i.e., for any $t \geq 0$

$$\varphi_t(\xi) = \mathbb{E}\{\mathrm{e}^{\mathrm{i}(\xi, X_t)}\} := \int_\Omega \mathrm{e}^{\mathrm{i}(\xi, X_t(\omega))} P_X(\mathrm{d}\omega),$$

where $(\cdot, \cdot)$ is the *dot* product in $\mathbb{R}^d$ and i is the imaginary unit.

It is not so hard (see Exercise 1.25) to check that any characteristic function $\varphi(\xi)$, $\xi \in \mathbb{R}^d$ has the following properties:

(a) $\varphi(\cdot)$ is continuous from $\mathbb{R}^d$ into $\mathbb{C}$ and $\varphi(0) = 1$,

(b) $\varphi(\cdot)$ is *positive definite* i.e., for any $k = 1, 2, \dots, \zeta_1, \dots, \zeta_k$ in $\mathbb{C}$ and $\xi_1, \dots, \xi_k$ in $\mathbb{R}^d$ we have

$$\sum_{i,j=1}^{k} \varphi(\xi_i - \xi_j) \zeta_i \bar{\zeta}_j \geq 0,$$

we refer to Shiryayev [227, Section 2.12, pp. 272–294] for a more detail. So, this is satisfied by the family of characteristic functions $\{\varphi_t : t \geq 0\}$. Moreover, since the Lévy process (P_X, X) has independent and stationary increments, see

properties (a) and (b), the one-parameter functions φ_t satisfies $\varphi_{t+s} = \varphi_t\varphi_s$, for any $t \geq s \geq 0$. From the property (c) follows that the function $t \mapsto \varphi_t(\xi)$ is continuous for any $\xi \in \mathbb{R}^d$ from $[0,\infty)$ into $\mathbb{C}$ (see Exercise 1.26). Thus the characteristic function are of the form $\varphi_t(\xi) = \exp[-t\psi(\xi)]$, for some non-negative and continuous function ψ with $\psi(0) = 0$. Actually the Lévy-Khintchine formula (see Protter [206, p. 32, Theorem 43] give a simple expression for the one-dimensional case

$$\psi(\xi) = \frac{\sigma^2}{2}\xi^2 - \mathtt{i}\alpha\xi + \int_{|x|\geq 1}(1 - \mathrm{e}^{i\xi x})\nu(\mathrm{d}x) + \int_{|x|<1}(1 - \mathrm{e}^{\mathtt{i}\xi x} + i\xi x)\nu(\mathrm{d}x),$$

where ν (called Lévy measure) is a Radon measure on $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ satisfying

$$\nu(\{|x| \geq 1\}) + \int_{|x|<1} |x|^2\nu(\mathrm{d}x) < +\infty,$$

and the constants σ, α are the characteristic parameters of the process i.e., uniquely determine the Lévy process. This generalizes to $\mathbb{R}^d$, where now σ^2 is a positive semi-definite quadratic $d \times d$ matrix, α, x and ξ are in $\mathbb{R}^d$ and the dot product is used, see Bertoin [21, Theorem I.1, pp. 13–15]. Moreover, the canonical filtration $\{\mathcal{F}(t) : t \geq 0\}$ associated with a Lévy process (P_X, X) is right-continuous, i.e. it satisfies the usual conditions, see Definition 1.7.

The converse can be established as follows. First, by means of the classic Bochner theorem (e.g. see Pallu de la Barrière [193, p. 157, Theorem 1] that for any non-negative and continuous function ψ with $\psi(0) = 0$ such that $\xi \mapsto \exp[-t\psi(\xi)]$ is positive definite for any $t \geq 0$, is indeed the characteristic function of some distribution), there exist a family of probability distribution $P(t, \mathrm{d}x)$ for any $t \geq 0$ whose characteristic function is precisely $\exp[-t\psi(\xi)]$. Now, based on the properties (a) and (b) of a Lévy process, the finite-dimensional distributions are determined by

$$\begin{aligned} &P(X(t_1) \in B_1\,,\, X(t_2) \in B_2\,,\, \dots\,,\, X(t_n) \in B_n) = \\ &\quad = P(X(t_1) \in B_1\,,\, X(t_i) - X(t_{i-1}) \in B_i - B_{i-1}\,,\, 2 \leq i \leq n) = \\ &\quad = P(t_1, B_1)P(t_2 - t_1, B_2 - B_1)\cdots P(t_n - t_{n-1}, B_n - B_{n-1})\,, \end{aligned}$$

for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$. Hence, Kolmogorov construction may be applied. The fact that the family of characteristic functions $\exp[-t\psi(\xi)]$ are continuous in t, locally uniformly in ξ implies that the process is continuous in probability i.e., condition (c) is satisfied. Therefore, if ν, σ and α are given then there exists a unique Lévy process with the prescribed characteristic parameters. Moreover, as proved in Protter [206, p. 21, Theorem 30] any Lévy process has a version which is *cad-lag* i.e., the paths $t \mapsto X_t(\omega)$ are right-continuous functions having left-hand limits from $[0,\infty)$ into E for any $\omega \in \Omega \setminus N$ with $P_X(N) = 0$. In other words, the Kolmogorov construction is valid on the space of cad-lag functions.

The transition function $P(t, x, A)$ of a Lévy process has the form of *convolutions semigroup*, i.e., a family of probability measures on $\mathbb{R}^d$ such that

(a) $\nu_s * \nu_t = \nu_{s+t}$ for any s, t

(b) $\nu_t \to \delta_0$ in the weak topology,

and

$$P(t,x,A) := \int_{\mathbb{R}^d} \mathbb{1}_A(x+y)\nu_t(\mathrm{d}y),$$

see Revuz and Yor [212, Propositon 2.18, pp. 96-97].

As seen in the classic Kolmogorov construction, the finite-dimensional distributions characterize the statistics of a process, but not its sample properties. If we are only interested in the statistics of a process, we may adopt another viewpoint to identify a process. Instead of given its finite-dimensional distributions, we may prescribe some time-evolution or dynamics.

Suppose that a transition (probability) function $P(s,x,t,A)$ (see Definition 1.28) and an initial (i.e. for $t = 0$) distribution ν are given. Then we can define an *absolute probability function* by

$$P_0(t,A) = \int P(0,x,t,A)\,\nu(\mathrm{d}x), \quad \forall A \in \mathcal{F}$$

and a family of finite-dimensional distributions for $t_1 < \cdots < t_n$ as follows

$$\begin{aligned} &P_{t_1,\ldots,t_n}(A_1 \times A_2 \times \cdots \times A_n) = \\ &\quad = \int_{A_1} P_0(t_1,\mathrm{d}x_1) \int_{A_2} P(t_1,x_1,t_2,\mathrm{d}x_2) \cdots \int_{A_n} P(t_{n-1},x_{n-1},t_n,\mathrm{d}x_n), \end{aligned}$$

for any $A_1, A_2, \ldots, A_n$ in $\mathcal{F}$. It is not hard to check (see Exercise 1.27) that the consistency condition (b') is satisfied.

Thus for a given transition function on $(E,\mathcal{F})$, with $E \subset \mathbb{R}^d$, we can use Kolmogorov construction to get a canonical process (i.e. a probability measure) on $(E^T, \mathcal{B}(E^T))$. Moreover, if the continuity condition (1.11) is satisfied, which is verifiable in term of the transition function and the initial probability i.e.,

$$\int_E |x-y|^\alpha P(s,x,t,\mathrm{d}y) \le C|t-s|^{1+\beta}, \quad \forall s,t \in [0,T],\ \forall x \in E,$$

for some positive constants α, β and C, then the construction can be performed over the space $C(T,E)$, with the local uniform convergence topology and the induced Borel σ-algebra. Similarly, the cad-lag condition (1.12) becomes

$$\int_E |x-y|^\alpha P(s,x,t,\mathrm{d}y) \le C|t-s|^{\frac{1}{2}+\beta}, \quad \forall s,t \in [0,T],\ \forall x \in E,$$

after some simplification (see Exercise 1.28).

If we denote by (P_X, X) the process constructed as above on the product space $(E^T, \mathcal{B}(E^T))$, with its canonical filtration $\{\mathcal{F}(t) : t \ge 0\}$, then we have the following property (derived from the Chapman-Kolmogorov identity)

$$P_X(X(t) \in \cdot \mid X(s)) = P(s,X(s),t,\cdot), \quad \forall t \ge s \tag{1.37}$$

So that $P(s,x,t,A)$ represents the conditional probability that the state of the process belongs to A at time t knowing that initially it was x at time s. This is referred to as the *Markov property*.

We only state now an useful result relative to the construction of standard Markov process, actually the so-called Markov-Feller processes, There are several way of construct a *good* Markov process from a given Markov (or sub-Markov) transition as in Definition 1.28 conditions (a), (b), (c) and (d). For instance we refer to Blumental and Getoor [28, Chapter 1, Theorem 9.4, pp. 46-51], Dynkin [70, Chapter III, Theorems 3.5, 3.6, pp. 81-92], Ethier and Kurtz [76, Chapter 4, Theorem 1.1, pp. 157-169], among others. We summarize these results in a non-homogeneous form as follows:

Theorem 1.32 (realization)**.** *Let $P(s,x,t,\mathrm{d}y)$ be a transition probability function satisfying properties* (a), (b), (c), (d) *of Definition 1.28 on a Polish space E. Then, for any given probability measure ν on E and any initial time $s \geq 0$, there exist a unique Markov process $(P, X_t, \mathcal{F}_t; t \geq s)$ satisfying condition (1.37) and such that X_s has ν as its distribution. If the transition probability function either* (1) *is Feller, i.e., the function $(s,x) \mapsto P(s,x,t,f)$ is continuous from $[0,t] \times E$ into $\mathbb{R}$ for any t in $(0,\infty)$ and any bounded continuous function f from E into $\mathbb{R}$, or* (2) *satisfies condition (1.36), then the Markov process $(P, X_t, \mathcal{F}_t; t \geq s)$ has cad-lag paths and right-continuous filtration. Moreover, if the function $(x,t) \mapsto P(s,x,t,f)$ is continuous from $[s,\infty) \times E$ into $\mathbb{R}$, for any s in $[0,\infty)$ and any bounded continuous function f from E into $\mathbb{R}$, then $(X_t, \mathcal{F}_t : t \geq s)$ is quasi-left-continuous, i.e., for any increasing sequence of stopping time $\{\tau_n : n = 1,2,\ldots\}$ almost surely strictly convergent to τ, $P(\tau_n \leq \tau_{n+1} < \tau < \infty, \tau > s) = 1$, then X_{τ_n} converges to X_τ a.s.* (or equivalently the σ-algebra $\mathcal{F}_\tau$ is the minimal σ-algebra containing the sequence of σ-algebra $\{\mathcal{F}_{\tau_n} : n = 1,2,\ldots\}$). Furthermore, if

$$\lim_{t \to s} \frac{1}{t-s} \sup_{x \in K} \{1 - P(s,x,t,B(x,\varepsilon))\} = 0,$$

for every $s \geq 0$ and x in E, then the Markov process $(P, X_t, \mathcal{F}_t; t \geq 0)$ has continuous paths. □

There are several ways of expressing the continuity condition (Feller's property) assumed in the previous theorem. Sometime, a transition probability $P(s,x,t,\mathrm{d}y)$ is called *stochastically continuous* if it satisfies

$$\lim_{t-s \to 0} P(s,x,t,\{y \in E : |y-x| < \delta\}) = 1, \quad \forall x \in E,$$

see Dynkin [70, Chapter 2]. This is equivalent to the continuity in probability of a realization of the Markov process and the Feller character used above. Moreover, a Markov process satisfying the above regularity on its transition function is called a *Markov-Feller process*. Actually, with the aid of the sample space $D([0,\infty),E)$ discussed later, we see that the so-called *realization* of the Markov-Feller process given in the previous theorem can be regarded as a probability measure P on $D([0,\infty),E)$, with $X_t(\omega) = \omega(t)$ the canonical process and $(\mathcal{F}_t : t \geq 0)$ its canonical filtration. Moreover, it satisfies the strong Markov property. At this point, a reading of Chapter 4 in Marcus and Rosen [166] could be of some help.

We may look at stochastic processes with complex values. In this context, we have the so-called second-order process, which are well adapted to dynamical systems.

Definition 1.33 (second-order process)**.** A (complete) probability measure P_X on $(\Omega, \mathcal{F})$, where $\Omega \subset \mathbb{C}^T$ and $\Omega \bigcap \mathcal{B}(\mathbb{C}^T) \subset \mathcal{F}$, together with a measurable mapping X from Ω into itself satisfying:

(a) $\mathbb{E}_X\{|X_t|^2\} < \infty$, for all t in T,

(b) $t \mapsto X_t$ from T into $L^2(P_X)$ is continuous

will be called a mean square continuous second-order process (P_X, X). □

This definition is independent of the sample path of the process, i.e., any version of the process should satisfy the conditions (a) and (b) above. Thus, as the case of processes continuous in probability, second-order processes cannot be regarded as random variables with values in some appropriate (sample) function space.

For a second-order process we can define the *mean* and *covariance* function:

$$\mu(t) := \mathbb{E}_X\{X_t\} \qquad R(t,s) := \mathbb{E}_X\{(X_t - \mu(t))\overline{(X_s - \mu(s))}\},$$

where the *over-bar* denotes the complex conjugate. Several properties can be discussed for this class of processes. For instance, we can mention that a second-order process is continuous in mean square if and only if its covariance function $R(t,s)$ is continuous. On the other hand, any continuous second-order process (P_X, X) with $T = [a,b]$ has a version of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\varphi_n(t)y_n(\omega),$$

where the convergence is in $L^2(P_X)$, uniformly for t in $[a,b]$, and $\{\varphi_n : n = 1,2,\dots\}$ are the orthonormal eigenfunctions and $\{\lambda_n : n = 1,2,\dots\}$ are the eigenvalues of

$$\int_a^b R(t,s)\varphi(s)\mathrm{d}s = \lambda\varphi(t), \qquad a \le t \le b$$

in $L^2(]a,b[)$ and $\{y_n : n = 1,2,\dots\}$ is an orthonormal system in $L^2(P_X)$ satisfying

$$y_x(\omega) = \sqrt{\lambda_n}\int_a^b \varphi_n(t)X_t(\omega)\mathrm{d}t.$$

We refer to Shiryayev [227, Section 2.10, pp. 260–272] for the discrete case, and to Wong [255, Chapter 3, pp. 74–138] for more details.

Related with the second-order processes are the so-called *processes with orthogonal increment*, which are defined as a (complete) probability measure P_X on $(\Omega, \mathcal{F})$, where $\Omega \subset \mathbb{C}^T$ and $\Omega \bigcap \mathcal{B}(\mathbb{C}^T) \subset \mathcal{F}$, together with a measurable mapping X from Ω into itself satisfying:

(a) $\mathbb{E}_X\{|X_t - X_s|^2\} < \infty$, for all t, s in T,

(b) $\mathbb{E}_X\{X_t\, X_s\} = 0$, for all t, s in T,

It can be proved that to each such a process corresponds a monotone non-decreasing functions F_X, uniquely determined up to an additive constant, satisfying

$$\mathbb{E}_X\{|X_t - X_s|^2\} = F_X(t) - F_X(s), \quad \forall t > s.$$

Moreover, the right-hand $X(t+)$ and the left-hand $X(t-)$ limits exists in the $L^2(P_X)$ sense, for every t in T, and $X(t-) = X(t) = X(t+)$, except for a countable set in T. Thus, processes with orthogonal increments can be normalized to be cad-lag in mean square sense. We refer to Doob [59, Chapter IX, pp. 425–451].

1.11 Examples of Markov processes

Two typical examples can be given, (a) the standard Wiener process (or Brownian motion) $(w(t), t \geq 0)$ and (b) the standard Poisson process $(p(t), t \geq 0)$. Both are the prototype of Lévy processes (see Definition 1.31). For the Wiener process, we take the Gauss kernel

$$p(t,x) = (2\pi t)^{-d/2} \exp(-\frac{|x|^2}{2t}), \quad t > 0,\ x \in \mathbb{R}^d \tag{1.38}$$

and consider the transition function

$$P_w(s,x,t,A) = \int_A p(t-s, y-x)\mathrm{d}y, \quad \forall t > s,\ A \in \mathcal{B}(\mathbb{R}^d),$$

the initial probability P_0 as the Dirac measure. Kolmogorov construction and path regularity apply in this case to generate a probability measure P_w in the space $C([0,+\infty), \mathbb{R}^d)$, called *Wiener measure*. Under P_w, the canonical process is a standard Wiener process. Certainly there several ways to construct a Brownian motion and a critical point is to show continuity of its paths. In general, a one-dimension standard Brownian motion is defined as a real valued stochastic process $\{B(t) : t \geq 0\}$ satisfying:

(a) $B(0) = 0$ and for $0 \leq s < t < \infty$, the difference $B(t) - B(s)$ is a normally distributed random variable with mean zero and variance $t - s$, i.e., for every α in $\mathbb{R}$,

$$P\big(B(t) - B(s) > \alpha\big) = \int_\alpha^{+\infty} \big(2\pi(t-s)\big)^{-1/2} \exp\big(-\frac{|x|^2}{2(t-s)}\big)\mathrm{d}x,$$

(b) for $0 \leq t_0 < t_1 < \cdots < t_n$, the family $\{B(t_0), B(t_k) - B(t_{k-1}),\ k = 1, \ldots, n\}$ is a set of independent random variables.

The above two properties characterize a *Brownian motion* (or Wiener process) as an indistinguishable stochastic process. Because it is well know that (e.g.

Chung [44, p. 145]) any Brownian motion has a version with continuous path (see 1.29), we work always with a continuous version, actually with the Wiener measure. A d-dimensional version is constructed by taking d independent one-dimensional Brownian motions. Note that the characteristic function is given by

$$\mathbb{E}\{\exp \mathtt{i}(\xi, B(t))\} = \exp\Big(-t\frac{|\xi|^2}{2}\Big), \qquad \forall t \geq 0,$$

with $\mathtt{i} = \sqrt{-1}$. There are many properties on the paths of a Wiener process $\{w(t) : t \geq 0\}$, we mention some of them (see Itô and McKean [113], Krylov [141, pp. 36-38])

(a) for any c non zero constant, $w(t+c) - w(c)$, $c^{-1}w(c^2t)$, $\forall t \geq 0$ and $tw(1/t)$, $\forall t > 0$ (and $= 0$ for $t = 0$) are also Wiener process;

(b) for any constants α in $(0, 1/2)$ and T in $(0, \infty)$ there exists a random variable $C(\omega)$ such that $\mathbb{E}\{|C|^p\} < \infty$ for any positive and finite p and

$$|w(t,\omega) - w(s,\omega)| \leq C(\omega)|t-s|^\alpha, \quad \forall t, s \in [0,T],\ \omega \in \Omega.$$

In particular, $|w(t,\omega)| \leq C(\omega)t^\alpha$ for $t \in [0,T]$ and $|w(t,\omega)| \leq C(\omega)t^{1-\alpha}$ for $t > T$.

(c) for any ω in $\Omega \setminus N$, with $P(N) = 0$ we have

$$\limsup_{s\downarrow 0} \frac{|w(t+s) - w(t)|}{\sqrt{-2s\ln s}} = 1, \quad \forall t > 0,$$
$$\limsup_{t\downarrow 0} \frac{w(t)}{\sqrt{2t\ln(-\ln t)}} = 1, \qquad \liminf_{t\downarrow 0} \frac{w(t)}{\sqrt{2t\ln(-\ln t)}} = -1,$$

which are referred as the *law of the iterated logarithm.*

(d) let $\{t_{0,n}, \dots, t_{k,n} : n = 1, 2, \dots\}$ be a sequence of partition of $[a,b] \subset [0,\infty)$ with mesh going to zero, then

$$\lim_{n\to\infty} \sum_{i=1}^{k} [w(t_i) - w(t_{i-1})]^2 = b - a, \quad a.s.$$

As a consequence, almost surely the sample paths $t \mapsto w(t,\omega)$ of a standard Wiener process are of unbounded variation on any interval. It is possible to construct a Wiener process based on a complete system of orthonormal function in L^2. For instance (Krylov [141, pp. 32]), let $\{\xi_n : n = 0, 1, \dots\}$ be a sequence of independent (standard) normal distributed random variables. Define the process

$$w_t^k(\omega) = \frac{1}{\sqrt{\pi}} t\,\xi_0(\omega) + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{N(k)} \xi_n(\omega) \frac{1}{n} \sin(nt).$$

Then there exists increasing sequence of positive integers $\{N(k) : k = 1, 2, \dots\}$ such that the sequence of processes w_t^k converges uniformly for $t \in [0, \pi]$ and

$\omega \in \Omega \setminus N$, $P(N) = 0$ to a Wiener process. Another typical expression is the locally uniformly convergent series

$$w_t(\omega) = \frac{\sqrt{2}}{\pi} \sum_{n=0}^{\infty} \xi_n \frac{\sin\big(t\pi(n+1/2)\big)}{n+1/2},$$

e.g., see Knight [133, Chapter 1].

Related to a Wiener process (or Brownian motion) is a Brownian motion with drift, defined by $x(t) := \sigma w(t) + bt$, for some constants σ and b, where each random variable $x(t)$ has a Gaussian distribution. Similarly, a *geometric* Wiener process is defined by

$$x(t) := \exp[\sigma w(t) + bt], \quad t \geq 0,$$

where now $x(t)$ has a log-normal distribution. On the other hand,

$$x(t) := w(t) + \sup_{0 \leq s \leq t} \{\max[w(s), 0]\}, \quad t \geq 0$$

defines a *reflected* Brownian motion on $[0, \infty)$, and $x(t) := w(t) - tw(1)$, $0 \leq t \leq 1$ defines a so-called *Brownian bridge*.

For a one-dimensional standard Poisson process with parameter $c > 0$, we define the transition function

$$P_p(s, x, t, A) = \exp\big[-c(t-s)\big] \sum_{k=0}^{\infty} \frac{[c(t-s)]^k}{k!} \mathbb{1}_A(x+k) \tag{1.39}$$

and we apply Kolmogorov construction. Here, the continuity condition (1.11) is not satisfied, but the process is continuous in probability (see property (c) of Definition 1.31 and Exercise 1.30). A Poisson process $\{p(t) : t \geq 0\}$, with parameter $c > 0$, is characterized by the following properties:

(a) $p(0) = 0$ and $0 \leq s < t < \infty$, the difference $p(t) - p(s)$ is a Poisson random variable with mean $c(t-s)$ i.e.,

$$P\big(p(t) - p(s) = n\big) = \big[c(t-s)\big]^n \exp\big[-c(t-s)\big]/n!, \quad n = 0, 1, \ldots;$$

(b) for $0 \leq t_0 < t_1 < \cdots < t_n$, the family $\{p(t_0), p(t_k) - p(t_{k-1}),\ k = 1, \ldots, n\}$ is a set of independent random variables.

Any Poisson process has a version with right continuous (and left hand limits) paths (see Chung [44, Theorem 3, p. 29]). Almost surely, the paths of a Poisson process are constant except for upward jumps of size one, of which there are finitely many in each bounded interval, but infinitely many in $[0, \infty]$. The times between successive jumps are independent and exponentially distributed with parameter c. Thus, if τ_n is the time between the n and the $(n+1)$ jumps, then $P(\tau_n > t) = \exp(-ct)$ for each $t \geq 0$. Actually, based on this last property, a Poisson process can be constructed as follows. Let $\{\tau_n : n = 1, 2, \ldots\}$ be a

sequence of independent exponentially distributed (with parameter c) random variables. Define the *counting process* as

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^{n} \tau_i \le t < \sum_{i=1}^{n+1} \tau_i \end{cases}$$

with values in $\{0, 1, 2, \dots\}$. It can be proved (see Protter [206, Chapter 1]) that $\{p(t) : t \ge 0\}$ is a Poisson process. Its characteristic function is given by

$$\mathbb{E}\{\exp(\mathtt{i}\xi p(t))\} = \exp\big[-tc(1 - \mathrm{e}^{\mathtt{i}\xi})\big], \qquad \forall t \ge 0.$$

Similarly, a *Compound-Poisson* process is given by

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ Z_n & \text{if } \sum_{i=1}^{n} \tau_i \le t < \sum_{i=1}^{n+1} \tau_i, \end{cases}$$

where $Z_n := \zeta_1 + \zeta_2 + \cdots + \zeta_n$, and $\{\zeta_n : n = 1, 2, \dots\}$ is a sequence of independent identically distributed (with distribution law ν) random variables, independent of the sequence $\tau_1, \dots, \tau_n$. Its transition function is given by the expression of the form

$$P_{cp}(s, x, t, A) = \exp\big[-c(t-s)\big] \sum_{k=0}^{\infty} \frac{[c(t-s)]^k}{k!} \nu^k(x + A), \tag{1.40}$$

where $\nu^k = \nu * \nu^{k-1}$ is the k measure-convolutions of ν, i.e., the distribution of the random variable $Z_k = \zeta_1 + \zeta_2 + \cdots + \zeta_k$. In general, if ν is a distribution in $\mathbb{R}^d$ with $\nu(\{0\}) = 0$ then a compound Poisson process $p_{cp}(t)$ in $\mathbb{R}^d$ has the characteristic function

$$\mathbb{E}\{\exp(\mathtt{i}\xi \cdot p_{cp}(t))\} = \exp\big[-tc(1 - \hat{\nu}(\xi))\big], \qquad \forall\, t \ge 0,$$

where $\hat{\nu}$ is the characteristic transform of ν, i.e.,

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^d} \exp(\mathtt{i}\xi \cdot x)\nu(\mathrm{d}x), \qquad \forall\, \xi \in \mathbb{R}^d.$$

In particular, if $d = 1$ and $\nu = \delta_1$ then we get an standard Poisson process as above, e.g., see Sato [220, p.18].

A more general viewpoint is to define a (standard or homogeneous) *Poisson measure* (or Poisson point process) $\{p(t, \cdot) : t \ge 0\}$ with Lévy (characteristic or intensity) measure $m(\cdot)$ by the conditions:

(a) $m(\cdot)$ is a Radon measure on $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$, i.e., $m(B) < \infty$ for any compact subset B of $\mathbb{R}^d_*$;

(b) $\{p(t, B) : t \ge 0\}$ is a Poisson process with parameter $m(B)$, for any compact subset B in $\mathbb{R}^d_*$ (with $p(t, B) := 0$ if $m(B) = 0$);

(c) if $B_1, B_2, \dots, B_n$ are disjoint Borel set in $\mathbb{R}^d_*$ then the Poisson processes $p(\cdot, B_1), p(\cdot, B_2), \dots, p(\cdot, B_n)$ are independent.

Note that the Lévy measure $m(\cdot)$ may have a singularity at the origin (at most) of order 2, i.e.,

$$\int_{\mathbb{R}^d_*} \big(|x|^2 \wedge 1\big) m(\mathrm{d}x) < \infty.$$

Clearly from (b) follows

$$\mathbb{E}\{p(t,B)\} = tm(B), \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*).$$

Also, Poisson measures can be represented by means of Dirac point mass (or atoms), i.e.,

$$p(t,B) = \sum_{0 \leq s \leq t} \mathbb{1}_{e(s) \in B}, \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*),$$

where $\{e(t) : t \geq 0\}$ is a *Poisson point process* with characteristic measure $m(\cdot)$, i.e., $\{e(t) : t \geq 0\}$ is a cad-lag (independent of each other) process taking values in $\mathbb{R}^d$, which is equal to 0 except for a countable number of times and its counting process $p(t,B)$ is a Poisson process. Equivalently, a (random) sequence $\{(e_n, \tau_n) : n \geq 1\}$ in $\mathbb{R}^d_* \times [0,\infty)$ such that its counting process

$$p(t,B) = \sum_n \mathbb{1}_{e_n \in B}\, \mathbb{1}_{0 \leq \tau_n \leq t}, \quad \forall t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*)$$

is a Poisson process. If $m(\cdot)$ is bounded, i.e., $m(\mathbb{R}^d_*) < \infty$ then the following expression

$$P_p(s,x,t,A) = \exp\big[-m(\mathbb{R}^d_*)(t-s)\big] \sum_{k=0}^{\infty} \frac{(t-s)^k}{k!} m^k(x+A),$$
$$m^k(B) = \int_{\mathbb{R}^d_*} m^{k-1}(z+B) m(\mathrm{d}z), \quad k = 1,2,\ldots,$$

gives the corresponding transition function for the Poisson measure process. Comparing with (1.40), we see that compound Poisson processes are particular cases of Poisson measure process. We refer to Bensoussan and Lions [17, Chapter 3, Section 5] Bertoin [21], Bremaud [32], Jacod and Shiryaev [117], Protter [206] and Sato [220] for a systematic discussion of the above statements.

Other less typical processes (but particular cases of the above) are (c) the Cauchy process $(c(t), t \geq 0)$ and (d) the deterministic process $(d(t), t \geq 0)$. For the one-dimensional Cauchy process, we define the transition function

$$P_c(s,x,t,A) = \int_A \frac{\pi(t-s)\mathrm{d}y}{(t-s)^2 + (y-x)^2}, \tag{1.41}$$

and we apply Kolmogorov construction. It can be proved that (see Exercise 1.31) the continuity condition (1.11) is not satisfied. Only a cad-lag version of this process can be constructed. This process can be thought of as the *trace* on the real line of trajectories of a two-dimensional Brownian motion, which moves by jumps.

For the deterministic process, we use an ordinary differential equation $[0,T]\times\mathbb{R}^d$, e.g. an initial valued problem

$$\dot{y}_{sx} = f(t, y_{sx}), \quad t > s, \quad y_{sx}(s) = x, \tag{1.42}$$

to define the transition function

$$p_d(s,x,t,A) = \mathbb{1}_A\big(y_{sx}(t)\big).$$

Certainly, the only deterministic model comparable with the previous examples is the case $y_{sx}(t) = (t-s)a + x$, for some constant a, which is homogeneous in time and space.

In this case we may use Kolmogorov construction with the continuity condition (1.11) satisfied. However, a direct approach show that the probability measure P_d constructed in the space $C([0,T],\mathbb{R}^d)$ is simply the Dirac measure with respect to $y_{sx}(t)$, i.e. for all $A \in \mathcal{B}(C)$

$$P_d(A) = \begin{cases} 1 & \text{if } y_{0x}(\cdot) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

where the initial probability is equal to δ_x.

Note that the Poisson process has values in a countable set and it is a typical example of the so-called (e) Markov pure jump process. Its (time homogeneous) transition density function, denoted by $p(t,i,j)$, should satisfy the Chapman-Kolmogorov identity

$$p(t-s,i,j) = \sum_k p(r-s,i,k)\, p(t-r,k,j), \quad \forall t > r > s,\ i,j.$$

Hence, differentiate either in s or in t and let either $r-s$ or $t-r$ approaches 0 to deduce either the *backward* equation

$$\begin{cases} \partial_s p(t-s,i,j) &= \displaystyle\sum_k \rho^+(s,i,k)\, p(t-s,k,j), \quad \forall t > s,\ i,j, \\ \rho^+(s,i,j) &= \displaystyle\lim_{r\to s} \partial_s p(r-s,i,j) \quad \forall s,\ i,j, \end{cases} \tag{1.43}$$

or the *forward* equation

$$\begin{cases} \partial_t p(t-s,i,j) &= \displaystyle\sum_k p(t-s,i,k)\, \rho^-(t,k,j), \quad \forall t > s,\ i,j, \\ \rho^-(t,i,j) &= \displaystyle\lim_{r\to t} \partial_t p(t-r,i,j) \quad \forall t,\ i,j, \end{cases} \tag{1.44}$$

The quantities $\rho^+(s,i,j)$ and $\rho^-(s,i,j)$ are the characteristic of the process, and clearly

$$\begin{cases} \rho^-(s,i,j) = -\rho^+(t,i,j) := \rho(i,j), \quad \text{independent of } s,t, \\ \rho(i,j) \geq 0, \quad \forall i \neq j, \qquad \rho(i,i) = -\displaystyle\sum_{j\neq i} \rho(i,j). \end{cases} \tag{1.45}$$

As in the deterministic example, instead of prescribing the transition function we give an infinity array of real numbers $\{\rho(i,j) : i,j = 1,\dots\}$ satisfying the above equation (1.45). These functions represent the *infinitesimal rate* at which i will jumps to j at time s. We can construct the transition function by solving the forward equation (1.44) with an initial condition, which is a linear system of ordinary differential equations if the i,j ranges over a finite set. We have to solve for $p(t,i,j)$ the equation

$$\begin{cases} \dot{p}(t,i,j) &= \displaystyle\sum_k p(t,i,k)\,\rho(k,j), \quad \forall t>0,\ i,j, \\ p(0,i,j) &= \delta_{i,j}, \end{cases} \tag{1.46}$$

where the dot means derivative in t and $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. It can be proved (see Exercise 1.32) that this equation (1.46) admits a unique solution which provides a transition function $p(t,i,j)$ as above, see Lamperti [149, Section 6.2, pp. 114-117]. The cad-lag condition (1.12) is satisfied, so that by means of Kolmogorov construction a *realization* of the Markov pure jump process exists in $D([0,T],\mathbb{R})$. Actually, a discrete arguments will provide a realization in $\{1,2,\dots\}^{[0,T]}$. For instance we refer to Chung [43] for a exhaustive treatment.

A so-called (Gaussian) *white noise* is generally understood in engineering literature as a family of independent centered Gaussian random variables $\{\xi(t) : t \geq 0\}$, i.e., mean $\mathbb{E}\{\xi(t)\} = 0$ and a covariance $R(t-s) = \mathbb{E}\{\xi(t)\xi(s)\}$ with constant spectral density, so that

$$\mathbb{E}\{\exp[-\mathtt{i}\xi(t)]\} = \frac{1}{2\pi}, \quad \forall t \geq 0.$$

Such a process has a spectrum on which all frequencies participate with the same intensity, hence a *white* spectrum (in analogy with the *white light* in optics, which contains all frequencies of visible light, uniformly). However, such a process does not exist in the traditional sense because we should have $R(t-s) = \delta(t-s)$, where δ is the Dirac's delta *generalized* function. All this can be mathematical interpreted in the Schwartz distribution sense.

Definition 1.34 (generalized process)**.** Denote by $\mathcal{S}(]0,\infty[)$ the space of rapidly decreasing infinite differentiable functions and by $\mathcal{S}'(]0,\infty[)$ its dual space, the space of temperate distributions on $]0,\infty[$. In the same way that a stochastic process can be considered as a random variable with values in $\mathbb{R}^{[0,\infty]}$, a *generalized stochastic process* is random variable with values in the Schwartz distribution space $(\mathcal{S}'(]0,\infty[), \mathcal{B}(\mathcal{S}'))$. □

Note that $\mathcal{S}(]0,\infty[)$ and $\mathcal{S}'(]0,\infty[)$ are reflexive and Fréchet (locally convex, complete and metrizable) spaces, e.g., see Schwartz [224], but only with a partial order. In particular, the concept of *a process adapted to a filtration* and *stopping time* need to be reviewed with generalized processes. Clearly, we may use $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}'(\mathbb{R}^d)$ instead of $\mathcal{S}(]0,\infty[)$ and $\mathcal{S}'(]0,\infty[)$.

Thus, a process $\{x(t) : t \geq 0\}$ can be regarded as a generalized process Φ, where

$$\langle \Phi_x(\omega), \varphi \rangle := \int_0^\infty x(t,\omega)\, \varphi(t) \mathrm{d}t, \quad \forall \varphi \in \mathcal{S}(]0,\infty[),\ \omega \in \Omega,$$

but it is clear that we loss the complete order on the index set. Also, a generalized (stochastic) process Φ is said to be Gaussian if for arbitrary linearly independent functions $\varphi_1, \varphi_2, \ldots, \varphi_n$ the n-dimensional random variable $(\Phi(\varphi_1), \Phi(\varphi_2), \ldots, \Phi(\varphi_n))$ is normally distributed. As in the classical sense, a generalized Gaussian process is uniquely determined by its means and covariance. An important advantage of a generalized process is the fact that its derivative always exists and is itself a generalized stochastic process. In particular if we start with a Wiener process $\{w(t) : t \geq 0\}$ consider the generalized derivative of its associated generalized process Φ_w i.e.

$$\langle \dot{\Phi}_w(\omega), \varphi \rangle := -\langle \Phi_w(\omega), \dot{\varphi} \rangle \quad \forall \varphi \in \mathcal{S}(]0,\infty[),\ \omega \in \Omega.$$

we obtain a Gaussian white noise (generalized) process. Similarly, from a Poisson process $\{p(t) : t \geq 0\}$ we get a Poisson white noise (generalized) process as the derivative of the generalized process associated with $\{\bar{p}(t) := p(t) - \mathbb{E}\{p(t)\} : t \geq 0\}$. For instance we refer to Arnold [7, pp. 50–56], Kallianpur and Karandikar [122] and Pallu de la Barrière [193, Chapter 7]. A comprehensive analysis on Markov (and Gausssian) processes can be found in the recent book by Marcus and Rosen [166].

1.12 Sample Spaces

At this point, the reader should be even more familiar with the topological aspect of real analysis. Perhaps some material like the beginning of the books by Billingsley [26], Pollard [199] and some points in Dudley [62] are necessary for the understanding of the next three sections.

Actually, we may look at an E-valued stochastic process $\{X(t) : t \in T\}$ as a random variable X with values in E^T endowed with the product Borel σ-algebra $\mathcal{B}^T(E)$ (generated by cylinder sets) Technically, we may talk about a random variable on a measurable space (without a given probability measure), however, the above Definition 1.23 assumes that a probability measure is given. If some information on the sample paths of the process is available (e.g., continuous paths) then the *big* space E^T and the *small* σ-algebra $\mathcal{B}^T(E)$ are adjusted to produce a suitable topological space $(\Omega, \mathcal{F})$ on which a probability measure can be defined.

When the index set T is uncountable, the σ-algebra $\mathcal{B}^T(E)$, $E \subset \mathbb{R}$ is rather small, since only a countable number of restrictions can be used to define a measurable set (see Exercise 1.33) so that a set of only one point $\{\omega\}$ is not measurable. This forces us to consider smaller sample spaces, where a topological structure is defined e.g., the space of continuous functions $C = C([0,\infty), E)$

from $[0,\infty)$ into E, with the uniform convergence over compact sets. The space $C([0,\infty),E)$ endowed with the natural metric

$$d_c(\omega,\omega') = \sum_{k=1}^{\infty} 2^{-k} \sup\{1 \wedge |\omega(t\wedge k) - \omega'(t\wedge k)| \quad : t \in [0,\infty)\}$$

becomes a complete separable metric space. Thus, the Borel σ-algebra on C coincides with the σ-algebra generated by the coordinate mappings (see Exercise 1.34).

Another typical example and perhaps the most commonly used sample space is the $D = D([0,\infty),E)$ the space of right continuous functions ω from $[0,\infty)$ into E having left limits. Note that any function in $D([0,\infty),E)$ is locally bounded and has at most countable many points of discontinuity. The space $D([0,\infty),E)$ can be endowed with a topology which makes it a complete separable metric space (see Exercise 1.35). This Skorokhod topology is given by the metric

$$d(\omega,\omega') = \inf\{p(\lambda) \vee \sum_{k=1}^{\infty} 2^{-k} q(\omega,\omega',\lambda,k) \quad : \lambda \in \Lambda\},$$

where Λ is the collection of strictly increasing functions λ mapping $[0,\infty)$ onto itself and such that

$$p(\lambda) = \sup\{|\ln(\lambda(s) - \lambda(t)) - \ln(s-t)| \quad : 0 \le t < s\}$$

is finite and

$$q(\omega,\omega',\lambda,k) = \sup\{1 \wedge |\omega(t\wedge k) - \omega'(\lambda(t)\wedge k)| \quad : t \in [0,\infty)\}.$$

We remark that the Skorokhod topology relative to $C([0,\infty),E)$ coincides with the locally uniform topology (see Exercise 1.36), so that C can be considered as a closed subspace of D. On the other hand, given an element ω in $D([0,\infty),E)$ and a positive number ε there exist times $0 = t_0 < t_1 < \cdots < t_n = 1/\varepsilon$ such that the oscillation of ω in each subinterval $[t_{i-1},t_i)$, $i = 1,\dots,n$ is not greater than ε, i.e., for ω_ε defined by $\omega_\varepsilon(t) = \omega(t_i)$ for any t in $[t_{i-1},t_i)$, we have $|\omega(t) - \omega_\varepsilon(t)| \le \varepsilon$. This is to say that any function in $D([0,\infty),E)$ can be approximated in the topology of $C([0,\infty),E)$ by right-continuous step functions, but it cannot be approximated in (the topology of) $D([0,\infty),E)$ by continuous functions. Clearly, the cad-lag functions endowed with the locally uniform convergence (i.e., D with the topology of C) is not a separable topological space. The interested reader is referred to, e.g., Billingsley [26, Chapter 3, pp. 109–153] for a comprehensive study. Sometime it is convenient to define the sample spaces $D(]-\infty,+\infty[,E)$ and $C(]-\infty,+\infty[,E)$, and even to assume that E is only a Polish space (i.e., a complete and separable metric space). Some extra difficulties appear when E is not locally compact.

Any continuous function f with a compact support in $[0,\infty)$ (or in $]0,\infty[$, if necessary) defines a linear functional on $D([0,\infty),E)$, namely

$$\langle f,\omega\rangle = \int_0^\infty f(t)\omega(t)\mathrm{d}t,$$

which results continuous (with the Skorokhod topology). Hence, the Hausdorff topology generated by those linear functional is weaker than the Skorokhod topology and makes D a Lusin space (note that D is not a topological vector space, the addition is not necessarily a continuous operation).

Recall that if S is a metric space then $\mathcal{B}(S)$ denotes the σ-algebra of Borel subsets of S, i.e. the smallest σ-algebra on S which contains all open subsets of S. In particular $\mathcal{B}(E)$, $\mathcal{B}(D)$ and $\mathcal{B}(C)$ are the Borel σ-algebras of the metric space E, $D([0,\infty),E)$ and $C([0,\infty),E)$, respectively. Sometimes we may use $\mathcal{B}$, when the metric space is known from the context. In particular, the Borel σ-algebra of $C = C([0,\infty),E)$ and $D = D([0,\infty),E)$ are the same as the σ-algebra generated by the coordinate functions $\{X_t(\omega) = \omega(t) : t\}$, i.e., a subset A of D belongs to $\mathcal{B}(D)$ if and only if $A \cap C$ belongs to $\mathcal{B}(C)$. Also, it is of common use the canonical right filtration (to be completed when a probability measure is given) $\bigcap_{s>t}\{\sigma\text{-algebra generated by } (X_r : r \le s)\}$. It can be proved that if $\{P_t : t \ge 0\}$ is a family of probability defined on $\mathcal{F}_t^0 = \sigma\{X_s : 0 \le s \le t\}$ such that the restriction of P_t to $\mathcal{F}_s^0$ coincides with P_s for every $s < t$, then there exists a probability P defined on $\mathcal{B}(D)$ such that P restricted to $\mathcal{F}_t^0$ agrees with P_t, e.g., see Bichteler [25, Appendix, Theorem A.7.1].

Definition 1.35 (continuous)**.** An E-valued, usually $E \subset \mathbb{R}^d$, *continuous* stochastic process is a probability measure P on $(C([0,\infty),E),\mathcal{B})$ together with a measurable mapping (P-equivalence class) X from $C([0,\infty),E)$ into itself. If the mapping X is not mentioned, we assume that it is the *canonical* (coordinate, projection or identity) mapping $X_t(\omega) = \omega(t)$ for any ω in $C([0,\infty),E)$, and in this case, the probability measure $P = P_X$ is called the *law* of the process. Similarly, a *right continuous having left-hand limits* stochastic process is a probability measure P on $(D([0,\infty),E),\mathcal{B})$ together with a measurable mapping X from $D([0,\infty),E)$ into itself. □

Note that a function X from $(C([0,\infty),E),\mathcal{B})$ into itself is measurable if and only if the functions $\omega \mapsto X(t,\omega)$ from $(C([0,\infty),E),\mathcal{B})$ into E are measurable for all t in $[0,\infty)$ (see Exercise 1.34). Since $C([0,\infty),E) \subset D([0,\infty),E)$ as a topological space with the same relative topology, we may look at a continuous stochastic process as probability measure on D with support in the closed subspace C.

Thus, to get a continuous (or cad-lag) version of a *general* stochastic process X (see Definition 1.23) we need to show that its probability law P_X has support in $C([0,\infty),E)$ (or in $D([0,\infty),E)$). On the other hand, separability of a general stochastic process can be taken for granted (see Theorem 1.19), after a suitable modification. However, for general stochastic processes viewed as a collection of random variables defined almost surely, a minimum workable assumption is (right or left) stochastic continuity (i.e., continuous in probability). Clearly, stochastic continuity cannot be stated in term of random variable having values in some functional space, but rather as a function on $[0,\infty)$ with values in some probability space, such as $L^p(\Omega,P)$, with $p \ge 0$.

When two or more cad-lag processes are given, we may think of having several probability measures (on the suitable space), say $P_1,\dots,P_n$, and we

canonical process $X(t) = \omega(t)$. However, sometimes it may be convenience to fix a probability measure e.g., $P = P_1$, with a canonical process $X = X_1$ as a reference, and consider all the other processes $P_2, \dots, P_n$ as either the probability measures $P_2, \dots, P_n$ on $(D, \mathcal{B})$ or as measurable mapping $X_2, \dots, X_n$, so that P_i is the image measure of P through the mapping X_i, for any $i = 2, \dots, n$. On the space $(D, \mathcal{B})$ we can also define two more *canonical* processes, the *pure jumps* process $\Delta X(t) = X(t) - X(t-)$, for $t > 0$ and the *left-limit* process

$$X(t-) = \begin{cases} X(0) & \text{if } t = 0, \\ \lim_{s \uparrow t} X(s) & \text{if } t > 0, \end{cases} \tag{1.47}$$

which may also be denoted by $X_-(t)$.

Processes X may be initially given in an abstract space $(\Omega, \mathcal{F}, P)$, but when some property on its sample path is given, such a continuity (see Definition 1.2) then we may look at X as a random variable taking values in a suitable topological space (e.g. C or D). Then by taking the image measure of P through X, we may really forget about the initial space $(\Omega, \mathcal{F}, P)$ and refer everything to the sample space, usually C or D.

Given a *filtered measurable space*, i.e. $\{\Omega, \mathcal{G}(t) : t \geq 0\}$ satisfying the usual conditions (see Definition 1.7) and a cad-lag stochastic process $\{P_X, X(t) : t \geq 0\}$ (where P_X is a probability on $(\Omega, \mathcal{G})$, with $\mathcal{G}(\infty) \subset \mathcal{G}$), we can assume that $X(t)$ is the canonical process $\omega(t)$ on $\Omega = D([0, \infty), E)$. If the process X is adapted then $\{\mathcal{G}(t) : t \geq 0\}$ can be viewed as a filtration on $D([0, \infty), E)$, subordinated to the *canonical* filtration $\{\mathcal{F}(t) : t \geq 0\}$, i.e., $\mathcal{G}(t) \subset \mathcal{F}(t)$, for all $t \geq 0$. Here the canonical σ-algebra $\mathcal{F}(t)$ is generated by the canonical process $\{\omega(s) : 0 \leq s \leq t\}$ and completed (when necessary) with respect to the probability measure P_X. Therefore, the process $\{P_X, X(t) : t \geq 0\}$ is progressively measurable with respect to the canonical filtration $\{\mathcal{F}(t) : t \geq 0\}$, i.e., the function $(\omega, s) \mapsto \omega(s)$ from $D([0, t], E) \times [0, t]$ into E is measurable with respect to the product σ-algebra $\mathcal{F}(t) \times \mathcal{B}([0, t])$, and P_X is a probability measure on $\{D([0, \infty), E), \mathcal{F}(\infty)\}$.

Note the following rather technical result. Let $\{X(t) : t \geq 0\}$ be a cad-lag process and $\{\mathcal{F}(t) : t \geq 0\}$ be its natural filtration (satisfying the usual conditions). If $\{Y(t) : t \geq 0\}$ is another process adapted to $\{\mathcal{F}(t) : t \geq 0\}$ (in short adapted to the process X) then there exists a family $\{\mathcal{Y}_t : t \geq 0\}$ of measurable functions $\mathcal{Y}_t : D([0, t), E) \to E$ such that $Y(t) = \mathcal{Y}_t\big(X(s) : s \in [0, t]\big)$, for every $t \geq 0$.

It is interesting to remark that $D([0, \infty), \mathbb{R}^d)$ is not a topological vector space, i.e., in the Skorokhod topology, we may have $\alpha_n \to \alpha$ and $\beta_n \to \beta$, but $\alpha_n + \beta_n$ is not converging to $\alpha + \beta$, unless α (or β) belongs to $C([0, \infty), \mathbb{R}^d)$. Moreover, the topology in $D([0, \infty), \mathbb{R}^d)$ is strictly stronger that the product topology in $D([0, \infty), \mathbb{R}^{d_1}) \times D([0, \infty), \mathbb{R}^{d_2})$, $d = d_1 + d_2$. The reader is referred to the book Jacod and Shiryaev [117, Chapter VI, pp. 288–347] for a comprehensive discussion.

1.13 Convergence of Processes

Since the concept of stochastic processes lead to the study of probability measures on a separable and complete metric space (also called *Polish space*). A good discussion on this subject can be founded in Ethier and Kurtz [76, Chapter 3, pp. 95–154] or Ikeda and Watanabe [110, Chapter 1, pp. 1–44]. We are going to state some of the key elements.

First consider $\mathcal{P}(\Omega)$ the family of probability measures on $(\Omega, \mathcal{B})$, where $\mathcal{B} = \mathcal{B}(\Omega)$ is the Borel σ-algebra on the Polish space Ω. This is also referred to as the family of Borel probability measures on Ω. The Prohorov metric on $\mathcal{P}(\Omega)$ is defined by

$$d(P, Q) = \inf\{\varepsilon > 0 : P(A) \leq Q(A^\varepsilon) + \varepsilon, \ \ \forall A \text{ closed in } \Omega\},$$

where $A^\varepsilon = \{\omega \in \Omega : \inf_{\omega' \in A} d_\Omega(\omega, \omega') < \varepsilon\}$, and $d_\Omega(\cdot, \cdot)$ is the metric on Ω. Thus $\mathcal{P}(\Omega)$ endowed with the above metric becomes a Polish space.

Denote by $C_b(\Omega)$ the space of real-valued continuous function on the Polish space (Ω, d_Ω) with the natural norm $\|f\| = \sup_\omega |f(\omega)|$. A sequence $\{P_n : n = 1, 2, \dots\}$ in $\mathcal{P}(\Omega)$ is said to *converge weakly* to P if

$$\lim_{n\to\infty} \int f(\omega)\, P_n(\mathrm{d}\omega) = \int f(\omega)\, P(\mathrm{d}\omega), \qquad \forall f \in C_b(\Omega).$$

In the previous condition, it is sufficient to take only functions f which are uniformly continuous. Actually, if we know that the limit measure is a probability then it is enough to satisfy the convergence for uniformly continuous functions with a bounded support, even more, if the space Ω is locally compact, then it suffices to use continuous functions with a compact support. The important point here is that the convergence in the Prohorov metric is equivalent to the above weak convergence.

A classic result so-called Skorokhod representation given some relation with the almost surely convergence.

Theorem 1.36 (Skorokhod). *Let $\{P_n : n = 1, 2, \dots\}$ be a sequence of probability measures on a Polish space Ω which converge weakly to P. Then in some probability space there exist random variables $X_n : n = 1, 2, \dots$ and X with values in Ω with distributions $P_n : n = 1, 2, \dots$ and P, respectively, such that $\lim_{n\to\infty} X_n = X$ almost surely.* □

One of the typical applications of this representation is the fact that the weak convergence is preserved by Borel mapping which is almost continuous with respect to the limiting measure (see Exercise 1.37).

Another point is the characterization of pre-compacts or relatively compact sets (i.e., with a compact closure) set in $\mathcal{P}(\Omega)$ with the weak convergence i.e., endowed with the Prohorov metric.

Theorem 1.37 (Prohorov). *A sequence of probability measures $\{P_n : n = 1, 2, \dots\}$ on a Polish space Ω has a weakly convergent subsequence if and only*

if the sequence is tight i.e, for any $\varepsilon, \delta > 0$ there exist $\omega_1, \dots, \omega_n$ in Ω such that

$$P_n(\bigcup_{i=1}^{n}\{\omega : d_\Omega(\omega_i, \omega) \le \delta\}) \ge 1 - \varepsilon$$

for all $n = 1, 2, \dots$. □

Usually, a family of probability measures $\{P_\alpha\}$ on Ω is said to be *tight* if for any $\varepsilon > 0$ there exists a compact set $K \subset \Omega$ such that $P_\alpha(K) \ge 1 - \varepsilon$ for any index α. Since a set is pre-compact in Ω if and only if it is *totally bounded* this is equivalent to the above condition.

In view of the above characterization of the weak convergence of measures, it is important to understand the structure of compact sets in the particular spaces $C([0,\infty), E)$ and $D([0,\infty), E)$, where (E, d_E) is a Polish space, in particular $E = \mathbb{R}^d$. Classic results applies to say that pre-compact sets are equivalent to totally bounded and equi-continuous sets. Thus a family $\{\omega_\alpha\}$ of functions in $C([0,\infty), E)$ is relatively compact if and only if

(a) for any $\delta > 0$ and rational $r \ge 0$, there exist $x_1, \dots, x_n$ in E such that for any index α we have

$$\omega_\alpha(r) \in \bigcup_{i=1}^{n}\{x : d_E(x_i, x) \le \delta\}$$

(b) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index α, and any t, s in $[0, T]$ we have $d_E(\omega_\alpha(t), \omega_\alpha(s)) < \delta$.

The fact that in (a) we require the condition to be satisfied only for rational is convenient for later. Now, for the space $D([0,\infty), E)$ we need to use the modulus of continuity $w(\omega, \delta, T)$ defined by

$$w(\omega, \delta, T) := \inf_{\{t_i\}} \sup_i \sup\{d_E(\omega(t), \omega(s)) : t_{i-1} \le s < t < t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_n = T$, with $t_i - t_{i-1} \ge \delta$ and $n \ge 1$. A shorter version of the modulus of continuity is given by the expression

$$w(\omega, \delta, T) := \sup_{0 \le t < T - \delta} \sup_{t \le s \le t' \le t + \delta} \{d_E(\omega(t'), \omega(s)) \wedge d_E(\omega(s), \omega(t))\},$$

where $\wedge$ means the minimum between numbers. Therefore, we replace (b) by the condition

(b1) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index α we have $w(\omega_\alpha, \delta, T) < \varepsilon$.

It is cleat that if $E = \mathbb{R}^d$ then (a) reduces to

(a1) for any rational $r \ge 0$, there $N > 0$ such that for any index α we have $|\omega_\alpha(r)| \le N$.

Theorem 1.38 (tight). *Let $X_1, X_2, \dots$ be a sequence of random variables with values in $D([0,\infty), E)$, with E a Polish space and $P_1, P_2, \dots$ be its associated probability law on $D([0,\infty), E)$. Then the sequence $P_1, P_2, \dots$ is tight (hence relatively compact) in $D([0,\infty), E)$ if an only if the following two conditions hold:*

(a') *for any $\varepsilon, \delta > 0$ and rational $r \geq 0$, there exist $x_1, \dots, x_k$ in E such that for any index n we have*

$$P_n(X_n(r) \in \bigcup_{i=1}^{k} \{x : d_E(x_i, x) \leq \delta\}) \geq 1 - \varepsilon,$$

(b') *for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have*

$$P_n(w(X_n, \delta, T) \geq \varepsilon) < \varepsilon.$$

Moreover, if the sequence is tight, then it is weakly convergent if and only its finite-dimensional distributions converge. □

It is clear that some obvious modifications should be done for a sequence of random variables in the space $C([0,\infty), E)$, i.e., re-defining $w(\omega, \delta, T)$ as $\sup\{d_E(\omega(t), \omega(s)) : |t-s| < \delta,\ s, t \in [0,T]\}$. In the space $C([0,\infty), \mathbb{R}^d)$, condition (b') simply becomes:

(b') for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P_n(\sup_{s,t\in[0,T],\ |t-s|<\delta} |X_n(t) - X_n(s)| \geq \varepsilon) < \varepsilon, \tag{1.48}$$

for any index n.

Naturally, the above result is useful to study cad-lag processes. It may be convenient to use Aldous' criterion in $D([0,\infty), \mathbb{R}^d)$, e.g., see Ethier and Kurtz [76, p. 137, Theorem 8.6, Chapter 3] or Liptser and Shiryayev [158, Section 6.3, pp. 515–519]. This is to replace condition (a') and (b') of the previous theorem with the following statement:

(a*) for any $\varepsilon > 0$ there exists $M > 0$ such that for any index n we have

$$P_n(|X_n(0)| \geq M) < \varepsilon, \tag{1.49}$$

(b*) for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have

$$\sup_{0<s\leq\delta} P_n(|X_n(\tau_n + s) - X_n(\tau_n)| \geq \varepsilon) < \varepsilon, \tag{1.50}$$

for any stopping time (relative to X_n) τ_n satisfying $0 \leq \tau_n \leq T$. The key facts here are that the sup is outside of the *integral* and that s is a (deterministic) number, so that $\tau_n + s$ becomes an optional time with respect to process $X_n(\cdot - s)$. Moreover, (b*) is equivalent to the following condition:

(b") for each $T > 0$ and for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any index n we have

$$P_n(|X_n(\theta_n) - X_n(\tau_n)| \geq \varepsilon) < \varepsilon, \tag{1.51}$$

for any stopping times (relative to X_n) θ_n and τ_n satisfying $0 \leq \theta_n \leq \tau_n \leq T$ and $\tau_n \leq \theta + \delta$.

If (P_n, X_n) is a sequence of homogeneous strong Markov processes in the canonical space $D([0,\infty), \mathbb{R}^d)$ with transition probability measure $P_n(x, t, \mathrm{d}y)$, and a sequence of stopping times τ_n, then the equality, with some $r > 0$,

$$\mathbb{E}\big\{\mathbb{E}\{|X_n(\tau_n + s) - X_n(\tau_n)|^r \,|\, X(\tau_n)\}\big\} = \\ = \int_{\mathbb{R}^d} \mathbb{E}\big\{|y - X(\tau_n)|^r P_n(X(\tau_n), s, \mathrm{d}y)\big\}$$

shows that Aldous' criterion conditions (b*) (1.50) is satisfied if

$$\limsup_{s\to 0} \sup_n \int_{\mathbb{R}^d} |y - x|^r P_n(x, s, \mathrm{d}y) = \\ = \limsup_{s\to 0} \sup_n \mathbb{E}\big\{|X_n(s) - X_n(0)|^r \,|\, X_n(0) = x\big\} = 0$$

which is a simple condition to verify. Moreover, the expression $|y - x|^r$ above could be replaced by $\gamma(|y - x|)$ with a strictly increasing continuous function $\gamma(\cdot)$ satisfying $\gamma(0) = 0$. For instance, the reader may check the book Bass [14, Chapter 34, pp 259–268].

The convergence of finite-dimensional distributions of a sequence $\{X_n : n \geq 1\}$ of $\mathbb{R}^d$-valued stochastic processes, means that for any finite number of times $t_1, \ldots, t_k$ we have

$$\lim_n \mathbb{E}\{h(X_n(t_1), \ldots, X_n(t_k))\} = \mathbb{E}\{h(X(t_1), \ldots, X(t_k))\},$$

for any continuous and bounded real-valued function h on $\mathbb{R}^k$. On the contrary to the convergence in $D([0,\infty), \mathbb{R}^d)$, no convergence condition on the paths is involved in this concept.

To control the modulus of continuity of a process $X(t)$, the following estimate is very useful. For any $\alpha, \beta > 0$ there exists a constant $C_0 = C_0(\alpha, \beta)$ such that

$$|f(t') - f(s')|^\alpha \leq C_0 \, |t' - s'|^\beta \int_0^T \mathrm{d}t \int_0^T \frac{|f(t) - f(s)|^\alpha}{|t - s|^{2+\beta}} \mathrm{d}s, \tag{1.52}$$

for any continuous function f on $[0, T]$ and any t', s' in $[0, T]$, see Da Prato and Zabczyk [52, Theorem B.1.5, pp. 311–316] or Stroock and Varadhan [241, Theorem 2.1.3, pp. 47–49]. Therefore, if for some constants p, q, $C > 0$ a process $X(t, \omega)$ satisfies

$$\mathbb{E}\big\{|X(t) - X(s)|^p\big\} \leq C|t - s|^{1+q}, \quad \forall t, s \in [0, T], \tag{1.53}$$

then by taken $p = \alpha$ and $\beta = r$ with $0 < r < q$ we deduce that there is another constant $C_0 = C_0(p, q, C, r)$ such that

$$\mathbb{E}\Big\{ \sup_{t,s\in[0,T],\ |t-s|<\delta} \{|X(t) - X(s)|^p\}\Big\} \le C_0\, \delta^r, \quad \forall \delta > 0. \tag{1.54}$$

Essentially, an estimate in L^∞ of the modulus of continuity is obtained based on an estimate in L^p. This is of particular interest for stochastic processes with continuous paths.

For cad-lag processes, a bound of the type: for every $0 \le t \le s \le t + \delta \le T$, and some positive constants C, p and q,

$$\mathbb{E}\Big\{ \big[|X(t+\delta) - X(s)| \wedge |X(s) - X(t)|\big]^p\Big\} \le C\delta^{1+q}, \quad \forall \delta > 0, \tag{1.55}$$

yields the estimate

$$\mathbb{E}\Big\{ \sup_{0\le t\le T-\delta} \sup_{t\le s\le s+\delta} \{|X(t+\delta) - X(s)| \wedge |X(s) - X(t)|\}^p\Big\} \le C_0\, \delta^r, \tag{1.56}$$

for every $\delta > 0$, any $0 < r < q$ and another constant $C_0 = C_0(p, q, C, r)$. The reader may consult the books Billingsley [26, Chapter 3, pp. 109–153] or Ethier and Kurtz [76, Chapter 3, pp. 95–154] for a complete discussion.

Sometime we have to use the space $B(\Omega)$ of all bounded and Borel measurable functions from the Polish space Ω into $\mathbb{R}^n$. The weak topology we need is the so-called *boundedly and pointwise* convergence i.e., a sequence of functions $\{f_n : n = 1, 2, \dots\}$ in $B(\Omega)$ converge boundedly and pointwise to f if $\sup_{n,\omega} |f_n(x)| < \infty$ and $f_n(x) \longrightarrow f(\omega)$ for every $\omega \in \Omega$. The (sequentially) closure of a set M in this topology is referred to as the bp-closure of M. A typical application of the Monotone Class Theorem (see Exercise 1.38) shows that the bp-closure of $C_b(\Omega)$ is the whole space $B(\Omega)$ i.e., it is bp-dense. Moreover, since Ω is separable, there exists a sequence $\{f_n : n = 1, 2, \dots\}$ of nonnegative continuous and bounded functions that span a bp-dense set in $B(\Omega)$. Note that this is not to say that any function in $B(\Omega)$ is a boundedly and pointwise limit of a sequence of function in $C_b(\Omega)$. Due to the probability measure, we prefer to use the Lebesgue space $L^\infty(\Omega, \mathcal{F}, P)$ instead of $B(\Omega)$, when ever is possible. The reader may consult the books Doob [59, Chapters VIII,...,X, pp. 123–177], Jacod and Shiryaev [117, Chapter VI, pp. 288–347], among other, for a complete discussion of convergence of measures and processes.

1.14 Existence of Probabilities

At this point, the reader may revise the some of the basic subjects treated in the book Malliavin [162]. In particular, a revision on measure theory, e.g., as in Kallenberg [121, Chapters 1 and 2, pp. 1–44], may be necessary.

Perhaps the Gaussian probability in $\mathbb{R}^n$ is the best well now situation, i.e., the probability space $(\mathbb{R}^n, \mathcal{B}^n, P_n)$, where $\mathcal{B}^n$ is the Borel σ-algebra in $\mathbb{R}^n$ and P_n is the probability measure given by

$$P_n(A) = (2\pi)^{-n/2} \int_A \exp\Big(-\frac{1}{2} \sum_{i=1}^n |x_i|^2\Big) \mathrm{d}x, \quad \forall A \in \mathcal{B}^n,$$

standard normal distribution or Gaussian with mean 0 and variance 1. However, the extension of this probability to the space $\mathbb{R}^\infty$ of all sequences of real numbers (with the product topology) is not so trivial. If X_n denotes the projection from $\mathbb{R}^\infty$ into $\mathbb{R}^n$, i.e., $X_n(x) = (x_1, \dots, x_n)$ the first n coordinates of x, then it is not immediate to establish the existence of a probability P defined on the Borel σ-algebra $\mathcal{B}^\infty$ such that $P(A) = P_n(X_n(A))$ for any A in $\mathcal{B}^\infty$. In general, the two points of interest here are the σ-algebra generated by the cylindrical sets $\{X_n^{-1}(A) : A \in \mathcal{B}^n,\, n \geq 1\}$ and the σ-additivity of P. Therefore, on the probability space $(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$, we may look at $\{X_n : n \geq 1\}$ as a sequence of Gaussian random variables which generates the σ-algebra $\mathcal{B}^\infty$. It is well know that the Hermit polynomials provide an orthonormal basis for the Hilbert space $L^2(\mathbb{R}^1, \mathcal{B}^1, P_1)$, however some tedious notation and details are needed to deduce an orthonormal basis for $L^2(\mathbb{R}^\infty, \mathcal{B}^\infty, P)$.

Now, our interest turns into the existence of probability measures, first in $\mathbb{R}^n$, next in separable Hilbert spaces and finally in Polish spaces, particularly in the space of tempered distributions. Thus, the discussion about the existence of a particular stochastic process with values in $\mathbb{R}^n$ becomes a discussion on the existence of probability measures on relatively large spaces, such as the Schwartz space of tempered distributions, where the Fourier transform can be used.

One way of constructing a probability measure is by prescribing its characteristic function (or its Fourier transform). In finite dimensional spaces we have the classical Bochner's Theorem stated as follow:

Theorem 1.39. *If $\Psi : \mathbb{R}^n \to \mathbb{C}$ is the characteristic function of a probability measure (space) $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$, i.e.,*

$$\Psi(\xi) = \int_{\mathbb{R}^n} \exp\big(\mathtt{i}(\xi, x)\big)\, P(\mathrm{d}x) = \mathbb{E}\big\{\exp\big(\mathtt{i}(\xi, \cdot)\big)\big\},$$

with $\mathtt{i} = \sqrt{-1}$, then (a) $\Psi(0) = 1$, (b) Ψ *is continuous and* (c) Ψ *is positive definite, i.e., for every natural number k, any ξ_i in $\mathbb{R}^n$ and any complex number z_i, $i = 1, \dots, k$ we have*

$$\sum_{i,j=1}^{k} \Psi(\xi_i - \xi_j) z_i \bar{z}_j \geq 0,$$

where $(\cdot, \cdot)$ denotes the scalar product in $\mathbb{R}^n$ and $\bar{z}$ is the conjugate of a complex number. Conversely, an arbitrary function $\Psi : \mathbb{R}^n \to \mathbb{C}$ satisfying the above properties (a), (b) *and* (c) *is the characteristic function of a probability measure P on $\mathbb{R}^n$.* □

This is also known as Bochner-Khintchin's Theorem, for instance, a complete proof can be find in Gnedenko [101, Section 39, pp. 289–293] or Jacob [115, Vol 1, Theorem 3.5.7, pp. 108–109].

Next, the (Schwartz) space of rapidly decreasing and smooth functions $\mathcal{S}(\mathbb{R})$ and its dual space of tempered distributions $\mathcal{S}'(\mathbb{R})$ is identified (via Hermite functions, i.e., given a sequence in s we form a function in $\mathcal{S}(\mathbb{R})$ by using the

terms as coefficients in the expansion along the orthonormal basis $\{\xi_n(x) : n \geq 1\}$, with

$$\xi_{n+1}(x) = \frac{\mathrm{e}^{-x^2/2}}{\pi^{1/4}\sqrt{n!}}\, p_n(\sqrt{2}x), \quad n = 1, 2, \dots,$$

where p_n is the Hermite polynomial of order n) with the Fréchet space of rapidly decreasing sequences

$$s = \big\{a = \{a_k\}_{k=0}^{\infty} \,:\, \lim_k k^m a_k = 0,\ \forall m = 1, 2, \dots\big\}.$$

This space is decomposed as $s = \bigcap_{m=0}^{\infty} s_m$ with s_m defined for every integer m as the space of all sequences $a = \{a_k\}_{k=0}^{\infty}$ satisfying

$$\|a\|_m = \Big[\sum_{k=0}^{\infty}(1 + k^2)^m |a_k|^2\Big]^{1/2} < \infty.$$

Its dual space is decomposed as $s' = \bigcup_{m=0}^{\infty} s'_m$, with $s'_m = s_{-m}$ and the natural paring between elements in s' and s (also between s'_m and s_m), namely,

$$\langle a', a\rangle = \sum_{k=0}^{\infty} a'_k a_k, \quad \forall a' \in s',\ a \in s.$$

Based on Bochner's result for finite dimensional spaces and Kolmogorov's extension, a probability measure with a prescribed characteristic function can be constructed in the space $\mathbb{R}^\infty$, the space of all sequences of real numbers. It takes some more effort (e.g., see Holden et al. [109, Appendix A, pp. 193–197]) to check that the probability measure is concentrated on the dual space s'. Indeed, use the continuity and the condition $\Psi(0) = 1$ to deduce that for any $\varepsilon > 0$ there exist $m > 0$ and $\delta > 0$ such that $\|a\|_m < \delta$ implies $|\Psi(a) - 1| < \varepsilon$, which yields

$$\int_{\mathbb{R}^\infty} \cos(\langle a', a\rangle)\, P(\mathrm{d}a') \geq 1 - \varepsilon - 2\delta^{-2}\|a\|_m^2, \quad \forall a \in s. \tag{1.57}$$

Now, for every sequence $b = \{b_k\}$, with $b_k > 0$ consider the (Gaussian) probability measure $\mu_{n,\sigma}$ on $\mathbb{R}^{n+1}$ defined by

$$\mu_{n,\sigma} = \prod_{k=0}^{n}(2\pi\sigma b_k)^{-1/2} \exp\big[-\frac{a_k^2}{2\sigma b_k}\big]\mathrm{d}a_k,$$

for any $\sigma > 0$. Recall that

$$\int_{\mathbb{R}^{n+1}} \cos(\langle a', a\rangle)\, \mu_{n,\sigma}(\mathrm{d}a) = \exp\big[-\frac{\sigma}{2}\sum_{k=0}^{n} b_k (a'_k)^2\big],$$
$$\int_{\mathbb{R}^{n+1}} \|a\|_m^2\, \mu_{n,\sigma}(\mathrm{d}a) = \sigma \sum_{k=0}^{n}(1 + k^2)^m b_k,$$

and integrate (1.57) with respect to $\mu_{n,\sigma}$ on $\mathbb{R}^{n+1}$ to get

$$\int_{\mathbb{R}^\infty} \exp\big[-\frac{\sigma}{2}\sum_{k=0}^{n} b_k(a'_k)^2\big]\,P(\mathrm{d}a') \ge 1-\varepsilon-2\delta^{-2}\sigma\sum_{k=0}^{n}(1+k^2)^m b_k.$$

Now, take $b_k=(1+k^2)^{-m-1}$ to have $\sum_{k=0}^{n}(1+k^2)^m b_k = C<\infty$, which imply, by means of the monotone convergence,

$$\int_{\mathbb{R}^\infty} \exp\big[-\frac{\sigma}{2}\|a'\|^2_{-m-1}\big]\,P(\mathrm{d}a') \ge 1-\varepsilon-2\delta^{-2}\sigma C.$$

Finally, let σ vanish to get $P(s'_{m+1})\ge 1-\varepsilon$, which proves that $P(s')=1$.

At this point, we can state the following version of a Bochner-Minlos theorem: On the space of test functions $\mathcal{S}(\mathbb{R})$ we give a functional Ψ which is positive definite, continuous and satisfies $\Psi(0)=1$, then there exists a (unique) probability measure P on the space of tempered distributions $\mathcal{S}'(\mathbb{R})$ having Ψ as its characteristic function, i.e.,

$$\Psi(\varphi)=\int_{\mathcal{S}'(\mathbb{R})}\exp\big(\mathtt{i}\langle\omega,\varphi\rangle\big)\,P(\mathrm{d}\omega)=\mathbb{E}\big\{\exp\big(\mathtt{i}\langle\cdot,\varphi\rangle\big)\big\},$$

where $\langle\cdot,\cdot\rangle$ denote the paring between $\mathcal{S}'(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$, i.e., the $L^2(\mathbb{R})$ inner product.

Certainly, this extends to multi-dimensional case, i.e., $\mathcal{S}(\mathbb{R}^d)$ and vector-valued functions $\mathcal{S}(\mathbb{R}^d;\mathbb{R}^n)$. Thus, we can state the following very useful result regarding the construction of a Lévy martingale measures:

Theorem 1.40 (Lévy noise)**.** *Let $\mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ be the space of tempered distributions in $\mathbb{R}$ with values in $\mathbb{R}^d$. Suppose that σ is a (real-valued) square $d\times d$ matrix and that π is a Radon measure in $\mathbb{R}^d$ satisfying*

$$\int_{\mathbb{R}^d}\big(|y|^2\wedge|y|\big)\pi(\mathrm{d}y)<\infty,\qquad \pi(\{0\})=0. \tag{1.58}$$

Then, there exists a unique probability measure P on $(\Omega,\mathcal{B})$, with $\Omega=\mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ and $\mathcal{B}=\mathcal{B}(\Omega)$ such that

$$\begin{cases}\mathbb{E}\big\{\exp\big[\mathtt{i}\langle\cdot,\varphi\rangle\big]\big\}=\exp\Big(-\dfrac{1}{2}\displaystyle\int_{\mathbb{R}}|\sigma\varphi(t)|^2\mathrm{d}t\Big)\times\\ \qquad\qquad\times\exp\Big(\displaystyle\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)}-1-\mathtt{i}(\varphi(t),y)\big]\pi(\mathrm{d}y)\Big),\end{cases} \tag{1.59}$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation with respect to P and $|\cdot|$ and $(\cdot,\cdot)$ are the Euclidean norm and scalar product, respectively. In particular, $\mathbb{E}\big\{\langle\cdot,\varphi\rangle\big\}=0$, and if also

$$\int_{\mathbb{R}^m}|y|^2\,\pi(\mathrm{d}y)<\infty, \tag{1.60}$$

then

$$\mathbb{E}\big\{\big|\langle\cdot,\varphi\rangle\big|^2\big\}=\int_{\mathbb{R}}\big|\sigma\varphi(t)\big|^2\mathrm{d}t+\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big|(\varphi(t),y)\big|^2\pi(\mathrm{d}y), \tag{1.61}$$

for any test function φ. □

Note that by replacing φ with $\lambda\varphi$, taking derivatives with respect to λ and setting $\lambda = 0$ we deduce the isometry condition (1.61), which yields an analogous equality for the scalar product $\mathbb{E}\{\langle\cdot,\varphi\rangle\,\langle\cdot,\psi\rangle\}$, with φ and ψ in $\mathcal{S}(\mathbb{R};\mathbb{R}^d)$. Clearly, from the calculation point of view, the Fourier transform for h in $\mathcal{S}(\mathbb{R}^d)$

$$\hat{h}(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} h(x)\mathrm{e}^{-\mathtt{i}(x,\xi)}\mathrm{d}x,$$

and its inverse

$$h(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} \hat{h}(\xi)\mathrm{e}^{\mathtt{i}(x,\xi)}\mathrm{d}\xi,$$

are useful to estimate

$$\left\{\begin{array}{l}\mathbb{E}\{\{h(\langle\cdot,\varphi_1\rangle,\dots,\langle\cdot,\varphi_d\rangle)\} = \\ \qquad\qquad = (2\pi)^{-d/2}\displaystyle\int_{\mathbb{R}^d}\hat{h}(\xi)\Psi(\xi_1\varphi_1+\ldots+\xi_d\varphi_d)\mathrm{d}\xi,\end{array}\right. \tag{1.62}$$

where Ψ is the characteristic function, i.e., the right-hand-side in (1.59).

Also, from the finite-dimensional case, we know that the functions

$$\exp\big(-|x|^2/2\big),\quad \exp\big(\mathrm{e}^{\mathtt{i}(x\cdot b)}-1\big),\quad \exp\big(-\mathtt{i}(x\cdot b)\big),$$

for b fixed, are characteristic functions of the Gaussian, the Poisson and the Dirac distributions. Therefore, any matrix $a = (a_{ij})$ of the form

$$a_{ij} = \exp\Big\{-|\zeta_i-\zeta_j|^2/2 + \mathrm{e}^{\mathtt{i}(\varsigma_i-\varsigma_j)}-1\Big\}$$

is a positive definite matrix. Thus, by approximating the integrals (by partial sums) in right-hand-side (called Ψ) of (1.59), we show that Ψ is indeed positive define.

Hence, we have constructed a d-dimensional *smoothed* (1-parameter) Lévy noise associated with (σ,π). Indeed, the canonical action-projection process, which is the natural paring

$$X(\varphi) = X(\varphi,\omega) = \langle\omega,\varphi\rangle,\quad \forall\varphi\in\mathcal{S}(\mathbb{R};\mathbb{R}^d),$$

can be regarded as a family of $\mathbb{R}$-valued random variables $X(\varphi)$ on the probability space $(\Omega,\mathcal{B}(\Omega),P)$, with $\Omega = \mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ and P as above. Clearly, this is viewed as a generalized process and the actual Lévy noise is defined by $\dot{X}(\varphi) = -\langle\omega,\dot{\varphi}\rangle$.

Considering the space $L^2(P)$ and the vector-valued space $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$ with the inner product defined by

$$\langle\varphi,\psi\rangle_{\sigma,\pi} = \int_{\mathbb{R}}\big(\sigma\varphi(t),\sigma\psi(t)\big)\mathrm{d}t + \int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}(\varphi(t),y)\,(\psi(t),y)\,\pi(\mathrm{d}y),$$

we can view $\varphi\mapsto X(\varphi,\cdot)$ as an isometry from $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$ into $L^2(P)$, initially defined on the test space $S(\mathbb{R};\mathbb{R}^d)$ and uniquely extended everywhere. Thus,

the expression $\langle \omega, \varphi \rangle$ makes sense almost surely (passing to the limit) for φ in $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^d)$. Now, for a given test function φ we denote by $\varphi_{i,t}$ the test function with only one non-zero component, namely, the i-component which is given by the expression $\mathbb{1}_{(0,t]}$, i.e., $\varphi_{i,t} = (0,\dots,\mathbb{1}_{(0,t]},\dots,0)$. Thus, a d-dimensional Lévy (martingale) process $\ell_i(t) := X(\varphi_{i,t})$ for $i = 1,2,\dots,d$ (with diffusion matrix $\sigma^*\sigma/2$ and Levy measure π) is almost sure defined. Indeed, because the scalar product is preserved, the stochastic process ℓ has orthogonal increments. Moreover, the linearity in φ and the *product* (or integral and exponential) form of the characteristic function (1.59) show that the random variable $\langle \cdot, \varphi \rangle$ is independent of $\langle \cdot, \psi \rangle$ as long as φ and ψ have disjoint support. Thus, the stochastic process $(\ell(t) : t \geq 0)$ is stationary with independent increments. The existence of a cad-lag version follows from the estimate

$$\begin{aligned}\mathbb{E}\big\{|\ell_i(s+r)-\ell_i(t)|^2|\ell_i(t)-\ell_i(s)|^2\big\} &= \\ = \mathbb{E}\big\{\big(\ell_i(s+r-t)\big)^2\big\}\,\mathbb{E}\big\{\big(\ell_i(t-s)\big)^2\big\} &\leq Cr^2,\end{aligned}$$

for any i, $0 \leq s \leq t \leq s+r \leq T$, any $T > 0$ and some positive constant $C = C_T$.

On the other hand, we can impose less restrictive assumptions on the Radon measure π, i.e., to separate the small jumps from the large jumps so that only assumption

$$\int_{\mathbb{R}^d} \big(|y|^2 \wedge 1\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0. \tag{1.63}$$

is needed. For instance, the Cauchy process in $\mathbb{R}^d$, where $\sigma = 0$ and the Radon measure π has the form

$$\int_{\mathbb{R}^d} \varphi(y)\pi(\mathrm{d}y) = \lim_{\varepsilon\to 0} \int_{|y|\geq\varepsilon} \varphi(y)|y|^{-d-1}\mathrm{d}y,$$

π does not integrate the function $\varphi(y) = |y|$, and

$$\begin{aligned}&\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1 - \mathtt{i}(\varphi(t),y)\mathbb{1}_{|y|\leq 1}\big]|y|^{-d-1}\mathrm{d}y\Big) = \\ &= \exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d} 2\big[\cos(\varphi(t),y) - 1\big]|y|^{-d-1}\mathrm{d}y\Big),\end{aligned}$$

replaces the second exponential in (1.59). Sometimes, we require a stronger (at the origin) integrability assumption on the Radon measure π, namely,

$$\int_{\mathbb{R}^m} \big(|y| \wedge 1\big)\pi(\mathrm{d}y) < \infty, \qquad \pi(\{0\}) = 0.$$

and the second exponential in (1.59) takes the form

$$\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)} - 1\big]\pi(\mathrm{d}y)\Big),$$

for instance, the case of the Γ-process in $\mathbb{R}^d$, $d = 1$ with parameters $c, \alpha > 0$, where $\sigma = 0$ and the measure π is given by

$$\int_{\mathbb{R}} \varphi(y)\pi(\mathrm{d}y) = \lim_{\varepsilon\to 0} c\int_{\varepsilon}^{\infty} \varphi(y)y^{-1}\mathrm{e}^{-\alpha y}\mathrm{d}y,$$

π does not have a finite mass, and

$$\exp\Big(c\int_{\mathbb{R}}\mathrm{d}t\int_0^\infty\big[\mathrm{e}^{\mathrm{i}\varphi(t)y}-1\big]y^{-1}\mathrm{e}^{-\alpha y}\mathrm{d}y\Big)$$

replaces the second exponential in (1.59).

The theory of martingales (see Section 3.4 later on) shows that the Lévy (martingale) process ℓ can be written as a continuous part (its Wiener process) and a purely discontinuous part (its Poisson jumps part). Alternatively, we can split the $\mathbb{R}^d$ space into $\mathbb{R}^n\times\mathbb{R}^m$, namely, $\omega=(\omega_{n\curvearrowleft},\omega_{\curvearrowright m})$ where $\omega_{n\curvearrowleft}$ and $\omega_{\curvearrowright m}$ are tempered distributions in $\mathbb{R}$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Thus if $\varphi(t)=(\varphi_{n\curvearrowleft}(t),\varphi_{\curvearrowright m}(t))$, where $\varphi_{n\curvearrowleft}(t)$ and $\varphi_{\curvearrowright m}(t)$ denote test functions in $\mathbb{R}$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, then $\langle\omega,\varphi\rangle=\langle\omega_{n\curvearrowleft},\varphi_{n\curvearrowleft}\rangle+\langle\omega_{\curvearrowright m},\varphi_{\curvearrowright m}\rangle$. Hence, we have a $(n+m)$-dimensional *smoothed* (1-parameter) Wiener-Poisson (Lévy) noise, i.e.,

$$X_{n\curvearrowleft}(\varphi,\omega):=\langle\omega_{n\curvearrowleft},\varphi_{n\curvearrowleft}\rangle,\qquad X_{\curvearrowright m}(\varphi,\omega):=\langle\omega_{\curvearrowright m},\varphi_{\curvearrowright m}\rangle,$$

the action-projection on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Clearly, $X_{n\curvearrowleft}$ provides a Wiener process independent of the Poisson martingale measure obtained from $X_{\curvearrowright m}$.

Therefore, by considering the vector-valued space $L^2_{\sigma,\pi}(\mathbb{R};\mathbb{R}^{n+m})$ where we have separate the first n components from the last m components, we can construct (almost sure defined) a n-dimensional Wiener process $w_i(t):=X(\varphi_{i,t})$ for $i=1,2,\dots,n$ (with diffusion matrix $\sigma^*\sigma/2$) and a m-dimensional Poisson martingale measure $q_i(t):=X(\varphi_{i,t})$ for $i=n+1,n+2,\dots,n+m$ (with Levy measure π, so that its jumps Δq_i form a Poisson point process). Indeed, the stochastic process

$$X_t=x+\big(w_1(t),\dots,w_n(t),q_1(t),\dots,q_m(t)\big),\quad\forall\,t\ge0,\,x\in\mathbb{R}^{n+m}\qquad(1.64)$$

(also denoted by X_t^x) has orthogonal increments, which implies that $(X_t:t\ge0)$ is stationary with independent increments, i.e., a Lévy process in law. To take a cad-lag version (which results continuous in the first n components) under assumption (1.60), we may use the estimates

$$\begin{aligned}&\mathbb{E}\big\{|w_i(t)-w_i(s)|^4\big\}=\mathbb{E}\big\{\big(w_i(t-s)\big)^4\big\}\le C|t-s|^2,\\&\mathbb{E}\big\{|q_j(s+r)-q_j(t)|^2|q_j(t)-q_j(s)|^2\big\}=\\&\qquad=\mathbb{E}\big\{\big(q_j(s+r-t)\big)^2\big\}\,\mathbb{E}\big\{\big(q_j(t-s)\big)^2\big\}\le Cr^2,\end{aligned}$$

for any i,j, $0\le s\le t\le s+r\le T$, any $T>0$ and some positive constant $C=C_T$. However, (for the Poisson point process) if only condition (1.63) holds then we can obtain suitable estimates using the equality (1.62). We have then described a way of constructing these processes.

Actually, the only properties used in Lévy's Theorem 1.40 is the fact that the complex-valued characteristic function Ψ is continuous (at zero suffices), positive definite and $\Psi(0)=1$. Indeed, this generalizes to separable Hilbert

spaces, e.g., see the book Da Prato and Zabczyk [51, Theorem 2.13, pp. 49–52], by adding an extra condition on Ψ. Recall that on a separable Hilbert space H, a mapping $S : H \to H$ is called a nuclear (or trace class) operator if for any (or some) orthonormal basis $\{e_i : i \geq 1\}$ in H the series $\sum_i |(Se_i, e_i)|$ is convergent. On the other hand, $\sigma : H \to H$ is called a Hilbert-Schmidt operator if for any (or some) orthonormal basis $\{e_i : i \geq 1\}$ in H the series $\sum_i (\sigma e_i, \sigma e_i)$ is finite.

Theorem 1.41 (Sazonov)**.** *A complex-valued function Ψ on a separable Hilbert space H is the characteristic function of a probability measure P on $(H, \mathcal{B}(H))$ if and only if* (a) *Ψ is continuous,* (b) *is positive definite,* (c) *$\Psi(0) = 1$ and satisfies the following condition:*

(d) *for every $\varepsilon > 0$ there exists a nonnegative nuclear (or trace class) operator S_ε such that each h in H with $(S_\varepsilon h, h) \leq 1$ yields $1 - \Re\{\Psi(h)\} \leq \varepsilon$.* □

Let $\sigma_i : H_0 \to H_0$ $(i = 1, 2)$ be two (symmetric) Hilbert-Schmidt operators on a separable Hilbert space H_0 with inner product $(\cdot, \cdot)_0$ and norm $|\cdot|_0$. Now, on the Hilbert space $H = L^2(\mathbb{R}, H_0^2)$, $H_0^2 = H_0 \times H_0$, consider the *characteristic* function

$$\begin{cases} \Psi(h_1, h_2) = \exp\Big(-\dfrac{1}{2} \displaystyle\int_{\mathbb{R}} |\sigma_1 h_1(t)|_0^2 \mathrm{d}t \Big) \times \\ \quad \times \exp\Big(\displaystyle\int_{\mathbb{R}} \mathrm{d}t \int_{H_0} \big[\mathrm{e}^{\mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0} - 1 - \mathtt{i}(\sigma_2 h_2(t), \sigma_2 u)_0 \big] \pi(\mathrm{d}u) \Big), \end{cases} \tag{1.65}$$

where π is a measure on $\mathcal{B}(H_0)$ such that

$$\int_{H_0} \big(|\sigma_2 u|_0^2 \wedge |\sigma_2 u|_0 \big) \pi(\mathrm{d}u) < \infty, \qquad \pi(\{0\}) = 0. \tag{1.66}$$

Under these assumptions the function Ψ is continuous on H, positive definite, $\Psi(0) = 1$ and the condition (d) of Theorem 1.41 is satisfied for a given $\varepsilon > 0$ with a trace class operator $S_\varepsilon : H \to H$ of the form

$$S_\varepsilon((b_k, b_\ell)e_j) = \begin{cases} (\sigma_1^* \sigma_1 b_k, \sigma_2^* \sigma_2 b_\ell) e_j & \text{if } j \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

for any $k, \ell = 1, \ldots,$ and for some $n = n(\varepsilon)$, where $\{e_j : j \geq 1\}$ is an orthonormal basis in Lebesgue space $L^2(\mathbb{R})$ and σ_i^* is the adjoint of σ_i, $i = 1, 2$, while $\{b_k : k \geq 1\}$ and $\{(b_k, b_\ell) : k, \ell \geq 1\}$ are orthonormal basis in the spaces H_0 and H_0^2, this means that,

$$\big(S_\varepsilon h, (b_k, b_\ell) e_j \big)_H = \int_{\mathbb{R}} \big[(\sigma_1 h_1(s), \sigma_1 b_k)_0 + (\sigma_2 h_2(s), \sigma_2 b_\ell)_0 \big] e_j(s) \mathrm{d}s,$$

for every $h = (h_1, h_2)$, with h_i in H_0, for any $k, \ell = 1, \ldots,$ and $j = 1, \ldots, n$ (otherwise, the left-hand term vanishes), where $(\cdot, \cdot)_H$ denotes the inner product in H.

Therefore Ψ is the characteristic function of a probability measure P on the Hilbert space H, i.e.,

$$\mathbb{E}\big\{ \mathrm{e}^{\mathtt{i}(h, \cdot)_H} \big\} = \Psi(h_1, h_2), \quad \forall h = (h_1, h_2) \in H,$$

where $(\cdot,\cdot)_H$ denotes the inner product in $H = L^2(\mathbb{R}, H_0^2)$. Hence a cad-lag version of a Lévy process on $\mathbb{R}$ or $[0,\infty)$ with parameters $(\sigma_1, \sigma_2, \pi)$ and values in H_0 is obtained as previously discussed in $\mathbb{R}^n \times \mathbb{R}^m$. Thus, the Lévy measure $\pi(\sigma_2^*\sigma_2)^{-1}$ is defined on the Hilbert space image $H_2 = \sigma_2^*\sigma_2(H_0)$ and the probability P can be considered on canonical sample space $\Omega = D([0,\infty), H_1 \times H_2)$ or $\Omega = D([0,\infty), H_1) \times D([0,\infty), H_2)$, with $H_1 = \sigma_1^*\sigma_1(H_0)$, where the canonical process $X(\omega) = \omega(t)$ has Ψ as its characteristic function. Clearly, a drift can be added and the parameters $(\sigma_1, \sigma_2, \pi)$ can be time-dependent with suitable assumptions.

The above arguments extend to the case of a countably Hilbertian space (of which a typical example is the space $\mathcal{S}(\mathbb{R}^d)$ of rapidly decreasing smooth functions with its dual $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions), where the role the Hilbert-Schmidt operators σ_i is better understood.

A *countably Hilbertian space* K is a separable Fréchet (i.e., complete locally convex topological) space where the topology is given by an increasing sequence $\{\|\cdot\|_n : n \geq 0\}$ of compatible (i.e., any Cauchy sequence in two norms and convergent to zero in one norm results convergent to zero also in the other norm) Hilbertian norms. Moreover, a space K is called *nuclear* if for any $n \geq 0$ there exists $m > n$ such that the canonical injection from K_m into K_n is Hilbert-Schmidt, where K_n denote the completion of K with the Hilbertian norm $\|\cdot\|_n$. Thus K_n is a sequence of decreasing Hilbert spaces and $K = \cap_n K_n$. Next, if we identify K_0 with its dual space K_0' (by Riezs' representation theorem) and we denote the dual space K_n' by K_{-n} (with its dual Hilbertian norm $\|\cdot\|_{-n}$, $n \geq 1$) then K_{-n} is a sequence of increasing Hilbert spaces, the dual space K' is sequentially complete and $K' = \cup_n K_{-n}$.

Theorem 1.42 (Minlos)**.** *A complex-valued function Ψ on a countably Hilbertian nuclear space K is the characteristic function of a probability measure P on the dual space $(K', \mathcal{B}(K'))$ if and only if Ψ is continuous at 0 in K, positive definite and $\Psi(0) = 1$.* □

Note that if K is a countably Hilbertian nuclear space then so is $\mathcal{S}(\mathbb{R}^d, K)$ (for instance, regarding $\mathcal{S}(\mathbb{R}^d, K)$ as the tensor product $\mathcal{S}(\mathbb{R}^d, K) = \mathcal{S}(\mathbb{R}^d) \otimes K$) and $K = \mathcal{S}(\mathbb{R}^d; \mathbb{R}^m)$ with $K' = \mathcal{S}'(\mathbb{R}^d; \mathbb{R}^m)$ is a typical example. Also $C([0,\infty), X)$ is a Fréchet space if X is so. However, $D([0,\infty), X)$ is a Polish (not a topological vector) space X is so. If $(\cdot,\cdot)$ is continuous inner product in a countably Hilbertian nuclear space K (i.e., the inner product is continuous in K_n for some n) and H is the Hilbert space completion of K with respect to $(\cdot,\cdot)$ then H is called *rigged Hilbert space* in K, and we have the triplet $K \subset H \subset K'$. Certainly, any K_n can be used as H, but this is not necessary in general.

On the other hand, a set A in $D([0,\infty), K')$ (resp. $C([0,\infty), K')$) is relatively compact if and only if one of the following conditions is satisfied:

(1) For any k in K the set $\{\langle \omega(\cdot), k\rangle : \omega \in A\}$ is relatively compact in $D([0,\infty), \mathbb{R})$ (resp. $C([0,\infty), \mathbb{R})$).

(2) For every $T > 0$ there exists n such that A_T the restriction of A to $D([0,T], \mathbb{R})$ (resp. $C([0,T], \mathbb{R})$) is relatively compact in $D([0,T], K_{-n})$ (resp.

$C([0,T],K_{-n}))$.

Clearly, any k in K defines a measurable map π_k from $D([0,\infty),K')$ (resp. $C([0,\infty),K')$) into $D([0,\infty),\mathbb{R})$ (resp. $C([0,\infty),\mathbb{R})$), $\pi_k(t,\omega)=\langle\omega,k\rangle$. Then a sequence $\{\mu_i : i\geq 1\}$ is tight in $D([0,\infty),K')$ (resp. $C([0,\infty),K')$) if and only if for every k in K the sequence $\{\mu_i\pi_k^{-1} : i\geq 1\}$ is tight as a Borel probability measure in $D([0,\infty),\mathbb{R})$ (resp. $C([0,\infty),\mathbb{R})$). Moreover, if for every $T>0$ there is n with the property that for every $\varepsilon>0$ there exists $M>0$ such that

$$\mu_i\big(\{\omega\in D([0,T],K') : \sup_{0\leq t\leq T}|\omega(t)|_{-n}\leq M\}\big)\geq 1-\varepsilon,$$

for every $i\geq 1$, then the sequence $\{\mu_i : i\geq 1\}$ regarded as Borel probability measure in $D([0,T],K_{-m})$ is tight, with $m\geq n$ such that the canonical injection from K_m into K_n (and so from K_{-n} into K_{-m}) is Hilbert-Schmidt.

Hence if $K\subset H_i\subset K'$, $i=1,2$ are two rigged Hilbert spaces then there is a probability measure P on $\mathcal{S}'(\mathbb{R}^n;H_1\times H_2)$ with characteristic function

$$\begin{cases}\mathbb{E}\big\{\exp\big(\mathtt{i}[(\varphi_1,\cdot)_1+(\varphi_1,\cdot)_2]\big)\big\}=\exp\Big(-\dfrac{1}{2}\displaystyle\int_{\mathbb{R}^n}|\varphi_1(t)|_1^2\mathrm{d}t\Big)\times\\ \qquad\times\exp\Big(\displaystyle\int_{\mathbb{R}}\mathrm{d}t\int_{H_2}\big[\mathrm{e}^{\mathtt{i}(\varphi_2(t),u)_2}-1-\mathtt{i}(\varphi_2(t),u)_2\big]\pi(\mathrm{d}u)\Big),\end{cases}\tag{1.67}$$

where π is a Radon measure on H_2 satisfying

$$\int_{H_2}\big(|u|_2^2\wedge|u|_2\big)\pi(\mathrm{d}u)<\infty,\qquad \pi(\{0\})=0,\tag{1.68}$$

and $(\cdot,\cdot)_i$, $|\cdot|_i$ denote the inner product and the norm in H_i, $i=1,2$. By comparison with (1.65) and (1.66) we see that the nuclear (or trace class) operators σ_1, σ_2 are really part of the Hilbert space where the Lévy process takes values. Moreover, the parameter t may be in $\mathbb{R}^d$ and a Lévy noise is realized as a generalized process.

For instance, the reader is referred to the book by Kallianpur and Xiong [123, Chapters 1 and 2, pp, 1–83] for details on most of the preceding definitions.

If the probability to be constructed is not space-homogeneous (i.e., it is non-stationary) then the canonical process $(X_t^x : t\geq 0)$ does not define a Markov process under P. Thus, if for each x in $\mathbb{R}^d$ we have a $d\times d$ square matrix $\sigma(x)$ and a Radon measure $\pi(x,\mathrm{d}y)$ in $\mathbb{R}^d$ as before, then for every function ψ in $L^2(\mathbb{R},\mathbb{R}^d)$, we can construct (assuming some condition on the x-dependency of σ and π) a probability measure $Q(\psi,\cdot)$ on $\Omega=\mathcal{S}'(\mathbb{R};\mathbb{R}^d)$ such that its characteristic function satisfies

$$\begin{cases}\displaystyle\int_\Omega\mathrm{e}^{\mathtt{i}\langle\omega,\varphi\rangle}Q(\psi,\mathrm{d}\omega)=\exp\Big(-\frac{1}{2}\int_{\mathbb{R}}|\sigma(\psi(t))\varphi(t)|^2\mathrm{d}t+\\ \qquad+\displaystyle\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}\big[\mathrm{e}^{\mathtt{i}(\varphi(t),y)}-1-\mathtt{i}(\varphi(t),y)\big]\pi(\psi(t),\mathrm{d}y)\Big),\end{cases}\tag{1.69}$$

Next, the expected Markov process is the $\mathbb{R}^d$-valued canonical process

$$X^x(t,\omega)=(X_i^x(t,\omega) : i=1,\dots,d),\quad X_i^x(t,\omega):=x_i+\langle\omega_i,\mathbb{1}_{(0,t]}\rangle$$

under the probability P^x, which is defined as the conditional probability

$$P^x\{\cdot \,|\, X^x\} = Q(X^x, \cdot).$$

Certainly, a drift and a killing terms can be added, and much more details are needed to complete this procedure. This involves the so called *pseudo-differential operators*, see the treatise by Jacob [115].

Related to the re-construction of probability is the following result (e.g., see Stroock and Varadhan [241, Theorem 1.3.5, pp. 34-36]). Let X be the canonical process in the canonical space either $C([0,\infty), E)$ or $D([0,\infty), E)$, where E is a Polish space. Assume $\{\tau_n : n \geq 0\}$ is a nondecreasing sequence of stopping times relative to the filtration $\{\mathcal{F}(t) : t \geq 0\}$, where $\mathcal{F}(t)$ and $\mathcal{F}(\infty)$ are the σ-algebras generated by $\{X(s) : 0 \leq s \leq t\}$ and $\{X(t) : t \geq 0\}$, respectively. Now, for each $n \geq 0$ let P_n be a probability measure defined on $\mathcal{F}(\tau_n)$. If $\lim_n P_n\{\tau_n \leq t\} = 0$ for every $t \geq 0$, and the probability P_{n+1} coincides with P_n on $\mathcal{F}(\tau_n)$ for any n, then there exists a probability measure P on $\mathcal{F}(\infty)$, which coincides with P_n on $\mathcal{F}(\tau_n)$ for every $n \geq 0$. Moreover, the same conclusion is true if $\mathcal{F}(t)$ is replaced by $\mathcal{F}(t+)$.

The reader interested in a guided tour to measure theoretic probability may take a look at the recent book by Pollard [200].

1.15 Exercises

This part is not intended as real exercises, but as guide and a complement to the previous sections, helping to clarify and specify some statements given previously. The reader may take a look at the book Chaumont and Yor [39].

Exercise 1.2. If $\mathcal{F}$ is a collection (or class) of subsets of Ω then $\mathcal{F}$ is called a π-system if for any A and B in $\mathcal{F}$ we have $A \cap B$ in $\mathcal{F}$ and a λ-system if Ω belongs to $\mathcal{F}$ and satisfies (a) for any A and B in $\mathcal{F}$ with $A \subset B$ we have $B \setminus A$ belongs to $\mathcal{F}$ and (b) for any monotone increasing sequence of sets $A_i \subset A_{i+1}$ we have $A = \bigcup_i A_i$ in $\mathcal{F}$. Prove the following versions of monotone class theorem: (1) If $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{G}$ is a π-system and $\mathcal{F}$ is a λ-system then the σ-field or σ-algebra $\sigma(\mathcal{G})$ generated by $\mathcal{G}$ is contained in $\mathcal{F}$. (2) If $\mathcal{F}$ is a π-system and $\mathcal{H}$ a linear space of functions from Ω into $\mathbb{R}$ such that $\mathbb{1}_\Omega$, $\mathbb{1}_A$ and $\varphi = \lim_i \varphi$ belong to $\mathcal{H}$, for every A in $\mathcal{F}$ and for any sequence φ_i in $\mathcal{H}$ such that $0 \leq \varphi_i \leq \varphi_{i+1}$, $\varphi_i(\omega) \to \varphi(\omega)$ and $\varphi(\omega)$ is finite for any ω, then $\mathcal{H}$ contains all $\sigma(\mathcal{F})$-measurable functions. □

Exercise 1.3. Let $\vee_n \mathcal{F}_n$ be the σ-algebra generated by sequence $\{\mathcal{F}_n : n \geq 1\}$ of sub σ-algebra in a probability space $(\Omega, \mathcal{F}, P)$. Use an argument of monotone class to show that for any set A in $\vee_n \mathcal{F}_n$ there exists a sequence $\{A_k : k \geq 1\}$ of sets in $\bigcup_n \mathcal{F}_n$ such that $P(A \setminus A_k)$ and $P(A_k \setminus A)$ converge to 0, e.g., see Kallenberg [121, Lemma 3.16, p. 54]. □

Exercise 1.4. Let x be a function from a measurable space $(\Omega, \mathcal{F})$ into a Polish space E (complete metric space), so X is a random variable with values in E.

Denote by $\mathcal{F}_x$ the σ-algebra generated by x. Use an argument of monotone class to show that any real valued random variable φ which is $\mathcal{F}_x$-measurable has the form $\varphi(\omega) = f(x(\omega))$, for some Borel function f from E into $\mathbb{R}$, e.g., see He et al. [105, Theorem 1.5, p. 5] or Kallenberg [121, Lemma 1.13, p. 7]. □

Exercise 1.5. Let $X(t,\omega)$ be a function from $T \times \Omega$ into $\mathbb{R}^d$, where T is a countable and $(\Omega, \mathcal{F})$ is a measurable space. Prove that the following statements are equivalent:

(a) The function $(t,\omega) \mapsto X(t,\omega)$ from is Borel measurable with respect to the product σ-algebra $\mathcal{B}(T) \times \mathcal{F}$.

(b) The function $\omega \mapsto X(\cdot,\omega)$ from Ω into $\mathcal{L}^0(T, \mathbb{R}^d)$.

Discuss possible extensions to the case where T is a Borel subset of $\mathbb{R}$, e.g. see Doob [60, Theorem 2.1.13 in p. 408]. □

Exercise 1.6. Prove properties (a) to (i) for optional or stopping times. □

Exercise 1.7. Calculate the mean and the covariance of random variables with a Gaussian, Poisson and exponential distributions. Moreover, show that if x is a Gaussian variable with variance r then the even moments can be calculate by recurrence, i.e., $\mathbb{E}\{|x|^{2n+2}\} = r(2n-1)\mathbb{E}\{|x|^{2n}\}$, for any integer $n \geq 1$. □

Exercise 1.8. Let $(\Omega, \mathcal{F})$ be a measurable space. Recall that a π-systems $\mathcal{F}_0$ is a subset of $\mathcal{F}$ which is stable under finite intersections, i.e., if A and B belongs to $\mathcal{F}_0$ then $A \cap B$ also belongs to $\mathcal{F}_0$. Also, we denote by $\sigma(\mathcal{F}_0)$ the minimal sub σ-algebra of $\mathcal{F}$ containing all the elements of $\mathcal{F}_0$, i.e. the σ-algebra generated by $\mathcal{F}_0$. Prove that if $\mathcal{H}$ and $\mathcal{G}$ are two sub σ-algebras which are generated by the π-systems $\mathcal{H}_0$ and $\mathcal{G}_0$, then $\mathcal{H}$ and $\mathcal{G}$ are independent if and only if $\mathcal{H}_0$ and $\mathcal{G}_0$ are independent, i.e., if and only if $P(H \cap G) = P(H)P(G)$ for any H in $\mathcal{H}_0$ and G in $\mathcal{G}_0$ (e.g., see the book by Bauer [15, Section 5.1, pp. 149–154]) □

Exercise 1.9. Prove that a linear combination of Gaussian random variables is also a Gaussian random variable. Calculate its mean and covariance, and check that all moments are finite. □

Exercise 1.10. Establish the existence for the conditional expectation on a given probability space $(\Omega, \mathcal{F}, P)$ for an integrable random variable x with respect to a sub σ-algebra $\mathcal{G}$ by two ways. Firstly (a) by means of the Radon-Nikodym theorem, i.e., on the measurable space $(\Omega, \mathcal{G})$ consider the probability measures $\nu(G) = \mathbb{E}\{x\mathbb{1}_G\}$ and $\mu(G) = \mathbb{E}\{\mathbb{1}_G\}$ satisfying $\nu \ll \mu$. Secondly (b) by means of the orthogonal projection π from the Lebesgue space $L^2(\Omega, \mathcal{F}, P)$ into the closed subspace $L^2(\Omega, \mathcal{G}, P)$, i.e., π satisfies $(x - \pi(x), y) = 0$, for any y in $L^2(\Omega, \mathcal{G}, P)$, where $(\cdot,\cdot)$ denotes the scalar product. □

Exercise 1.11. Let x, y be real random variables on a complete probability space. If z is a random variable with values in some Polish space E then prove that the relation $x = \mathbb{E}\{y \mid z\}$ is characterized by the condition $\mathbb{E}\{y\varphi(z)\} = \mathbb{E}\{x\varphi(z)\}$, for all $\varphi : E \longrightarrow \mathbb{R}$ which is bounded and continuous. Moreover, if E is locally compact, then the class of continuous function with compact support is sufficient to characterized the conditional expectation. □

Exercise 1.12. Prove properties (a) to (g) of the conditional expectation. □

Exercise 1.13. Prove properties (h) to (m) of the conditional expectation. □

Exercise 1.14. Let $\mathcal{G}$ be the σ-algebra generated by a sequence $\{x_i : i = 1, 2, \ldots\}$ of measurable functions from $(\Omega, \mathcal{F})$ into $(\mathbb{R}, \mathcal{B})$, and x be an integrable random variable in the complete probability space $(\Omega, \mathcal{F}, P)$ with values in $\mathbb{R}^n$. Use an argument of *monotone class* to show that $\mathbb{E}\{x \mid \mathcal{G}\} = 0$ if and only if $\mathbb{E}\{f(x_1, x_2, \ldots, x_m)\, x\} = 0$ for any m and any bounded continuous function f from $\mathbb{R}^n$ into $\mathbb{R}$, e.g. see Yong and Zhou [261, Proposition 1.12 in p. 13]. □

Exercise 1.15. Prove that if x is in $L^1(\Omega, \mathcal{F}, P)$ then the family of elements in $L^1(\Omega, \mathcal{F}, P)$, defined by $\{y = \mathbb{E}\{x \mid \mathcal{G}\} : \mathcal{G}$, is a sub σ-algebra of $\mathcal{F}\}$ is uniformly integrable. Indeed use Jensen's inequality to establish that $kP\{|y| > k\} \leq \mathbb{E}\{|x|\}$, for any k, and in view of

$$\int_{|y|>k} |y(\omega)| P(\mathrm{d}\omega) \leq \int_{|y|>k} |x(\omega)| P(\mathrm{d}\omega),$$

the desired result follows. □

Exercise 1.16. On a probability space $(\Omega, \mathcal{F}, P)$, let x be a real random variable independent of a sub σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and f be a bounded Borel measurable function in $\mathbb{R}^2$. Define $f_1(\eta) = \mathbb{E}\{f(x, \eta)\}$. Prove that f_1 is Borel measurable and $f_1(y) = \mathbb{E}\{f(x, y) \mid \mathcal{G}\}$ almost surely. □

Exercise 1.17. Let $\mathcal{G}$ be a finitely-generated σ-algebra, i.e., $\mathcal{G} = \sigma[F_1, \ldots, F_n]$. First, prove that also $\mathcal{G}$ can be expressed as $\sigma[G_1, \ldots, G_m]$, where the sets $G_1, \ldots, G_m$ are pairwise disjoint and minimal in the sense that any proper subset of $\{G_1, \ldots, G_m\}$ does not generate $\mathcal{G}$. Actually, $\{G_1, \ldots, G_m\}$ is a partition and the set G_i are called *atoms* of $\mathcal{G}$, which has exactly 2^m elements. Second, gives an explicit expression of $P\{A \mid \mathcal{G}\}(\omega)$ in term of the family of sets $G_1, \ldots, G_m$. Third, if X is a simple random variable (i.e., having a finite number of values, say $x_1, \ldots, x_m$ with $P\{X = x_i\} > 0$ and $\sum_i P\{X = x_i\} = 1$) then show that $\sigma(X)$ (i.e., the minimal σ-algebra for which X is measurable) is finitely-generated, calculate $P\{A \mid X = x_i\}$, for $i = 1, \ldots, m$ and consider the function $x \mapsto P(x, A)$ defined as $P(x, A) = P\{A | X = x_i\}$ if $x = x_i$ for some $i = 1, \ldots, m$, and $P(x, A) = P(A)$ otherwise. Fourth, show that the expression $P(X, A)$ is a regular conditional probability of A given X, i.e., for any A measurable set we have $P\{A \mid X\} = P(X, A)$ almost surely, see Remark 1.17. □

Exercise 1.18. Let $\{X(t) : t \geq 0\}$ be a (separable) stochastic process on the probability space $(\Omega, \mathcal{F}, P)$ with valued into $\mathbb{R}$. Prove that if X is either right or left continuous in probability then any dense set Q on $[0, \infty)$ is separant. □

Exercise 1.19. On a probability space $(\Omega, \mathcal{F}, P)$, let $\{X_n(t) : t \geq 0\}$, $n \geq 1$ be a sequence of families of random variables and $\mathcal{F}_n(t)$ be the σ-algebra generated by the random variables $\{X_n(s) : 0 \leq s \leq t\}$ and all sets of measure zero in $\mathcal{F}$. Assume that $X_n(t)$ converges in probability to $X(t)$, for every $t \geq 0$. Prove that

$$\lim_{n\to\infty} \mathbb{E}\{y \mid \mathcal{F}_n(t)\} = \mathbb{E}\{y \mid \mathcal{F}(t)\}, \quad \forall t \geq 0,$$

for every integrable random variable y such that the above limit exists in probability. □

Exercise 1.20. Let $\mathcal{H}$ be a σ-algebra of a probability space $(\Omega, \mathcal{F}, P)$. Define $\mathcal{I}$ the collection of all sets in $\mathcal{F}$ independent of $\mathcal{H}$. Prove that $\mathcal{I}$ is closed under monotone union and intersection. Deduce that $\mathcal{I}$ is the σ-algebra $\mathcal{H}^\perp$ generated by all sets in $\mathcal{F}$ independent of $\mathcal{H}$. □

Exercise 1.21. Let $X = (X_t : t \geq 0)$ be a family of random variables with values in a complete separable metric (Polish) space E defined on a probability space $(\Omega, \mathcal{F}, P)$. Assume that X satisfies the Markov property 1.35 and suppose that X (regarded as a process) is right-continuous in probability, i.e., for every $\varepsilon > 0$ and $t \geq 0$ there exists $\delta > 0$ such that $P\{|X_s - X_t| \geq \varepsilon\} < \varepsilon$ for every s in $(t, t+\delta)$. Prove that X satisfies the Markov property as in Definition 1.27 with the natural filtration $(\mathcal{F}_t : t \geq 0)$, i.e., the minimal increasing family of σ-algebra satisfying the usual conditions such that X_t is $\mathcal{F}_t$-measurable for every $t \geq 0$. □

Exercise 1.22. Let X be a $\mathbb{R}^d$-valued adapted stochastic process in $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and suppose that for some transition function $p(s, x, t, \mathrm{d}y)$ we have

$$P\{h(X_t) \in B \mid \mathcal{F}_s\} = \int_{\mathbb{R}^d} h(y) p(s, X_s, t, \mathrm{d}y), \quad \forall t > s \geq 0,$$

almost surely, for every $t > s \geq 0$, and any real-valued continuous and bounded function h. Verify that X is indeed a Markov process. Which other classes of functions h could be used? How about processes taking values in some topological space E instead of $\mathbb{R}^d$? □

Exercise 1.23. Let $X = (X_t : t \geq 0)$ be a (strong) Markov process with values in a complete separable metric (Polish) space E defined on a probability space $(\Omega, \mathcal{F}, P)$, and with transition probability function $p(s, x, t, B)$, $t > s \geq 0$, x in E and B in $\mathcal{B}(E)$. If necessary, assume that for every B the mapping $(s, x, t) \mapsto p(s, x, t, B)$ is measurable. Define

$$\dot{p}((s, x), t, \mathrm{d}r \times \mathrm{d}e) := \delta_{s+t}(\mathrm{d}r)\, p(s, x, s+t, \mathrm{d}e), \quad \forall t > 0,$$

for all (s, x) in $\dot{E} := [0, \infty) \times E$, and where $\delta_s(\mathrm{d}r)$ is the Dirac unit mass in $[0, \infty)$ concentrated at $r = s$. Prove that $\dot{X} := (t, X_t)$ is a (strong) homogeneous Markov process with values in $\dot{E}$ and with transition probability function $\dot{p}((s, x), t, B)$, $t > 0$, (s, x) in $\dot{E}$ and B in $\mathcal{B}(\dot{E})$. □

Exercise 1.24. Prove that if X is a d-dimensional stochastic process with independent and stationary increments (see properties (a) and (b) of Lévy processes) then

$$\mathbb{E}\{\varphi(X(s) - X(t))\psi(X(t))\} = \mathbb{E}\{\varphi(X(s) - X(t))\}\, \mathbb{E}\{\psi(X(t))\},$$
$$\mathbb{E}\{\varphi(X(s) - X(t))\} = \mathbb{E}\{\varphi(X(s - t))\},$$

for any continuous function from $\mathbb{R}^d$ into $\mathbb{R}$ with compact support. □

Exercise 1.25. Prove that any characteristic function possesses the properties (a) and (b) of continuity and positive define stated in the text. □

Exercise 1.26. Verify that the characteristic functions of a Lévy process have exponential form. Discuss to what this correspond on the finite-dimensional distributions viewpoint. □

Exercise 1.27. Verify that the consistency properties are satisfied for a family of finite-dimensional distributions constructed from a transition function $P(s,x,t,A)$, an initial time t_0 and probability P_0. □

Exercise 1.28. Prove that for a Lévy process (P_X, X) the continuity condition (1.11) reduces to

$$\mathbb{E}_X\{|X_h|^\alpha\} \le Ch^{1+\beta} \qquad \forall h > 0,$$

for a some positive constants α, β and C. Similarly the cad-lag condition (1.12) can be expressed by

$$\mathbb{E}_X\{|X_h|^\alpha\} \le Ch^{\frac{1}{2}+\beta} \qquad \forall h > 0,$$

for a some positive constants α, β and C. Extend this result to processes generated by a transition function i.e. to Markov processes. □

Exercise 1.29. By means of the finite-dimensional distributions proves that the Wiener process satisfies the continuity condition (1.11) so that its paths are continuous. □

Exercise 1.30. Show that the continuity condition (1.11) is not satisfied for the Poisson process but a direct calculation proves that it is continuous in probability, see property (c) of Definition 1.31. □

Exercise 1.31. Proved that the cad-lag condition (1.12) is satisfied for the Cauchy process i.e. there exist positive constants α, β and C such that

$$\int_E |x-y|^\alpha P(s,x,t,\mathrm{d}y) \le C|t-s|^{\frac{1}{2}+\beta}, \quad \forall s,t \in [0,T],\ \forall x \in \mathbb{R}$$

for ant $T > 0$. □

Exercise 1.32. Let $\{\rho(t,i,j) : i,j = 1,\dots,n\}$ be a family of continuous functions from $[0,\infty)$ into $\mathbb{R}$ satisfying

$$\rho(t,i,j) \ge 0, \quad \forall i \ne j, \qquad \rho(t,i,i) = -\sum_{j\ne i} \rho(t,i,j).$$

Consider the n-dimensional system of ordinary differential equations

$$\dot{p}_s(t,i,j) = \sum_k p_s(t,i,k)\,\rho(t,k,j), \quad \forall t > s,\ i,j.$$

where the dot means derivative in t and p_s is the fundamental solution, i.e. it satisfies $p_s(s,i,j) = \delta_{i,j}$.

(a) First, assume that the data are constants i.e., $\rho(t,i,j) = \rho_(i,j)$ and denote by ρ the square matrix $(\rho_(i,j) : i,j = 1,\dots,n)$. Show that in this case, the fundamental solution $p_s(t,i,j) = \exp[-(t-s)\rho_d]\exp[(t-s)(\rho-\rho_d)]$, where ρ_d is the diagonal matrix with the coefficients $(\rho_{i,i} : i = 1,\dots,n)$. Conclude that all entries of $p_s(t,i,j)$ are non-negative and that each line adds (sum in j) to one.

(b) Extend the previous conclusion to the general case where the data ρ may depend on t.

(c) Assume the data are constant in t as in (a) and define $\lambda = -\inf_i \rho_i, i > 0$. Let Y_n be a Markov chain with transition probability $P(Y_n = j \,|\, Y_{n-1} = i) = \rho_{i,j}/\lambda$ if $i \neq j$ for $n = 1,2,\dots,$ and let $\tau_1,\tau_2,\dots$ be a sequence of independent identically distributed exponentially random variables with parameter λ i.e., $P(\tau_i > t) = \exp(-\lambda t)$, which is independent of $(Y_0,Y_1,\dots)$. Prove that $X_t = Y_n$ for t in the stochastic interval $[\![T_n, T_{n+1}[\![$, where $T_0 = 0$ and $T_n = \tau_1+\tau_2+\cdots+\tau_n$, gives a realization of the pure jumps Markov with the above infinitesimal rates (see Durrett [67, p. 250, Example 2.1]).

(d) Discuss the case of the double sequence $\{\rho(t,i,j) : i,j = 1,2,\dots\}$ of continuous functions. □

Exercise 1.33. Let $\mathcal{B}^T(\mathbb{R})$ be the product σ-algebra (i.e., generated by the cylinder sets), which may be smaller that $\mathcal{B}(\mathbb{R}^T)$ (the minimal σ-algebra containing all open sets in $\mathbb{R}^T$, endowed with the product topology). Prove that a typical set in $\mathcal{B}^T(\mathbb{R})$ has the form

$$\{\omega \in \mathbb{R}^T : (\omega(t_1),\omega(t_2),\dots) \in A\}$$

where $A \in \mathcal{B}(\mathbb{R}^{\{1,2,\dots\}})$ and $(t_1,t_2,\dots)$ is a sequence in $\mathbb{R}$. Verify that a singleton (i.e., a set of only one element) belongs to $\mathcal{B}(\mathbb{R}^T)$, but does not belong to $\mathcal{B}^T(\mathbb{R})$ if the index set T is uncountable. □

Exercise 1.34. Let $\mathcal{F}$ be the σ-algebra generated by the coordinate random variables $\omega \mapsto \omega(t)$ from $C([0,\infty),E)$ into E, where E is a complete separable metric space and t ranges over a dense set Q of $[0,\infty)$. Prove that $\mathcal{F} = \mathcal{B}$, where $\mathcal{B}$ is the (Borel) σ-algebra generated by open all sets in $C([0,\infty),E)$. Conclude that a function X from $(C([0,\infty),E),\mathcal{B})$ into itself is measurable if and only if the functions $\omega \mapsto X(t,\omega)$ from $(C([0,\infty),E),\mathcal{B})$ into E are measurable for all t in $[0,\infty)$. Prove the same result for the space $(D([0,\infty),E),\mathcal{B})$, where now $\mathcal{B}$ is the (Borel) σ-algebra generated by open all sets in $D([0,\infty),E)$. □

Exercise 1.35. For a function x from $[0,T]$ into $\mathbb{R}$ which have only discontinuities of first class we define two modulii of continuity $w(x,h)$ and $w'(x,h)$ one by

$$w(x,h,T) : \quad = \quad \inf_{\{t_i\}} \sup_i \sup\{|x(t) - x(s)| : t_{i-1} \leq s < t < t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \cdots < t_n = T$,

with $t_i - t_{i-1} \geq h$ and $n \geq 1$, and the other by

$$w'(x,h) : = \sup\{\sqrt{|x(t+h)-x(s)|\,|x(s)-x(t)|} : \\ : -h < t < s < t+h < T+h\},$$

where we have extended $x(t) = x(0)$ for $t < 0$ and $x(t) = x(T)$ for $t > T$. Prove that for any x in $D([0,T],E)$ we have $w(x,h) \longrightarrow 0$ and $w'(x,h) \longrightarrow 0$ as $h \to 0$. By means of the above modulii of continuity, give a characterization of pre-compact sets in the the space $D([0,T],E)$ endowed with the Skorokhod topology. □

Exercise 1.36. Show that the space $D([0,+\infty],E)$ complete with the locally uniform convergence [i.e., the topology in $C([0,+\infty],E)$], but is not separable. On the other hand, show that $C([0,+\infty],E)$ is a closed subspace of $D([0,+\infty],E)$. □

Exercise 1.37. Let X be a Borel measurable function form Ω into itself, where (Ω, d_Ω) be a separable and complete metric space. Suppose that $\{P_n : n = 1,2,\dots\}$ is a sequence of probability measures on Ω which converges weakly to P. Define $\{Q_n : n = 1,2,\dots\}$ and Q as the image measures through the mapping X of $\{P_n : n = 1,2,\dots\}$ and P. Prove that if X is P-almost surely continuous then the sequence $\{Q_n : n = 1,2,\dots\}$ converges weakly to the measure Q. □

Exercise 1.38. Let Ω be a complete metric space. Use a monotone class argument to show that the smallest class of functions M in $B(\Omega)$ satisfying:

(a) if $\{f_n\}$ is a sequence in M boundedly and pointwise convergent to f then f belongs to $B(\Omega)$,

(b) if A is open in Ω then the characteristic (or indicator) function $\mathbb{1}_A$ belongs to M,

(c) if f and g are in M then $\alpha f + \beta g$ is in M for any constant α and β,

is actually $B(\Omega)$. □

Chapter 2

Semigroup Theory

This is mainly a service Chapter, it complement the previous part in Probability with the *analytic* counterpart. Again the focus is the *connection* between Markov processes and semigroups on function spaces. The first reading may be a little hard, since only references to proofs are given. The reader should be prepared for a *crash introduction* to the area of semigroup theory applied to partial differential equations, starting at a simple level of functional analysis. There are several sources for the semigroup theory, the reader may consult Engel and Nagel [74] and Pazy [196] for a general introduction to the semigroup theory for linear operator, while in Lunardi [160] and Taira [244] a full treatment of analytic semigroups can be found. An approach via pseudo-differential operators can be found in Jacob [115], while Dirichlet form techniques are developed in Fukushima et al. [92]. Also Fattorini [78] and Tanabe [246] have a complete discussion of evolution equations, while Hille and Phillips [108], Kato [128] and Yosida [262] are classical references.

On the other hand, stochastic processes and partial differential equations interplay very deeply, besides semigroup theory, there are many other well established exchanges between probability and analysis, as discussed in the books Bass [12, 13] with great generality. The interested reader may enlarge this point of view by consulting, for instance, the books Doob [60], Durrett [66], Dynkin [70], Pinsky [198], Port and Stone [202], Stroock and Varadhan [241], among others.

2.1 Functions and Operators

Let X and Y be two (complex) Banach spaces. First recall that a linear operator $A\colon X \to Y$ is *continuous* (or equivalently *bounded*) if A maps bounded sets of X into bounded sets of Y. Also, A is called *compact* if it maps bounded set of X into relatively compact sets of Y. Moreover, we denote by $L(X,Y)$ the Banach space of linear and bounded operators from X into Y, endowed with the operator norm $\|A\| = \sup\{\|Ax\| : \|x\| \leq 1\}$. In the case of $X = Y$ we

simply set $L(X) = L(X, X)$ and we speak of operators in X. If K is a linear compact operator between two Banach spaces X and Y, then its transpose or *dual* mapping K' is also a linear compact operator from Y' into X', the dual spaces of Y and X, respectively. If a linear operator $A\colon \mathcal{D}(A) \subset X \to Y$ is defined on a subspace $\mathcal{D}(A)$ of X and is bounded, we could extend the operator A to the closure $\overline{\mathcal{D}(A)}$. In the case of a linear unbounded operator $A\colon \mathcal{D}(A) \subset X \to Y$, A is said to be *closed* if its graph $\{(x, y) \in X \times Y : y = Ax,\ x \in \mathcal{D}(A)\}$ is closed in $X \times Y$. Also, an operator $A\colon \mathcal{D}(A) \subset X \to Y$ is called *closable* if the closure of its graph defines a closed operator.

• *Remark* 2.1. An operator $A\colon \mathcal{D}(A) \subset X \to Y$ is closed if and only if for any sequence $\{x_n\} \subset \mathcal{D}(A)$ converging to x, such that also the sequence $\{Ax_n\}$ is convergent to some y, we have that $x \in \mathcal{D}(A)$ and $y = Ax$. Also, an operator $A\colon \mathcal{D}(A) \subset X \to Y$ is closable if and only if for any sequence $\{x_n\} \subset \mathcal{D}(A)$ converging to 0, such that also the sequence $\{Ax_n\}$ is convergent to some y, we have that $y = 0$. Any bounded operator $T \in L(X)$ is a closed operator. □

Now recall that if A is a linear closed (possibly unbounded) operator in X, the *resolvent set* $\rho(A)$ of A is the set of complex values λ for which $\lambda I - A$ is invertible i.e., $(\lambda I - A)^{-1}$ exists and is a bounded linear operator in X. The *spectrum* $\sigma(A)$ of A is the complement in $\mathbb{C}$ of $\rho(A)$. The family $R(\lambda, A) = (\lambda I - A)^{-1}$, $\lambda \in \rho(A)$ of bounded linear operators is called the *resolvent operators* of A.

It is easy to show the following identity (*resolvent identity*): for any $\lambda, \mu \in \rho(A)$ we have

$$(\lambda I - A)^{-1} - (\mu I - A)^{-1} = (\mu I - \lambda)(\lambda I - A)^{-1}(\mu I - A)^{-1}. \tag{2.1}$$

Fixed a point $\mu \in \rho(A)$, for all λ of the disc of the complex plane centered in μ given by $|\lambda - \mu|\, \|(\mu I - A)^{-1}\| < 1$, we have that the series

$$\sum_{k=0}^{\infty} (\mu I - \lambda)^k (\mu I - A)^{-k-1}.$$

is (operator norm) convergent to $(\lambda I - A)^{-1}$. Thus, the resolvent set $\rho(A)$ is open and the function $\lambda \mapsto R(\lambda, A)$ is (operator norm) analytic thanks to the previous identity. Moreover, since $\|R(\lambda, A)\| \geq 1/\mathrm{dist}(\lambda, \sigma(A))$ we deduce that $\|R(\lambda, A)\| \longrightarrow \infty$ as λ approaches $\sigma(A)$. Hence $\rho(A)$ is the *maximal* domain of analyticity of $R(\lambda, A)$.

• *Remark* 2.2. It is noteworthy to observe that if we have a family of bounded operators $\{R(\lambda) : \lambda \in \mathcal{O}\}$, with $\mathcal{O}$ subset of $\mathbb{C}$, and such that they verify the identity (2.1), then the family is the resolvent of a unique densely defined closed operator A if and only if the null space $\mathcal{N}(R(\lambda)) = \{0\}$ and the range $\mathcal{R}(R(\lambda))$ is dense in H. Obviously $\mathcal{O} \subset \rho(A)$, see Pazy [196, p. 36]. □

Theorem 2.3 (Riesz-Schauder)**.** *Let K be compact operator from a Banach space into itself. Then the following properties hold:*

(a) *The spectrum $\sigma(K)$ of K is either finite or a countable set accumulating*

only at 0.

(b) *Every non-zero element of* $\sigma(K)$ *is an eigenvalue of finite multiplicity i.e., we have* $nul(\lambda I - K) = nul(\lambda I - K') < +\infty$, *for all* $\lambda \neq 0$

(c) *For any* $\lambda \neq 0$, *the operator* $\lambda I - K$ *is one-to-one if and only if it is onto.* □

2.1.1 Adjoint Operators

If X and Y are (complex) Hilbert spaces with inner products $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_Y$, then, given a densely defined operator $A\colon \mathcal{D}(A) \subset X \to Y$, the *adjoint operator* (instead of dual mapping) A^* is defined by $A^*y = z$, for all those $y \in Y$ such that $|(y, Ax)_Y| \leq C\,|x|$ and, Riesz' theorem yields the existence of a unique $z \in X$ such that $(y, Ax)_Y = (z, x)_X$. The set of all such y, denoted by $\mathcal{D}(A^*)$, is a vector space that can be the only null vector. If A is a densely defined and closable operator then its adjoint A^* is also densely defined and closed, and its double adjoint A^{**} is the closure (i.e., its minimum closed extension) of A. Thus densely defined and closed (or closable) operators are key candidates for the adjoint concept. In the case of $X = Y = H$, where H is a Hilbert space with inner product $(\cdot,\cdot)$, we say that $A\colon \mathcal{D}(A) \subset H \to H$ is *symmetric* if $(Ax, y) = (x, Ay)$ for every x, y in $\mathcal{D}(A)$ where $\mathcal{D}(A)$ is *dense* in H (otherwise it is sometimes called Hermitian). For a densely defined operator A, the orthogonal complement of the range is the null space of the adjoint, i.e., $\mathcal{R}(A)^\perp = \mathcal{N}(A^*)$ and if $\mathcal{R}(A)$ is closed then $\mathcal{R}(A) = \mathcal{N}(A^*)^\perp$, i.e., the equation $Ax = y$ has a solution if and only if y belongs to $\mathcal{N}(A^*)^\perp$.

In other words a densely defined operator A is symmetric if and only if $A \subset A^*$, that is A^* is an extension of A. If A is a densely defined (closed) symmetric operator on complex Hilbert H then $\sigma(A) \subset \mathbb{R}$. More specifically

(1) (Ax, x) is real for every x in $\mathcal{D}(A)$,

(2) all eigenvalues (i.e., all complex values λ such that $\lambda I - A$ is not one-to-one) are real,

(3) eigenvectors (i.e., $x \neq 0$ such that $Ax = \lambda x$) corresponding to distinct eigenvalues are orthogonal,

(4) the continuous spectrum (i.e., all complex values λ such that $\lambda I - A$ is one-to-one and the range is dense, but the inverse defined on the range is not continuous) is also real.

Summarizing, if A is a symmetric operator we have

$$A \subset \bar{A} = A^{**} \subset A^*,$$

where $\bar{A}$ is the closure of A. In the case of $A = A^*$ (this implies the equality $\mathcal{D}(A) = \mathcal{D}(A^*)$ of the domains) the operator is called *self-adjoint*. Any self-adjoint extension S of a symmetric operator satisfies

$$A \subset \bar{A} = A^{**} \subset S \subset A^*.$$

If also A^* is symmetric or $\bar{A}$ is self-adjoint, we have a unique self-adjoint extension, in this case A is called *essentially self-adjoint* and $\mathcal{D}(A)$ a *core*.

A typical self-adjoint operator is the *orthogonal projection* over a closed subspace. It can be proved (von Neumann Theorem) that if A is a densely defined closed operator between two Hilbert spaces H and K then the operator A^*A is self-adjoint on H, the domain $\mathcal{D}(A^*A)$ is a core of A and $B = (I+A^*A)^{-1}$ exists as a bounded linear operator on H. Also, B is symmetric and positive, $\|B\| \leq 1$ and $\|AB\| \leq 1$.

Note that the operator $A\phi = \mathtt{i}\phi'$ defined in $H = L^2(0,1)$ and domain $\mathcal{D}(A) = H_0^1(0,1)$ is closed and symmetric, but it is not self-adjoint, because $A^*\phi = \mathtt{i}\phi'$ but $\mathcal{D}(A^*) = H^1(0,1)$. If the zero-boundary conditions $\phi(0) = \phi(1) = 0$ defining the domain $\mathcal{D}(A) = H_0^1(0,1)$ are replaced by periodic condition $\phi(0) = \phi(1)$ (actually, $\phi(0) = c\phi(1)$, for some constant c, suffices) then A becomes self-adjoint. On the other hand, if the boundary conditions are only $\phi(0) = 0$ then A is not longer a symmetric operator and its resolvent set $\rho(A) = \mathbb{C}$, i.e., the spectrum set is empty. So, a densely defined closed and symmetric operator is not automatically self-adjoint, and the domain plays an important role, i.e., unless the domain is implicitly given, we are referring to the couple $(\mathcal{D}(A), A)$.

An easy test to determine when a closed symmetric operator is self-adjoint is given in term of the *deficiency indices* γ^- and γ^+, defined as the dimension of the subspace $\mathcal{R}(A - \mathtt{i}I)^\perp$ and $\mathcal{R}(A + \mathtt{i}I)^\perp$, respectively. The statement is as follows: the closure of densely defined symmetric operator A, (i.e., A^{**}) is self-adjoint if and only if the deficiency indices are both zero, i.e., $\gamma^- = \gamma^+ = 0$. In this case, the compression (or residual) spectrum of A (i.e., all complex values λ such that $\lambda I - A$ is one-to-one and the inverse defined on the range is not continuous, but the range is dense) is empty, and so any complex value λ with a non-zero imaginary part is in the resolvent set $\rho(A)$.

If the operator $A : \mathcal{D}(A) \subset H \to H$ is self-adjoint, there exists a (unique spectral) family of projection operators $\{E(\lambda), \lambda \in \mathbb{R}\}$ on H, which is increasing (if $\lambda < \lambda'$ then $E(\lambda) \leq E(\lambda')$ or equivalent $E(\lambda')E(\lambda) = E(\lambda')$) and right-continuous $E(\lambda + \varepsilon)x \to E(\lambda)x$ as $\varepsilon \to 0^+$, for every x in H, with the limits $E(-\infty) = 0$ and $E(+\infty) = I$, and commute with A, such that

$$Ax = \int_{\mathbb{R}} \lambda \, \mathrm{d}E(\lambda)x, \quad \forall x \in \mathcal{D}(A).$$

If also A is non-negative, $A \geq 0$, then

$$Ax = \int_0^\infty \lambda \, \mathrm{d}E(\lambda)x, \qquad A^{1/2}x = \int_0^\infty \sqrt{\lambda} \, \mathrm{d}E(\lambda)x, \quad \forall x \in \mathcal{D}(A).$$

If A is also bounded the integral is only over some interval $[m, M)$. Similarly, if the operator $A : \mathcal{D}(A) \subset H \to H$ is *skew-adjoint* i.e., $A^* = -A$, (or equivalent $A = \mathtt{i}B$ for some self-adjoint operator B) then the spectral representation takes the (obvious) form

$$Ax = \int_{\mathbb{R}} \mathtt{i}\lambda \, \mathrm{d}E(\lambda)x, \quad \forall x \in \mathcal{D}(A).$$

The previous spectral decomposition applies to a bigger class of operator, so called *normal* N (i.e., with the property $N^*N = NN^*$). In particular, if $U : H \to H$ is an *unitary* operator, i.e., $U^* = U^{-1}$ then

$$Ux = \int_{[0,2\pi)} \mathrm{e}^{\mathrm{i}\lambda} \, \mathrm{d}E(\lambda)x, \quad \forall x \in H.$$

In the case of compact self-adjoint operator $T\colon H \to H$ the Hilbert-Schmidt theorem affirms that there exists a sequence of non-zero real eigenvalues $\{\mu_i\}_{i=1}^N$ with N equal to the rank of T, such that $|\mu_i|$ is monotone non-increasing, and if $N = \infty$ then $\mu_i \to 0$. Moreover, if each eigenvalue is repeated in the sequence according to its multiplicity, then there exists an orthonormal set $\{e_i\}_{i=1}^N$ of corresponding eigenfunctions, i.e., $Te_i = \mu_i e_i$, which is complete in the range $\mathcal{R}(T)$ and the following representation

$$Tx = \sum_{i=1}^{N} \mu_i (x, e_i) e_i, \quad \forall x \in H$$

hold true. Similarly, if A is a densely defined self-adjoint operator on H having a compact resolvent $R(\lambda, A)$ then the previous spectral decomposition applies to $T = R(\lambda, A)$, and $\lambda_i = \lambda - \mu_i^{-1}$ is an eigenvalue for A,

$$Ax = \sum_{i=1}^{N} \lambda_i (x, e_i) e_i, \quad \forall x \in \mathcal{D}(A)$$

is also valid. For instance, the reader is referred to either the introductory books Engel and Nagel [74, Chapter IV], Renardy and Rogers [211, Chapter 7], or to the classic books Kato [128, Chapter 5], Reed and Simon [210, Chapter VIII], Riesz and Nagy [213, Chapter IX], Yosida [262, Chapter VIII].

2.1.2 Fredholm Operators

We continue with a definition of a class of operator frequently found in what follows.

Definition 2.4 (Fredholm)**.** A linear mapping A between two Banach spaces X and Y is called a *Fredholm* operator if it satisfies the following five conditions:

(a) the domain $\mathcal{D}(A)$ of A is dense in X,

(b) A is a closed operator,

(c) the null space $\mathcal{N}(A)$ of A has finite dimension,

(d) the range $\mathcal{R}(A)$ is closed in Y,

(f) the range $\mathcal{R}(A)$ has finite co-dimension.

The index $\mathrm{ind}(A)$ of A is defined as the dimension of the null space $\mathcal{N}(A)$ of A [also called *nullity* of A and denoted by $\mathrm{nul}(A)$] minus the co-dimension of the range $\mathcal{R}(A)$ of A [also called *deficiency* of A and denoted by $\mathrm{def}(A)$]. □

The Fredholm Alternative theorem affirms that if $K : X \longrightarrow X$ is a compact operator and I denotes the identity operator, then $I - K$ and its transpose $I - K'$ are Fredholm operators, the nullity of $I - K$ and $I - K'$ is the same and $\text{ind}(I - K) = 0$.

There is a nice characterization of Fredholm operators as follows. Let A be a closed linear operator from X into Y with domain $\mathcal{D}(A)$ dense in X. Then A is a Fredholm operator if and only if there exist a bounded linear operator $S : Y \longrightarrow X$ and compact operators $P : X \longrightarrow X$, $Q : Y \longrightarrow Y$ such that

$$\text{(a) } SA = I - P \text{ on } \mathcal{D}(A) \text{ and (b) } AS = I - Q \text{ on } S^{-1}(\mathcal{D}(A)).$$

Actually, the operator S in (a) and in (b) may be different, but then it can be taken to be the same. Moreover, $\mathcal{R}(P) = \mathcal{N}(A)$ and the dimension of $\mathcal{R}(Q)$ is equal to the co-dimension of $\mathcal{R}(A)$

Let us mention some other important properties in this context. The composition preserves Fredholm operators i.e., if A and B are Fredholm operators, A from X into Y and B from Y into X, then BA is also a Fredholm operator and $\text{ind}(BA) = \text{ind}(B) + \text{ind}(A)$. A compact perturbation preserves Fredholm property i.e., if A is a Fredholm operator form X into Y and K is a compact linear operator, then $A + K$ is a Fredholm operator and $\text{ind}(A + K) = \text{ind}(A)$. Transposition preserves Fredholm operators i.e., if A is a Fredholm operator form X into Y and Y is reflexive, then the transpose operator A' from Y' into X' is a Fredholm operator and $\text{ind}(A') = -\text{ind}(A)$.

Another usual form of the Fredholm Alternative theorem is the following assertion: Let K be a compact operator from a Banach space X into itself. Then either (a) for each f in X the equation $u - Ku = f$ has a unique solution or else (b) the homogeneous equation $u - Ku = 0$ has non-trivial solutions. In addition, if (a) holds then the inverse operator $(I - K)^{-1}$ is continuous, and if (b) holds then the space of solutions of the homogeneous equation is finite dimensional, and the non-homogeneous equation $u - Ku = f$ has a solution if and only if f is orthogonal to the null space of $I - K'$.

As we will see, typical candidates for Fredholm operators are obtained from the study of partial differential equations with some boundary conditions, for instance we refer to the classic books of Courant and Hilbert [49], Dunford and Schwartz [63], Kato [128], Wloka [254] for a complete treatment and to Brezis [35] and Renardy and Rogers [211] for a comprehensive introduction.

2.1.3 Rigged Hilbert Spaces

Sometime we need to use topological vector spaces, in particular when dealing with stochastic partial differential equations, the notion of Hilbert-Schmidt operators becomes important as well as nuclear spaces. The interested reader is referred to the standard books, e.g., Conway [48], Dunford and Schwartz [63, Section XI.6, pp. 1009–1044], Gelfand and Vilenkin [96, Section I.4, pp. 103–127], Yosida [262].

Let H and K be separable Banach spaces, H' and K' be their dual spaces and $L(H, K)$ denote the separable Banach space of all linear bounded operators

from H into K endowed with the usual supremum norm

$$\|T\|_{L(H,K)} = \|T\| = \sup\big\{\|Th\|_K \,:\, h \in H,\ \|h\|_H \le 1\big\}.$$

An operator T in $L(H,K)$ is called *nuclear* operator if there exist two sequences $\{k_i\}$ in K and $\{h_j'\}$ in H' representing T, i.e., such that

$$Th = \sum_{i=1}^{\infty} k_i \langle h_i', h\rangle, \quad \forall h \in H, \quad \text{and} \quad \sum_{i=1}^{\infty} \|k_i\|_K \, \|h_i'\|_{H'} < \infty,$$

where $\langle\cdot,\cdot\rangle$ denotes the pairing duality. The set of all nuclear operator from H to K is denoted by $L_1(H,K)$ and forms a separable Banach space with the norm

$$\|T\|_{L_1(H,K)} = \|T\|_1 = \inf\Big\{\sum_{i=1}^{\infty} \|k_i\|_K \, \|h_i'\|_{H'}\Big\},$$

where the infimum is taken over all possible sequences representing T. In particular, we write $L(H) = L(H,H)$ and $L_1(H) = L_1(H,H)$. Clearly, for any T in $L_1(H,K)$, R in $L(H)$ and S in $L(K)$ we have RT and TS in $L_1(H,K)$, and $\|RT\|_1 \le \|R\|\,\|T\|_1$ and $\|TS\|_1 \le \|T\|_1\,\|S\|$. Moreover, any nuclear operator is a compact operator.

If $\{e_i\}$ is any complete orthonormal sequence (or basis) in a separable Hilbert space H, then the expression

$$\mathrm{Tr}(T) = \sum_{i=1}^{\infty} (Te_i, e_i)_H \tag{2.2}$$

define the trace of any element T in $L_1(H)$, independently of the particular choice of the basis. Moreover, $|\,\mathrm{Tr}(T)| \le \|T\|_1$ and $\mathrm{Tr}(ST) = \mathrm{Tr}(TS) \le \|T\|_1\,\|S\|$ for any S in $L(H)$ and T in $L_1(H)$. Furthermore, a nonnegative operator T in $L(H)$ is nuclear if and only if for some basis the series (2.2) is finite, and in this case $\mathrm{Tr}(T) = \|T\|_1$.

In a Hilbert space, nuclear operators are also called *trace-class* operators. Moreover, if H and K are two separable Hilbert spaces with basis $\{e_i\}$ and $\{f_j\}$, respectively, then a linear bounded operator T from H into K is called a *Hilbert-Schmidt* operator if the series

$$\|T\|_2 = \sum_{i=1}^{\infty} |Te_i|_K^2 = \sum_{i,j=1}^{\infty} |(Te_i, f_j)_K|^2 = \sum_{j=1}^{\infty} |T^* f_j|_H^2$$

are finite, where $|\cdot| = |\cdot|_H$ and $(\cdot,\cdot) = (\cdot,\cdot)_K$ denote the norm and the inner product in the corresponding space, and T^* is the adjoint operator of T. Certainly, the number $\|T\|_2$ is independent of the particular choice of basis. The set of all Hilbert-Schmidt operators from H into K, denoted by $L_2(H,K)$, becomes a separable Hilbert space with the inner product

$$(S,T)_2 = (S,T)_{L_2(H,K)} = \sum_{i=1}^{\infty} (Se_i, Te_i)_K,$$

which yields the norm $\|T\|_2$ are previously defined. Again we write $L_2(H) = L_2(H,K)$. Clearly, for any T in $L_2(H,K)$, R in $L(H)$ and S in $L(K)$ we have RT and TS in $L_2(H,K)$, and $\|RT\|_2 \le \|R\|\,\|T\|_2$ and $\|TS\|_2 \le \|T\|_2\,\|S\|$. Moreover, any Hilbert-Schmidt operator is compact, and for any S in $L_2(H,K)$ and T in $L_2(K,H)$ we have TS in $L_1(K)$ with $\|ST\|_1 \le \|S\|_2\,\|T\|_2$. In particular, if T is in $L_2(H)$ then $TT^* = T^*T$ is nonnegative and belongs to $L_1(H)$. For instance, see Da Prato and Zabczyk[51, Appendix C, pp. 415-419].

Recall that a Fréchet space is a complete metrizable locally convex topological vector space (i.e., the topology is given by a sequence of seminorms). For instance, the space $\mathcal{S}(\mathbb{R}^d)$ of rapidly decreasing smooth functions (the Schwartz space) and its dual, the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$.

Definition 2.5. A separable Fréchet space Φ expressed as a decreasing countable intersection $\cap_{n=0}^{\infty}\Phi_n$ of Hilbert spaces Φ_n is called a *countably Hilbertian space* if Φ_m is continuously and densely embedded in Φ_n for any $m > n$. The space Φ is said to be *nuclear* if for any n there exits $m > n$ such that the inclusion from Φ_m into Φ_n is a Hilbert-Schmidt operator. Any Hilbert space $H \subset \Phi_0$ with a inner product $(\cdot,\cdot)_H$ continuous in the topology of a nuclear countably Hilbertian space Φ is called a *rigged Hilbert space*. □

The essential element of a countably Hilbertian space is the expression of the Hilbertian norm $\phi \mapsto \|\phi\|_n = \sqrt{(\phi,\phi)_n}$, defined for any n and ϕ in Φ, for which Φ_n is the completion of Φ under $\|\cdot\|_n$. Since the inclusion from Φ_m into Φ_n is continuous and dense, for any $m > n \ge 0$, any Cauchy sequence in the m-norm is a Cauchy sequence in the n-norm, and if it converges to zero in the n-norm then it converges to zero in the m-norm.

Identifying the dual space Φ_0' with Φ_0 by Riesz representation, the dual space Φ' is sequentially complete and $\Phi' = \cup_{n=0}^{\infty}\Phi_{-n}$, where $\Phi_{-n} = \Phi_n'$ is the dual space of Φ_n. By means of a Baire category argument, we can show that (1) Φ is dense in Φ' and (2) any continuous seminorm in Φ is actually a continuous seminorm in some Φ_n. Hence, for every rigged Hilbert space there exists $n \ge 0$ such that

$$\Phi \subset \cdots \subset \Phi_{n+1} \subset H \subset \Phi_n \subset \cdots \subset \Phi_1 \subset \Phi_0 = \Phi_0' \subset \Phi_{-1} \subset \cdots \subset \Phi',$$

where all inclusions are continuous and dense. The norm and the inner product in the dual space Φ_{-n} is denoted by $\|\cdot\|_{-n}$ and $(\cdot,\cdot)_{-n}$, respectively. Clearly, because the dual (anti-dual, if working with complex numbers) of Φ_{-n} is Φ_n, the topological dual (anti-dual) space Φ'' of Φ' coincides with Φ, i.e., the space of Φ is reflexive.

A typical example of a nuclear countably Hilbertian space is the Schwartz space $\mathcal{S} = \mathcal{S}(\mathbb{R}^d)$ of rapidly decreasing functions in $\mathbb{R}^d$. Its dual space $\mathcal{S}' = \mathcal{S}'(\mathbb{R}^d)$ is the space of tempered distributions. The topology in $\mathcal{S}$ is given by the family of seminorms

$$\|\varphi\|_{n,\alpha} = \sup_x \left\{|x|^n |D_x^\alpha \varphi(x)|\right\},$$

for any $n \geq 0$ and any (derivative) multi-index $\alpha = (\alpha_1, \cdots, \alpha_d)$ of order $|\alpha| = \alpha_1 + \cdots + \alpha_d$ in the variable x. The Hilbert space $\mathcal{S}_n$ is the Sobolev space $H^n(\mathbb{R}^d)$ with the Hilbertian norm

$$\|\varphi\|_n = \Big\{ \sum_{|\alpha| \leq n} \int_{\mathbb{R}^d} |D_x^\alpha \varphi(x)|^2 \mathrm{d}x \Big\}^{1/2}, \quad n = 0, 1, \cdots$$

and a rigged Hilbert space is any Sobolev space $H^s(\mathbb{R}^d)$ with possible fractionary exponent $s \geq 0$. Certainly, the dual spaces are the Sobolev spaces with negative integer exponents.

Similarly, for a smooth d-dimensional domain $\mathcal{O}$, the space $\Phi = C_0^\infty(\mathcal{O})$ of functions differentiable of any order and vanishing at the boundary is a nuclear countably Hilbertian space and Φ_n is the Solobev space $H_0^n(\mathcal{O})$. Moreover, the space $C^\infty(\overline{\mathcal{O}})$ of smooth functions up to the boundary is a nuclear countably Hilbertian space. Indeed, if $\overline{\mathcal{O}} = \cap_{n \geq 0} \mathcal{O}_n$, where $\mathcal{O}_n$ is a smooth domain with $\overline{\mathcal{O}}_{n+1} \subset \mathcal{O}_n$, then $C^\infty(\overline{\mathcal{O}}) = \cap_{n=1}^\infty H_0^n(\mathcal{O}_n)$. Note that the injection from $H^0(\mathcal{O}) = L^2(\mathcal{O})$ is not dense in the dual space of $H^1(\mathcal{O})$. On the contrary, the space $\mathcal{D}(\mathcal{O})$ of smooth functions with compact support endowed with the inductive topology is not a countably Hilbertian space.

Clearly, any separable Hilbert space H is a countably Hilbertian space if we take all $\Phi_n = H$, which is an example of almost no importance. However, only a finite-dimensional Hilbert space is a nuclear countably Hilbertian space with $\Phi_n = H$, for any $n \geq 0$. Initially, rigged Hilbert spaces (also called Gelfand triples) are defined as a Hilbert space H satisfying $\Phi \subset H \subset \Phi'$, with continuous and dense inclusion. Nevertheless, we required the countably Hilbertian space Φ to be nuclear to avoid trivialities. On the other hand, we may define a *rigged Banach space* B by asking $\Phi \subset B \subset \Phi'$ and $\Phi \subset B' \subset \Phi'$ (all inclusions are dense and continuous), i.e., explicitly $\Phi_n \subset B \subset \Phi_{-n}$ and $\Phi_n \subset B' \subset \Phi_{-n}$, for some $n \geq 0$. The typical candidates for rigged Banach spaces are the Lebesgue spaces $L^p(\mathcal{O})$ and the Sobolev spaces $W_0^{n,p}(\mathcal{O})$, $1 < p < \infty$, $n \geq 0$.

A rigged Hilbert space can be constructed from a Hilbert space H and a Hilbert-Schmidt operator T in $L_2(H)$ satisfying $\mathrm{Ker}(T^*) = \{0\}$, by setting $\Phi_0 = H$ and $\Phi_n = T^n(H)$ with the Hilbertian norm $\|h\|_n = \|T^n h\|_0$. In this case, the inclusion from Φ_{n+1} into Φ_n is (besides being continuous and dense) a Hilbert-Schmidt operator. Note that $(Th, k) = 0$, for every h in H, implies that $T^* k = 0$, i.e., the image $T(H)$ is dense in H.

On the other hand, for a given rigged Hilbert space H we have $\Phi_{n+1} \subset H \subset \Phi_n$ and for some $m > n$ the inclusion, denoted by T, from Φ_m into Φ_n is Hilbert-Schmidt operator, i.e., we may consider T as an element in $L_2(H)$, and reconstruct Φ from H and T. Thus, the concept of rigged Hilbert space becomes clear as a Hilbert space and a Hilbert-Schmidt operator with all the rich structure behind.

Essentially any concept developed on rigged Hilbert spaces can be extended in a natural way to dual nuclear countably Hilbertian spaces. One of the common point is to use an orthonormal basis in H with elements in Φ, e.g., the eigenfunctions of the underlying Hilbert-Schmidt operator. For instance, the reader is referred to Kallianpur and Xiong [123, Chapter 1, pp. 1-43].

2.1.4 Integrals and Series

If $u(t)$ is a continuous (sometimes call strongly continuous) function from an real interval $[a, b]$ into a Banach space X such that

$$\int_a^b \|u(t)\| \mathrm{d}t < +\infty$$

then the Riemann integral of $u(t)$ over $[a, b]$ can be defined just as in the case of real valued function. In this case we say that the function is (strongly) integrable on $[a, b]$, and by the triangle inequality we have

$$\left\| \int_a^b u(t) \mathrm{d}t \right\| \le \int_a^b \|u(t)\| \mathrm{d}t.$$

Similarly, it is defined the concept of (strong) differentiability.

It should be clear by now that in most cases we are interested in finding the inverse of a given operator, so to conclude this section let us mention the *method of continuity* used in various linear boundary problems arising in the theory of partial differential equations, in particular parabolic and elliptic equations, e.g., Gilbarg and Trudinger [100], Ladyzhenskaya and Uraltseva [148], Ladyzhenskaya, Solonnikov and Uraltseva [147], Lieberman [153].

Theorem 2.6 (method of continuity)**.** *Let X be a Banach space and Y be a normed vector space. Suppose that A_0 and A_1 are two bounded linear operators from X into Y and there exists a constant $C > 0$ such that, denoting by $A_\theta = \theta A_1 + (1-\theta)A_0$,*

$$\|x\|_X \le C \|A_\theta\, x\|_Y, \quad \forall x \in X, \tag{2.3}$$

for every θ in $[0, 1]$. Then the operator A_0 is surjective if and only if A_1 is surjective. □

Actually, (2.3) implies that A_t is injective for any $t \in [0, 1]$. Hence, if for some θ, A_θ is also surjective, then A_θ^{-1} exists and (2.3) implies that $\|A_\theta^{-1}\| \le C$. Now, if $|t - \theta| < \delta = C^{-1}\|A_1 - A_0\|^{-1}$, it is easy to see that

$$A_t^{-1} = (I - (t-\theta)A_\theta^{-1}(A_1 - A_0))^{-1} A_\theta^{-1} = (I + T + T^2 + T^3 + \cdots)\, A_\theta^{-1}$$

where $T = (t-\theta)A_\theta^{-1}(A_1 - A_0) \in L(X)$ has norm strictly less than 1. Usually, (2.3) is some a priori estimate and A_0 is a *simple* operator which is known to be invertible.

A (strongly) continuous operator valued function $S(t)$ is a function defined on an interval $[a, b]$ with values in $L(X)$ satisfying for any point t_0

$$\lim_{t \to t_0} \|S(t)x - S(t_0)x\| = 0, \quad \forall x \in X.$$

Continuous in norm at the point t_0 means

$$\lim_{t \to t_0} \|S(t) - S(t_0)\| = 0.$$

in the $L(X)$ norm, also referred to as the operator norm. Similarly, we define the concept of differentiability, e.g., $S(t)$ is (strongly) differentiable at the point t_0 if there exist an operator $\dot{S}(t_0)$ in $L(X)$ such that

$$\lim_{t\to t_0} \left\| \Big(\frac{S(t) - S(t_0)}{t - t_0} \Big) x - \dot{S}(t_0)x \right\| = 0, \quad \forall x \in X.$$

Notice that Leibniz formula can be extended to strongly or norm differentiable functions.

As in the case of real valued series we can define the *Exponential* function by the series

$$\exp(tA) = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n, \quad \forall t \in \mathbb{R},$$

where A is a bounded linear operator in X. This series converges in the norm of the space $L(X)$ and enjoys the following properties: (a) a bound on the norm

$$\| \exp(tA)\| \leq \exp(t\|A\|),$$

(b) multiplication property

$$\exp(tA)\exp(sA) = \exp[(t+s)A], \quad \forall t, s \in \mathbb{R},$$

and (c) the function $\exp(tA)$ is norm differentiable on R and we have

$$\lim_{t\to 0} \left\| \frac{\exp[(t+s)A] - \exp(sA)}{s} - A\exp(sA) \right\| = 0, \quad \forall s \in \mathbb{R}.$$

Perhaps, the reader interested in functional analysis for stochastic processes may consult the recent book by Bobrowski [29].

2.2 Continuous Semigroups

Complete proofs of most of what we are going to discuss here can be found in Pazy [196] and Yosida [262].

Definition 2.7 (semigroup)**.** A one-parameter family $\{S(t) : t \geq 0\}$ of bounded linear operators from a Banach space X into itself is called a *semigroup of class (C_0)* or simply a (continuous) semigroup if it satisfies

(a) $S(t+s) = S(t)S(s), \quad \forall t, s \geq 0,$
(b) $\lim_{t\downarrow 0} \|S(t)x - x\| = 0, \quad \forall x \in X.$

It is called a *contraction* if $\|S(t)\| \leq 1$ for every $t \geq 0$. □

The key condition (a) is called the semigroup property. Also, it is clear that from (b) we must have $S(0) = I$. Sometimes, the family $\{S(t) : t \geq 0\}$ is referred to as a *strongly continuous semigroup*. If the property (b) is replaced by the stronger property $\lim_{t\downarrow 0} \|S(t) - I\| = 0$, i.e. the convergence is in the operator norm $L(X)$, then we speak of uniformly continuous semigroup. However, the only uniformly continuous semigroups are the exponential functions $S(t) = \mathrm{e}^{At}$, for some bounded operator $A \in L(X)$.

• *Remark* 2.8. A first property of a semigroup of class (C_0) is the fact that there exist constants M and ω such that

$$\|S(t)\| \leq M \exp(\omega t),$$

and that the function $t \mapsto S(t)x$ is continuous (in X) for t in $[0, \infty[$. We may change the norm in the space X to

$$\|x\|_0 := \inf_{t\geq 0} \|\exp(-\omega t)\, S(t)x\|,$$

which yields the same topology, i.e.,

$$\|x\| \leq \|x\|_0 \leq M\|x\|, \quad \forall x \in X.$$

Clearly, the new semigroup $\bar{S}(t) := \exp(-\omega t)\, S(t)$ is a contraction with the new norm, i.e., $\|\bar{S}(t)\|_1 \leq 1$, e.g., see Ahmed [1]. □

The *infinitesimal generator* A of a semigroup of class (C_0) is defined as the linear (possibly unbounded) operator

$$Ax = \lim_{t\downarrow 0} \frac{S(t)x - x}{t},$$

whose domain $\mathcal{D}(A)$ is the set of x in X for which the above limit exists (in the norm of the space X).

• *Remark* 2.9. It can be shown that the domain $\mathcal{D}(A)$ is a dense vector subspace of X and that the infinitesimal generator A is a closed operator. □

A characterization of a contraction semigroup of class (C_0) is given by

Theorem 2.10 (Hille-Yosida)**.** *A linear (unbounded) operator A on a Banach space X is the infinitesimal generator of a contraction semigroup of class (C_0) $\{S(t) : t \geq 0\}$ if and only if*
(a) *A is closed and $\mathcal{D}(A)$ is dense in X,*
(b) *$(\lambda I - A)$ is invertible for every $\lambda > 0$ and $\|(\lambda I - A)^{-1}\| \leq \lambda^{-1}$.* □

A similar result holds for a semigroup of class (C_0) satisfying $\|S(t)\| \leq M \exp(\omega t)$, provided we replace the condition (b) by the estimate

$$\|(\lambda I - A)^{-n}\| \leq M(\lambda - \omega)^{-n},$$

valid for any $\lambda > \omega$ and $n = 1, 2, \ldots$

An important tool in the proof of the above results is the so-called Yosida approximations of A

$$A_\lambda := \lambda A R(\lambda, A) = \lambda A(\lambda I - A)^{-1},$$

satisfy $A_\lambda x \to Ax$ as $\lambda \to +\infty$ for every x in X, and the fact that the resolvent operator can be expressed as

$$R(\lambda, A)x = \int_0^\infty \exp(-\lambda t) S(t)x \mathrm{d}t,$$

and $\lambda R(\lambda, A)x \to x$ as $\lambda \to +\infty$ for every x in X.

An exponential formula similar to the Yosida approximations is given by the "finite difference" approximations for $h > 0$

$$S_h(t) = \mathrm{e}^{tA(h)}, \qquad A(h)x = h^{-1}[S(h)x - x], \quad x \in X,$$

which satisfies $S_h(t)x \to S(t)x$, for any $x \in X$ and uniformly on bounded set in t.

Since a semigroup $S(t)$ of class (C_0) is uniquely determinate by its infinitesimal generator A, we use the notation $S(t) = \mathrm{e}^{tA}$ or $\exp(tA)$, where the operator A satisfies (a) and (b) of Hille-Yosida Theorem. Moreover, the inversion of the Laplace transform holds, i.e., if $\|\mathrm{e}^{tA}\| \le M\mathrm{e}^{\omega t}$ then for any $\gamma > \max\{\omega, 0\}$ and any $x \in \mathcal{D}(A)$ we have

$$\int_0^t \mathrm{e}^{sA} x \mathrm{d}s = \frac{1}{2\pi \mathtt{i}} \int_{\gamma - \mathtt{i}\infty}^{\gamma + \mathtt{i}\infty} \lambda^{-1} \mathrm{e}^{\lambda t} R(\lambda, A) x \mathrm{d}\lambda, \tag{2.4}$$

where the integral on the right converges uniformly in bounded interval on t.

If X' is the dual of a Banach space X and if $x \in X$, then by the Hahn-Banach theorem that there exists an element in the dual $x' \in X'$ such that

$$\langle x', x\rangle = \|x\|^2 = \|x'\|^2, \tag{2.5}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality paring.

Definition 2.11 (dissipative)**.** A linear (unbounded) operator A is *dissipative* if for every x in $\mathcal{D}(A)$ there is an element in the dual $x' \in X'$ such that $\Re\langle x', Ax\rangle \le 0$[1] and equality (2.5) hold. □

The above definition is equivalent to the following condition

$$\|(\lambda I - A)x\| \ge \lambda \|x\|, \quad \forall x \in \mathcal{D}(A),\ \lambda > 0,$$

which may be used as the initial definition. On the other hand, it is clear that if X is a Hilbert space then an operator is dissipative if and only if $\Re\langle x, Ax\rangle \le 0$ for any x in $\mathcal{D}(A)$.

It is proved in Pazy [196, Theorems 4.5 and 4.6, pp. 15–17] that if A is a dissipative operator in a Banach space X then

[1] Here $\Re(\cdot)$ denotes the real part of a complex number.

(a) if for some $\lambda_0 > 0$ the range $\mathcal{R}(\lambda_0 I - A) = X$ then $\mathcal{R}(\lambda I - A) = X$ for all $\lambda > 0$,

(b) if A is closable then its closure $\bar{A}$ is also dissipative,

(c) if $\mathcal{D}(A)$ is dense in X then A is closable,

(d) if for some $\lambda > 0$ the range $\mathcal{R}(\lambda I - A) = X$ and X is a reflexive space then $\mathcal{D}(A)$ is dense in X.

The sub-differential $\partial\|x\|$ of $\|\cdot\|$ (the norm of a real Banach space X) at x is defined as follows

$$\partial\|x\| = \{x' \in X' \ : \ \|x + y\| - \|x\| \geq \langle x', y\rangle, \quad \forall y \in X\},$$

where X' is the dual space of X. We can show that for $x = 0$ we have $\partial\|x\| = \{x' \in X' \ : \ \|x'\| \leq 1\}$, while for $x \neq 0$ we get $\partial\|x\| = \{x' \in X' \ : \ \|x'\| = 1$ and $\langle x', x\rangle = \|x\|\}$. Thus, a map A from $\mathcal{D}(A)$ into X is dissipative if and only if for any x, y in $\mathcal{D}(A)$ there exists z' in $\partial\|x - y\|$ such that $\langle z, Ax - Ay\rangle$. This concept of dissipative maps is very useful for non-linear operators.

Most of the typical semigroup examples are defined on function spaces; the infinitesimal generator is often defined by analytic expressions that make sense only for smooth functions. This makes hard to verify condition (a) of Hille-Yosida Theorem 2.10. Hence, another characterization of a semigroup of class (C_0) is given by

Theorem 2.12 (Lumer-Phillips)**.** *Let A be a linear operator with dense domain $\mathcal{D}(A)$ in a Banach space X. We have:*

(a) *If A is dissipative and there is $\lambda_0 > 0$ such that the range $\mathcal{R}(\lambda_0 I - A)$ is dense in X, then A is closable and its closure $\bar{A}$ is the infinitesimal generator of a contraction semigroup of class (C_0) on X.*

(b) *If A is the infinitesimal generator of a contraction semigroup of class (C_0) on X then range $\mathcal{R}(\lambda I - A)$ is X for all $\lambda > 0$. and A is dissipative. Moreover the condition used in the Definition 2.11 is valid for any dual element x' satisfying (2.5).* □

Notice that if the linear operator A is such that $A - cI$ is dissipative, for some constant c, then Lumer-Phillips' theorem may be applied to $A - cI$, generating a contraction semigroup $S_c(t)$. Hence, A generates a (so-called quasi-contraction) semigroup written as $S(t) = \mathrm{e}^{ct} S_c(t)$.

Let $\{S(t) : t \geq 0\}$ be a semigroup of class (C_0) on a Banach space X, with infinitesimal generator A defined in $\mathcal{D}(A)$. Its *dual semigroup* $\{S'(t) : t \geq 0\}$ is defined on the dual Banach space X' by

$$\langle S'(t)x', x\rangle := \langle x', S(t)x\rangle, \qquad \forall x \in X, \ x' \in X', \ t > 0.$$

It is clear that the semigroup property (a) of Definition 2.7 is satisfied by the one-parameter family $\{S(t) : t \geq 0\}$, however the strong continuity (b) is not necessarily conserved. Similarly, starting from the infinitesimal generator A, the dual operator A' can be defined by

$$\langle A'x', x\rangle := \langle x', Ax\rangle, \qquad \forall x \in \mathcal{D}(A), \ x' \in \mathcal{D}(A'),$$

where $x' \in \mathcal{D}(A')$ if $A'x'$ defined as above, a priori on $\mathcal{D}(A)$, can be extended to an element of X'. Because A is closed and densely defined, then it dual A' is uniquely determined as an unbounded closed linear operator in X'. If the Banach space X is reflexive, then the domain $\mathcal{D}(A')$ of A' is dense e.g., see Goldberg [102, p. 56, Theorem II.3.14]. Actually, it can be proved that the restriction of the dual semigroup $\{S'(t) : t \geq 0\}$ to Y' the closure (in X') of the $\mathcal{D}(A')$ is a semigroup of class (C_0) on a Banach space Y', whose infinitesimal generator is the restriction of A' to the set of x' such that $A'x' \in Y'$, see Pazy [196, p. 39, Theorem 10.4].

When the one-parameter family $\{S(t)\}$ of bounded linear operators from a Banach space X into itself are defined for any $t \in \mathbb{R}$ and satisfies the conditions of Definition 2.7 we say that it forms a groups of bounded operators of class (C_0). It is possible to prove that if $\{S(t) : t \geq 0\}$ is a semigroup of class (C_0) and for any $t > 0$ the inverse $S^{-1}(t)$ exists and is a bounded operator then $\{S(t)\}$ can be embedded in a (C_0) group. Actually, it is enough to know that for some $t_0 > 0$ the operator $S(t_0)$ has a bounded inverse to conclude that $S(t)$ is actually a group.

As mentioned early, for an operator A densely defined on a (complex) Hilbert space H, the adjoint operator A^* is defined by the relation $(Ax, y) = (x, A^*y)$ for any x in $\mathcal{D}(A)$ and y in $\mathcal{D}(A^*)$, where the domain of A^* is the subspace of elements y in H such that for some constant $C = C_y$ we have $|(Ax, y)| \leq C\|x\|$ for any x in $\mathcal{D}(A)$. Thus, a densely defined operator A on a Hilbert space H is called *self-adjoint* if $A^* = A$ and *skew-adjoint* if $A^* = -A$. A bounded operator U is called *unitary* if $U^* = U^{-1}$. A strongly continuous semigroup $\{S(t) : t \geq 0\}$ defined in a Hilbert space H is called *symmetric* if each operator $S(t)$ is symmetric, for every $t \geq 0$. It can be proved (see Pazy [196, Corollary 10.6]) that the generator A of a strongly continuous semigroup is self-adjoint if and only if the semigroup is symmetric.

The following result is very important for the generation of a group of unitary operators, see Engel and Nagel [74, p. 89, Theorem 3.24] and Pazy [196, p. 41, Theorem 10.8], where a comprehensive study on semigroups can be found.

Theorem 2.13 (Stone)**.** *A densely defined operator A on a Hilbert space generates a unitary group $S(t)$ if and only if A is skew-adjoint or equivalent if an only if $\mathtt{i}A$ is self-adjoint, where $\mathtt{i} = \sqrt{-1}$.* □

If the infinitesimal generator A is given a priori then a natural notation for the semigroup $S(t)$ is the expression $\exp(tA)$ or e^{tA}, even if the operator A is only closable densely defined in the Banach space X. Sometimes, it is necessary to combine to infinitesimal generators A and B. For instance, if A and B commute, i.e., $\exp(tA)\exp(tB) = \exp(tB)\exp(tA)$ then $\exp[t(A+B)] = \exp(tA)\exp(tB)$ as expected. In general the following result, know as Trotter product formula (see Taylor [249, Vol 2, Chapter 11, Theorem A.1, pp. 381–385]) is useful.

Theorem 2.14 (Trotter)**.** *Let A and B be two infinitesimal generators of contraction semigroups of class (C_0), denoted by $\exp(tA)$ and $\exp(tB)$, on a Banach space X. If the closure of $A + B$ is the infinitesimal generator of contraction*

semigroup of class (C_0) *denoted* $\exp[t(A+B)]$, *then*

$$\exp[t(A+B)]x = \lim_{n\to\infty} [\exp(\tfrac{t}{n}A)\ \exp(\tfrac{t}{n}B)]^n x$$

for any x *in* X. □

Typical examples of semigroups come from evolution equations. For instance, hyperbolic equations generates groups of bounded linear operators and parabolic equations generates only semigroups.

2.3 Analytic Semigroups

This is an important class of semigroups which carries most of the key properties of a parabolic equation. Roughly speaking, the properties (a) and (b) of Definition 2.7 are also valid for t in a sector in the complex numbers of the positive real numbers, i.e., $-a_0 < \arg(t) < a_0$, for some $a_0 > 0$. However, most of the properties are better seen on the infinitesimal generator. We refer to the books by Lunardi [160] and by Taira [244] for a complete discussion and proofs. Note that unless explicitly stated, $\|\cdot\|$ denotes the norm in the Banach space where "$\cdot$" belongs to.

Definition 2.15 (sectorial)**.** A linear operator A with domain $\mathcal{D}(A)$ in a Banach space X is said to be *sectorial* if there exist constants $\omega \in \mathbb{R}$, $\theta \in]\pi/2, \pi]$ and $M > 0$ such that

(a) $S_{\omega,\theta} := \{\lambda \in \mathbb{C} : \lambda \neq \omega,\ |\arg(\lambda - \omega)| < \theta\} \subset \rho(A)$,

(b) $\|R(\lambda, A)\| \leq \dfrac{M}{|\lambda - \omega|}, \quad \forall \lambda \in S_{\omega,\theta}$,

where $\rho(A)$ is the resolvent set of A and $R(\lambda, A) := (\lambda I - A)^{-1}$ is the corresponding resolvent operator. □

The linear operator A may not be densely defined, i.e., $\mathcal{D}(A)$ is not necessarily dense in X, but the fact that the resolvent set of A is not empty implies that A is closed, so that $\mathcal{D}(A)$ endowed with the graph norm $\|x\|_{\mathcal{D}(A)} := \|x\| + \|Ax\|$, is a Banach space. If we insist in having a dense domain for A, we may replace X by the closure of $\mathcal{D}(A)$ in X.

For a given sectorial operator A we can define a family of linear bounded operators $\{\mathrm{e}^{tA} : t \geq 0\}$ in X by means of the integral

$$\mathrm{e}^{tA} := \frac{1}{2\pi \mathtt{i}} \int_{\omega+\gamma(r,\eta)} \mathrm{e}^{t\lambda} R(\lambda, A)\mathrm{d}\lambda, \quad t > 0, \tag{2.6}$$

where $r > 0$, $\eta \in]\pi/2, \theta[$, and $\gamma(r, \eta)$ is the curve $\{\lambda \in \mathbb{C} : |\arg(\lambda)| = \eta,\ |\lambda| \geq r\} \bigcup \{\lambda \in \mathbb{C} : |\arg(\lambda)| \leq \eta,\ |\lambda| = r\}$, oriented counterclockwise. Since $\eta > \pi/2$ and the function $\lambda \mapsto \mathrm{e}^{t\lambda} R(\lambda, A)$ is holomorphic in $S_{\omega,\theta}$, the definition of e^{tA} makes sense and it is independent of the choice of r and η. To complete the notation we define e^{0A} as the identity operator. Then, based on the following properties, we refer to $\{\mathrm{e}^{tA} : t \geq 0\}$ as the *analytic semigroup* generated by the sectorial operator A.

Theorem 2.16. *Let A be a sectorial operator with domain $\mathcal{D}(A)$ in Banach space X. Then the family of linear bounded operators $\{e^{tA} : t \geq 0\}$ in X defined by (2.6) satisfies the following properties:*

(a) *The semigroup property holds, i.e., $e^{(t+s)A} = e^{tA}e^{sA}$, for any $t, s \geq 0$.*

(b) *The function $t \mapsto e^{tA}x$ from $[0, +\infty[$ into X is continuous at t=0 if and only if x belongs to the closure of $\mathcal{D}(A)$ in X, which implies that condition (b) of the Definition 2.7 is satisfied if the $\mathcal{D}(A)$ is dense in X.*

(c) *The mapping $t \mapsto e^{tA}$ from $]0, +\infty[$ into $L(X)$ has an analytic extension in the sector $\{\lambda \in \mathbb{C} : \lambda \neq 0, \ |\arg(\lambda)| < \theta - \pi/2\}$. Moreover, for any $x \in X$, $t > 0$ and $n = 1, 2, \ldots$, the element $e^{tA}x$ belongs to $D(A^n)$ and $e^{ntA}x = A^n e^{tA}x$.*

(d) *There constants $M_0, M_1, \ldots$, such that*

$$\|t^k(\omega I - A)^k e^{tA}\|_{L(X)} \leq M_k e^{\omega t},$$

where ω is the constant in Definition 2.15. □

It is proved in Lunardi [160, p. 43, Proposition 2.1.11] that if A is a linear operator defined on $\mathcal{D}(A) \subset X$ such that the resolvent set $\rho(A)$ contains a half plane $\{\lambda \in \mathbb{C} \ : \ \Re(\lambda) \geq \omega\}$, and

$$\|\lambda R(\lambda, A)\|_{L(X)} \leq M, \quad \text{if } \Re(\lambda) \geq \omega, \tag{2.7}$$

for some $\omega \in \mathbb{R}$ and $M > 0$, then A is sectorial.

If the constant ω in Definition 2.15 of the sectorial operator A is strictly negative, then $\{e^{tA} : t \geq 0\}$ is a semigroup of contractions. Moreover, fractional powers of A can be defined by the integral formula

$$(-A)^{-\alpha} := \frac{1}{2\pi \mathtt{i}} \int_{\gamma(r,\eta)} (-\lambda)^{-\alpha} R(\lambda, A) \mathrm{d}\lambda, \quad \alpha \in]0, 1[, \tag{2.8}$$

where $\gamma(r, \eta)$ is the curve used in (2.6). It is easy to check that $(-A)^{-\alpha}$ is one-to-one, which allows to define $(-A)^{\alpha}$ as the inverse of $(-A)^{-\alpha}$. The operators $(-A)^{\alpha}$ are called *fractional powers* of $-A$ and its domain $D((-A)^{\alpha})$ provided interpolations spaces. A key representation is the following

$$(-A)^{\alpha} e^{tA} = \frac{1}{2\pi \mathtt{i}} \int_{\gamma(r,\eta)} e^{t\lambda} (-\lambda)^{\alpha} R(\lambda, A) \mathrm{d}\lambda, \quad t > 0, \tag{2.9}$$

More details can be found in Lunardi [160, Chapter 2, pp. 35–67] and Pazy [196, Chapter 2, pp. 42–75].

• *Remark* 2.17. The following representations of fractional powers don't use complex integration, see Pazy [196, Section 2.6]

$$(-A)^{\alpha} x = \frac{\sin(\pi\alpha)}{\pi} \int_0^{\infty} t^{\alpha-1} A(tI - A)^{-1} x \, \mathrm{d}t, \quad \forall x \in \mathcal{D}(A),$$
$$(-A)^{-\alpha} = \frac{\sin(\pi\alpha)}{\pi} \int_0^{\infty} t^{-\alpha} (tI - A)^{-1} \mathrm{d}t, \quad \text{in } H,$$
$$(-A)^{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} t^{-\alpha} e^{tA} \mathrm{d}t, \quad \text{in } H,$$

for α in $(0, 1)$. □

An important subclass of analytic semigroups are the so called *variational*. This semigroups are defined on a Hilbert space H, instead of a Banach space X, and generated as follows.

Definition 2.18 (variational)**.** A linear operator A with domain $\mathcal{D}(A)$ in a Hilbert space H is *variational* if there exist a Hilbert space V densely embedded in H and a continuous bilinear form $a : V \times V \to \mathbb{R}$ such that

$$a(v,v) \geq \alpha \|v\|_V - \beta \|v\|_H, \quad \forall\, v \in V$$

for some constants $\alpha > 0$ and $\beta \geq 0$, and the conditions

$$\begin{cases} v \in \mathcal{D}(A) \ \text{ iff } \ |a(v,h)| \leq M_v \|h\|_H, & \forall\, h \in H, \\ a(u,v) = \langle -Au, v \rangle, & \forall\, u \in \mathcal{D}(A),\ v \in V. \end{cases}$$

determine the domain and the operator. □

The lower and upper bounds (coercivity and continuity) of the bilinear form $a(\cdot,\cdot)$ yield $\|\lambda x - Ax\|_H \geq c_0 |\lambda - \beta|$ if $\Re(\lambda) > \beta$, for a constant c_0 depending only the ratio M/α. On the other hand, Lax-Milgram Theorem can be used to solve the equation $Ax + \lambda x = y$ in x for any y in H. Hence A is a sectorial operator, as proved in Tanabe [246] (see also Jacob [115, Vol. I, Section 4.7, Theorem 4.7.10]), i.e.,

Theorem 2.19. *If A is a variational operator in a Hilbert space H satisfying the above definition, then A generate an analytic semigroup $\{\mathrm{e}^{tA} : t \geq 0\}$ such that $\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{\beta t}$, for all $t \geq 0$. Moreover, the following* energy *estimate holds*

$$\int_0^t a(\mathrm{e}^{sA} v, \mathrm{e}^{sA} v) \mathrm{d}s = \frac{1}{2} \big[\|v\|_H^2 - \|\mathrm{e}^{tA} v\|_H^2 \big], \quad \forall\, t \geq 0,\ v \in V.$$

On the other hand, if $\{S(t) : t \geq 0\}$ is an analytic semigroup on a Hilbert space H then there exists a variational operator A satisfying $S(t) = \mathrm{e}^{tA}$. □

For instance, the reader may check the books Brezis [34] and Zeidler [264] regarding monotone operators.

Given an analytic semigroup $\{S(t) : t \geq 0\}$ on a Hilbert space H, we may define a family of bilinear forms

$$a^{(t)}(u,v) = \frac{1}{t}(u - S(t)u, v), \quad \forall t > 0$$

and

$$\mathcal{D}(a) := \big\{ u \in H : \sup_{t>0} a^{(t)}(u,u) < \infty \big\}$$

to prove that

$$a(u,v) := \lim_{t \to 0} a^{(t)}(u,v), \quad \forall u,\, v \in \mathcal{D}(a),$$

e.g., see Fukushima et al. [92], Ma and Röckner [161] for a full discussion on Dirichlet forms.

This notion is related to the so-called *monotone* operators in Hilbert spaces and its application to *control theory*, e.g. see Zabczyk [263, Part IV, Chapter 1, pp. 176–205].

Let us give a perturbation result for analytic semigroups, e.g., see Jacob [115, Vol I, Theorem 4.4.5, p. 323],

Theorem 2.20. *Let* $\{\mathrm{e}^{tA} : t \geq 0\}$ *be an analytic semigroup on a Banach space* $(X, \|\cdot\|_X)$. *Suppose* $(\mathcal{D}(B), B)$ *is a closed linear operator with* $\mathcal{D}(B) \subset \mathcal{D}(A)$ *and such that for every* $\varepsilon > 0$ *there is* $\beta(\varepsilon)$ *satisfying*

$$\|Bx\|_X \leq \varepsilon \|Ax\|_X + \beta(\varepsilon)\|x\|_X, \quad \forall x \in \mathcal{D}(A).$$

Then $(\mathcal{D}(A), A + B)$ *generates an analytic semigroup* $\{\mathrm{e}^{t(A+B)} : t \geq 0\}$. □

The proof is based on the fact that $\|B(\lambda I - A)^{-1}\|_{L(X)} < 1$, so $I - B(\lambda I - A)^{-1}$ is invertible, for $|\lambda|$ sufficiently large within a sector.

2.4 Resolvent Properties

The notation $R(\lambda, A) := (\lambda I - A)^{-1}$ (used until know to denote the *resolvent* operator) stress the dependency on the infinitesimal generator A of the given C_0 semigroup $\{S(t) : t \geq 0\}$ on a Banach space X. However, the resolvent operators can be constructed independent of A, as the Laplace transform of the semigroup, i.e.,

$$R(\lambda)x := \int_0^\infty \mathrm{e}^{-\lambda t} S(t)x \, \mathrm{d}t, \tag{2.10}$$

which enjoys several characteristic properties used in the following

Definition 2.21 (resolvent)**.** A one-parameter family $\{R(\lambda) : \lambda > 0\}$ of linear bounded operators from the Banach space X into itself is called a *regular contracting resolvent* or *strongly continuous contraction resolvent* if it satisfies

(a) $R(\lambda) - R(\nu) = (\nu - \lambda)R(\lambda)R(\nu), \quad \forall \lambda, \nu > 0,$

(b) $\|\lambda R(\lambda)\| \leq 1, \quad \forall \lambda > 0,$

(c) $\lim_{\lambda \to \infty} \|\lambda R(\lambda)x - x\| = 0, \quad \forall x \in X.$

Here, (a) is the resolvent equations, (b) means that the operator $\lambda R(\lambda)$ *(not* $R(\lambda)$*)* is a contraction mapping on X, and the regularity condition (c) is the equivalent of the *strong continuity*. The resolvent is called *closed* if there is a bounded operator denoted by $R(0)$ such that (a) holds for any $\lambda \geq 0$. □

For any strongly continuous contraction resolvent $\{R(\lambda) : \lambda > 0\}$ on a Banach space X there exist a unique closed non-positive and densely define

operator $(\mathcal{D}(A), A)$ on X, called the *generator of the resolvent* such that $R(\lambda) = (\lambda I - A)^{-1}$. As mentioned later, the operator A can be define in term of the resolvent.

Given a strong continuous semigroup of contraction $\{S(t) : t \geq 0\}$ as in Definition 2.7 the map $t \mapsto S(t)x$ is continuous for any x in X so that the Laplace transform (2.10) makes sense and satisfies the conditions of a regular contracting resolvent $\{R(\lambda) : \lambda > 0\}$. The converse of this statement also holds, i.e., any regular contracting resolvent is the Laplace transform of a strong continuous semigroup of contraction, actually we may write

$$S(t)x := \lim_{n\to\infty} \left(\tfrac{n}{t} R(\tfrac{n}{t})\right)^n x, \quad \forall x \in X,$$

see Dellacherie and Meyer [58, Theorem XIII.4, pp. 87–89].

A resolvent $\{R(\lambda) : \lambda > 0\}$ is always infinitely differentiable for any $\lambda > 0$. For a contracting resolvent $\{R(\lambda) : \lambda > 0\}$, (i.e., when only (a) and (b) of the Definition 2.21 is satisfied) we may define the *regularity space* R and *co-regularity space* C by the conditions

$$\begin{cases} R := \{x \in X \ : \ \lim_{\lambda\to\infty} \lambda R(\lambda)x = x\} \quad \text{and} \\ C := \{\lim_{\lambda\to\infty} \lambda R(\lambda)x = 0\} \end{cases} \tag{2.11}$$

which are closed subspaces of the initial Banach space X. As mentioned in Definition 2.21, if $R = X$ the resolvent is called *regular*, and now it is called *co-regular* if $C = X$. In view of the resolvent equation, the image (or range) space $R(\lambda)X$ is independent of λ and its closure produce the regularity space R. Thus R an invariant subspace. Hence, by restricting the definition of the resolvent $\{R(\lambda) : \lambda > 0\}$ to the Banach subspace R, we get a regular contracting resolvent. Hence, in what follows we assume that the given resolvent is regular, i.e., $R = X$, without any loss of generality.

Since $R'(\lambda)$, defined by $R'(1/\lambda) := \lambda[I - \lambda R(\lambda)]$, is a resolvent satisfying $\|\lambda R'(\lambda)\| \leq 2$ we see that the image (or range) space $\big(I - \lambda R(\lambda)\big)X$ is independent of λ and its closure produce the co-regularity space C. Thus, the behavior of $R(\lambda)$ as λ vanishes can be deduced from the behavior as λ goes to infinity.

The generator of the resolvent $\{R(\lambda) : \lambda > 0\}$ is the linear (possible unbounded) operator A with domain $\mathcal{D}(A)$ defined by the condition

$$x \in \mathcal{D}(A) \ \text{ and } \ Ax = y \quad \text{ if and only if } \quad \lim_{\lambda\to\infty} \lambda[\lambda R(\lambda) - I]x = y.$$

Similarly, the co-generator V with domain $\mathcal{D}(V)$ defined by the condition

$$x \in \mathcal{D}(V) \ \text{ and } \ Vx = y \quad \text{ if and only if } \quad \lim_{\lambda\to 0} R(\lambda)x = y.$$

If the resolvent is closed by a bounded linear operator $R(0)$ then $R(0) - R(\lambda) = \lambda R(\lambda)R(0)$ for any $\lambda > 0$. Thus $[I - \lambda R(\lambda)][I + \lambda R(0)] = I$ and therefore $\mathcal{D}(V) = [I - \lambda R(\lambda)]X = X$, i.e., the resolvent is co-regular and $V = R(0)$. The

co-generator V may be called the (strong) *potential*. The various weak versions of this potential form part of the so-called *potential theory* and *ergodic theory*.

Since the resolvent equation (a) of Definition 2.21 implies that $R(\lambda)$ is injective and $R(\lambda)X$ is included in the domain of the generator A. Actually, $R(\lambda)$ is an isomorphism from X onto $\mathcal{D}(A)$, for any $\lambda > 0$, and we have $x \in \mathcal{D}(A)$ and $Ax = y$ if and only if form some $\lambda > 0$ we have $x, y \in X$ and $x = R(\lambda)(\lambda x - y)$. This give a characterization of the domain $\mathcal{D}(A)$ without involving a passage to the limit. The generator A coincides with the infinitesimal generator A defined for a strongly continuous semigroup, see Dellacherie and Meyer [58, Chapters XI–XVI], Rogers and Williams [214, Section III.5, pp. 303–320] for a comprehensive treatment.

Sometimes, the Definition 2.21 of resolvent treats each property (a), (b) and (c) independently, e.g., see Engel and Nagel [74, Section III.4, pp. 205–218] or Pazy [196, Section 1.9, pp. 36–38]. In this context, given a subset of the complex plane, a one-parameter family $\{R(\lambda) : \lambda \in \Lambda\}$ of bounded linear operators from the Banach space X into itself is called a *pseudo-resolvent* in Λ if (a) is satisfied for any λ and ν in Λ. Next, (b) is relaxed as follows: there exists an unbounded sequence $\{\lambda_n\}$ in Λ such that

$$\text{(b')} \quad \|\lambda_n R(\lambda_n)\| \le M, \quad \forall n,$$

for some constant $M > 0$. In particular, the resolvent are no longer contractions. Similarly, (c) is relaxed, the limit is taken within the unbounded sequence $\{\lambda_n\}$, i.e., there exists a sequence $\{\lambda_n\}$ in Λ such that $|\lambda_n| \to \infty$ and

$$\text{(c')} \quad \lim_n \|\lambda R(\lambda)x - x\| = 0, \quad \forall x \in X.$$

There are two typical results for a pseudo-resolvent $\{R(\lambda) : \lambda \in \Lambda\}$:

(1) The null space $\mathcal{N}(R(\lambda))$ and the range $\mathcal{R}(R(\lambda))$ are independent of λ, and $\mathcal{N}(R(\lambda))$ is a closed subspace of X.

(2) If condition (c') holds (or equivalently, the range $\mathcal{N}(R(\lambda))$ is dense in X and (b') is satisfied), then there exits a unique densely defined closed linear operator A in X such that $R(\lambda) = (\lambda I - A)^{-1}$.

As an example, suppose that $S(t)$ satisfies the semigroup property $S(t + s) = S(t)S(s)$, but it is not a strongly continuous semigroup (i.e., we do not necessarily have $S(t)f \to f$ as $t \to 0$). However, its Laplace transform

$$R(\lambda)f := \int_0^\infty \mathrm{e}^{-\lambda t} S(t)f \mathrm{d}t, \quad \lambda \in \Lambda$$

may be one-to-one (injective) in X, e.g., when X is Banach space of continuous functions. Because $R(\lambda)$ satisfies the resolvent identity, there exits a unique closed operator A such that $R(\lambda) = (\lambda I - A)^{-1}$, which is called the *generator* of $\{S(t) : t \ge 0\}$.

2.5 Bilinear Forms

Another way to construct a (C_0) semigroup of operators is by means of bilinear forms (e.g., see Reed and Simon [210, Vol. I, p. 276].

A bilinear form $a(u,v)$ on a *real* Hilbert space H, with inner or scalar product $(\cdot,\cdot)$, is a map $a\colon \mathcal{D}(a)\times\mathcal{D}(a)\to\mathbb{R}$, where $\mathcal{D}(a)$, called the *domain* of the form, is a dense subspace of H, such that $a(u,v)$ is separately linear in each variable. In the sequel we will often use the word "form" for "bilinear form".

Together with a form $a(u,v)$ we consider the following two bilinear forms

$$a_{\text{sym}}(u,v)=\frac{a(u,v)+a(v,u)}{2},\qquad a_{\text{ant}}(u,v)=\frac{a(u,v)-a(v,u)}{2},$$

respectively the *symmetric* and the *antisymmetric* part. Associated to a bilinear form is the *quadratic form* $q(u)=a(u,u)$ with $u\in\mathcal{D}(a)$, i.e., the bilinear form evaluated on the diagonal, and we have the following *polarization* identity:

$$a_{\text{sym}}(u,v)=\frac{1}{4}\Big(a(u+v,u+v)-a(u-v,u-v)\Big),$$

that shows how it is possible to recover the symmetric part of a form by knowing the associated quadratic form. Obviously, the form is called symmetric if we have $a(u,u)=a_{\text{sym}}(u,u)$ or equivalently $a_{\text{ant}}(u,u)=0$.

• *Remark* 2.22. Given a bilinear form $a(\cdot,\cdot)$ on $\mathcal{D}(a)\subset H$ we can associate an (possibly unbounded) operator $A\colon \mathcal{D}(A)\subset\mathcal{D}(a)\to H$ by the formula $a(u,v)=(-Au,v)$ for every v in $\mathcal{D}(a)$ and any u in $\mathcal{D}(A)$, where the domain $\mathcal{D}(A)$ is defined by all u in $\mathcal{D}(a)$ such that, for some constant K_u, we have $|a(u,v)|\le K_u\,|v|$ for all v in $\mathcal{D}(a)$ (recall that $\mathcal{D}(a)$ is dense in H), hence we can define $-Au$ by means of Riesz' theorem. □

2.5.1 Bounded Forms

If the bilinear form $a(u,v)$ satisfies the following property: there exist a constant $C>0$ such that

$$|a(u,v)|\le C|u|\,|v|,\qquad \forall u,v\in\mathcal{D}(a),\tag{2.12}$$

then the form is called continuous (or bounded) in H. It can be extended in a unique way to a form, that we indicate still by a, defined on the whole H. In this case the previous remark takes the following form: there exists a unique bounded operator $-A$ such that $a(u,v)=(-Au,v)$ for any $u,v\in H$.

The following result is very important in the variational formulation of partial differential equations,

Theorem 2.23 (Lax-Milgram)**.** *Let $a(\cdot,\cdot)$ be a continuous (or bounded, i.e., (2.12) holds) bilinear form defined on a Hilbert space H with inner product $(\cdot,\cdot)$, satisfying the following coercivity assumption*

$$|a(v,v)|\ge\kappa|v|^2,\qquad \forall v\in H.\tag{2.13}$$

Then for any $f \in H$ there exists a unique $u \in H$ such that

$$a(u,v) = (f,v), \qquad \forall v \in H.$$

Moreover, if the form $a(\cdot,\cdot)$ is also symmetric the previous solution u is characterized by the property

$$\tfrac{1}{2}a(u,u) - (f,u) = \min_{v\in H}\left\{\tfrac{1}{2}a(v,v) - (f,v)\right\}.$$

Furthermore, the mapping $T : f \mapsto u$ from the H into itself is a linear homomorphism satisfying

$$\kappa\|u\|_H \le \|T^{-1}u\|_H \le C\|u\|_H, \qquad \forall u \in H,$$

where C and κ are the constants in (2.12) and (2.13). □

For a given element f in the dual space H', we may consider the problem of finding u in H such that

$$a(u,v) = \langle f, v\rangle, \qquad \forall v \in H.$$

Again, the mapping $f \mapsto u$ is a linear homeomorphism from H' onto H.

2.5.2 Positive Forms

A bilinear form (not necessarily bounded) is said *positive definite*, or only *positive*, or *nonnegative*, if $a(u,u) \ge 0$ for all $u \in \mathcal{D}(a)$. If also $a(u,u) = 0$ implies $u = 0$ then it is called *strictly positive*.

Given a positive form $a(u,v)$ we will consider, for any $\alpha > 0$ the following forms

$$(u,v)_\alpha = a_{\text{sym}}(u,v) + \alpha\,(u,u), \tag{2.14}$$

which are scalar products on $\mathcal{D}(a)$ and all the norms $\|u\|_\alpha$ induced are equivalent, see e.g. Ma and Röckner [161, Section I.2, p. 15]. Hence there exists a unique completion V of $\mathcal{D}(a)$ with respect to any of the previous norms, and we have $V \subset H$. Clearly, V is a Hilbert space with respect to any of the previous norms: the topology doesn't change also if the scalar product depends on α. Thus, a positive form is *closed* if $\mathcal{D}(a) = V$, i.e., $\mathcal{D}(a)$ is complete (or a Hilbert space) endowed with any norm $\|\cdot\|_\alpha$, $\alpha > 0$.

• *Remark* 2.24. It is easy to check that a positive form is closed if and only if whenever $u_n \in \mathcal{D}(a)$, $u_n \to u$ in H and $a(u_n - u_m, u_n - u_m) \to 0$, as $n, m \to \infty$, then $u \in \mathcal{D}(a)$ and $a(u_n - u, u_n - u) \to 0$. Hence the property to be closed is relative only to the symmetric part of a positive form. □

• *Remark* 2.25. Note that the symmetric part $a_{\text{sym}}(u,v)$ of a positive form satisfies $a_{\text{sym}}(u,v)^2 \le a_{\text{sym}}(u,u)a_{\text{sym}}(v,v)$, for every u and v in $\mathcal{D}(a)$. Hence we have

$$|a_{\text{sym}}(u,v)| \le \|u\|_\alpha\|v\|_\alpha, \quad \forall u,v \in \mathcal{D},$$

i.e., the symmetric part of a positive form is continuous in $\mathcal{D}(a)$, endowed with any norm $\|\cdot\|_\alpha$, or simply it is bounded in V. If the form is *strictly* positive, then the symmetric part itself $a_{\text{sym}}(u,v)$ is a scalar product. □

• *Remark* 2.26. A generalization of positive form is the notion of *semibounded* form (also called *lower bounded* or *bounded below* form), i.e. $a(u,u) \geq -M\,|u|^2$ for some $M > 0$ and all $u \in \mathcal{D}(a)$. Many results for positive forms hold true also for semibounded forms. Also, note that the *symmetry* of the form is not required in the above definition. □

We have (e.g., see Reed and Simon [210, Vol I, Thm VIII.15])

Theorem 2.27 (symmetric)**.** *Given a closed, positive and symmetric form $a(u,v)$ there is a unique (non positive and densely defined) self-adjoint operator $A\colon \mathcal{D}(A) \subset \mathcal{D}(a) \subset H \to H$ such that $a(u,v) = (-Au,v)$ for $u \in \mathcal{D}(A)$. Conversely, given a non positive, self-adjoint operator A, the form $a(u,v) = ((-A)^{1/2}u, (-A)^{1/2})$ with $u,v \in \mathcal{D}(a) = \mathcal{D}((-A)^{1/2})$ is a closed, positive and symmetric form.* □

• *Remark* 2.28. A closed, positive and symmetric form $a(u,v)$ is always a continuous (bounded) form in $V = \mathcal{D}(a)$. □

For the non-symmetric case we have

Theorem 2.29 (non symmetric)**.** *Let $a(\cdot,\cdot)$ be a closed and positive bilinear form on $\mathcal{D}(a) \subset H$, which is continuous in $\mathcal{D}(a) = V$ with the norm $\|\cdot\|_1$ as in (2.14) for $\alpha = 1$, i.e.,*

$$|a(u,v)| \leq M\|u\|_1\|v\|_1, \quad \forall u,v \in \mathcal{D}(a), \tag{2.15}$$

for some constant M. Then there exits a unique (densely defined) operator $A\colon \mathcal{D}(A) \subset \mathcal{D}(a) \subset H \to H$, such that $a(u,v) = (-Au,v)$ for any $u \in \mathcal{D}(A)$ and $v \in \mathcal{D}(a)$. Moreover, A is the infinitesimal generator of a strongly continuous contraction semigroup. Conversely given a strongly continuous contraction semigroup, with infinitesimal generator A, there exists a minimal closed and positive bilinear form $a(\cdot,\cdot)$, on $\mathcal{D}(a) \subset H$, such that $\mathcal{D}(A) \subset \mathcal{D}(a)$, satisfying (2.15). □

• *Remark* 2.30. The continuity condition (2.15) is also called weak sectorial in Ma and Röckner [161]. Actually, considering the operator A extended to the complexification of H, it generates an analytic semigroup. In view of Remark 2.25, the bound (2.15) is equivalent to

$$|a_{\text{ant}}(u,v)| \leq M\|u\|_1\|v\|_1, \quad \forall u,v \in \mathcal{D},$$

which clearly is a condition on the antisymmetric part, in other words, it gives control of the antisymmetric part with the symmetric part. □

• *Remark* 2.31. The previous theorem states that A is the infinitesimal generator of a strongly continuous contraction semigroup. Hence for $\alpha > 0$, $\alpha I - A$ has a bounded inverse $(\alpha I - A)^{-1}$ in H such that $\alpha\|(\alpha I - A)^{-1}\|_{L(H)} \leq 1$ and, denoting $u = (\alpha - A)^{-1}f$ with f in H, we have that u is the unique element in $\mathcal{D}(a)$ such that $a(u,v) + \alpha(u,v) = (f,v)$ for every v in $\mathcal{D}(a)$. □

2.5.3 Closable Forms

The usual way to define a positive closed form is first to define it on a suitable domain $\mathcal{D}$ and then trying to extend the form to a closed form. This procedure requires the form to closable.

A positive form is *closable* if whenever $u_n \in \mathcal{D}$, $u_n \to 0$ in H and $a(u_n - u_m, u_n - u_m) \to 0$, as $n, m \to \infty$, then $a(u_n, u_n) \to 0$.

Also this property of closability is a condition on the symmetric part. In fact, in the case of a closable positive symmetric form $a(u, v)$, it is easy to check that the form has a smallest closed extension to $V = \overline{\mathcal{D}}$, where $\overline{\mathcal{D}}$ is the completion of $\mathcal{D}$ with respect to the norm $\|\cdot\|_1$.

Another application is the following (see Reed and Simon [210, Vol II, Thm X.23])

Theorem 2.32 (Friedrichs extension)**.** *Let A be a positive symmetric operator and let $a(u, v) = (Au, v)$ for every $u, v \in \mathcal{D}(A)$. Then $a(\cdot, \cdot)$ is a closable symmetric positive bilinear form and if $\hat{a}(\cdot, \cdot)$ denotes its smallest closed symmetric positive bilinear extension form then there exists a unique positive self-adjoint operator $\hat{A}$, which is an extension of A, such that $\hat{a}(u, v) = (\hat{A}^{1/2}u, \hat{A}^{1/2}v)$ for every $u, v \in \mathcal{D}(\hat{A}^{1/2})$.* □

The closability of the form is due mainly to the fact that the operator A is a symmetric operator. Clearly, Theorem 2.27 is used to obtain the extension $\hat{A}$.

Let us consider the following example taken from Reed and Simon [210, p. 280]: let $H = L^2(0, 1)$ and define the operator $Au = -u''$ on $\mathcal{D}(A) = C_0^\infty(0, 1)$, i.e., infinite differentiable functions with compact support in $(0, 1)$. This operator is a positive, symmetric and $\mathcal{D}(A^*) = H^2(0, 1)$, the classic Sobolev space of order 2. The closure A^{**} has the domain $\mathcal{D}(A^{**}) = \{u \in H^2(0, 1) : u(0) = u(1) = u'(0) = u'(1) = 0\}$. The Friedrichs (self-adjoint) extension $\hat{A}$ has the domain $\mathcal{D}(\hat{A}) = H_0^1(0, 1) \cap H^2(0, 1)$. There are many other self-adjoint extensions as $\mathcal{D}(A_{a,b}) = \{u \in H^2(0, 1) : au(0) + u'(0) = bu(1) + u'(1) = 0\}$.

On the other hand, there are positive symmetric forms which are not closable, e.g., take $H = L^2(\mathbb{R})$, $\mathcal{D}(a) = C_0^\infty(\mathbb{R})$ and $a(u, v) = u(0)\, v(0)$.

• *Remark* 2.33. Let $a(\cdot, \cdot)$ be a positive definite bilinear form with domain $\mathcal{D}(a)$, not necessarily closed, but satisfying the sector (or continuity) condition (2.15). It is clear that the bound (2.15) holds for $a_1(\cdot, \cdot) = a(\cdot, \cdot) + (\cdot, \cdot)$ instead of $a(\cdot, \cdot)$ with another constant M'. Thus, the inequality

$$a_1(u, u) \le a_1(u - v, u - v) + M' \sqrt{\|u - v\|_1 \, \|u\|_1} + a_1(u, v),$$

where $\|\cdot\|_1 = \sqrt{a_1(\cdot, \cdot)}$, shows that if $\{v_n : n \ge 1\}$ is a Cauchy sequence in the graph norm $\|\cdot\|_1$, $v_n \to 0$ in H and $a(v, v_n) \to 0$ for every v in $\mathcal{D}(a)$, then $a(v_n, v_n) \to 0$. Therefore, following Theorems 2.29 and 2.32, for a given closed non positive definite densely defined operator A with domain $\mathcal{D}(A) \subset H$, we define the bilinear form $a(u, v) = (-Au, v)$ for any u and v in $\mathcal{D}(a) = \mathcal{D}(A)$. If $a(\cdot, \cdot)$ satisfies the sector (or continuity) condition (2.15) then the bilinear form $a(\cdot, \cdot)$ is closable, and its closure is defined on $\overline{\mathcal{D}(a)} = V$, with $\mathcal{D}(A) \subset V \subset H$,

and A is the infinitesimal generator of a strongly continuous analytic contraction semigroup, see Theorem 2.19 on variational generators. □

Thus, in the non symmetric case, if the closable form $a(u,v)$ satisfies also condition (2.15), or in other words it is (weak) sectorial, then it is possible to define its smallest closed form on $V = \overline{\mathcal{D}}$ that extends $a(\cdot,\cdot)$.

Given a positive form $a(u,v)$ with (dense) domain $\mathcal{D}(a)$, let us suppose there exists a Hilbert space V, such that $\mathcal{D}(a) \subset V \subset H$, where the inclusion of V in H is supposed continuous and $\mathcal{D}(a)$ is supposed dense in V. We will call the form *coercive* (with respect to V), or V-elliptic as Lions-Magenes [155, p. 217], if

$$a(u,u) \geq m\|u\|_V^2 - \lambda|u|^2, \quad \forall u \in \mathcal{D}(a), \tag{2.16}$$

for some positive constant $m, \lambda > 0$. If a coercive positive form is continuous on V, i.e., there exists a constant $M > 0$ such that

$$|a(u,v)| \leq M\|u\|_V\|v\|_V, \quad \forall u,v \in \mathcal{D}(a), \tag{2.17}$$

then it is closable. Hence it is possible to extend the form to all V in such a way the form is a positive closed form that verifies (2.17). Notice that a bilinear form satisfying (2.16) is not necessarily positive, only $a(\cdot,\cdot) + \lambda(\cdot,\cdot)$ is actually positive.

• *Remark* 2.34. If $a(\cdot,\cdot)$ is a closable, positive and continuous bilinear form with domain $\mathcal{D}(a) \subset H$, then we can extend the definition of $a(\cdot,\cdot)$ to a Hilbert space $V \subset H$ with continuous and dense inclusion, actually V is the completion of $\mathcal{D}(a)$ with the any of the graph norm induced by the inner product (2.14), e.g., $(\cdot,\cdot)_1 = (\cdot,\cdot) + a_{\text{sym}}(\cdot,\cdot)$ which becomes the inner product $(\cdot,\cdot)_V$ in V. Thus, $a(\cdot,\cdot) + \alpha(\cdot,\cdot)$ with $\alpha > 0$ is a continuous and coercive bilinear form in V so that Lax-Milgram Theorem 2.23 can be used, i.e., for any f in the dual space V' there is a unique u in V such that

$$a(u,v) + \alpha(u,v) = \langle f, v\rangle, \qquad \forall v \in V.$$

Setting $u = R_\alpha f$, we have a linear homeomorphism from V' into V. In particular, if f belong to H and $\langle f, v\rangle = (f,v)$ the family of operators $R_\alpha : H \to H$, $\alpha > 0$ is a resolvent with $\alpha\|R_\alpha\|_{L(H)} \leq 1$, i.e., a contraction resolvent. If $a(u,v) = (-Au,v)$ for every u,v in $\mathcal{D}(A)$ then $R_\alpha = (\alpha I - A)^{-1}$ and $\mathcal{D}(A) = R_\alpha(H)$. Since

$$\alpha R_\alpha f - f = R_\alpha[\alpha f - (\alpha I - A)f] = R_\alpha A f,$$

which yields $|\alpha R_\alpha f - f| \leq |Af|/\alpha$. This proves that the resolvent operators are strongly continuous, i.e., $|\alpha R_\alpha f - f| \to 0$ as $\alpha \to 0$ for every f in H. □

• *Remark* 2.35. Let $(\mathcal{D}(A), A)$ be a non positive closed densely defined operator on H. Define the bilinear form $a(u,v) := (-Au,v)$ for any u and v in $\mathcal{D}(A)$ and assume that it satisfies a weak sectorial condition, namely,

$$|(Au,v)| \leq M\|u\|_1\|v\|_1, \quad \forall u,v \in \mathcal{D}(A), \tag{2.18}$$

with $\|\cdot\|_1$ is the norm corresponding to the scalar product

$$(u,v)_1 := \tfrac{1}{2}[(-Au,v)+(u,-Av)]+(u,v),$$

in particular this hold if A is symmetric. Then, according to Remark 2.33, the form $a(\cdot,\cdot)$ is closable and its closure is defined in the Hilbert space V, the completion of $\mathcal{D}(A)$ with the norm $\|\cdot\|_V := \|\cdot\|_1$. Denote by $R_\alpha := (\alpha I - A)^{-1}$ the (Lax-Milgram) resolvent operators given in Remark 2.34. Now, define the family of bilinear forms

$$a^{(\alpha)}(u,v) = \alpha(u - \alpha R_\alpha u, v), \quad \forall \alpha > 0$$

and

$$\mathcal{D}(a) := \big\{u \in H : \sup_{\alpha>0} a^{(\alpha)}(u,u) < \infty\big\}.$$

Then $\mathcal{D}(a) = V$ and

$$a(u,v) := \lim_{\alpha\to\infty} a^{(\alpha)}(u,v), \quad \forall u,\, v \in \mathcal{D}(a),$$

e.g., see Fukushima et al. [92], Ma and Röckner [161]. □

• *Remark* 2.36. In the case of a complex Hilbert space the theory of bilinear forms can be extended to the so called *sesquilinear* forms. Recall that a sesquilinear form is a map $a(\cdot,\cdot)\colon \mathcal{D}(a)\times\mathcal{D}(a)\to\mathbb{C}$, which is linear in the first variable and antilinear([2]) in the second variable (analogously to a complex scalar product). □

2.6 Abstract Cauchy Problem

Given a linear (not necessarily bounded) operator A, with a dense domain $\mathcal{D}(A)$ on a Banach space X and an initial value x in X, consider the following linear evolution equation in X

$$\begin{cases} \dot{u}(t) &= Au(t), \quad \forall\, t \geq 0, \\ u(0) &= x, \end{cases} \tag{2.19}$$

where the *dot* means derivative in t or time derivative. This is an *initial value problem* (IVP) or the *Cauchy problem* relative to the operator A.

The Cauchy problem is *well posed* if we have a unique solution (existence and uniqueness) and continuous dependency of initial data i.e.,

(a) for any x in $\mathcal{D}(A)$ there exists exactly one strongly differentiable function $u(\cdot)$ on the set $[0,+\infty[$, satisfying (2.19), and

(b) if $\{x_n : n = 1,2,\dots\}$ is a sequence in $\mathcal{D}(A)$ converging to 0 in X, then for all $t \geq 0$ the corresponding solutions $u_n(t)$ of (2.19) converge to 0.

[2] A antilinear map is also said *semilinear*. The prefix *sesqui* derives from the latin language, it means "one and a half". In fact sesquilinear describes the fact that the map is linear in the first variable and semilinear in the second one.

The semigroup theory was built on the above Cauchy problem. The next step is to consider a non homogeneous Cauchy problem in a Banach space, namely

$$\begin{cases} \dot{u}(t) &= Au(t) + f(t), \quad \forall\, t \geq 0, \\ u(0) &= x, \end{cases} \tag{2.20}$$

where $f(\cdot)$ is a given function. Here it is necessary to specify the class of functions $f(\cdot)$ that we will allow, which in turn will determine where the solution will belongs. We will make use of the following Banach spaces, with their natural norms and for some T in $(0, +\infty)$,

$C([0,T], X)$ or $C([0,T], D_A)$ (strongly) continuous functions on $[0,T]$ with values in X or D_A, where D_A is the Banach space $\mathcal{D}(A)$ endowed with the graph norm $x \mapsto \|x\| + \|Ax\|$ associated with the operator A,

$C^1([0,T], X)$ (strongly) continuously differentiable functions on $[0,T]$ with values in X, $L^p(]0,T[, X)$, $1 \leq p \leq +\infty$ Lebesgue space of p-integrable functions on $]0,T[$ with values in X, and similarly with D_A instead of X,

$W^{1,p}(]0,T[, X)$, $1 \leq p \leq +\infty$ defined as the closure of $C^1([0,T], X)$ in the Sobolev norm $f \mapsto \|f\|_{L^p} + \|\dot{f}\|_{L^p}$ as a subspace of $L^p(]0,T[, X)$. Notice that the space $W^{1,p}(]0,T[, X)$ has a dense and continuous inclusion into $C([0,T], X)$.

Thus, a *strict solution* in $L^p(]0,T[, X)$ [resp. in $C([0,T], X)$] is a function u that belong to $W^{1,p}(]0,T[, X) \cap L^p(]0,T[, D_A)$ [resp. in $C^1([0,T], X) \cap C([0,T], D_A)$ and fulfills equation (2.20). On the other hand, a *weak solution* is a function u in $C([0,T], X)$ such that for every φ in $\mathcal{D}(A')$,

$$\langle \varphi, u(t) \rangle = \langle \varphi, x \rangle + \int_0^t \langle A'\varphi, u(s) \rangle \mathrm{d}s + \int_0^t \langle \varphi, f(s) \rangle \mathrm{d}s,$$

where A' is the dual operator of A.

We refer to Da Prato and Zabczyk [51, Appendix A] for more details.

Theorem 2.37. *Let A be the infinitesimal generator of a semigroup $\{S(t) : t \geq 0\}$ of class (C_0) on X, and let f be in $L^1(]0,T[, X)$. Then for any x in $\mathcal{D}(A)$ there exists a unique weak solution u of the IVP (2.20) and the variation of constant formula*

$$u(t) = S(t)x + \int_0^t S(t-s) f(s) \mathrm{d}s, \quad \forall\, t \in [0,T], \tag{2.21}$$

holds. This is called a mild *solution.* □

By means of the Yosida approximations $A_n = nA(nI - A)^{-1}$ of A, we see that

$$u_n(t) = \exp(tA_n)x + \int_0^t \exp[(t-s)A_n] f(s) \mathrm{d}s, \quad \forall\, t \in [0,T]$$

is the unique solution in $W^{1,1}(]0,T[, X)$ of an IVP similar to (2.20), with A_n instead of A, and moreover $u_n \to u$ in $C([0,T], X)$.

Theorem 2.38. *Let A be the infinitesimal generator of a semigroup $\{S(t) : t \geq 0\}$ of class (C_0) on X, and let x be in $\mathcal{D}(A)$. If f belongs to the space $W^{1,p}(]0,T[,X)$, $1 \leq p \leq +\infty$, [resp. to $L^p(]0,T[,D_A)$], then the IVP (2.20) has a unique strict solution u in $L^p(]0,T[,X)$ [resp. in $C([0,T],X)$], which is given by the formula (2.21) and moreover u belongs to $C^1([0,T],X) \cap C([0,T],D_A)$ [resp. $W^{1,p}(]0,T[,X) \cap C([0,T],D_A)$].* □

More properties can be found in Fattorini [78]. There are several results about regularity of the solution, for instance, if f is in $C^\alpha([0,T],X)$, for some $0 < \alpha < 1$ (i.e., Hölder continuous function), then the solution u of the IVP (2.20) belongs to $C^1([0,T],X)$. Results of maximal regularity involve the use of fractional powers operator and interpolation spaces, see Lunardi [160] for example.

In some case, it is necessary to consider an initial value problem of the form

$$\begin{cases} \dot{u}(t) &= A(t)u(t) + f(t), \quad \forall\, t \in [s,T], \\ u(s) &= x, \end{cases} \tag{2.22}$$

which is usually referred to as an *evolution equation*. Here. $\{A(t) : t \geq 0\}$ is a one-parameter family of densely defined operators in a Banach space X, and $0 \leq s \leq t \leq T$.

To study this type of evolution systems or equations (2.22) a *two-parameters semigroup* is used, namely $\{U(s,t) : 0 \leq s \leq t \leq T\}$ is a family of linear and bounded operators in X satisfies

(a) $U(s,s) = I$, $\quad U(s,r)U(r,t) = U(s,t)$, $\forall 0 \leq s \leq r \leq t \leq T$,

(b) $(s,t) \mapsto U(s,t)x$ is continuous for any $0 \leq s \leq t \leq T$ and any $x \in X$.

This involves the concept of *stable family of infinitesimal generators*, see Friedman [89, Part 2, pp. 91–197], Pazy [196, Chapter 5, pp. 126–182], Tanabe [246].

This technique can be used to study also *quasi-linear* equations (e.g., a compact perturbation of a linear equation). Moreover, a natural extensions to the case of *monotone operators* and *variational inequalities* (see Bensoussan and Lions [16, 17], Brezis [34], Duvaut and Lions [68], Friedman [91]).

For instance, in Da Prato and Zabczyk [52, Proposition 5.5.6, pp. 75–79] an abstract result relative to the non-linear equation

$$\begin{cases} \dot{u}(t) &= Au(t) + F(t,u(t)), \quad \forall\, t \geq 0, \\ u(0) &= x, \end{cases} \tag{2.23}$$

is proved. If (1) the (linear unbounded) operator A generates a semigroup $\{S(t) : t \geq 0\}$ of class (C_0) on X, satisfying $\|S(t)\| \leq \mathrm{e}^{\omega t}$ for any $t \geq 0$ and some $\omega \geq 0$, and (2) F is continuous from $[0,T] \times X$ into X and $A + F(t,\cdot) - \eta I$ is dissipative for any t in $[0,T]$ and some $\eta \geq 0$, then there exist one and only one mild solution u for any given x in X. Moreover, it is actually a *strong solution* i.e., there is a sequence $\{u_n\}$ of functions in $C^1([0,T],X) \cap C([0,T],D_A)$ such that $u_n \to u$ and $\dot{u} - Au_n - F(\cdot,u_n) \to 0$ in $C([0,T],X)$. It is clear that if $F(t,x)$ is continuous in t and uniformly Lipschitz continuous in x then condition (2) is satisfied.

2.7 Applications to PDE

In general, the theory of (analytic) semigroups applies to time-independent *parabolic PDE* whilst an operator group can be constructed for some time-independent *hyperbolic PDE*. In all cases, the starting point is a deep knowledge of the unbounded operator to be used as infinitesimal generator, i.e., to apply the semigroup theory to partial differential equations the starting point is to establish several key estimates on *elliptic PDE*. On the other hand, starting from a semigroup built from the theory of parabolic PDE, all results concerning infinitesimal generator can be applied to elliptic PDE.

Besides the classic references like Ladyzhenskaya and Uraltseva [148] and Gilbarg and Trudinger [100] for a deep treatment of second order elliptic equations, the reader may consult (among others) the books Hellwig [106] and Landis [150] for a graduate student level approach and recently the book Evans [77].

2.7.1 Parabolic Operators

First consider second-order elliptic operator A in the Hilbert (Lebesgue) space $L^2(\Omega)$, on a smooth domain Ω of $\mathbb{R}^d$. Thus, we consider a bilinear form $a(u,v)$ defined by

$$a(u,v) = \int_\Omega \left[\sum_{i,j=1}^{d} (a_{ij}\partial_i u + a^j u)\partial_j v + \sum_{i=1}^{d} (a_i \partial_i u + a_0 u)v \right] \mathrm{d}x \ ,$$

for any u, v in the Hilbert space $H^1(\Omega)$ (functions whose first-order derivatives belong to L^2), where the coefficients satisfy

$$a_{ij},\ a^j,\ a_i,\ a_0 \ \in L^\infty(\Omega) \ \text{ and}$$
$$\exists\, \mu > 0 \ \text{ such that } \ \sum_{i,j=1}^{d} a_{ij}\xi_i\xi_j \geq \mu|\xi|^2 \ \text{ a.e. in } \Omega,\ \forall \xi \in R^d \ .$$

Thus $a(\cdot,\cdot)$ is a continuous and coercive bilinear form on $H^1(\Omega)$, i.e., there exist constants M_0, μ_0, λ_0 depending only on the L^∞ bounds of the coefficients and on the ellipticity constant μ, such that

$$|a(u,v)| \leq M_0 \|u\| \, \|v\| \,, \qquad \forall\, u, v \in H^1(\Omega) \ ,$$
$$a(u,u) + \lambda_0 |u|^2 \geq \mu_0 \|u\|^2 \,, \qquad \forall\, u \in H^1(\Omega) \ ,$$

where $\|\cdot\|$ and $|\cdot|$ denotes the (Hilbert) norm in $H^1(\Omega)$ and $L^2(\Omega)$, respectively. For a fixed u in $H^1(\Omega)$, we may consider the mapping $v \mapsto a(u,v)$ as a continuous linear functional, denoted be $a(u,\cdot)$, on $H^1(\Omega)$, or in $H_0^1(\Omega)$, (closure of functions with support on Ω) or in general in $H_0^1(\Omega \cup \Gamma)$ (closure of functions with support on $\Omega \cup \Gamma$, with Γ a part of the boundary $\partial\Omega$).

If the functions a_{ij} and a^j are differentiable and $(\cdot,\cdot)$, $(\cdot,\cdot)_{\partial\Omega}$ denote the inner products in $L^2(\Omega)$, $L^2(\partial\Omega)$, respectively, then an integration by parts plus a density argument show that

$$a(u,v) = (Lu, v) + (Bu, v)_{\partial\Omega} \,, \qquad \forall\, u \in H^2(\Omega) \,, \quad v \in H^1(\Omega),$$

where $H^2(\Omega)$ is the Hilbert space of functions whose derivatives up to the 2 belong to L^2, and L is the elliptic second order differential operator in divergence form given by

$$Lu = -\sum_{i,j=1}^{d} \partial_j(a_{ij}\partial_i u + a^j u) + \sum_{i=1}^{d} a_i \partial_i u + a_0 u \ ,$$

and B is the boundary first order differential operator given by

$$Bu = \sum_{i=1}^{d} b_i \partial_i u + b_0 u \ ,$$

where

$$b_i = \sum_{j=1}^{d} a_{ij} n_j \ , \qquad b_0 = \sum_{j=1}^{d} a^j n_j \ ,$$

and $n_j(x),\ j = 1, \dots, d$ are the components of the exterior unit normal vector to Ω at the point x in $\partial\Omega$. Notice that some degree of smoothness of the boundaries $\partial\Omega$ and Γ are necessary to establish the above Green formula, for instance piecewise C^1 boundary.

It is worthwhile to notice that the definition of the principal part of L uses only the *symmetric part* of the matrix (a_{ij}), i.e., we could define L with $(a_{ij} + a_{ji})/2$ instead of (a_{ij}). However, even the non-symmetric part of matrix (a_{ij}) is involved into the definition of the boundary operator B or the whole bilinear form $a(\cdot,\cdot)$. Actually, B is called the co-normal boundary operator associated with L if in the above (formal) expression the matrix (a_{ij}) is symmetric.

Thus, as a distribution in Ω, we may identify $a(u,\cdot)$ with the expression Lu, i.e., $a(u,\varphi) = \langle Lu, \varphi\rangle$ for any test function φ in $\mathcal{D}(\Omega)$. Then, as a distribution in the whole space $\mathbb{R}^d$ (a priori with support in $\overline{\Omega}$), we may identify $a(u,\cdot) - \langle Lu, \cdot\rangle$ with Bu, with support in $\partial\Omega$, i.e., $a(u,\varphi) - \langle Lu, \varphi\rangle_\Omega = \langle Bu, \varphi\rangle_{\partial\Omega}$ for any test function φ in $\mathcal{D}(\mathbb{R}^d)$. Thus, the mappings L and B are linear and bounded operators from $H^1(\Omega)$ into its dual space.

Considering the bilinear form $a(\cdot,\cdot)$ on the space $V := H_0^1(\Omega \cup \Gamma)$, we are on the assumption of Theorem 2.19, and therefore $Au := a(u,\cdot)$ is a variational infinitesimal generator. Again if the functions a_{ij} and a^j are differentiable then $A = -L$, otherwise the expression is only formal. The domain $\mathcal{D}(A)$ of A is the closure in the $H^2(\Omega)$ norm of smooth functions v satisfying the boundary conditions $v = 0$ on $\partial\Omega \setminus \Gamma$ and $Bv = 0$ on Γ. When Γ is a proper part of the boundary $\partial\Omega$ and Γ is connected with $\partial\Omega \setminus \Gamma$ several technical difficulties appear. We refer to Baiocchi and Capelo [8], Chen and Wu [40], Evans [77], Edmunds and Evans [71], Kinderlehrer and Stampacchia [131], Lieberstein [154], Necas [186], Oleinik and Radkevic[192], Schechter [221], Troianiello [250], Wloka [254], among others, for a comprehensive treatment, while specific oblique boundary conditions can be found, e.g. in Popivanov and Palagachev [201], Yanushaskas [257].

The reader may compare the semigroup theory with a more direct approach used to study second order parabolic equations, e.g. see Eidelman [72], Ladyzhenskaya et al. [147] and Lieberman [153], among others.

This Hilbert space theory generalize to bilinear forms of any even order. Essentially, based on the so-called Gårding's inequality, the same arguments can be followed, e.g., Friedman [89].

The L^p theory is more involved, several steps are necessary. The starting point is a second order (uniformly) elliptic differential operator L of the form

$$\begin{cases} L(x,\partial_x)\varphi(x) = -\sum_{i,j=1}^{d} a_{ij}(x)\partial_{ij}\varphi(x) + \\ \qquad\qquad + \sum_{i=1}^{d} a_i(x)\partial_i\varphi(x) + a_0(x)\varphi(x)\,, \end{cases} \tag{2.24}$$

and a boundary first order operator B of the form

$$B(x,\partial_x)\varphi(x) = -\sum_{i=1}^{d} b_i(x)\partial_i\varphi(x) + b_0(x)\varphi(x)\,, \tag{2.25}$$

where Ω is a domain with C^2 boundary and the coefficients satisfy

$$\begin{cases} \sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \mu|\xi|^2, \quad \forall\,\xi\in\mathbb{R}^d\,, \quad x\in\Omega\,, \\ a_{ij}\in C^0(\Omega), \qquad a_i,\, a_0\in L^\infty(\Omega), \\ \sum_{i,j=1}^{d} a_{ij}n_in_j = \sum_{i=1}^{d} b_in_i \ \text{ in } \ \partial\Omega, \qquad b_i,\, b_0\in C^1(\Omega), \end{cases} \tag{2.26}$$

When $\Omega = \mathbb{R}^d$ the second-order coefficients are uniformly continuous (and bounded) and certainly, for Dirichlet boundary conditions, the boundary operator B is not involved.

Consider L as an unbounded operator in $L^p(\Omega)$, with either Dirichlet boundary conditions or oblique B-boundary conditions. First, a priori elliptic (Agmon-Douglis-Nirenberg) estimates are obtained for (uniformly) elliptic differential operator of the following type: for any $1 < p < \infty$ there is a positive constant $C = C_p$ depending only on p, μ, the bounds of the coefficients a_{ij}, a_i, a_0, the modulus of continuity of a_{ij} and the domain Ω such that

$$\begin{cases} \|u\|_{2,p} \leq C\Big[\|Lu\|_{0,p} + \|u\|_p\Big], \quad \forall u\in W^{2,p}(\Omega), \\ \text{satisfying either} \quad u = 0 \text{ on } \partial\Omega \quad \text{or} \quad Bu = 0 \text{ on } \partial\Omega, \end{cases} \tag{2.27}$$

where $\|\cdot\|_p$ is the norm in $L^p(\Omega)$, $W^{2,p}(\Omega)$ is the Banach (Sobolev) space of functions whose derivatives up to the 2 belong to L^p, with the natural norm $\|\cdot\|_{2,p}$. When $\Omega = \mathbb{R}^d$, the same a priori bounds hold for u in $W^{2,p}(\mathbb{R}^d)$.

Next, based on the above a priori estimate applied to the (complex) elliptic operator $L(x,\partial_x)+\mathrm{e}^{\mathrm{i}\theta}\partial_t^2$ in $n+1$ variables (x,t), with θ in $(-\pi/2,\pi/2)$, we can deduce that for some constants $C_p>0$ and $\omega_p>0$ the following estimate holds

$$\|\partial_x^2 u\|_p + |\lambda|^{1/2}\|\partial_x u\|_p + |\lambda|\|u\|_p \le C_p\|\lambda u + Lu\|_p, \tag{2.28}$$

for every u belonging to $W^{2,p}(\Omega)$ satisfying either $u=0$ on $\partial\Omega$ or $Bu=0$ on $\partial\Omega$, and for any λ with $\Re(\lambda)\ge\omega_p$. Hence, the operator $A:=-L$ with domain $\mathcal{D}(A)$ defined as the Sobolev space $W^{2,p}(\Omega)$ with one of the boundary conditions either $u=0$ on $\partial\Omega$ or $Bu=0$ on $\partial\Omega$, generates an analytic semigroup in $L^p(\Omega)$.

Once the a priori estimates have been obtained, the above argument applies to Hölder space $C^\alpha(\overline{\Omega})$, $0<\alpha<1$ and to some extend to $C^1(\overline{\Omega})$, $C^0(\overline{\Omega})$, $L^1(\Omega)$ and $L^\infty(\Omega)$, e.g., Lunardi [160, Chapter 3, pp. 69–119].

If the operator L is given in divergence form, as in the beginning of this subsection, then it is possible to establish the fact that A is an infinitesimal generator of an analytic semigroup in $L^p(\Omega)$ directly, without using the deep result about a priori estimates, see Pazy [196, Theorem 3.6, pp. 215–219].

2.7.2 Coercive Forms

Let D be domain (the closure of D is the closure of its interior) in the Euclidean d-dimensional space $\mathbb{R}^d$ and m be a nonnegative Radon measure on D, i.e., finite on compact subsets K of D. Let $\mathcal{D}$ be a subspace of $C^\infty(\bar{D})$ (restriction of function in $C_0^\infty(\mathbb{R}^d)$ to $\bar{D}$) which contains $C_0^\infty(D)$. Now, suppose that the bilinear form

$$((u,v)) := \sum_{i=1}^d \int_D (\partial_i u)(\partial_i v)\mathrm{d}m + \int_D uv\mathrm{d}m, \quad \forall u,v\in\mathcal{D}, \tag{2.29}$$

is closable in the Hilbert space $H=L^2(D,m)$, i.e., the differentiation ∂_i (initially defined on $\mathcal{D}$) is a closable operator. For instance, if $m(\mathrm{d}x)=\rho(x)\mathrm{d}x$, with $\rho\ge0$ locally integrable in D then we need the following condition: there exists a set N of zero Lebesgue measure such that for every x in $D\setminus N$ satisfying $\rho(x)>0$ there exists $\varepsilon>0$ such that $\operatorname{ess\,sup}\{\rho(y):|y-x|<\varepsilon\}>0$. In particular, this holds if ρ is lower semicontinuous.

The norm and the inner product in $L^2(D,m)$ are denoted by $|\cdot|$ and $(\cdot,\cdot)$, respectively. Since $((\cdot,\cdot))$ is closable, the closure $\bar{\mathcal{D}}$ of $\mathcal{D}$ under the norm $\|\cdot\|$, induced by the inner product $((\cdot,\cdot))$, is a Hilbert space, e.g., if $\mathcal{D}=C_0^\infty(D)$ or $\mathcal{D}=C^\infty(\bar{D})$ then we get either $H_0^1(D,m)$ or $H^1(D,m)$.

Let $\{a_{ij}:i,j=1,\dots,d\}$ be a (uniformly) positive definite matrix of m-integrable functions in D which is sectorial (not necessarily symmetric), i.e., for every x in $D\setminus N$ with $m(N)=0$, the following inequalities hold

$$\sum_{i,j=1}^d \xi_i\xi_j\, a_{ij}(x) \ge c\sum_{i=1}^d \xi_i^2, \quad \forall\xi\in\mathbb{R}^d, \tag{2.30}$$

and

$$\Big|\sum_{i,j=1}^{d}\xi_i\eta_j\,a_{ij}(x)\Big|\le C\Big(\sum_{i,j=1}^{d}\xi_i\xi_j\,a_{ij}(x)\Big)^{1/2}\Big(\sum_{i,j=1}^{d}\eta_i\eta_j\,a_{ij}(x)\Big)^{1/2},\quad(2.31)$$

for every ξ and η in $\mathbb{R}^d$, and for some constants $C, c > 0$. Clearly, if the non-symmetric part of (a_{ij}) is bounded, i.e., there exists $C > 0$ such that $|a_{ij}(x) - a_{ji}(x)| \le C$, for m-almost every x, then we find that the sector condition (2.31) is satisfied, after using the coerciveness assumption (2.30).

The following bilinear form

$$a(u,v) := \sum_{i,j=1}^{d}\int_D a_{ij}(\partial_i u)(\partial_j v)\mathrm{d}m,\quad \forall u,v\in\mathcal{D}$$

is coercive in $L^2(D,m)$ and satisfies a sectorial condition, i.e.,

$$a(u,u)+c|u|^2\ge c\|u\|^2,\quad \forall u\in\mathcal{D},$$

and

$$|a(u,v)|\le C\sqrt{a(u,u)}\,\sqrt{a(v,v)},\quad \forall u,v\in\mathcal{D}.$$

where the constants C and c are the same as in (2.30) and (2.31).

To check that $a(\cdot,\cdot)$ is closable, let $\{u_n : n \ge 1\}$ be a sequence in $\mathcal{D}$ such that $u_n \to 0$ in $L^2(D,m)$ and $a(u_n - u_k, u_n - u_k) \to 0$ as $n,k\to\infty$. Because $a(\cdot,\cdot)$ is coercive, we have $\|u_n\|\to 0$ which yields $\partial_i u_k \to 0$ in m-measure. Hence, by means of Fatou lemma and the coerciveness assumption (2.30), we have

$$a(u_n,u_n)\le \liminf_k a(u_n-u_k,u_n-u_k),$$

which can be made as small as desired, provided n is sufficiently large. This shows that $a(u_n,u_n)\to 0$, i.e., $(a(\cdot,\cdot),\mathcal{D})$ is a positive closable bilinear form on $L^2(D,m)$. We can consider its closure defined on the Hilbert space V, which is the closure of $\mathcal{D}$ under the norm $\|\cdot\|_V$, induced by the symmetric part of the form $a(\cdot,\cdot)+(\cdot,\cdot)$. Note that $V\subset\bar{\mathcal{D}}$, and that $a(\cdot,\cdot)$ is not necessarily continuous on $H^1(D,m)$ or in $\bar{\mathcal{D}}$, but it satisfies a sectorial condition in V. It is clear that the above technique remains valid if $D\subset\mathbb{R}^{d_1}$, $d_1\ge d$, and the summation in conditions (2.29), (2.30) and (2.31) are kept only up to d, so that $a(\cdot,\cdot)$ may be *degenerate* in (d_1-d) variables.

Therefore, we have a positive closed (coercive) and continuous bilinear form $a(\cdot,\cdot)$ on a Hilbert space $V\subset H = L^2(D,m)$. Then, the unique operator A satisfying $a(u,v) = (-Au,v)$, for every u in $\mathcal{D}(A)$ and v in V, is densely defined. Moreover, A generates a strongly continuous (analytic) semigroup of contractions on $L^2(D,m)$. Actually, as mentioned later, $\{\mathrm{e}^{tA} : t\ge 0\}$ is also a Markov semigroup.

On the other hand, let b_0, b_i and b^j, for $i,j = 1,\dots,d$, be functions in $L^1(D,m)$. Consider the (first order) bilinear form

$$b(u,v) := \sum_{i=1}^{d}\int_D(\partial_i u)b_i v\mathrm{d}m+\sum_{j=1}^{d}\int_D ub^j(\partial_j v)\mathrm{d}m+\int_D ub_0 v\mathrm{d}m,$$

in $L^2(D,m)$. It is clear that $b(\cdot,\cdot)$ by itself cannot satisfy a sectorial condition in general. However, if all coefficients b_0, b_i and b^j belongs to $L^\infty(D,m)$, i.e., they are bounded, then there exists a constant $C>0$ such that

$$|b(u,v)| \le C\big[\|u\|\,|v| + |u|\,\|v\| + |u|\,|v|\big], \quad \forall u,v\in\mathcal{D},$$

proving that $a(\cdot,\cdot)+b(\cdot,\cdot)$ is coercive (although not necessarily positive) and satisfies a sectorial condition. Therefore, $a(\cdot,\cdot)+b(\cdot,\cdot)$ generates a strongly continuous (analytic) semigroup of contractions on $L^2(D,m)$. Certainly, this includes the case where D is open, $m(\mathrm{d}x)=\mathrm{d}x$ is the Lebesgue measure. Moreover, if $D=\bar D$ is closed then we may consider measures (or densities) on the boundary ∂D of D to study non-homogeneous boundary conditions.

On the other hand, assuming that

$$\mu(\varphi) := \int_D \big[2b_0\varphi + (b_i+b^i)\partial_i\varphi\big]\mathrm{d}m \ge 0, \quad \forall\varphi\in\mathcal{D},\ \varphi\ge 0,$$

i.e., $2b_0 - \sum_i \partial_i(b_i+b^i) \ge 0$ if m is the Lebesgue measure and $\mathcal{D}=C_0^\infty(D)$, then the identity

$$2b(u,u) = \int_D \big[(b_i+b^i)\partial_i u^2 + 2b_0u^2\big]\mathrm{d}m = 2\int_D u^2\mathrm{d}\mu$$

show that the (first order) bilinear form $b(\cdot,\cdot)$, initially defined on $\mathcal{D}$, is a positive. Moreover, even when the coefficients b_i and b^j are not necessarily bounded (and b_0 is only bounded from below), if the Radon measure μ has a density with respect to m then $a(\cdot,\cdot)+b(\cdot,\cdot)$ is closable in $L^2(D,m)$, by means of an arguments similar to the one used with $a(\cdot,\cdot)$. Furthermore, if the initial Radon measure m is the Lebesgue measure then the assumption that the (first order) coefficients b_i and b^j belong to $L^d(D)$, $D\subset\mathbb{R}^d$ and b_0 is in $L^{d/2}(D)$ suffices to prove that the sum $a(\cdot,\cdot)+b(\cdot,\cdot)$ is a closable and sectorial form. Alternatively, if it is known a priori that $a(\cdot,\cdot)+b(\cdot,\cdot)$ is positive and sectorial then we could use Remark 2.33 to show that it is also closable.

For instance, the reader is refer to the books Fukushima et al. [92], Jacob [115, Vol I], Ma and Röckner [161] for Dirichlet forms and to Baiocchi and Capelo [8], Bensoussan and Lions [16], Brezis [34], Friedman [91], Kinderlehrer and Stampacchia [131], Lions and Magenes [155], for variational inequalities techniques.

2.7.3 Hyperbolic Operators

The starting point is an (uniformly) elliptic second order differential operator $L(x,\partial_x)$, either in divergence form or not in divergence form, in a domain Ω of $\mathbb{R}^d$ and first order boundary differential operator $B(x,\partial_x)$. In the previous setting, we have the Sobolev space $W^{1,p}(\Omega)$ or $H^1(\Omega)$ with Dirichlet or B-oblique boundary conditions, which is now denoted by V. Also we have the Lebesgue space $L^p(\Omega)$ or $L^2(\Omega)$, which is now denoted by H. Thus, two Banach (or Hilbert) spaces $V\subset H$, with dense image. The Banach space V carries the

boundary conditions and the operator L is densely defined on H, actually with domain $W \subset V$, where W is either $W^{2,p}(\Omega) \cap V$ or $H^2(\Omega) \cap V$. A typical example is the wave equation where $L = -\Delta$, with Δ being the Laplacian operator. To treat the hyperbolic equation $\partial_t^2 u(x,t) + L(x,\partial_x)u(x,t) = 0$ plus initial and boundary conditions, we consider the matrix operator A defined by

$$A := \begin{pmatrix} 0 & 1 \\ -L & 0 \end{pmatrix}$$

so that the hyperbolic equation becomes $\partial_t u(t) = Au(t)$, where now $u(t)$ is the column vector with components $u_1(t) = u(x,t)$ and $u_2(t) = \partial_t u(x,t)$. By setting $u_1(t) = -\lambda w_1 - w_2$ and $u_1(t) = \lambda w_2 - L w_1$, we reduce the equation $\lambda u - Au = f$ to the one-dimensional equation $\lambda^2 w_i + L w_i = f_i$, which is solved as in the previous subsection for $\Re(\lambda^2) > \omega$, in particular for $\lambda > \sqrt{\omega}$. Hence, the matrix operator A is an infinitesimal generator of a (C_0) semigroup in the space $V \times H$. Actually, the above argument applies to the matrix operator $-A$, so that e^{tA} is a (C_0) group in $V \times H$. It is clear that depending on the smoothness assumptions on the coefficients of the differential operators $L(x,\partial_x)$ and $B(x,\partial_x)$, the above technique can be used in the Hölder space $C^{1,\alpha}(\overline{\Omega}) \times C^{\alpha}(\overline{\Omega})$.

Another typical case is the Schrödinger equation in the whole space $\mathbb{R}^d$ of the type $-\mathtt{i}\partial_t u(x,t) + L_0(x,\partial_x)u(x,t) + V(x)u(x,t) = 0$ where $L_0(x,\partial_x)$ is a symmetric (uniformly) elliptic second order operator (with real coefficients) (e.g., the Laplacian Δ) and V is a (real-valued) function called the potential. By setting $A := -\mathtt{i}L_0 - \mathtt{i}V$ with domain $\mathcal{D}(A) := H^2(\mathbb{R}^d)$, we get a skew-adjoint operator A, provided the potential function $V(x)$ is real and belongs to $L^p(\mathbb{R}^d)$, with $p > d/2$. Thus by means of Stone's Theorem, the operator A is the infinitesimal generator of a (C_0) group of unitary operator e^{tA}, we refer to Pazy [196, pp. 219–225] for details.

Also, symmetric hyperbolic systems can be studied. For instance, for d symmetric $n \times n$ matrices $a^k(x) = (a^k_{ij}(x))$, with x in $\mathbb{R}^d$, consider the (vector) operator

$$Au(x) = \sum_k a^k(x)\partial_k u(x), \quad \forall x \in \mathbb{R}^d,$$

where $u = (u_i)$ is a (vector-value) function in $L^2(\mathbb{R}^d;\mathbb{R}^n)$. Assuming that the coefficients of a^k are bounded and Lipschitz continuous, by means of an integration by parts we obtain

$$\begin{aligned} (u, Au) &= \sum_{ijk} \int_{\mathbb{R}^d} u_i(x) a^k_{ij}(x) \partial_k u_j(x) \mathrm{d}x = \\ &= -\frac{1}{2} \sum_{ijk} \int_{\mathbb{R}^d} u_i(x) \big[\partial_k a^k_{ij}(x)\big] u_j(x) \mathrm{d}x, \end{aligned}$$

proving that $A - cI$ is dissipative, for some constant c sufficiently large. Hence, Lumber-Phillips' theorem applies if we consider A as an unbounded operator on $L^2(\mathbb{R}^d;\mathbb{R}^n)$ with domain $\mathcal{D}(A)$, which contains the Sobolev space $H^1(\mathbb{R}^d;\mathbb{R}^n)$.

Some details are needed to completely identify the close extension of the operator A. Thus, $A-cI$ is an infinitesimal generator of a semigroup of contractions $S_c(t)$, so that A generates the semigroup $S(t) = \mathrm{e}^{ct}S_c(t)$. The interested reader may consult, for instance, the book Renardy and Rogers [211, Section 11.3, pp. 407–411] or in general the book Engel and Nagel [74] for a comprehensive treatment oriented to evolution equations.

2.8 Markov-Feller Semigroups

Let E be a locally compact Polish (i.e., complete separable metric) space and $\mathcal{B}(E)$ be the σ-algebra of Borel subsets of E. Usually, $E \subset \mathbb{R}^d$, is an smooth domain. Moreover, in this section we may even take E to be a locally compact Hausdorff space with countable basis. As before, $B(E)$ denotes the space of Borel measurable and bounded functions from E into $\mathbb{R}^d$, endowed with the supremum norm $\|\cdot\|$. Recall that a function f *vanishes at infinity* if for any $\varepsilon > 0$ there is a compact set K such that $|f(x)| < \varepsilon$ for any $x \in E \setminus K$. Also, $C_b(E)$, respectively $C_0(E)$, stands for the space of continuous functions which are bounded, respectively vanishing at infinity. It is clear that if the whole space $E = K$ is compact then $C_b(K) = C_0(K)$, in this case we use the notation $C(K)$. The Riesz representation theorem states that any bounded linear functional on $C(K)$ may be uniquely written in the form

$$\mu(f) := \int_K f(x)\,\mu(\mathrm{d}x), \quad \forall f \in C(K),$$

where μ is a regular bounded (signed) measure on $E = K$ (recall that regular means that for any measurable subset A of E and for every $\varepsilon > 0$ there exits an open set $\mathcal{O}$ and a closed set F, with $F \subset A \subset \mathcal{O}$ and $\mu(\mathcal{O} \setminus F) < \varepsilon$), see Dunford and Schwartz [63, p. 265, Theorem IV.6.3]. Moreover, if E is a locally compact Polish space, we may construct its one-point compactification $\bar{E} = E \cup \{\infty\}$ where open sets in $\bar{E}$ are of the open sets in E and the sets of the form $\{\infty\} \cup (E \setminus K))$ for any compact subsets K of E. Any regular bounded measure μ in $\bar{E}$ satisfying $\mu(\{\infty\} = 0)$, have the property that for any $\varepsilon > 0$ there exits a compact subset K of E such that $\mu(E \setminus K) < \varepsilon$, which usually refer to as μ being *tight* in E. The Banach space $C_0(E)$ of continuous functions on E vanishing at infinity can be identified with the Banach space of continuous function on $\bar{E}$ satisfying the zero boundary condition at infinity $f(\infty) = 0$. Thus, any bounded linear functional on $C_0(E)$ can be uniquely represented by a regular bounded (signed) measure (tight) in E, e.g., see Folland [85, Chapter 7] or Malliavin [162, Chapter II]. Therefore, we assume that the base space E is such that bounded linear nonnegative on $B(E)$ are uniquely given by (regular) bounded measures on E.

Definition 2.39 (Markov)**.** Let E be a locally compact Hausdorff space with countable basis and denote by $\mathcal{B}(E)$ its Borel σ-algebra.

(1) A one-parameter family $\{S(t) : t \geq 0\}$ of bounded linear operators from the

Banach space $B(E)$ into itself is called a *(sub-)Markov semigroup* with *(sub-)Markovian kernels* $\{P(t,x,A) : t \geq 0,\ x \in E,\ A \in \mathcal{B}(E)\}$ given by

$$S(t)f(x) = \int_E f(y)\, P(t,x,\mathrm{d}y), \quad \forall f \in B(E),$$

if it satisfies

(a) $S(t+s) = S(t)S(s), \quad \forall t,s \geq 0,$

(b) $S(t)f(x) \geq 0, \quad \forall t \geq 0,\ x \in E \quad$ if $\quad f(x) \geq 0,\ \forall x \in E,$

(c) $S(t)\mathbb{1}_E(x) \leq 1, \quad \forall t \geq 0,\ x \in E.$

or equivalently

(a)$'$ for each $s,t \geq 0$, x in E and A in $\mathcal{B}(E)$ we have

$$P(s+t,x,A) = \int_E P(s,x,\mathrm{d}y)\, P(t,y,A),$$

which is referred to as the Chapman-Kolmogorov identity.

(b)$'$ for each t and x the function $A \mapsto P(t,x,A)$ is a (non-negative) measure on $\mathcal{B}(E)$ with $P(t,x,E) \leq 1$ and $P(0,x,\{x\}) = 1$,

(c)$'$ for each t and A in $\mathcal{B}(E)$ the function $x \mapsto P(t,x,A)$ is a Borel measurable,

(2) It is called a *transition function* if for every A in $\mathcal{B}(E)$ the mapping $(t,x) \mapsto P(t,x,A)$ is jointly Borel measurable in $[0,\infty) \times E$.

(3) It is called *stochastically continuous* if

$$\lim_{t \to 0} P(t,x,\mathcal{O}) = 1,$$

for every x in E and any open neighborhood $\mathcal{O}$ of x.

(4) It satisfies the (pointwise) *Feller property* (respectively, *strong Feller property*) if for every $t > 0$ the function $x \mapsto S(t)f(x)$ is continuous at each point of continuity of the function f (respectively, at each point x). □

Here, we have denoted by $\mathbb{1}_A$ the characteristic function of the subset A, i.e., $\mathbb{1}_A(x) = 1$ if x belongs to A and zero otherwise. It is clear that condition (a) is the usual semigroup property, condition (b) is the *weak maximum principle* and inequality (c) is a *normalization* condition. Actually, condition (c) can be replaced by the equality $S(t)\mathbb{1}_E(x) = 1$, for any $t \geq 0$, $x \in E$, without any lost of generality, by using the one-point compactification. This give rise to the distinction between Markov and sub-Markov semigroups.

If the base space E is not locally compact, then we normally add the condition

(d) $A \mapsto S(t)\mathbb{1}_A(x) \quad$ is σ-additive on $\quad \mathcal{B}(E), \quad \forall t \geq 0,\ x \in E$

to the definition of a Markov semigroup. This condition is automatically satisfied if E is a locally compact Polish space. We refer to a Markov semigroup $S(t)$ or to a Markov kernels $P(t,x,\cdot)$ indistinctly.

In general, a Markov semigroup is not strongly continuous in $B(E)$, even if it satisfies the above Feller property. Moreover, a joint measurability condition is needed to define the resolvent operators

$$R_\lambda f(x) := \int_0^\infty \mathrm{e}^{-\lambda t} S(t) f(x) \mathrm{d}t, \quad \forall x \in E,\ \lambda > 0,$$

as a mapping from $B(E)$ into itself. This is precisely the condition (2) in Definition 2.39, i.e., a Markov transition function.

If E is a Polish space, the Markov semigroup $\{S(t) : t \geq 0\}$ or its Markov kernels $P(t, x, \cdot)$ is stochastically continuous if and only if

$$\lim_{t \to 0} P(t, x, B(x, \delta)) = 1, \quad \forall x \in E,\ \delta > 0,$$

where $B(x, \delta)$ is the ball of center x and radius δ in E. Notice that the above Feller property refers to the *space* variable x, whilst stochastically continuous involves the *time* variable t. Even if the base space E is not locally compact, it is proved in Da Prato and Zabczyk [52, p. 13] that a Markov semigroup $\{S(t) : t \geq 0\}$ is stochastically continuous if and only if $S(t)f(x)$ converges to $f(x)$ as $t \to 0$, for any $x \in E$ and any function f which is either (a) bounded and continuous or (b) bounded and uniformly continuous or (c) bounded and Lipschitz continuous.

It is clear that a stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$ is (Borel) measurable, i.e, $(t, x) \mapsto S(t)f(x)$ is jointly Borel measurable in $[0, \infty) \times E$, i.e, a Markov transition function. Thus we can use the general results in Dellacherie and Meyer [58, Section XIV.24, pp. 169–172]) to construct a cad-lag realization of the associated Markov (strong Markov, since it is stochastically continuous) Markov process as described in Chapter 1. Note that a systematic study on analytic methods for Markov diffusion semigroup can be found in Bertoldi and Lorenzi [22].

2.8.1 Feller Semigroups

A good way is to consider the semigroup restricted to the space $C_b(E)$ or $C_0(E)$ and impose the Feller property.

Definition 2.40 (Feller)**.** Let E be a Polish space. Then a one-parameter family $\{S(t) : t \geq 0\}$ of bounded linear operators from a closed subspace C of the Banach space $C_b(E)$ (e.g., $C_0(E)$ or the whole space $C_b(E)$) into itself is called a *Feller semigroup* if it satisfies

(a) $S(t + s) = S(t)S(s), \quad \forall t, s \geq 0,$

(b) $0 \leq S(t)f \leq 1, \quad \forall t \geq 0 \quad \text{if} \quad 0 \leq f \leq 1,$

(c) $\lim_{t \downarrow 0} \|S(t)f - f\| = 0, \quad \forall f \in C.$

Actually, a Feller semigroup need not to be *strongly continuous* as indicated by condition (d), usually only a *weakly continuous* condition, namely

$$\text{(e)} \quad \lim_{t \downarrow 0} S(t)f(x) = f(x), \quad \forall x \in E, \quad \forall f \in C$$

is required. However, since our base space E is locally compact (see Dellacherie and Meyer [58, Theorem XIII.19, pp. 98–99]) these two conditions are equivalent. □

Roughly speaking, a *Markov semigroup* is semigroup associated with some Markov process and a *Feller semigroup* (or Markov-Feller semigroup or Feller-Dynkin semigroup) is a Markov semigroup which satisfies the *Feller property*. Actually, most of the key results on Markov theory requires a stochastically continuous Markov semigroup.

The measure theory ensures that any Feller semigroup in either $C_b(E)$ or $C_0(E)$ can be extended to be a Markov semigroup in $B(E)$. It is clear that for a Feller semigroup, the condition (e) above on *weakly continuous* is the equivalent of *stochastically continuous* for a Markov semigroup. Clearly, in a locally compact Polish space, a stochastically continuous Markov semigroup is indeed a Feller semigroup as mentioned in the above definition.

The following result (on locally compact Polish space E) is taken from Taira [243, Chapter 9, Section 2, pp. 333–340],

Theorem 2.41 (Markov-Feller)**.** *Let $\{S(t) : t \geq 0\}$ be a Markov semigroup in $B(E)$, which leaves invariant the subspace $C_0(E)$ (i.e., $S(t)f \in C_0(E)$, $\forall t > 0$, $f \in C_0(E)$) and is uniformly stochastically continuous (i.e., the continuity condition in t at 0 holds uniformly on compact set in x). Then the restriction of $S(t)$ to $C_0(E)$ is a Feller semigroup if and only if the following property is satisfied*

(L) *For any $T, \varepsilon > 0$ and any compact $C \subset E$ there exists another compact set $K = K(T, \varepsilon, C)$ of E such that $P(t, x, C) < \varepsilon$, for any $t \in [0, T]$ and any $x \in E \setminus K$,*

where $P(t, x, \cdot)$ is the Markov transition function associated with $\{S(t) : t \geq 0\}$. □

Since a Feller semigroup $S(t)$ on $C_0(E)$ is strongly continuous, the general (contraction) semigroup theory applies to characterize the infinitesimal generator A of $S(t)$. The extra property involved in term of the resolvent operator $R(\lambda, A) = (\lambda I - A)^{-1}$ can be formulated as

$$f \in C_0(E), \ f \geq 0 \Longrightarrow R(\lambda, A)f \geq 0, \tag{2.32}$$

which is referred to as the *weak Maximum Principle*. In connection with this, we mention the following result, see Taira [243, Chapter 9, Section 3, pp. 340–349],

Theorem 2.42. *Let E be a compact Polish space and A be a linear operator (not necessarily bounded) in $C(E) = C_b(E) = C_0(E)$ densely defined on the*

domain $\mathcal{D}(A)$. Assume that for some $\lambda \geq 0$ the range $\mathcal{R}(\lambda I - A)$ of $\lambda I - A$ is dense in $C(E)$. If the weak Maximum Principle is satisfied in the following sense

(wMP) *If u belongs to $\mathcal{D}(A)$ and it takes a positive maximum at the point x_0 in E then $Au(x_0) \leq 0$,*

then the closure $\bar{A}$ of the operator A is the infinitesimal generator of a Feller semigroup in the space $C(E)$.

Since a Feller semigroup $\{S(t) : t \geq 0\}$ has the property that $(t, x) \mapsto S(t)f(x)$ is Borel measurable, the Laplace transform (2.10) defines the resolvent operators $\{R(\lambda) : \lambda > 0\}$ from $B(E)$ into itself. The specific properties of a Markov semigroup $\{S(t) : t \geq 0\}$ (on $B(E)$ Borel bounded functions on E) as in Definitions 2.39 or 2.40 become

$$\begin{cases} \text{(a)} \;\; R(\lambda)f \geq 0, \quad \forall f \geq 0, \quad \lambda > 0, \\ \text{(b)} \;\; \lambda R(\lambda)\mathbb{1}_E(x) \leq 1, \quad \forall x \in E. \end{cases} \tag{2.33}$$

If the initial data is a contracting resolvent (see (a) and (b) of Definition 2.21) then by reducing the space $B(E)$ to the regularity space R, we may construct its associated (strongly continuous) semigroup on R.

For instance, the reader may consult the books Jacob [115, Vol I, Chapter 4] and Taira [243, Chapters 9 and 10] for a detailed presentation on Feller semigroups, among other topics.

2.8.2 Markov Process Realization

Until know, we have mentioned several results concerning the construction of a Markov process starting from a transition probability function. To understand better the relation between Markov processes and Markov semigroups, we discuss the main steps and difficulties to construct a realization of a given Markov semigroup $S(t)$ with transition function $P(t, x, \mathrm{d}y)$ on a compact base space $\bar{E}$, the one-point compactification of E, assumed locally compact and $P(t, x, \{\infty\}) := 1 - P(t, x, E)$, so that $P(t, x, \bar{E}) = 1$. Intuitively, given any initial position x in $\bar{E}$ at time 0, the probability distribution for a location of the stochastic process X at a future time $t > 0$ is $P(t, x, \mathrm{d}y)$. Thus to construct a separable version (see Definition 1.1) of the stochastic process X, we need to construct a measure on the space of *paths* $\bar{E}^I$, where I is a countable dense set in $(0, \infty)$, for instance I is the set of positive rational numbers. Since $\bar{E}$ is compact and I countable, the product topology in $\bar{E}^I$ produces a compact metrizable space. In view of Riesz representation theorem, to construct a probability measure on $\bar{E}^I$, it suffices to construct a positive linear functional $\mathbb{E}^x$ from the space $C(\bar{E}^I)$ of real-valued continuous functions to $\mathbb{R}$. Naturally, first we define $\mathbb{E}^x$ on the subspace $C_c(\bar{E}^I)$ of cylindrical functions, consisting of continuous functions that depend on only finitely many factors of $\bar{E}^I$, i.e., functions on $C(\bar{E}^I)$ of the form $\psi(\omega) = F(\omega(t_1), \omega(t_2), \ldots, \omega(t_n))$, $t_1 < t_2 < \cdots < t_n$, for

some $n \geq 1$, where F is a continuous function in $\bar{E}^n$ and t_i belongs to I for $i = 1, 2, \ldots, n$. We set

$$\begin{cases} \mathbb{E}^x(\psi) := \displaystyle\int P(t_1, x_1 - x, \mathrm{d}x_1) \int P(t_2 - t_1, x_2 - x_1, \mathrm{d}x_2) \ldots \\ \quad \ldots \displaystyle\int F(x_1, x_2, \ldots, x_n)\, P(t_n - t_{n-1}, x_n - x_{n-1}, \mathrm{d}x_n), \end{cases} \tag{2.34}$$

which is well defined in view of the semigroup or Chapman-Kolmogorov identity, see Section 1.10). Hence, $\mathbb{E}^x : C_c(\bar{E}^I) \to \mathbb{R}$ is a positive linear functional satisfying $\mathbb{E}^x(\mathbb{1}) = 1$. By the Stone-Weierstrass theorem on polynomial approximations, the subspace $C_c(\bar{E}^I)$ is dense and then $\mathbb{E}^x$ has a unique continuous extension to $C(\bar{E}^I)$, with the same properties. Then, we have a unique probability measure P_x on $C(\bar{E}^I)$ such that

$$\mathbb{E}^x(\psi) = \int_{\bar{E}} \psi(\omega) P_x(\mathrm{d}\omega)$$

and (2.34) holds on $C_c(\bar{E}^I)$. To go further in this construction, we need some regularity in the variable t, i.e., the function

$$\rho(\varepsilon, \delta) := \sup_x \int_{|y-x| \geq \varepsilon} P(\delta, x, \mathrm{d}y)$$

as δ goes to zero for any fixed ε. If we want to get a probability measure P_x on the space $C([0, \infty), \bar{E})$ of continuous functions from $[0, \infty)$ into $\bar{E}$, then we need to show first that the set $C_{\mathrm{lu}}(I, \bar{E})$, of uniformly locally continuous functions (which are extended uniquely to continuous functions from $[0, \infty)$ into $\bar{E}$) is a Borel subset of $\bar{E}^I$, which contains the support of P_x. Actually, this is the *hard* point in the construction of the measure P_x. To this purpose, the set $C_{\mathrm{lu}}(I, \bar{E})$ can be expressed as a countable intersection of a countable union closed sets, namely

$$C_{\mathrm{lu}}(I, \bar{E}) = \bigcap_{k,\ell=1}^{\infty} \bigcup_{n=1}^{\infty} F'(k, 1/\ell, 1/n),$$

where $F'(k, \varepsilon, \delta)$ is the complement of the set

$$\begin{aligned} F(k, \varepsilon, \delta) &:= \bigcup \big\{ E(a, b, \varepsilon) \ : \ 0 \leq a < b \leq a + \delta \leq k \big\}, \\ E(a, b, \varepsilon) &:= \big\{ \omega \in \bar{E}^I \ : \ \exists\, t, s \in I \cap [a, b],\ |\omega(t) - \omega(s)| > 2\varepsilon \big\}. \end{aligned}$$

Since the complement of $E(a, b, \varepsilon)$ is closed in $\bar{E}^I$, the set $F(k, \varepsilon, \delta)$ is open and therefore $C_{\mathrm{lu}}(I, \bar{E})$ is a Borel set. Next, assuming

$$P_x\{E(a, b, \varepsilon)\} \leq 2\rho(\varepsilon/2, b - a)), \tag{2.35}$$

we have

$$P_x\{F(k, \varepsilon, \delta)\} \leq 2\frac{k}{\delta}\rho(\varepsilon, \delta))$$

and if $\rho(\varepsilon,\delta)/\delta$ goes to zero as δ goes to zero, we can deduce the equality $P_x\{C_{\rm lu}(I,\bar{E})\} = 1$. To obtain the estimate (2.35), we may express the set $E(a,b,\varepsilon)$ as an increasing limit of sets of the form $A = \{\omega : \exists\, i,j$ such that $|\omega(t_i)-\omega(t_j)| > 2\varepsilon$, for some ε,δ,n and $t_1 < t_2 < \cdots < t_n$ in I, with $t_n - t_1 \leq \delta$. Each set A of the above form is contained in the union $B \cup (C_1 \cap D_1)\ldots(C_n \cap D_n)$, where $B := \{\omega : |\omega(t_n)-\omega(t_1)| > \varepsilon\}$, $C_i := \{\omega : |\omega(t_n)-\omega(t_i)| > \varepsilon\}$ and $D_i := \{\omega : |\omega(t_1)-\omega(t_i)| > 2\varepsilon$ and $|\omega(t_1)-\omega(t_j)| \leq 2\varepsilon,\ \forall j \leq i-1\}$. It is clear that $P_x(B) \leq \rho(\varepsilon,\delta)$ and $P_x(C_i) \leq \rho(\varepsilon,\delta)$. Since C_i is independent of D_i, we have $P_x(C_i \cap D_i) = P_x(C_i)\,P_x(D_i)$, which allow us to conclude, see Taylor [249, Vol 2, Chapter 11, pp. 303–307].

The point here is that under the assumption $\rho(\varepsilon,\delta)/\delta \to 0$ as $\delta \to 0$, we are able to construct the probability measure in the sample space of continuous paths $C([0,\infty),\bar{E})$, e.g. the Wiener measure. In some cases, e.g., the (compound) Poisson measure, we have only $[\rho(\varepsilon,\delta)]^2/\delta \to 0$ as $\delta \to 0$ so that the construction in the sample space $C([0,\infty),\bar{E})$ fails. Then, the sample space of right-continuous (with left-hand limits) paths $D([0,\infty),\bar{E})$ is used, see Section 1.12. In this case, the set $E(a,b,\varepsilon)$ is re-defined as

$$E(a,b,\varepsilon) := \big\{\omega \in \bar{E}^I : \exists\, t,s,r \in I, \text{ such that } |\omega(t)-\omega(s)| \wedge |\omega(s)-\omega(r)| > 2\varepsilon \quad a \leq r < s < t \leq b\big\}.$$

Using the fact that the two events $\{|\omega(t)-\omega(s)| > 2\varepsilon\}$ and $\{|\omega(s)-\omega(r)| > 2\varepsilon\}$ are actually of the previous form and independent of each other, we notice that estimate (2.35) is modified as follows

$$P_x\{E(a,b,\varepsilon)\} \leq 2[\rho(\varepsilon/2, b-a))]^2. \tag{2.36}$$

Naturally, instead of the subspace $C_{\rm lu}(I,\bar{E})$ we use the subset $D_{\rm lu}(I,\bar{E})$ of $\bar{E}^I$ composed by the restriction to I of functions in $D([0,\infty),\bar{E})$. Some more detailed analysis is needed to effectively give a realization of the Markov process in the sample space $D([0,\infty),\bar{E})$. The interested reader may take a look at Jacob [115, Vol III, Chapter 3] for a more complete overview on Feller semigroups and processes.

2.8.3 Pointwise Continuous Semigroups

First, notice that for a given strongly continuous semigroup in a Banach space, the weak infinitesimal generator denoted by $\bar{A}$ is not a genuine extension of the strong infinitesimal generator A, indeed their domain of definition are the same and they agree on it. Moreover, a weakly continuous semigroup in a Banach space is actually strongly continuous, see Pazy [196, Chapter 2, Section 2, pp. 42–44].

Thus, one way to proceed is to consider the weak-star topology in $B(E)$, i.e., boundedly pointwise convergence. Hence, the notion of *pointwise continuous* semigroup (also called weakly continuous) and *weak-star infinitesimal generator* (also called weak infinitesimal generator) are necessary, see Dynkin [70]. Given a stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$, we restrict our

attention to the subspace $B^0(E)$ of real bounded Borel functions f on E such that the map $t \mapsto S(t)f(x)$ is continuous for any x in E. It is clear that $B^0(E)$ contains $C_b(E)$ and it is invariant under $S(t)$ for any $t \geq 0$. Thus $\{S(t) : t \geq 0\}$ is a pointwise continuous semigroup on $B^0(E)$, i.e., besides (a), (b) and (c) of Definition 2.39 it also satisfies the condition $S(t)f(x) \to f(x)$ for any x in E and any f in $B^0(E)$. Then, the *weak-star* infinitesimal generator $\bar{A}$ can be (densely) defined on $B^0(E)$ be means of the boundedly pointwise convergence, i.e., $\bar{A}f = g$ if and only if $[S(t)f - f]/t$ converges (boundedly pointwise) to g, this means

$$\sup_{t>0} \frac{|S(t)f(x) - f(x)|}{t} \leq C, \quad \forall x \in E,$$

for some constant $C = C_f > 0$ and

$$\lim_{t\to 0} \frac{S(t)f(x) - f(x)}{t} = g(x), \quad \forall x \in E,$$

where necessarily $g = \bar{A}f$ belongs to $B(E)$.

This approach is more relevant when the base space E is not locally compact Polish space, i.e., E may be an infinite dimensional Hilbert space endowed with the weak or strong topology. For instance, as in [172], suppose that a (strong) homogeneous Markov process $y(t, x)$ is know (e.g., via stochastic differential equations) and then, a semigroup is define as follow

$$\Phi_\alpha(t)h(x) := E\{\mathrm{e}^{-\alpha t} h(y(t, x))\}, \tag{2.37}$$

for any $\alpha > 0$, on the space of $C_b(X)$ of real (uniformly) continuous and bounded functions, where X is an open subset (or the closure of an open subset) in a separable Banach space. Sometimes, we are required to consider the semigroup $\{\Phi_\alpha(t) : t \geq 0\}$ on a space with unbounded functions, e.g., $C_p^0(X)$ be the space of real uniformly continuous functions on any ball and with a growth bounded by the norm to the $p \geq 0$ power, in another words, the space of real functions h on X such that $x \mapsto h(x)(1 + |x|^2)^{-p/2}$ is bounded and locally uniformly continuous, with the weighted sup-norm

$$\|h\| = \|h\|_{C_p^0} := \sup_{x\in X}\{|h(x)|(\lambda + |x|^2)^{-p/2}\}, \tag{2.38}$$

with $\lambda > 0$. Suppose that the Markov process $y(t, x)$, defined on some probability space $(\Omega, \mathcal{F}, P)$, satisfies the conditions:

(1) $x \mapsto y(t, x)$ is locally uniformly continuous (in x), locally uniformly continuous for t in $[0, \infty)$, i.e., for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any x, $\bar{x}$ in X satisfying $|x - \bar{x}| < \delta$, $|x| \leq 1/\varepsilon$ and $|\bar{x}| \leq 1/\varepsilon$ we have

$$P\Big\{ \sup_{0\leq t\leq 1/\varepsilon} |y(t, x) - y(t, \bar{x})| \geq \varepsilon \Big\} < \varepsilon. \tag{2.39}$$

(2) $t \mapsto y(t, x)$ is locally uniformly continuous (in t), for any x in X, (actually in a dense subset suffices) i.e., for any x in X and for any $\varepsilon > 0$ there is a $\delta > 0$

such that

$$P\Big\{ \sup_{0\le t\le\delta} \sup_{0\le s\le 1/\varepsilon} |y(t+s,x)-y(s,x)| \ge \varepsilon \Big\} < \varepsilon. \tag{2.40}$$

(3) For any $p>0$ there are positive constants α_0 and λ sufficiently large such that the following estimate

$$E\{\sup_{t\ge 0} \mathrm{e}^{-\alpha_0 t}(\lambda+|y(t,x)|^2)^{p/2}\} \le C_p\,(\lambda+|x|^2)^{p/2}, \quad \forall t\ge 0,\ x\in\mathcal{O} \tag{2.41}$$

holds, with some $C_p\ge 1$ and $C_p=1$ if the sup is removed in the left-hand side. Here we are using the notation $|\cdot|$ for either the Euclidean norm or the norm in the Banach space containing X.

It is clear that (2.41) plays a role only when X is unbounded and that the closure of an open subset, say $\bar{X}$ could be used instead of X in all that follows. The associate semigroup $\Phi_\alpha(t)$ is not necessarily a *strongly continuous* semigroup on $C_b(X)$ nor on $C_p^0(X)$. Actually, we have in mind $X=\mathbb{R}^d$ (i.e., an stochastic ODE where the above conditions are easily verified and $\Phi_\alpha(t)$ is strongly continuous) but these conditions apply also for more general situations (stochastic PDE), such as the stochastic Navier-Stokes equation, e.g. Menaldi and Sritharan [175, 176].

It is clear that $C_b(X)\subset C_q^0(X)\subset C_p^0(X)$ for any $0\le q<p$. Then for any $\alpha\ge 0$, the (linear) semigroup $(\Phi_\alpha(t),\, t\ge 0)$ with an α-exponential factor is a weak-star continuous Markov semigroup in the space $C_p(X)$, i.e.,

$$\begin{cases} \Phi_\alpha(t+s)=\Phi_\alpha(t)\Phi_\alpha(s), \quad \forall\, s,t\ge 0,\\ \|\Phi_\alpha(t)h\|\le\|h\|, \quad \forall\, h\in C_{bp}(X),\\ \Phi_\alpha(t)h(x)\to h(x) \ \text{ as } \ t\to 0, \quad \forall\, h\in C_p^0(X),\\ \Phi_\alpha(t)h(x)\ge 0, \quad \forall\, h\ge 0, \quad h\in C_p^0(X). \end{cases} \tag{2.42}$$

This follows immediately from the conditions (2.39), (2.40) and (2.41) imposed on the Markov process $y(t,x)$.

Since the semigroup is not strongly continuous, we cannot consider the *strong* infinitesimal generator as acting on a dense domain in $C_p^0(X)$. However, this Markov semigroup $\{\Phi_\alpha(t) : t\ge 0\}$ may be considered as acting on real Borel functions with p-polynomial growth, which is a Banach space with the norm (2.38) and is denoted by $B_p(X)$. It is convenient to define the family of semi-norms on $B_p(X)$

$$p_0(h,x) := E\Big\{ \sup_{s\ge 0} |h(y(s,x))|\, \mathrm{e}^{-\alpha_0 s}\Big\}, \quad \forall x\in X, \tag{2.43}$$

where 2α, $2p$ and λ satisfy the estimate (2.41), and when $p=0$ we may take $\alpha_0=0$. If a sequence $\{h_n\}$ of equi-bounded functions in $B_p(X)$ satisfies $p_0(h_n-h,x)\to 0$ for any x in X, we say that $h_n\to h$ boundedly pointwise relative to the above family of semi-norms. In view of (2.40), it is clear that $p_0(\Phi_\alpha(t)h-h,x)\to 0$ as $t\to 0$, for any function h in $C_p^0(X)$ and any x in X.

Let us defined $B^0_p(X)$ be the subspace of functions $\bar h$ in $B_p(X)$ such that the mapping $t \mapsto \bar h[y(t,x)]$ is almost surely continuous on $[0,\infty)$ for any x in X and satisfies

$$\lim_{t\to 0} p_0(\Phi_\alpha(t)\bar h - \bar h, x) = 0, \quad \forall\, x \in X. \tag{2.44}$$

where $p_0(\cdot,\cdot)$ is the semi-norm given by (2.43). This is the space of functions (uniformly) continuous over the random field $y(\cdot,x)$, relative to the family of semi-norms (2.43), and it is independent of α, as long as (2.41) holds. Hence, we may consider the semigroup on the Banach space $B^0_p(X)$, endowed with the norm (2.38). The *weak-star* infinitesimal generator $\bar A_\alpha$ with domain $\mathcal{D}_p(\bar A_\alpha)$ (as a subspace of $B^0_p(X)$) is defined by the boundedly pointwise limit $[\Phi_\alpha(t)h - h]/t \to \bar A_\alpha h$ as $t \to 0$, relative to the family of semi-norms (2.43). Also, it is clear that $p_0(\Phi_\alpha(t)\bar h, x) \le p_0(\bar h, x)$ for any $t \ge 0$, $\bar h$ in $B^0_p(X)$ and x in X. We include the proof of the following results for the sake of completeness,

Proposition 2.43 (density)**.** *If assumptions* (2.39), (2.40) *and* (2.41) *hold, then* $C^0_p(X) \subset B^0_p(X)$, *the semigroup* $\{\Phi_\alpha(t) : t \ge 0\}$ *leaves invariant the space* $B^0_p(X)$, *and for any function* $\bar h$ *in* $B^0_p(X)$, *there is an equi-bounded sequence* $\{\bar h_n\}$ *of functions in* $\mathcal{D}_p(\bar A_\alpha)$ *satisfying* $p_0(\bar h_n - \bar h, x) \to 0$ *for any* x *in* X.

Proof. Indeed, since any function h in $C_p(X)$ is such that $x \mapsto h(x)\,(\lambda + |x|^2)^{-q/2}$, $q > p$, is uniformly continuous for x in X, we may use the estimate (2.41) to reduce the proof of the property (2.44) to the following condition

$$\lim_{t\to 0} P\Big\{ \sup_{0\le s\le T} |y(t+s,x) - y(s,x)| \Big\} = 0, \quad \forall\, x \in X,\ T > 0, \tag{2.45}$$

which follows from (2.40). This verifies the fact that $C^0_p(X) \subset \mathcal{B}^0_p(X)$.

Next, from the strong Markov property we deduce

$$\begin{aligned} p_0(\Phi_\alpha(t)\bar h, x) &= E\Big\{ \sup_{s\ge 0} E\{|\bar h[y(t+s,x)]|\mathrm{e}^{-\alpha_0(t+s)} \,|\, y(t,x)\}\mathrm{e}^{-(\alpha-\alpha_0)t}\Big\} \\ &\le E\Big\{ \sup_{s\ge 0} |\bar h[y(t+s,x)]|\mathrm{e}^{-\alpha_0(t+s)}\Big\} = p_0(\bar h, x), \end{aligned}$$

for any x in $\mathcal{O}$ and $t \ge 0$. Therefore,

$$\begin{aligned} p_0(\Phi_\alpha(r+t)\bar h - \Phi_\alpha(t)\bar h, x) &= p_0(\Phi_\alpha(t)[\Phi_\alpha(r)\bar h - \bar h], x) \le \\ &\le p_0(\Phi_\alpha(r)\bar h - \bar h, x), \end{aligned}$$

which proves that the space $B^0_p(\mathcal{O})$ is invariant under the semigroup.

Finally, to approximate a function $\bar h$ in $B^0_p(\mathcal{O})$ by regular functions, we can define the sequence $\{\bar h_n\ n = 1,2,\dots\}$ by

$$\bar h_n(x) := n\int_0^\infty \mathrm{e}^{-nt}\Phi_\alpha(t)\bar h(x)\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t} E\big\{\bar h(y(\frac{t}{n},x))\mathrm{e}^{-\alpha(\frac{t}{n})}\big\}\mathrm{d}t,$$

and apply the Markov property to get

$$\big|E\big\{\sup_{s\geq 0}[\bar{h}_n(y(s,x)) - \bar{h}(y(s,x))]\mathrm{e}^{-\alpha_0 s}\big\}\big| \leq$$
$$\leq \int_0^\infty \mathrm{e}^{-t}\big[E\big\{\sup_{s\geq 0}|\bar{h}(y(s+\frac{t}{n},x))\mathrm{e}^{-\alpha(\frac{t}{n})} - \bar{h}(y(s,x))|\mathrm{e}^{-\alpha_0 s}\big\}\big]\mathrm{d}t.$$

Thus, from the estimates (2.40) and (2.41) we deduce

$$\lim_{n\to\infty}\big|E\big\{\sup_{s\geq 0}[\bar{h}_n(y(s,x)) - \bar{h}(y(s,x))]\mathrm{e}^{-\alpha_0 s}\big\}\big| = 0,$$

for any fixed x in X. □

A clear consequence of the above results is that given $\alpha > 0$, $p \geq 0$, λ sufficiently large to ensure (2.41), and a function $\bar{h}$ in $B_p^0(\mathcal{O})$, there is another function $\bar{u}$ in $\mathcal{D}_p(\bar{A}_\alpha)$ such that $-\bar{A}_\alpha\bar{u} = \bar{h}$, where the solution admits the explicit representation

$$\bar{u} = \int_0^\infty \Phi_\alpha(t)\bar{h}\,\mathrm{d}t. \tag{2.46}$$

The right-hand side is called the *weak-star* resolvent operator and is denoted by either $R_\alpha := (-\bar{A}_\alpha)^{-1}$ or $R_\alpha := (\alpha I - \bar{A}_0)^{-1}$. For any $\alpha > \alpha_0$ we obtain

$$\|\Phi_\alpha(t)\bar{h}\| \leq \mathrm{e}^{-(\alpha-\alpha_0)t}\,\|\bar{h}\|, \qquad p_0(\Phi_\alpha(t)\bar{h},x) \leq \mathrm{e}^{-(\alpha-\alpha_0)t}p_0(\bar{h},x), \tag{2.47}$$

for any $t \geq 0$, and

$$\|R_\alpha\bar{h}\| \leq \frac{1}{\alpha-\alpha_0}\,\|\bar{h}\|, \qquad p_0(R_\alpha\bar{h},x) \leq \frac{1}{\alpha-\alpha_0}\,p_0(\bar{h},x), \tag{2.48}$$

for any x in X and where the norm $\|\cdot\|$ and the semi-norms $p_0(\cdot,x)$ are given by (2.38) and (2.43), respectively. Notice that $\alpha_0 = 0$ for $p = 0$, and it is clear that for any $\bar{h} \leq h$ (pointwise) we have $R_\alpha\bar{h} \leq R_\alpha h$, which is a weak form of the maximum principle.

Limiting the operator to the space $C_u(X)$ of bounded uniformly continuous functions, we find the so-called π-semigroups as proposed in Priola [205]. When the $\Phi_\alpha(t)$ is a strongly continuous Markov-Feller semigroup (typically an stochastic ODE) the weak version of the semigroup is of limited importance, since the domain of the infinitesimal generator is dense (in norm) in the space $C_p^0(\mathcal{O})$ of locally uniformly continuous functions with a growth bounded by the p-power of the norm. In general, we only have a weakly continuous Markov-Feller semigroup (typically stochastic PDE) and this weak version is very useful.

2.8.4 Invariant Distribution

Let E be a (locally compact) Polish space and $\{S(t) : t \geq 0\}$ be a stochastically continuous Markov semigroup on Banach space $B(E)$ of all bounded Borel real-valued functions on E, with Markov transition function $\{P(t,x,\cdot) : t \geq 0,\ x \in$

$E\}$,

$$S(t)f(x) = \int_E f(y)P(t,x,\mathrm{d}y), \quad \forall t \geq 0,\ x \in E.$$

We begin with the following

Definition 2.44 (invariant)**.** A probability measure μ on the Borel σ-algebra $\mathcal{B}(E)$ is called an *invariant distribution* or *invariant probability measure* of the Markov semigroup $\{S(t) : t \geq 0\}$ if

$$\int_E S(t)f\mathrm{d}\mu = \int_E f\mathrm{d}\mu,$$

or in term of the kernels

$$\int_E \mu(\mathrm{d}x)\int_E f(y)P(t,x,\mathrm{d}y) = \int_E f(x)\mu(\mathrm{d}x),$$

for every f in $B(E)$ and every $t > 0$. □

Notice that if an invariant distribution μ exits then the Markov semigroup satisfies $S(t)\mathbb{1} = \mathbb{1}$ or equivalently $P(t,x,E) = 1$, for every $t \geq 0$ and x in E, i.e, the semigroup has to be Markov, not sub-Markov, see Definition 2.39.

If $\{S(t) : t \geq 0\}$ is also a Markov-Feller semigroup then, besides $S(t) : B(E) \to B(E)$, we have $S(t) : C \to C$ for some closed subspace C of $C_b(E)$, the space of bounded continuous real-valued functions on E, e.g., C could be continuous functions vanishing at infinity or uniformly continuous functions or the whole space $C_b(E)$. Then $\{S(t) : t \geq 0\}$ is strongly continuous in C and the infinitesimal generator $(\mathcal{D}(A), A)$ is densely defined on C and

$$S(t)f - f = \int_0^t AS(s)f\mathrm{d}s = \int_0^t S(s)Af\mathrm{d}s,$$
$$\int_E [S(t)f - f]\mathrm{d}\mu = \int_E A\Big(\int_0^t S(s)f\mathrm{d}s\Big)\mathrm{d}\mu,$$

for any probability measure μ. This proves that μ is an invariant distribution if and only if

$$\int_E Af\mathrm{d}\mu = 0, \quad \forall f \in \mathcal{D}(A),$$

provided $\{S(t) : t \geq 0\}$ is a Markov-Feller semigroup.

The following result give a condition for the existence of an invariant probability measure, see Doob, Khasminskii, Krylov-Bogoliubov theorems in Da Prato and Zabczyk [52, Chapters 3 and 4],

Theorem 2.45 (existence)**.** *Let $\{P(t,x,\cdot) : t \geq 0,\ x \in E\}$ be a stochastically continuous Markov transition function on a Polish space E. If the family of time-average probabilities $\{R(t,x,\cdot) : t \geq 0,\ x \in E\}$,*

$$R(t,x,\cdot) := \frac{1}{t}\int_0^t P(s,x,\cdot)\mathrm{d}s,$$

is tight for t in $[t_0,\infty)$, for some $x = x_0$ and $t_0 > 0$, then there exists an invariant distribution μ. □

Recall that a family of probabilities $\{R(t,x_0,\cdot) : t \geq t_0\}$ is tight when for every $\varepsilon > 0$ there exits a compact subset $K = K_\varepsilon$ of E such that $R(t,x_0,K) \geq 1-\varepsilon$, for every $t \geq t_0$. Thus, any weak limit as $t \to \infty$ of the time-average probabilities is an invariant probability measure.

As it was defined early, $\{S(t) : t \geq 0\}$ is strongly Feller if $S(t_0) : B(E) \to C$, for some $t_0 > 0$. Also, a Markov transition function $\{P(t,x,B) : t \geq 0,\ x \in E,\ B \in \mathcal{B}(E)\}$ is called *irreducible* if there exist $t_0 > 0$ such that for every non-empty open subset $\mathcal{O}$ of E and for any x in $\mathcal{O}$ we have $P(t_0,x,\mathcal{O}) > 0$. Furthermore, it is called *regular* if there exists $t_0 > 0$ such that all transition probabilities $\{P(t_0,x,\cdot) : x \in E\}$ are mutually equivalent.

Theorem 2.46 (uniqueness)**.** *Let μ be an invariant distribution of a stochastically continuous Markov transition function $\{P(t,x,\cdot) : t \geq 0,\ x \in E\}$ on a Polish space E. If it is strongly Feller and irreducible then it is also regular, the invariant distribution is unique and*

(1) *for any x in E and B in $\mathcal{B}(E)$ we have $P(t,x,B) \to \mu(B)$ as $t \to \infty$,*

(2) *there exists $t_0 > 0$ such that all probabilities measures $\{P(t,x,\cdot) : t \geq t_0,\ x \in E\}$ are equivalent to μ.* □

A set B in $\mathcal{B}(E)$ is called invariant with respect to a stochastically continuous Markov transition function $\{P(t,x,B) : t \geq t_0,\ x \in E,\ B \in \mathcal{B}(E)\}$ having an invariant probability measure μ if except in a set of μ-measure zero, $P(t,\cdot,B) = \mathbb{1}_B$, for every $t > 0$. Then an invariant probability measure μ is called *ergodic* if the only invariant sets have μ measure 0 or 1, i.e., if $P(t,\cdot,B) = \mathbb{1}_B$ μ-a.s. implies $\mu(B) = 0$ or $\mu(B) = 1$. It can be proved that an invariant distribution μ is ergodic if and only if the time-average commute with the space average, i.e.,

$$\frac{1}{T}\int_0^T \mathrm{d}t \int_0^T f(y)P(t,\cdot,\mathrm{d}y) \to \int_E f(y)\mu(\mathrm{d}y) \ \text{ in } \ L^2(E,\mu),$$

as $T \to \infty$, for every f in $L^2(E,\mu)$.

Sometimes a stronger convergence than (1) in Theorem 2.46 is necessary, e.g., exponential convergence. Based on Doob's ergodicity Theorem on a compact space E, the so-called Doeblin's condition, namely, there exist $t_0 > 0$ and $\delta > 0$ such that

$$P(t_0,x,B) - P(t_0,y,B) \leq 1-\delta, \quad \forall x,y \in E,\ B \in \mathcal{B}(E), \tag{2.49}$$

imply the existence of a unique invariant probability measure μ and the exponential convergence

$$\Big|\int_E f(y)P(t,x,\mathrm{d}y) - \int_E f(y)\mu(\mathrm{d}y)\Big| \leq C\mathrm{e}^{-\omega t}\sup_{y\in E}|f(y)|,$$

for some positive constants C and ω, and for every x in E, as long as E is compact.

Typical conditions to ensure the tightness of the probability measures needed in Theorem 2.45 are given in term of the existence of Liapunov functions, see Khasminskii [130]. For instance, if there exists a function φ in $\mathcal{D}(A)$ such that $\varphi \geq 0$ and satisfying $\varphi(x) \to -\infty$ as $|x| \to \infty$ (which means that for every $m > 0$ there is a compact set $K = K_m$ of E such that $\varphi(x) < -m$ for every x in $E \setminus K$) the family $\{P(t, x_0, \cdot) : t \geq 0,\, x \in E\}$ is tight for every x_0 fixed. The existence of a Liapunov function satisfying $\varphi(x) \to +\infty$ as $|x| \to \infty$ and $A\varphi - \alpha\varphi \leq C$ for some positive constants α and C, yields the uniqueness of the invariant probability measure.

If μ is an invariant distribution then Jensen's inequality yields

$$\Big| \int_E f(y) P(t, x, \mathrm{d}y) \Big|^p \leq \int_E |f(y)|^p P(t, x, \mathrm{d}y) = \int_E |f(y)|^p \mu(\mathrm{d}y),$$

for every p in $[1, \infty)$. The stochastically continuous Markov semigroup $\{S(t) : t \geq 0\}$ can be extended to a strongly continuous semigroup of contractions in the Lebesgue spaces $L^p(E, \mu)$. Moreover, any other probability measure ν which is equivalent to μ (i.e., ν is absolutely continuous with respect to μ and conversely) can be used to extend the semigroup to $L^p(E, \nu)$.

In a finite-dimensional setting, let D be a domain in $\mathbb{R}^d$ (i.e., the closure of the interior of D is equal to its closure) and let $(A, \mathcal{D})$ be a linear operator defined on a linear sub-space $\mathcal{D}$ of $C^\infty(\bar{D})$ containing $C_0^\infty(D)$. Assume that there exist a probability measure μ on D such that

$$Av \in L^1(D, \mu) \quad \text{and} \quad \int_D Av(x)\, \mu(\mathrm{d}x) = 0, \quad \forall v \in \mathcal{D}.$$

Then, we want to find a unique extension of $(A, \mathcal{D})$ which generates a strongly continuous Markov semigroup $\{T(t) : t \geq 0\}$ in $L^p(D, \mu)$, $1 \leq p < \infty$, having μ as an invariant measure. Several conditions are given in the literature to ensure this construction, e.g., see Stannat [235] and references there in.

2.9 Dirichlet Forms

Of particular interest is the Dirichlet form theory, e.g., see the books Fukushima et al. [92], Silverstein [228], and for a non symmetric extension we refer to Ma and Röckner [161].

Consider a Hausdorff space X and a σ-finite nonnegative Borel measure m on X (i.e., defined on the Borel σ-algebra $\mathcal{B}(X)$, which is assumed to be also generated by all the continuous functions on X), such that $L^2(X, m)$ is a (real) separable Hilbert space. In most of the cases, X is a locally compact separable Hausdorff space and m is a measure finite on compact sets and strictly positive on each non empty open set. Let us denote $\phi(x) = (0 \vee x) \wedge 1$. Clearly, if $u \in L^2(X, m)$ then $\phi(u)$ is still in $L^2(X, m)$, so that ϕ is in a sense a *cut-off* function.

A bounded linear operator $T\colon L^2(X, m) \to L^2(X, m)$ is called *Markovian* if $0 \leq Tu \leq 1$ m-a.e. whenever $0 \leq u \leq 1$ m-a.e., i.e. $0 \leq T(\phi(u)) \leq 1$ m-a.e. for any $u \in L^2(X, m)$.

Denote by $C_0^0(X)$ the Frèchet space of all real-valued continuous functions with compact support in X and by $C_0(X)$ the Banach space of all real-valued continuous functions vanishing at infinity (i.e., $C_0(X)$ is the closure of $C_0^0(X)$ in $C_b(X)$, the Banach space of continuous and bounded functions). As previously mentioned, starting with a closable positive and continuous (or bounded) bilinear form $a(\cdot,\cdot)$ with domain $\mathcal{D}(a) \subset H$, we can extend the form to a Hilbert space V (the completion of $\mathcal{D}(a)$ in a graph norm) where $a_\alpha(\cdot,\cdot) = a(\cdot,\cdot)+\alpha(\cdot,\cdot)$, with $\alpha > 0$ is a coercive and continuous bilinear form, i.e., there exist constant $M \geq \kappa > 0$, $\kappa = \kappa_\alpha$ such that

$$\begin{cases} a(u,u) + \alpha|u|^2 \geq \kappa\|u\|, & \forall u \in V, \\ |a(u,v)| \leq M\|u\|\,\|v\|, & \forall u,v \in V, \end{cases} \tag{2.50}$$

where $\|\cdot\| = \|\cdot\|_{\mathrm{V}}$ denotes the norm in V, i.e., $\|\cdot\|_1$, and $|\cdot| = \|\cdot\|_H$ is the norm in H. Recall that if $a(\cdot,\cdot)$ is a closed positive symmetric form then $a_\alpha(\cdot,\cdot)$ is coercive and continuous form. The continuity of the form $a(\cdot,\cdot)$ is only needed in the non-symmetric case.

Let us consider a closed, positive and continuous bilinear form $a(\cdot,\cdot)$ on $L^2(X,m)$, i.e., $a(\cdot,\cdot)$ is defined on a Hilbert space V, which is densely and continuously included in $H = L^2(X,m)$ and condition (2.50) is satisfied.

Definition 2.47 (Dirichlet form)**.** A closed, positive and continuous bilinear form $a(\cdot,\cdot)$ defined on $\mathcal{D}(a) \subset H$, $H = L^2(X,m)$ is called a *Dirichlet form* if $\phi(u) \in \mathcal{D}(a)$ for any $u \in \mathcal{D}(a)$ and

$$a\big(u+\phi(u), u-\phi(u)\big) \geq 0 \quad \text{and} \quad a\big(u-\phi(u), u+\phi(u)\big) \geq 0, \tag{2.51}$$

for every u in $\mathcal{D}(a)$, with $\phi(t) = (0 \vee t) \wedge 1$. Moreover, a Dirichlet form is called *regular* if $C_0^0(X) \cap \mathcal{D}(a)$ is dense in $\mathcal{D}(a)$ with the graph norm $\|u\|_1$ and dense in $C_0(X)$. □

Usually, to verify that a particular bilinear form $a(\cdot,\cdot)$ with domain $\mathcal{D}(a)$ satisfies the above condition (2.51) we replace the cut-off function $\phi(t) = (0 \vee t) \wedge 1$ by smooth functions $\phi_\varepsilon(t)$, for $\varepsilon > 0$, satisfying

$$\begin{aligned} &\phi_\varepsilon : \mathbb{R} \to [-\varepsilon, 1+\varepsilon], \qquad \phi_\varepsilon(t) = 1, \quad \forall t \in [0,1], \\ &0 \leq \phi_\varepsilon(t) - \phi_\varepsilon(t') \leq t - t', \quad \forall t < t'. \end{aligned}$$

Then, condition (2.51) is satisfied if $\phi_\varepsilon(u) \in \mathcal{D}(a)$ for any $u \in \mathcal{D}(a)$ and

$$\liminf_{\varepsilon\to 0} a\big(u \pm \phi_\varepsilon(u), u \mp \phi_\varepsilon(u)\big) \geq 0,$$

for every u in $\mathcal{D}(a)$, When $a(\cdot,\cdot)$ is symmetric, condition (2.51) becomes

$$\phi(u) \in \mathcal{D}(a), \quad a\big(\phi(u), \phi(u)\big) \leq a(u,u), \tag{2.52}$$

or

$$\limsup_{\varepsilon\to 0} a\big(\phi_\varepsilon(u), \phi_\varepsilon(u)\big) \leq a(u,u),$$

for every u in $\mathcal{D}(a)$.

The Dirichlet form is a particular case of positive, closed, bilinear form. Hence, there is a unique contraction operator semigroup associated with a Dirichlet form and, when extended to the complexification of H, i.e. the (classes of) functions of $L^2(X,m)$ are complex valued, the semigroup is analytic. Actually we have

Theorem 2.48 (symmetric case)**.** *Given a symmetric Dirichlet form $a(u,v)$ there exists a unique (non positive and densely defined) self-adjoint operator $A\colon \mathcal{D}(A) \subset \mathcal{D}(a) \subset L^2(X,m) \to L^2(X,m)$ such that $a(u,v) = (-Au,v)$ for u in $\mathcal{D}(A)$ and the corresponding semigroup $T_t = \mathrm{e}^{At}$ is Markovian. Conversely, given a non positive, self-adjoint operator A on $L^2(X,m)$, such that $T_t = \mathrm{e}^{At}$ is Markovian, the form $a(u,v) = ((-A)^{1/2}u, (-A)^{1/2}v)$ with u,v in $\mathcal{D}(a) = \mathcal{D}((-A)^{1/2})$ is a symmetric Dirichlet form.* □

For the non symmetric case we have

Theorem 2.49 (non symmetric case)**.** *Given a Dirichlet form $a(u,v)$ there is a unique contraction semigroup $T_t = \mathrm{e}^{At}$, such that T_t and T_t^* are Markovian and that $a(u,v) = (-Au,v)$ for u in $\mathcal{D}(A)$. Conversely, given a a contraction semigroup $T_t = \mathrm{e}^{At}$, such that T_t and T_t^* are Markovian, the form $a(u,v) = (-Au,v)$ with u,v in $\mathcal{D}(A)$ is closable and the minimal extension is a Dirichlet form.* □

A typical example of symmetric Dirichlet form is the following. Let $P(t,x,A)$ be a Markov transition function, with S_t the corresponding Markov semigroup, such that:

$$\int_X S_t u\, v \,\mathrm{d}m = \int_X u\, S_t v \,\mathrm{d}m \tag{2.53}$$

for any non negative measurable u,v and any t. If

$$\lim_{t\to 0} S_t u(x) = u(x), \quad m - \text{a.e.}$$

for any u in $C_0^0(X)$, then S_t is strongly continuous in $L^2(X,m)$. Hence it is a Markovian semigroup, and in view of Theorem 2.49, there is a corresponding Dirichlet form. A Markov process with a transition function $P(t,x,A)$ on the states space X, where its associated semigroup $\{S_t : t \geq 0\}$ satisfies (2.53) is called a symmetric Markov process (with respect to the measure m), i.e., the operators $S_t : H \to H$, with $H = L^2(X,m)$, are symmetric.

There are also relations with the Markov process theory. Fukushima et al. [92, Theorem 6.2.1] proved the following result for symmetric forms, and its extension to non symmetric forms can be found in Ma and Röckner [161].

Theorem 2.50. *Given a (symmetric) regular Dirichlet form on $L^2(X,m)$, there exists a (symmetric) Markov process (actually a Hunt process), whose Dirichlet form is the given one.* □

We have also the uniqueness of the Markov (Hunt) process in the following sense of equivalence between Markov processes. First let us define the concept of *exceptional set*: a subset $N \subset X$ is called *exceptional* if there exists a Borel set $\tilde{N} \supset N$ such that $P_m(\sigma_{\tilde{N}} < +\infty) = 0$, where $\sigma_A = \inf\{t > 0 : X_t \in A\}$ and P_m is the product measure of transition functions and m as initial measure. Any exceptional set has null m-measure and, with the potential theory language, a set is exceptional if and only if it has capacity zero.

Among the exceptional sets, it is important to define, as *properly exceptional sets*, those sets N that are Borel, $m(N) = 0$ and $X \setminus N$ are "invariant" for the process X_t, that is $P_x(\tilde{\Omega}) = 1$ for any $x \in X \setminus N$, where

$$\tilde{\Omega} = \{\omega : X_t(\omega) \in X \setminus N,\ X_{t-}(\omega) \in X \setminus N,\ \forall t \geq 0\}.$$

In other words, it is possible to restrict the process X_t to $\tilde{\Omega}$ without changing the statistical properties of X_t.

Two m-symmetric Markov (Hunt) processes are *equivalent* if there is one common properly exceptional set, outside which their transition functions coincide and both processes are associated to the same Dirichlet form.

• *Remark* 2.51. Comparing with the result of Stroock and Varadhan [240], we obtain basically the same type of results. However, we have to observe that the martingale method is valid also in directions forbidden to the method of Dirichlet form (e.g. second order differential operators not in divergence form). On the other hand the method of Dirichlet forms works also in infinite dimensions or on topological space of fractal nature. □

Perhaps, the following setting is a typical example of *jump* Dirichlet form. Let (X, ρ) be a locally compact (separable complete metric) Polish space (i.e., essentially, a Borel subset of $\mathbb{R}^d$) and m be a positive Radon measure on X with full support, i.e., m is a nonnegative measure defined on the Borel σ-algebra $\mathcal{B}(X)$, finite on compact sets and strictly positive on non-empty open sets, with $\mathrm{supp}(m) = X$. Let $j(x, \cdot)$ be a kernel in X, i.e., (a) for every B in $\mathcal{B}(X)$ the function $x \mapsto j(x, B)$ is measurable and (b) for every x in X the mapping $B \mapsto j(x, B)$ is a nonnegative measure in $\mathcal{B}(X)$. The kernel $j(x, \cdot)$ is such that

$$\int_K m(\mathrm{d}x) \int_X \big(\rho^2(x, \xi) \wedge 1\big) j(x, \mathrm{d}\xi) < \infty, \tag{2.54}$$

for every compact subset K of X, and there exists another kernel denoted by $j^*(x, \cdot)$ satisfying the above integrability condition and

$$\int_A j(x, B)\, m(\mathrm{d}x) = \int_B j^*(x, A)\, m(\mathrm{d}x), \quad \forall A, B \in \mathcal{B}(X).$$

The j^* kernel is the m-adjoint of j kernel, and clearly, whenever $j = j^*$, the kernel is called m-symmetric. It is clear that $j(x, \mathrm{d}\xi)\, m(\mathrm{d}x)$ and $j^*(x, \mathrm{d}\xi)\, m(\mathrm{d}x)$ can be regarded as Radon measures (denoted by j_m and j_m^*) on $X \times X$ with support outside of the diagonal set $\{(x, y) \in X \times X : x = y\}$. Now, define the positive bilinear form

$$a(u, v) := \int_{X \times X} [u(x) - u(y)][v(x) - v(y)]\, j_m(\mathrm{d}x, \mathrm{d}y),$$

on $L^2(X,m)$, with domain $\mathcal{D}(a) := \{u \in L^2(X,m) : a(u,u) < \infty\}$. Note that if the measure m is bounded then $\mathcal{D}(a)$ contains all Lipschitz continuous functions. Since j_m is a nonnegative measure, the sectorial estimate

$$|a(u,v)| \le \sqrt{a(u,u)}\sqrt{a(v,v)}, \quad \forall u,v \in \mathcal{D}(a)$$

is granted from the definition. On the other hand, setting $\Gamma_\varepsilon = \{(x,y) \in K \times K : \rho(x,y) > \varepsilon\}$, for any $\varepsilon > 0$, we have

$$\int_{\Gamma_\varepsilon} [u(x) - u(y)]^2 \, j_m(\mathrm{d}x, \mathrm{d}y) \le 2 \int_{\Gamma_\varepsilon} [u^2(x) + u^2(y)] \, j_m(\mathrm{d}x, \mathrm{d}y)$$

and

$$\int_{\Gamma_\varepsilon} [u^2(x) + u^2(y)] \, j_m(\mathrm{d}x, \mathrm{d}y) = \\ = \int_K u^2(x) \, m(\mathrm{d}x) \int_K \mathbb{1}_{\{\rho(x,\xi)>\varepsilon\}} [j(x, \mathrm{d}\xi) + j^*(x, \mathrm{d}\xi)].$$

Hence, in view of condition (2.54), we deduce that $u = 0$ m-a.e. implies $a(u,u) = 0$. By means of this last property, we can shows that $a(\cdot,\cdot)$ is closable and indeed a (regular) Dirichlet form if $\mathcal{D}(a)$ is dense in $L^2(X,m)$. Alternatively, let $a^\varepsilon(\cdot,\cdot)$ be as $a(\cdot,\cdot)$ but integrating only on the region Γ_ε instead of $X \times X$, and define the operator

$$Au = \lim_{\varepsilon \to 0} A_\varepsilon u, \quad A_\varepsilon u(x) = \int_{\rho(x,\xi)>\varepsilon} [u(\xi) - u(x)] \, [j(x, \mathrm{d}\xi) + j^*(x, \mathrm{d}\xi)],$$

with domain $\mathcal{D}(A) \subset L^2(X,m)$, where $\mathcal{D}(A)$ is the set of all u in $\mathcal{D}(a)$ such that the above limit Au exits in $L^2(X,m)$. Since A_ε is a bounded operator, $a^\varepsilon(u,v) = (-A_\varepsilon u, v)$, for every u, v in $L^2(X,m)$, and

$$a(u,v) = (-Au, v) = \lim_{\varepsilon \to 0} (-A_\varepsilon u, v), \quad \forall u \in \mathcal{D}(A), \ v \in \mathcal{D}(a),$$

the argument of Remark 2.33 shows that $a(\cdot,\cdot)$ is closable, if the domain $\mathcal{D}(A)$ is dense in $L^2(X,m)$. This is the case when X is a domain D on R^d, so that $C_0^\infty(D) \subset \mathcal{D}(A)$. Moreover, by means of smooth cut-off functions ϕ_ε we can prove that $a(\cdot,\cdot)$ satisfies (2.51) or equivalently, that A and A_ε satisfy the weak maximum principle.

• *Remark* 2.52. It is clear that the previous example of jump Dirichlet form can be complemented with diffusion terms, as in Subsection 2.7.2. □

To study time-dependent coefficients, we use spaces of the type $L^2(0,T;V)$ or $L^2(\mathbb{R},V)$ and adapting this concept lead to the so-called *generalized Dirichlet forms*, which sill have corresponding generators, strongly continuous contraction semigroups and resolvents. These are (mainly non-symmetric) bilinear forms represented as the sum of a coercive and continuous bilinear form $a_0(\cdot,\cdot)$ defined in Hilbert space $\mathcal{D}(a_0)$, densely and continuously included in $\mathcal{H} = L^2(X,m)$, (i.e., conditions (2.16) and (2.17) in Subsection 2.5.3) and a

linear (unbounded) operator $(\mathcal{D}(A_1), A_1)$ in $L^2(X,m)$, which is a small perturbation in the sense that $(\mathcal{D}(A_1), A_1)$ and its adjoint $(\mathcal{D}(A_1^*), A_1^*)$ generate a pair of strongly continuous semigroups of contractions e^{tA_1} and $\mathrm{e}^{tA_1^*}$, such that e^{tA_1} and $\mathrm{e}^{tA_1^*}$ can be extended (as a strongly continuous semigroup, not necessarily of contractions) to the dual space $\mathcal{D}'(a_0)$, where the infinitesimal generators are denoted by $(\mathcal{D}_1, A_1)$ and $(\mathcal{D}_1^*, A_1^*)$. Thus the generalized bilinear forms is defined by the expressions

$$a(u,v) := \begin{cases} a_0(u,v) - \langle A_1 u, v\rangle, & \forall u \in \mathcal{D}(a_0)\cap\mathcal{D}_1,\ v\in\mathcal{D}(a_0), \\ a_0(u,v) - \langle A_1^* v, u\rangle, & \forall u \in \mathcal{D}(a_0),\ v\in\mathcal{D}(a_0)\cap\mathcal{D}_1^*. \end{cases}$$

where $\langle\cdot,\cdot\rangle$ denotes the dual pairing between $\mathcal{D}'(a_0)$ and $\mathcal{D}(a_0)$. Note that $\mathcal{V} = \mathcal{D}(a_0)\cap\mathcal{D}_1$ and $\mathcal{V}_* = \mathcal{D}(a_0)\cap\mathcal{D}_1^*$ are Hilbert spaces, and we have $\mathcal{D}(a_0)\subset\mathcal{H}\subset\mathcal{D}'(a_0)$, $\mathcal{V}\subset\mathcal{H}\subset\mathcal{V}'$, and $\mathcal{V}_*\subset\mathcal{H}\subset\mathcal{V}_*'$, all with continuous and dense inclusions, where $\mathcal{H} = L^2(X,m)$ has been identified with its dual space $\mathcal{H}'$. Moreover, for the above setting, the bilinear form $a(\cdot,\cdot)$ and its adjoint are such that the conclusions of Lax-Milgram Theorem 2.23 holds on the Hilbert space $\mathcal{V}$ and $\mathcal{V}_*$. Some specific conditions (similar to Definition 2.47) are imposed on the bilinear form $a(\cdot,\cdot)$ to be called a generalized Dirichlet form.

Perhaps a typical example is as follows:

$$a_0(u,v) := \int_{\mathbb{R}} a_0^{(t)}(u(t),v(t))\mathrm{d}t, \quad \forall u,v\in\mathcal{D}(a_0) := L^2(\mathbb{R},V),$$

where $a_0^{(t)}(\cdot,\cdot)$ is a family of coercive and continuous (i.e., satisfying a sector condition) bilinear forms (uniformly in t) defined in a common domain V (which is a subspace of a Hilbert space H), and the unbounded linear operator A_1 is the time-derivative ∂_t with initial domain $H^1(\mathbb{R},H)$ and extended to the space

$$\mathcal{D}_1 := \big\{v\in L^2(\mathbb{R},H) : \partial_t v\in L^2(\mathbb{R},V')\big\}.$$

Clearly, V is itself a Hilbert space densely and continuously included in $H = L^2(X_0,m_0)$, $\mathcal{V} = \mathcal{V}_* = L^2(\mathbb{R},V)\cap\mathcal{D}_1$ and the reference space $\mathcal{H} = L^2(X,m)$ is given with $X = \mathbb{R}\times X_0$ and $m = \mathrm{d}t\times m_0$. Certainly, there are may other interesting examples, the interested reader is referred to Stannat [236, 234, 235] and references there in.

2.10 Dynamical Systems

The concept of dynamical system is related to Markovian systems, which in turn are related with Markov processes and semigroups. A quadruple $(\Omega,\mathcal{F},\mathbb{P},\theta)$ is called a *dynamical system* if $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space and $\theta = (\theta_t,\ t\in\mathbb{R})$ is a group of invertible, measurable transformation from Ω into itself, preserving the probability measure $\mathbb{P}$, i.e.,

$$\mathbb{P}(\theta_t A) = \mathbb{P}(A) \quad \forall A\in\mathcal{F},\ t\in\mathbb{R}.$$

The group $\theta = (\theta_t,\, t \in \mathbb{R})$ induces a group of linear transformations $(U_t,\, t \in \mathbb{R})$, either on the real Hilbert space $\mathcal{H} := L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$ or complex Hilbert space $\mathcal{H}_{\mathbb{C}} := L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{C})$, by the expression

$$U_t\xi(\omega) := \xi(\theta_t\omega), \quad \forall \xi \in \mathcal{H}_{\mathbb{C}},\ \omega \in \Omega,\ t \in \mathbb{R}.$$

The dynamical system is *strongly mixing* if

$$\lim_{t\to\infty} \mathbb{P}(\theta_t A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Let E be a Polish space, denote by $\mathcal{E} := \mathcal{B}(E)$ the Borel subsets of E, by $B(E)$ the bounded and Borel functions with real (or complex) values, by $\mathcal{M}_1(E)$ the (convex) set of all probability measures on $(E, \mathcal{E})$, and by $B(x, \delta)$ the open ball in E with center x and radius $\delta > 0$. A Markovian transition function is a family $(P_t(x, A),\, t \geq 0,\, x \in E,\, A \in \mathcal{E})$ of transformation satisfying:

(a) for each $t \geq 0$ and $x \in E$ the mapping $A \mapsto P_t(x, A)$ is a probability measure,

(b) for each $t \geq 0$ and $A \in \mathcal{E}$ the function $x \mapsto P_t(x, A)$ is measurable with respect to $\mathcal{E}$,

(c) for each $x \in E$ and $A \in \mathcal{E}$, $P_0(x, A) = \mathbf{1}_A(x)$,

(d) the identity

$$P_{t+s}(x, A) = \int_E P_t(x, \mathrm{d}y) P_s(y, A), \quad \forall t, s \geq 0,\ x \in E,\ A \in \mathcal{E},$$

holds true. A Markovian transition function defines a semigroup of linear operators $(P_t,\, t \geq 0)$ on the space $B(E)$ by the formula

$$P_t h(x) := \int_E h(y) P_t(x, \mathrm{d}y), \quad \forall x \in E,\ h \in B(E),$$

which is called the Markovian semigroup associated with the given transition function. Now, either the Markovian semigroup $(P_t,\, t \geq 0)$, or the Markovian transition function $(P_t(x, A),\, t \geq 0, x \in E, A \in \mathcal{E})$, is called *stochastically continuous* if

$$\lim_{t\to 0} P_t(x, B(x, \delta)) = 1, \quad x \in E,\ \delta > 0.$$

This is equivalent to the condition $P_t h(x) \to h(x)$, for every x in E and h continuous, or uniformly continuous or Lipschitz continuous. The dual semigroup $(P_t^*,\, t \geq 0)$ on $\mathcal{M}_1(E)$ is defined by

$$P_t^*\mu(A) := \int_E P_t(x, A)\mu(\mathrm{d}x), \quad \forall t \geq 0,\ A \in \mathcal{E},$$

and a measure ν is called *invariant* if $P_t^*\nu = \nu$, for every $t \geq 0$.

Next, given a Markovian semigroup $(P_t,\, t \geq 0)$ with an invariant measure ν, a canonical dynamical system $(\Omega, \mathcal{F}, \mathbb{P}_\nu, \theta)$ can be constructed as described below, with $\Omega := E^{\mathbb{R}}$, the space of all functions from $\mathbb{R}$ into E, $\mathcal{F} := \mathcal{E}^{\mathbb{R}}$

the product Borel σ-algebra, $\theta_t\omega(s) := \omega(t+s)$ the translation group, $\mathbb{P}_\nu$ the Kolmogorov extension of the transition function and the invariant measure, where the canonical process $X(t) := \omega(t)$, with t in $\mathbb{R}$, will be Markovian with transition function $(P_t(x, A),\, t \geq 0,\, x \in E,\, A \in \mathcal{E})$, and stationary with a stationary distribution ν.

As mentioned in previous sections, the Kolmogorov extension is constructed as follows, for any cylindrical set $C = \{\omega \in E^{[0,\infty)} : \omega(t_1) \in A_1, \omega(t_2) \in A_2, \dots, \omega(t_n) \in A_n\}$, where $0 \leq t_1 < \cdots < t_n$, $A_1, \dots, A_n$ in $\mathcal{E}$, we define

$$\mathbb{P}_x(C) := \int_{A_1} P_{t_1}(x, \mathrm{d}x_1) \int_{A_2} P_{t_2-t_1}(x_2 - x_1, \mathrm{d}x_2) \dots \\ \dots \int_{A_n} P_{t_n-t_{n-1}}(x_n - x_{n-1}, \mathrm{d}x_n),$$

which can be extended to a unique probability $\mathbb{P}_x$ on $(E^{[0,\infty)}, \mathcal{E}^{[0,\infty)})$. Similarly, define the probability measure on the cylindrical set C by

$$\mathbb{P}_\nu(C) := \int_{A_1} \nu(\mathrm{d}x_1) \int_{A_2} P_{t_2-t_1}(x_2 - x_1, \mathrm{d}x_2) \dots \\ \dots \int_{A_n} P_{t_n-t_{n-1}}(x_n - x_{n-1}, \mathrm{d}x_n),$$

which can be extended to a unique probability $\mathbb{P}_\nu$ on $(E^{\mathbb{R}}, \mathcal{E}^{\mathbb{R}})$, again by means of Kolmogorov extension theorem. It is clear that for every Γ in $\mathcal{E}^{\mathbb{R}}$ and $\varepsilon > 0$ there is a cylindrical set C such that

$$\mathbb{P}_\nu(\Gamma \setminus C) + \mathbb{P}_\nu(C \setminus \Gamma) \leq \varepsilon.$$

If $\mathcal{F}_t$ is the σ-algebra generated by the canonical processes $(X(s),\, s \leq t)$ for every t in $\mathbb{R}$, then we have $\mathbb{P}_\nu$ almost surely

$$\mathbb{P}_\nu(X(t+s) \in A \mid \mathcal{F}_t) = \mathbb{P}_\nu(X(t+s) \in A \mid \sigma(X(t)) = \\ = P_s(X(t), A),$$

for every A in $\mathcal{E}$, and $s \geq 0$. Or even more general,

$$\mathbb{P}_\nu(X(t+\cdot) \in \Gamma \mid \mathcal{F}_t) = \mathbb{P}_\nu(X(t+\cdot) \in \Gamma \mid \sigma(X(t)) = \mathbb{P}_{X(t)}(\Gamma),$$

for every Γ in $\mathcal{E}^{[0,\infty)}$.

The group $\theta = (\theta_t,\, t \in \mathbb{R})$ of invertible and measurable transformation from $\Omega = E^{\mathbb{R}}$ into itself, defined by

$$\theta_t\omega(\cdot) := \omega(\cdot + t), \quad \forall \omega \in \Omega,\ t \in \mathbb{R},$$

is the group of translations. Since ν is invariant, the canonical process $X(t, \omega) := \omega(t)$, t in $\mathbb{R}$ is stationary, i.e.,

$$\mathbb{P}_\nu(X \in \theta_t\Gamma) = \mathbb{P}_\nu(X \in \Gamma), \quad \forall t \in \mathbb{R},\ \Gamma \in \mathcal{F},$$

where

$$\theta_t\Gamma := \{\omega \in \Omega : \theta_t^{-1}\omega \in \Gamma\},$$

and the transformation θ_t preserve the measure $\mathbb{P}_\nu$.

Notice that the canonical process X in $(\Omega, \mathcal{F}, \mathbb{P}_\nu)$ is stochastically continuous if and only if the Markovian transition function $(P_t(x, A),\, t \geq 0,\, x \in E,\, A \in \mathcal{E})$ is stochastically continuous in the above sense. Moreover, under this condition, the group $(U_t,\, t \in \mathbb{R})$ of linear transformations on the (real of complex) Hilbert space $\mathcal{H}_\nu := L^2(\Omega, \mathcal{F}, \mathbb{P}_\nu)$ is continuous, i.e.,

$$\lim_{t\to s} U_t\xi = U_s\xi, \quad \forall \xi \in \mathcal{H}_\nu,\ s \in \mathbb{R}.$$

On the other hand, other conditions (see Theorem 2.41) are necessary to show that support of the probability $\mathbb{P}_\nu$ is indeed $C_0(E)$, the space of continuous functions vanishing at infinity. The reader is referred to the book Da Prato and Zabczyk [52] for a comprehensive treatment.

A Markovian semigroup $(P_t,\, t \geq 0)$ on $B(E)$ with an invariant measure ν has a unique extension to a semigroup of nonnegative contraction operators in $L^p(E, \mathcal{F}, \nu)$, with $p = 1, 2, \infty$. Moreover, the mean ergodic theorem holds, i.e., for every h in $L^p(E, \mathcal{F}, \nu)$, with $p = 1, 2$, the average limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P_k h = h^*$$

exists in $L^p(E, \mathcal{F}, \nu)$, and

$$P_1 h^* = h, \qquad \int_E h(x)\nu(\mathrm{d}x) = \int_E h^*(x)\nu(\mathrm{d}x),$$

see Yosida [262, Theorems XIII.1.1 and XIII.1.2, pp. 381–382]. It is clear that $(P_t,\, t \geq 0)$ is also a C_0-semigroup in $L^2(E, \mathcal{F}, \nu)$, whenever it is stochastically continuous.

2.11 Integro-differential Operators

We are interested in integro-differential operators associated with diffusion processes with jumps, see Gikhman and Skorokhod [99, p. 245] and Bensoussan and Lions [17, p. 178]. For a comprehensive treatment on (elliptic/parabolic) integro-differential operators, we refer to the books Garroni and Menaldi [93, 94].

A Radon measure $M(x, \mathrm{d}z)$ on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$, for any x in $\mathbb{R}^d$, determines this operator. Depending on the assumptions on the singularity at the origin of the Lévy kernel $M(x, \mathrm{d}z)$ we may classify these integro-differential operators. The expression

$$I_1\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi] M_1(\cdot, \mathrm{d}z), \tag{2.55}$$

with

$$\int_{|z|<1} |z| M_1(\cdot, \mathrm{d}z) < \infty$$

define an integro-differential operator of order 1, since, in view of the mean value theorem, the expression (2.55) makes sense for bounded continuously differentiable functions. However, a form

$$I_2\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi - z \cdot \nabla\varphi] M_2(\cdot, \mathrm{d}z), \tag{2.56}$$

with

$$\int_{|z|<1} |z|^2 M_2(\cdot, \mathrm{d}z) < \infty$$

gives an integro-differential operator of order 2, since, by Taylor's formula, the expression (2.56) makes sense for bounded twice-continuously differentiable functions. On the other hand,

$$I_0\varphi = \int_{\mathbb{R}^d_*} [\varphi(\cdot + z) - \varphi] M_0(\cdot, \mathrm{d}z), \quad \text{with} \quad \int_{\mathbb{R}^d_*} M_0(\cdot, \mathrm{d}z) < \infty \tag{2.57}$$

provides a bounded (or order 0) integral (or non-local) operator, since the expression (2.57) makes sense for bounded functions.

Note that in the definitions (2.55) and (2.56) of the operators I_1 and I_2, we may replace the region of integration $\{|z| < 1\}$ by one of the form $\{|z| < r\}$, for any $r > 0$. The interesting part is the singularity at the origin, i.e., *small jumps*. On the other hand, in the definition (2.57) of the operator I_0 we may use $\{|z| \geq r\}$, for any $r > 0$, as the region of integration, instead of the whole space $\mathbb{R}^d_*$. Here the interest is on the integrability at infinity, i.e., *large jumps*.

Let us present some typical examples. First, an operator of order 0,

$$I_0\varphi = \lambda[\varphi(\cdot + \zeta) - \varphi],$$

for some constants $\lambda > 0$, $\zeta \in \mathbb{R}^d_*$. Here the Lévy kernel is $M_0(x, \cdot) = \lambda\delta_\zeta$, where δ_ζ denotes the Dirac measure at ζ. Second, two examples of order 1,

$$I_1\varphi = \int_{|z|<1} [\varphi(\cdot + z) - \varphi] |z|^{-d} \mathrm{d}z,$$

where the Lévy kernel $M_1(x, \mathrm{d}z) = \mathbb{1}_{|z|<1}|z|^{-d}\mathrm{d}z$, and

$$I_1\varphi = \sum_{n=1}^{\infty} \lambda_n [\varphi(\cdot + \zeta_n) - \varphi],$$

with

$$\lambda_n \geq 0, \quad \sum_{n=1}^{\infty} \lambda_n = \infty, \quad \sum_{n=1}^{\infty} \lambda_n |\zeta_n| < \infty.$$

Here $\zeta_n \to 0$ as $n \to \infty$ and the Lévy kernel $M_1(x,\cdot) = \sum_{n=1}^{\infty} \lambda_n \delta(\cdot - \zeta_n)$. Next, two examples of order 2,

$$I_0\varphi + I_2\varphi = \int_{R_*^d} [\varphi(\cdot + z) - \varphi - z \cdot \nabla\varphi \mathbb{1}_{|z|<1}]|z|^{-d-1}\mathrm{d}z,$$

where the Lévy kernels are $M_0(x,\mathrm{d}z) = \mathbb{1}_{|z|>1}|z|^{-d-1}\mathrm{d}z$ and $M_2(x,\mathrm{d}z) = \mathbb{1}_{|z|<1}|z|^{-d-1}\mathrm{d}z$, and

$$I_2\varphi = \sum_{n=1}^{\infty} \lambda_n[\varphi(\cdot + \zeta_n) - \varphi - \zeta_n \cdot \nabla\varphi], \quad \text{with}$$
$$\lambda_n \geq 0, \quad \sum_{n=1}^{\infty} \lambda_n(1 + |\zeta_n|) = \infty, \quad \sum_{n=1}^{\infty} \lambda_n|\zeta_n|^2 < \infty.$$

where the Lévy kernel $M_2(x,\cdot) = \sum_{n=1}^{\infty} \lambda_n \delta(\cdot - \zeta_n)$. Notice that in all examples, the Lévy kernels $M(x,\mathrm{d}z)$ are independent of x.

Working with operators of the type (2.55) or (2.56), we see that the Lévy kernel $M_1(\cdot,\mathrm{d}z)$ and $M_2(\cdot,\mathrm{d}z)$ can be approximated by bounded kernels of the form

$$M_{i,\varepsilon}(\cdot,\mathrm{d}z) = \mathbb{1}_{|z|\geq\varepsilon} M_i(\cdot,\mathrm{d}z), \quad i = 1,2.$$

We see that as ε goes to 0, the integro-differential operators (2.55) or (2.56) are limits of bounded non-local operators of the type (2.57).

Definition 2.53 (order γ)**.** We say that an integro-differential operator $I = I_\gamma$ is (a) of order $\gamma = 0$ (or bounded) if

$$I\varphi = \int_{\mathbb{R}_*^d} [\varphi(\cdot + z) - \varphi]M(\cdot,\mathrm{d}z), \quad \text{with} \int_{\mathbb{R}_*^d} M(\cdot,\mathrm{d}z) < \infty,$$

(b) of order γ in $(0,1]$ if

$$I\varphi = \int_{\mathbb{R}_*^d} [\varphi(\cdot + z) - \varphi]M(\cdot,\mathrm{d}z),$$
$$\text{with} \int_{|z|<1} |z|^\gamma M(\cdot,\mathrm{d}z) + \int_{|z|\geq 1} M(\cdot,\mathrm{d}z) < \infty,$$

(c) of order γ in $(1,2]$ if

$$I\varphi = \int_{\mathbb{R}_*^d} [\varphi(\cdot + z) - \varphi - z \cdot \nabla\varphi \mathbb{1}_{|z|<1}]M(\cdot,\mathrm{d}z),$$
$$\text{with} \int_{|z|<1} |z|^\gamma M(\cdot,\mathrm{d}z) + \int_{|z|\geq 1} M(\cdot,\mathrm{d}z) < \infty.$$

In all cases, γ is also referred to as the order of the Lévy kernel $M(\cdot,\mathrm{d}z) = M_\gamma(\cdot,\mathrm{d}z)$. □

Notice that the order γ of an integro-differential operator does not (completely) characterize the behaviour of the singularity of the Lévy kernel $M(\cdot, \mathrm{d}z)$. Actually, the most significant values are $\gamma = 0$ (where the operator is bounded), $\gamma = 1$ (where the expression used to define operator changes), and in general $\gamma = 2$. The use of "order" of the operator may be questionable, since an operator of order γ is also an operator of order γ', for any $\gamma \leq \gamma' \leq 2$. For the sake of simplicity we use the expression "of order γ" instead of "of order at most γ".

It is hard to track the dependency on the variable x of the Lévy kernel $M(x, \mathrm{d}z)$ to ensure that the integro-differential operator I acts on Lebesgue (Sobolev) and Hölder spaces. We will make precise how the variable x intervenes on the Lévy kernel $M(x, \mathrm{d}z)$, allowing enough flexibility to include modulation of the amplitude (or intensity) of jumps (well adapted for stochastic differential equations, see Gikhman and Skorokhod [99, p. 215]) and the density (or size) of jumps (better adapted for the martingale problem theory, see Bensoussan and Lions [17, p. 251]).

A priori the integro-differential operator is defined for functions $\varphi(x)$, with x in the whole space $\mathbb{R}^d$. However, we want to consider equations on a domain $\overline{\Omega}$ of $\mathbb{R}^d$, with either Dirichlet or Neumann boundary conditions, and even with oblique boundary conditions. We then need to localize the operator into $\overline{\Omega}$, e.g., by extending the data φ onto $\mathbb{R}^d \setminus \overline{\Omega}$. Thus $I\varphi$ becomes $I\tilde{\varphi}$, where $\tilde{\varphi}$ is a suitable extension of φ (defined only on $\overline{\Omega}$) to the whole space $\mathbb{R}^d$. The extension depends on the boundary value problem under consideration, which has a probabilistic interpretation. For instance, it is natural to use the zero-extension to study homogeneous Dirichlet boundary conditions. This corresponds to stopping the diffusion process with jumps (in the whole space $\mathbb{R}^d$) at the first exit time of the domain $\overline{\Omega}$. It is clear that the zero-extension will present some extra difficulties, e.g., if φ belongs to $W_0^{1,p}(\Omega) \cap W^{2,p}(\Omega)$ then the zero-extension $\tilde{\varphi}$ belongs to $W_0^{1,p}(\mathbb{R}^d)$ but in general, it is not an element of $W^{2,p}(\mathbb{R}^d)$.

As seen later, to treat the homogeneous Neumann (or oblique) boundary conditions, we will use a condition on the jumps (namely, no jumps outside of $\overline{\Omega}$) that will make the extension unnecessary, i.e., any extension $\tilde{\varphi}$ of φ will produce the same value for $I\tilde{\varphi}$.

2.11.1 The Epsilon-estimates

We need to describe the dependency of the variable x in the Lévy kernel $M(x, \mathrm{d}z)$. Suppose that there exist a σ-finite measure space $(F, \mathcal{F}, \pi)$, two Borel measurable functions $j(x, \zeta)$ and $m(x, \zeta)$ from $\mathbb{R}^d \times F$ into $\mathbb{R}^d_*$ and $[0, \infty)$, respectively, such that

$$M(x, A) = \int_{\{\zeta : \mathtt{j}(x,\zeta) \in A\}} m(x, \zeta)\pi(\mathrm{d}\zeta), \tag{2.58}$$

for any Borel measurable subset A of $\mathbb{R}^d_*$. The functions $\mathtt{j}(x, \zeta)$ and $m(x, \zeta)$ are called the *jump size (or amplitude)* and the *jump density (or intensity)*, respectively. The conditions (2.55), (2.56) or (2.57) on the singularity at the origin of the Lévy kernel $M(x, \mathrm{d}z)$ will be assumed to hold uniformly in x, so

that for some measurable function $\bar{\jmath}(\zeta)$ from F into $(0,\infty)$ and some constant $C_0 > 0$ we have

$$\begin{cases} |\mathtt{j}(x,\zeta)| \le \bar{\jmath}(\zeta), \qquad 0 \le \mathtt{m}(x,\zeta) \le 1, \\ \displaystyle\int_{\{\bar{\jmath}<1\}} [\bar{\jmath}(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{\bar{\jmath}\ge 1\}} \pi(\mathrm{d}\zeta) \le C_0, \end{cases} \tag{2.59}$$

where $0 \le \gamma \le 2$ is the order of the Lévy kernel. Actually, we may allow $0 \le m(x,\zeta) \le C$ if we re-define the measure $\pi(\mathrm{d}\zeta)$.

Thus, for any smooth function φ the integro-differential operator has the form

$$\begin{cases} I\varphi &= \displaystyle\int_F [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta) = \\ &= \displaystyle\int_0^1 \mathrm{d}\theta \int_F \mathtt{j}(\cdot,\zeta)\cdot\nabla\varphi(\cdot+\theta\mathtt{j}(\cdot,\zeta))\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta), \end{cases} \tag{2.60}$$

for $0 \le \gamma \le 1$ and

$$\begin{cases} I\varphi &= \displaystyle\int_{\{\bar{\jmath}<1\}} [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi - \mathtt{j}(\cdot,\zeta)\cdot\nabla\varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta) + \\ &\quad + \displaystyle\int_{\{\bar{\jmath}\ge 1\}} [\varphi(\cdot + \mathtt{j}(\cdot,\zeta)) - \varphi]\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta), \end{cases} \tag{2.61}$$

for $1 < \gamma \le 2$, where the first term can be rewritten as

$$\int_0^1 (1-\theta)\mathrm{d}\theta \int_{\{\bar{\jmath}<1\}} \mathtt{j}(\cdot,\zeta)\cdot\nabla^2\varphi(\cdot+\theta\mathtt{j}(\cdot,\zeta))\mathtt{j}(\cdot,\zeta)\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta).$$

In order to study this integro-differential operator as acting on Lebesgue (Sobolev) spaces, we will need to perform a change of variables. Assume that the jump amplitude function $\mathtt{j}(x,\zeta)$ is continuously differentiable in x for any fixed ζ, and that there exist a constant $c_0 > 0$ such that for any x, x' and $0 \le \theta \le 1$ we have

$$c_0|x-x'| \le |(x-x') + \theta[\mathtt{j}(x,\zeta) - \mathtt{j}(x',\zeta)]| \le c_0^{-1}|x-x'|. \tag{2.62}$$

This implies that the change of variables $X = x + \theta\mathtt{j}(x,\zeta)$ is a diffeomorphism of class C^1 in R^d, for any θ in $[0,1]$ and ζ in F. Moreover, the Jacobian of the transformation satisfies

$$c_1^{-1} \le \det[I_d + \theta\nabla\mathtt{j}(x,\zeta)] \le C_1, \tag{2.63}$$

for any x, ζ, θ and some constants $C_1, c_1 \ge 1$. Here I_d is the identity matrix in $\mathbb{R}^d$, $\nabla\mathtt{j}(x,\zeta)$ is the matrix of the first partial derivatives in x, and $\det[\cdot]$ denotes the determinant of a matrix.

In order to study the integro-differential operator in the Hölder space C^α, we also need Hölder continuity of the amplitude and density of jumps. For some exponent $0 < \alpha < 1$ we assume that there exist a measurable function (again

denoted by) $\bar{\jmath}(\cdot)$ from F into $(0,\infty)$ and some constant $M_0 > 0$ such that for any x, x' and ζ we have

$$\begin{cases} |\mathtt{j}(x,\zeta) - \mathtt{j}(x',\zeta)| \le \bar{\jmath}(\zeta)|x-x'|^\alpha, \\ |\mathtt{m}(x,\zeta) - \mathtt{m}(x',\zeta)| \le M_0|x-x'|^\alpha, \\ \displaystyle\int_{\{\bar{\jmath}<1\}} [\bar{\jmath}(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{\bar{\jmath}\ge 1\}} \pi(\mathrm{d}\zeta) \le M_0. \end{cases} \tag{2.64}$$

Let $\mathcal{O}$ be a bounded subset of $\mathbb{R}^d$ and set $\mathcal{O}_\varepsilon = \{x \in \mathbb{R}^d : x = y + z,\ y \in \mathcal{O},\ |z| < \varepsilon\}$. Due to the non-local character of the integro-differential operator I we need a function φ to be defined in a neighborhood of the closure $\overline{\mathcal{O}}$ to consider $I\varphi$ in $\mathcal{O}$. Thus, we define the support of I as the closed subset $\overline{\mathcal{O}}_I$ of $\mathbb{R}^d$, where

$$\overline{\mathcal{O}}_I = \overline{\bigcup\{x + \operatorname{supp} M(x,\cdot) \,:\, x \in \overline{\mathcal{O}}\}} \tag{2.65}$$

and $\operatorname{supp} M(x,\cdot)$ means the support of the Lévy kernel (or measure)

$$M(x,B) = \int_{\mathtt{j}(x,\zeta)\in B} \mathtt{m}(x,\zeta)\pi(\mathrm{d}\zeta)\ , \quad B \subset \mathbb{R}^d_* \text{ measurable Borel}\ .$$

Proposition 2.54 (ε-estimates)**.** *If the integro-differential operator I has the form (2.60) or (2.61), and conditions (2.59) and (2.62) are satisfied then for every $\varepsilon > 0$ there exists constants C and $C(\varepsilon)$ depending only on ε, the dimension d, the bounds C_0 and c_1 of conditions (2.59) and (2.63) such that*

$$\|I\varphi\|_{L^p(\mathcal{O})} \le C\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)}, \quad \text{if } \gamma = 0,$$

$$\|I\varphi\|_{L^p(\mathcal{O})} \le \varepsilon\|\nabla\varphi\|_{L^p(\mathcal{O}_\varepsilon)} + C(\varepsilon)\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)}, \quad \text{if } 0 < \gamma \le 1$$

and, if $1 < \gamma \le 2$, then

$$\|I\varphi\|_{L^p(\mathcal{O})} \le \varepsilon\|\nabla^2\varphi\|_{L^p(\mathcal{O}_\varepsilon)} + C(\varepsilon)\Big[\|\varphi\|_{L^p(\overline{\mathcal{O}}_I)} + \|\nabla\varphi\|_{L^p(\mathcal{O})}\Big],$$

for $1 \le p \le \infty$. Moreover, if we also assume the Hölder condition (2.64) on the coefficients, then the above estimates are valid with the C^α-norm instead of the L^p-norm, and in this case the constants C and $C(\varepsilon)$ depends also on the bounds M_0 of assumption (2.64). □

At this point, it should be clear that the integro-differential operator I is naturally non-local, i.e., we need to use functions defined on the whole space $\mathbb{R}^d$. So, a direct approach to consider I as acting on functions φ defined only on a (proper) domain Ω of $\mathbb{R}^d$, is to extend first φ to the whole space. Thus, denoting by $\tilde{\varphi}$ a suitable extension of φ, we have $I\varphi = I\tilde{\varphi}$, by definition. However, if we assume that

$$\mathtt{m}(x,\zeta) \ne 0 \quad \text{implies} \quad x + \theta\mathtt{j}(x,\zeta) \in \overline{\Omega},\ \forall\theta \in [0,1], \tag{2.66}$$

valid for any (x,ζ) in $\overline{\Omega}\times F$, then we see that the value $I\varphi = I\tilde{\varphi}$ is independent of the extension $\varphi \mapsto \tilde{\varphi}$ used. Indeed, notice that $I\tilde{\varphi}$ is always defined as the limit $I_\varepsilon\tilde{\varphi}$, where the Lévy kernel of I_ε is $M_\varepsilon(\cdot,\mathrm{d}z) = \mathbb{1}_{(|z|>\varepsilon)}M(\cdot,\mathrm{d}z)$. Condition (2.66) means that all jumps from $\overline{\Omega}$ are within $\overline{\Omega}$. Hence, under this condition (2.66), we may consider $I\varphi$ without any reference to the extension used for its proper definition (included estimates on its norm).

From the stochastic process viewpoint, some action should be taken when the jumps are outside of the region under consideration, e.g., we may stop or reflect the jumps, so that condition (2.66) will be eventually satisfied for the actual (or modified) jumps. So that in general, this will take care of Dirichlet, Neumann and oblique boundary conditions. However, for homogeneous Dirichlet boundary conditions problems, we have a natural *zero-extension* which corresponds to stopping the stochastic process at the first exit time of $\overline{\Omega}$. The problem with zero-extension is that a function φ in $W^{2,p}(\Omega)\cap W_0^{1,p}(\Omega)$ gives a zero-extension φ^o in $W^{1,p}(\mathbb{R}^d)$, but the first order derivative may be discontinuous across the boundary $\partial\Omega$. To overcome this difficulty, we need to impose some integrability conditions on the functions

$$\mathtt{m}_\Omega(x,\zeta) = \mathbb{1}_{(x+\mathtt{j}(x,\zeta)\notin\Omega)}\mathtt{m}(x,\zeta),$$

$$\mathtt{m}^1_\Omega(x,\zeta) = \mathtt{j}(x,\zeta)\mathtt{m}_\Omega(x,\zeta)$$

as seen below. Indeed, let denote by I_Ω the integro-differential operator corresponding to the density m_Ω, i.e. for a smooth function v in the whole space $\mathbb{R}^d$ and for $1<\gamma\le 2$ we have

$$\begin{aligned} I_\Omega v(x) &= \lim_{\varepsilon\to 0}\int_{\bar{\jmath}\ge\varepsilon}[v(x+j(x,\zeta))-v(x)]\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta) \\ &\quad -\lim_{\varepsilon\to 0}\int_{\varepsilon\le\bar{\jmath}<1}\nabla v(x+\mathtt{j}(x,\zeta))\cdot\mathtt{m}^1_\Omega(x,\zeta)\pi(\mathrm{d}\zeta). \end{aligned}$$

Thus, we can write $Iv = I_\Omega v + (I-I_\Omega)v$ and, for any x in Ω, the first term (i.e., $I_\Omega v$) reduces to only one integral and the second term vanishes if the function v vanishes in Ω. Hence, if φ^o and $\tilde{\varphi}$ are two extensions to the whole space of a given function in Ω (e.g., φ^o the zero-extension and $\tilde{\varphi}$ a smooth extension) then $v=\varphi^o-\tilde{\varphi}$ vanishes in Ω. Assuming v smooth (which may not be the case!) we may use the previous argument to see that $I\varphi^o = I\tilde{\varphi} + I_\Omega(\varphi^o-\tilde{\varphi})$. Thus, we have the following *localization* of the operator I, by imposing the above *vanishing* property for non-smooth functions.

Definition 2.55 (localization)**.** Let Ω be a bounded domain in $\mathbb{R}^d$ with smooth boundary, e.g., $C^{2+\alpha}$, and let I be the integro-differential operator given by (2.60) or (2.61) of order γ in $[0,2]$. For a smooth function φ defined on $\overline{\Omega}$ we denote by φ^o the zero-extension to whole space $\mathbb{R}^d$ and by $\tilde{\varphi}$ a smooth (say $C^{2+\alpha}$) extension. Under the conditions (2.59) we define the *localization* of I (to the domain Ω) as $I\varphi^o = I\tilde{\varphi} + I_\Omega(\varphi^o-\tilde{\varphi})$, where the operator I_Ω is given by

$$I_\Omega(\varphi^o-\tilde{\varphi}) = -\int_{\{\zeta\in F\,:\,x+\mathtt{j}(x,\zeta)\notin\Omega\}}\tilde{\varphi}(\cdot+\mathtt{j}(\cdot,\zeta))\mathtt{m}(\cdot,\zeta)\pi(\mathrm{d}\zeta),$$

with the above notation. □

If I has the form (2.61) of order γ in $(1,2]$, then we assume that for some γ_1 in $[1,\gamma]$ there exist a measurable function $\lambda_1(\zeta)$ and a constant $K_1>0$ such that for every x in Ω and ζ with $x+\mathtt{j}(x,\zeta)$ belonging to $\mathbb{R}^d\setminus\Omega$,

$$\begin{cases} |\mathtt{j}(x,\zeta)|\mathtt{m}(x,\zeta)\le \mathrm{d}^{1-\gamma_1}(x,\partial\Omega)\lambda_1(\zeta),\\ \displaystyle\int_{\bar\jmath<1}\lambda_1(\zeta)\pi(\mathrm{d}\zeta)\le K_1, \end{cases} \tag{2.67}$$

where $\mathrm{d}(x,\partial\Omega)$ denotes the distance from x to the boundary $\partial\Omega$, and $\bar\jmath(\zeta)$ is the function in (2.59). Notice that if x is in Ω but $x+\mathtt{j}(x,\zeta)$ is not in Ω then

$$|\mathtt{j}(x,\zeta)|\mathtt{m}(x,\zeta)=|\mathtt{j}(x,\zeta)|^\gamma|\mathtt{j}(x,\zeta)|^{1-\gamma}\mathtt{m}(x,\zeta)\le$$
$$\le|\mathtt{j}(x,\zeta)|^\gamma\mathrm{d}^{1-\gamma}(x,\partial\Omega)\mathtt{m}(x,\zeta).$$

The function $\lambda(\zeta)=\sup_x|\mathtt{j}(x,\zeta)|^\gamma m(x,\zeta)$ is bounded by $[\bar\jmath(\zeta)]^\gamma$, which is integrable in view of assumption (2.59). This show that condition (2.67) is always satisfied with $\gamma_1=\gamma$. This γ_1 in $[1,\gamma]$ is called the *boundary order* of I (and of the Lévy Kernel) with respect to the bounded domain Ω.

In Hölder spaces, we need to assume that the function $m_\Omega(x,\zeta)=m(x,\zeta)$ only if $x+\mathtt{j}(x,\zeta)\notin\Omega$ and zero otherwise, satisfies the following inequalities for any x, x' and ζ

$$\begin{cases} \displaystyle\int_F\big(\bar\jmath_\Omega(\zeta)\wedge1\big)^{1-\alpha}\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta)\le M_1,\\ \displaystyle\Big|\int_{\bar\jmath<1}\mathtt{j}(x,\zeta)\mathtt{m}_\Omega(x,\zeta)\pi(\mathrm{d}\zeta)-\\ \qquad\displaystyle-\int_{\bar\jmath<1}\mathtt{j}(x',\zeta)\mathtt{m}_\Omega(x',\zeta)\pi(\mathrm{d}\zeta)\Big|\le M_1|x-x'|^\alpha,\\ \displaystyle\int_F\big(\bar\jmath_\Omega(\zeta)\wedge1\big)|\mathtt{m}_\Omega(x,\zeta)-\mathtt{m}_\Omega(x',\zeta)|\pi(\mathrm{d}\zeta)\le M_1|x-x'|^\alpha, \end{cases} \tag{2.68}$$

where the function $\bar\jmath(\zeta)$ is as in assumption (2.59), $\bar\jmath_\Omega(\zeta)=\sup\{|\mathtt{j}(x,\zeta)|:x\in\Omega,\ x+\mathtt{j}(x,\zeta)\notin\Omega\}$, the constant M_1 is positive and the exponent α is the same as in condition (2.64).

We modified Proposition 2.54 as follows.

Proposition 2.56 (ε-loc-estimates)**.** *If the integro-differential operator I has the form (2.60) or (2.61), and conditions (2.59), (2.62) and (2.67) are satisfied then for any smooth function φ which vanishes on the boundary $\partial\Omega$ we have the following estimates:*
(1) *if $\gamma=0$ and $1\le p\le\infty$ then*

$$\|I\varphi\|_{L^p(\Omega)}\le C\|\varphi\|_{L^p(\Omega)},$$

(2) *if $0<\gamma\le1$ and $1\le p\le\infty$ then*

$$\|I\varphi\|_{L^p(\Omega)}\le\varepsilon\|\nabla\varphi\|_{L^p(\Omega)}+C(\varepsilon)\|\varphi\|_{L^p(\Omega)},$$

(3) *if* $1 < \gamma_1 \le \gamma \le 2$ *and* $1 \le p < d/(\gamma_1 - 1)$ *or if* $\gamma_1 = 1$ *and* $1 \le p \le \infty$ *then*

$$\|I\varphi\|_{L^p(\Omega)} \le \varepsilon \|\nabla^2 \varphi\|_{L^p(\Omega)} + C(\varepsilon)\Big[\|\nabla\varphi\|_{L^p(\Omega)} + \|\varphi\|_{L^p(\Omega)}\Big],$$

where $\varepsilon > 0$ *is arbitrary and the constant* C *and the function* $C(\varepsilon)$ *depend only on* d, γ_1, Ω *and the bounds in conditions* (2.59) *and* (2.62). *Moreover, if* $\gamma > 1 - \alpha$, (2.64) *and* (2.68) *are satisfied, then we have the following estimates:*
(1) *if* $\gamma = 0$ *then*

$$\|I\varphi\|_{C^\alpha(\overline{\Omega})} \le C\|\varphi\|_{C^\alpha(\overline{\Omega})},$$

(2) *if* $0 < \gamma \le 1$ *then*

$$\|I\varphi\|_{C^\alpha(\overline{\Omega})} \le \varepsilon \|\nabla\varphi\|_{C^\alpha(\overline{\Omega})} + C(\varepsilon)\|\varphi\|_{C^\alpha(\overline{\Omega})},$$

(3) *if* $1 < \gamma \le 2$ *then*

$$\|I\varphi\|_{C^\alpha(\overline{\Omega})} \le \varepsilon \|\nabla^2 \varphi\|_{C^\alpha(\overline{\Omega})} + C(\varepsilon)\Big[\|\nabla\varphi\|_{C^\alpha(\overline{\Omega})} + \|\varphi\|_{C^\alpha(\overline{\Omega})}\Big],$$

where $\varepsilon > 0$ *is arbitrary and the constant* C *and the function* $C(\varepsilon)$ *depend only on* d, Ω *and the bounds in conditions* (2.59), (2.64), (2.67) *and* (2.68). □

2.11.2 A Priori Estimates

The starting point of the L^p theory is a second order (uniformly) elliptic differential operator L of the form (2.24) and a boundary first order operator B of the form (2.25) where Ω is a domain with C^2 boundary and the coefficients satisfy (2.26). When $\Omega = \mathbb{R}^d$ the second-order coefficients are uniformly continuous (and bounded) and certainly, for Dirichlet boundary conditions, the boundary operator B is not necessary.

Consider $L - I$ as an unbounded operator in $L^p(\Omega)$, with either Dirichlet boundary conditions or B-oblique boundary conditions. Mixed boundary conditions can be used as long as the boundary $\partial\Omega$ is composed by two smooth (closed and disjointed) portions Γ and $\partial\Omega \setminus \Gamma$ on which Dirichlet and B-oblique boundary conditions are imposed. Unless Ω is the whole space $\mathbb{R}^d$, the integro-differential operator I need to be *localized* and assumptions should be such that the ε-estimates hold. For instance, besides hypotheses (2.59) and (2.62), if (homogeneous) Dirichlet boundary conditions are used then we need to impose also (2.67) with boundary order γ_1 such that $(\gamma_1 - 1)p < d$. However, for (homogeneous) B-oblique boundary conditions we need to impose (2.66), i.e., the localization is trivial since no jumps outside $\overline{\Omega}$ are allowed.

Set $A := I - L$, based on the ε-estimates of the Proposition 2.56 and the Agmon-Douglis-Nirenberg estimates (2.27) for (uniformly) elliptic differential operator L, we deduce that for any $1 < p < \infty$ there is a positive constant $C = C_p$ depending only on p, μ, the bounds of the coefficients a_{ij}, a_i, a_0, the

modulus of continuity of a_{ij}, the domain Ω and the bounds in the assumptions on I, such that

$$\begin{cases} \|u\|_{2,p} \le C\Big[\|Au\|_{0,p} + \|u\|_p\Big], \quad \forall u \in W^{2,p}(\Omega), \\ \text{satisfying either} \quad u = 0 \text{ on } \partial\Omega \quad \text{or} \quad Bu = 0 \text{ on } \partial\Omega, \end{cases} \tag{2.69}$$

where $\|\cdot\|_p$ is the norm in $L^p(\Omega)$, $W^{2,p}(\Omega)$ is the Banach (Sobolev) space of functions whose derivatives up to the 2 belong to L^p, with the natural norm $\|\cdot\|_{2,p}$. When $\Omega = \mathbb{R}^d$, the same a priori bounds hold for u in $W^{2,p}(\mathbb{R}^d)$.

Therefore, we deduce that for some constants $C_p > 0$ and $\omega_p > 0$ the following estimate holds

$$\|\partial_x^2 u\|_p + |\lambda|^{1/2}\|\partial_x u\|_p + |\lambda|\|u\|_p \le C_p\|Au - \lambda u\|_p, \tag{2.70}$$

for every u in $W^{2,p}(\Omega)$ satisfying either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, and for any λ with $\Re(\lambda) \ge \omega_p$. Hence, the (elliptic) integro-differential operator A with domain $\mathcal{D}(A)$ defined as the Sobolev space $W^{2,p}(\Omega)$ with one of the boundary conditions either $u = 0$ on $\partial\Omega$ or $Bu = 0$ on $\partial\Omega$, generates an analytic semigroup in $L^p(\Omega)$.

Once a priori estimates have been obtained, the above argument applies to Hölder space $C^\alpha(\overline{\Omega})$, $0 < \alpha < 1$ and to some extend to $C^1(\overline{\Omega})$, $C^0(\overline{\Omega})$, $L^1(\Omega)$ and $L^\infty(\Omega)$.

2.11.3 Maximum Principles

In order to apply the theory of Markov-Feller semigroups we need to establish the *maximum principle* for (elliptic) integro-differential operators. There are several versions (depending on regularity imposed on the solution) of the maximum principle valid for elliptic second-order differential operators of the form (2.24). Moreover, the type of maximum principle we need to obtain a Markov-Feller semigroup is of a global character and related to an equation of the form

$$\begin{cases} Iu - Lu = f \text{ in } \Omega \\ u = 0 \text{ on } \mathbb{R}^d \setminus \Omega, \end{cases} \tag{2.71}$$

and

$$\begin{cases} Iu - Lu = f \text{ in } \Omega \\ Bu = 0 \text{ on } \partial\Omega, \end{cases} \tag{2.72}$$

or even mixed boundary conditions. The maximum principle is formally stated as follows: Let u be a function satisfying (2.71) or (2.72) with $f \ge 0$ then $u \ge 0$. Certainly, the function space where u belongs and the assumptions on the coefficients of the operators L and I determine the meaning of the above equations.

The interested reader should consult the books Garroni and Menaldi [93, 94] for a comprehensive study on second-order integro-differential problems,

and Portenko [203] and Skubachevskii [231], among others, for more general boundary conditions.

For unbounded domains Ω an extra conditions of the type $a_0(x) \geq c > 0$ for some positive constant c is necessary to prevent ergodic situations. Really, to generate a Markov-Feller semigroup $S(t)$ satisfying $S(t)\mathbb{1} = \mathbb{1}$ for every $t \geq 0$ we need $a_0(x) = 0$, for any x, otherwise, we have a sub-Makovian Feller semigroup.

To conclude, let us mention that the analytic semigroup generated under the conditions of the previous section, is also a Feller-Markov semigroup in $C^0(\overline{\Omega})$.

2.12 Green and Poisson Functions

This is a short comment on (elliptic/parabolic) integro-differential operators with oblique boundary conditions as discussed in the books by Garroni and Menaldi [93, 94]. For instance, for a comprehensive analysis on the oblique boundary conditions for parabolic second-order differential equations we can see Tsuchiya [251, 252], and for Wentzell boundary conditions we can check Menaldi and Tubaro [177], and even more general type of boundary conditions can be found in the books Portenko [203] and Skubachevskii [231], among others.

The modern (analytic) semigroup theory is a powerful method to treat many problems. Perhaps a classic application is to study parabolic equations from elliptic equations, where starting from a priori (elliptic) estimates the whole theory of parabolic equations can be developed. For elliptic and parabolic equations there are (classic) direct arguments based on so called Green and Poisson functions. Essentially, the inverse of an integro-differential problem is a Fredholm operator of an integral type, and in the case of parabolic equations this is related with a Markov process. Actually, the density probability transition function of a Markov process is the Green functions and the so-called *local time* is related with the Poisson function.

Let $L-I$ be an elliptic integro-differential operator as in the previous section. Given three functions $f(x,t)$, $\psi(x,t)$ and $\varphi(x)$ defined for $x \in \Omega$ and $t \in [0,T]$, we consider the second order integro-differential parabolic equation

$$\begin{cases} \partial_t u + Lu = Iu + f \quad \text{in } \Omega \times (0,T]\,, \\ u = \varphi \quad \text{on } \Omega \times \{0\}\,, \\ Bu = 0 \quad \text{on } \partial\Omega \times [0,T]\,, \end{cases} \tag{2.73}$$

with *homogeneous* oblique boundary conditions, and

$$\begin{cases} \partial_t v + Lv = Iv \quad \text{in } \Omega \times (0,T]\,, \\ v = 0 \quad \text{on } \Omega \times \{0\}\,, \\ Bv = \psi \quad \text{on } \partial\Omega \times [0,T]\,, \end{cases} \tag{2.74}$$

with *non homogeneous* oblique boundary conditions.

Definition 2.57 (Green/Poisson function)**.** A measurable function $G(x,t,\xi)$ defined in $\overline{\Omega} \times (0,T] \times \Omega$ and locally integrable in (t,ξ) is called a *Green function*

for the parabolic second order integro-differential operator $L - I$ in Ω, with oblique boundary conditions given by first order differential operator B on $\partial\Omega$ if for any smooth functions $f(\xi,\tau)$ and $\varphi(\xi)$ with compact supports in $\Omega\times(0,T]$ and Ω, respectively, the potential function

$$u(x,t) = \int_0^t \mathrm{d}\tau \int_\Omega G(x,t-\tau,\xi) f(\xi,\tau)\mathrm{d}\xi + \int_\Omega G(x,t,\xi)\varphi(\xi)\mathrm{d}\xi$$

is either a *classic solution*, i.e., in the space $C^{2,1}(\Omega\times(0,T])\cap C^{1,0}(\partial\Omega\times(0,T])$ or a *strong solution*, i.e., in the space $W_p^{2,1}(\Omega\times(0,T))$ of the problem (2.73) with homogeneous oblique boundary conditions. Similarly, the *Poisson function* is a measurable function $P(x,t,\xi)$ defined in $\overline{\Omega}\times(0,T]\times\partial\Omega$ and locally integrable in (t,ξ) such that the potential function

$$v(x,t) = \int_0^t \mathrm{d}\tau \int_{\partial\Omega} P(x,t-\tau,\xi)\psi(\xi,\tau)\mathrm{d}\xi$$

is either a *classic solution*, i.e., in the space $C^{2,1}(\Omega\times(0,T])\cap C^{1,0}(\partial\Omega\times[0,T])$ or a *strong solution*, i.e., in the space $W_p^{2,1}(\Omega\times(0,T))$ of the problem (2.74) with non homogeneous oblique boundary conditions, for any smooth function $\psi(\xi,\tau)$ with a compact support in $\partial\Omega\times(0,T]$. □

The differential part of the Green function G_L is the piece of the Green function due to the differential operator L, i.e., the solution u of the equation

$$\begin{cases} \partial_t u + Lu = f \quad \text{in } \Omega\times(0,T]\,, \\ u = \varphi \quad \text{on } \Omega\times\{0\}\,, \\ Bu = 0 \quad \text{on } \partial\Omega\times[0,T]\,, \end{cases}$$

with *homogeneous* oblique boundary conditions, is given by the expression

$$u(x,t) = \int_0^t \mathrm{d}\tau \int_\Omega G_L(x,t-\tau,\xi) f(\xi,\tau)\mathrm{d}\xi + \int_\Omega G_L(x,t,\xi)\varphi(\xi)\mathrm{d}\xi\,,$$

for any smooth functions $f(\xi,\tau)$ and $\varphi(\xi)$ with compact supports in $\Omega\times(0,T]$ and Ω. Actually, in view of the estimates on G_L, the above representation formula remains valid for a more general class of functions, either in the Hölder space $C^{2+\alpha,1+\alpha/2}(\overline{\Omega}\times(0,T])$, $0<\alpha<1$ or in the Sobolev space $W_p^{2,1}(\Omega\times(0,T))$, $1<p<\infty$.

The following results are found in Solonnikov [232, 233] and Ivasišen [114]

Theorem 2.58. *Let Ω be a bounded domain in $\mathbb{R}^d$ with its boundary $\partial\Omega$ of class $C^{1,\alpha}$, with $0<\alpha<1$, and L and B be the operators as above, satisfying (2.26). Then the strong Green function for the parabolic second order differential operator $\partial_t + L$ in $\Omega\times(0,T]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega\times[0,T]$ exists and satisfies the following estimate*

$$|\nabla^\ell G_L(x,t,\xi)| \le C t^{-(d+\ell)/2}\exp(-c|x-\xi|^2/t)\,,$$

for every (x,t,ξ) in $\overline{\Omega}\times(0,T]\times\Omega$, for any $\ell=0,1,2$ and some positive constants C and c. Moreover, if the boundary $\partial\Omega$ is of class $C^{2,\alpha}$ and the lower order coefficients a_i are in $C^{\alpha}(\overline{\Omega})$ and the boundary coefficients b_i are in $C^{1+\alpha}(\partial\Omega)$, then G_L is the classic Green function and enjoys the estimates

$$\begin{cases} |\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(y,t,\xi)|\le M|x-y|^\alpha t^{-(d+\ell+\alpha)/2}\times \\ \qquad\qquad \times[\exp(-m|x-\xi|^2/t)+\exp(-m|y-\xi|^2/t)]\,, \\ |\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(x,s,\xi)|\le M|t-s|^{\alpha/2}\times \\ \qquad\qquad \times[t^{-(d+\ell+\alpha)/2}\exp(-m|x-\xi|^2/t)+ \\ \qquad\qquad +s^{-(d+\ell)+\alpha/2}\exp(-m|x-\xi|^2/s)]\,, \end{cases}$$

for every x,y in $\overline{\Omega}$, s,t in $(0,T]$ and ξ in Ω, and

$$\begin{cases} |\nabla^\ell G_L(x,t,\xi)-\nabla^\ell G_L(x,t,\eta)|\le M|\xi-\eta|^\alpha t^{-(d+\ell+\alpha)/2}\times \\ \qquad\qquad \times[\exp(-m|x-\xi|^2/t)+\exp(-m|x-\eta|^2/t)]|\,, \end{cases}$$

for every x in $\overline{\Omega}$, t in $(0,T]$ and ξ,η in Ω, for any $\ell=0,1,2$ and some positive constants C, c, M and m. In all estimates, the constants C, c, M and m depend only on the bounds imposed on the coefficients (of the differential operators L and B) throughout the various assumptions and, on the domain $\Omega\times(0,T]$. Estimates similar to the above hold for the Poisson function. □

Let G_L be the Green function associated with the differential operator L. To construct the Green function G associated with the integro-differential operator ∂_t+L-I, we solve a Volterra equation

$$\begin{cases} \text{either} & \text{find } Q_I \text{ such that} & Q_I=Q_L+Q_L\star Q_I\,, \\ \text{or} & \text{find } G \text{ such that} & G=G_L+G_L\star IG\,, \end{cases} \tag{2.75}$$

with the relations $Q_L=IG_L$ and $G=G_L+G_L\star Q_I$. Recall that the bullet $\star$ means the kernel-convolution, i.e., for any $\varphi(x,t,y,s)$ and $\psi(x,t,y,s)$

$$(\varphi\star\psi)(x,t,y,s):=\int_0^T \mathrm{d}\tau\int_{\mathbb{R}^d}\varphi(x,t,z,\tau)\psi(z,\tau,y,s)\mathrm{d}z,$$

and, in particular for any $\varphi(x,t,y)$ and $\psi(x,t,y)$,

$$(\varphi\star\psi)(x,t,y):=\int_0^t \mathrm{d}\tau\int_{\mathbb{R}^d}\varphi(x,\tau,z)\psi(z,t-\tau,y)\mathrm{d}z,$$

for every $t>0$, x and y in $\mathbb{R}^d$. Actually, we express Q_I as the following series

$$Q_I=\sum_{n=1}^{\infty}Q_n\,,\qquad Q_0=Q_L,\quad Q_n=Q_L\star Q_{n-1}\,,\ n\ge 1, \tag{2.76}$$

where the convergence is in the sense of following Green spaces.

To estimate the Green function of the integro-differential operator $\partial_t + L - I$ we consider a number of semi-norms used to define the Green function spaces in the domain $\Omega \times (0,T]$. For any kernel $\varphi(x,t,\xi)$, with $x,\xi \in \Omega$, $t \in (0,T]$, $k \geq 0$ and $0 < \alpha < 1$, we define

$$C(\varphi,k) = \inf\{C \geq 0 \ : \ |\varphi(x,t,\xi)| \leq Ct^{-1+(k-d)/2}, \ \forall x,t,\xi\}, \tag{2.77}$$

$$K(\varphi,k) = K_1(\varphi,k) + K_2(\varphi,k), \tag{2.78}$$

$$\left\{ \begin{array}{l} K_1(\varphi,k) = \inf\{K_1 \geq 0 : \\ \qquad\qquad : \displaystyle\int_\Omega |\varphi(x,t,\xi)|\mathrm{d}\xi \leq K_1 t^{-1+k/2}, \ \forall x,t\}, \\ K_2(\varphi,k) = \inf\{K_2 \geq 0 : \\ \qquad\qquad : \displaystyle\int_\Omega |\varphi(x,t,\xi)|\mathrm{d}x \leq K_2 t^{-1+k/2}, \ \forall t,\xi\}, \end{array} \right. \tag{2.79}$$

$$M(\varphi,k,\alpha) = M_1(\varphi,k,\alpha) + M_2(\varphi,k,\alpha) + M_3(\varphi,k,\alpha), \tag{2.80}$$

$$\left\{ \begin{array}{l} M_1(\varphi,k,\alpha) = \inf\{M_1 \geq 0 : |\varphi(x,t,\xi) - \varphi(x',t,\xi)| \leq \\ \qquad\qquad \leq M_1|x-x'|^\alpha t^{-1+(k-d-\alpha)/2}, \ \forall x,x',t\}, \end{array} \right. \tag{2.81}$$

$$\left\{ \begin{array}{l} M_2(\varphi,k,\alpha) = \inf\{M_2 \geq 0 : |\varphi(x,t,\xi) - \varphi(x,t',\xi)| \leq \\ \qquad\qquad \leq M_2|t-t'|^{\alpha/2}[t^{-1+(k-d-\alpha)/2} \vee t'^{-1+(k-d-\alpha)/2}], \\ \qquad\qquad\qquad\qquad\qquad\qquad \forall x,t,t',\xi\}, \end{array} \right. \tag{2.82}$$

$$\left\{ \begin{array}{l} M_3(\varphi,k,\alpha) = \inf\{M_3 \geq 0 : |\varphi(x,t,\xi) - \varphi(x,t,\xi')| \leq \\ \qquad\qquad \leq M_3|\xi-\xi'|^\alpha t^{-1+(k-d-\alpha)/2}, \ \forall x,t,\xi,\xi'\}, \end{array} \right. \tag{2.83}$$

$$\left\{ \begin{array}{l} N(\varphi,k,\alpha) = N_1(\varphi,k,\alpha) + N_2(\varphi,k,\alpha) + \\ \qquad\qquad + N_3(\varphi,k,\alpha) + N_4(\varphi,k,\alpha), \end{array} \right. \tag{2.84}$$

$$\left\{ \begin{array}{l} N_1(\varphi,k,\alpha) = \inf\{N_1 \geq 0 : \displaystyle\int_\Omega |\varphi(x,t,\xi) - \varphi(x',t,\xi)|\mathrm{d}\xi \leq \\ \qquad\qquad \leq N_1|x-x'|^\alpha t^{-1+(k-\alpha)/2}, \ \forall x,x',t,s\}, \end{array} \right. \tag{2.85}$$

$$\left\{ \begin{array}{l} N_2(\varphi,k,\alpha) = \inf\{N_2 \geq 0 : \displaystyle\int_\Omega |\varphi(x,t,\xi) - \varphi(x,t',\xi)|\mathrm{d}\xi \leq \\ \qquad\qquad \leq N_2|t-t'|^{\alpha/2}[t^{-1+(k-\alpha)/2} \vee t'^{-1+(k-\alpha)/2}], \\ \qquad\qquad\qquad\qquad\qquad\qquad \forall x,t,t'\}, \end{array} \right. \tag{2.86}$$

$$\begin{cases} N_3(\varphi,k,\alpha) = \inf\{N_3 \geq 0 : \displaystyle\int_\Omega |\varphi(x,t,\xi) - \varphi(x,t',\xi)|\mathrm{d}x \leq \\ \qquad \leq N_3|t-t'|^{\alpha/2}[t^{-1+(k-\alpha)/2} \vee t'^{-1+(k-\alpha)/2}], \\ \qquad\qquad\qquad\qquad\qquad\qquad \forall t,t',\xi\}, \end{cases} \tag{2.87}$$

$$\begin{cases} N_4(\varphi,k,\alpha) = \inf\{N_4 \geq 0 : \displaystyle\int_\Omega |\varphi(x,t,\xi) - \varphi(x,t,\xi')|\mathrm{d}x \leq \\ \qquad \leq N_4|\xi-\xi'|^{\alpha} t^{-1+(k-\alpha)/2},\ \forall t,\xi,\xi'\}, \end{cases} \tag{2.88}$$

$$R(\varphi,k,\alpha) = R_1(\varphi,k,\alpha) + R_2(\varphi,k,\alpha), \tag{2.89}$$

$$\begin{cases} R_1(\varphi,k,\alpha) = \inf\{R_1 \geq 0 : \\ \qquad : \displaystyle\int_\Omega |\varphi(Z,t,\xi) - \varphi(Z',t,\xi)|J_\eta(Z,Z')\mathrm{d}z \leq \\ \qquad \leq R_1\eta^\alpha t^{-1+(k-\alpha)/2},\ \forall Z,Z',t,\xi \text{ and } \eta>0\}, \end{cases} \tag{2.90}$$

$$\begin{cases} R_2(\varphi,k,\alpha) = \inf\{R_2 \geq 0 : \\ \qquad : \displaystyle\int_\Omega |\varphi(x,t,Z) - \varphi(x,t,Z')|J_\eta(Z,Z')\mathrm{d}z \leq \\ \qquad \leq R_2\eta^\alpha t^{-1+(k-\alpha)/2},\ \forall x,t,Z,Z' \text{ and } \eta>0\}, \end{cases} \tag{2.91}$$

where the change of variables $Z(z)$ and $Z'(z)$ are diffeomorphisms of class C^1 in $\mathbb{R}^d$, and the Jacobian

$$J_\eta(Z,Z') = |\det(\nabla Z)| \wedge |\det(\nabla Z')| \tag{2.92}$$

if $|Z-Z'| \leq \eta$ and Z, Z' belong to $\overline{\Omega}$, and vanishing otherwise, here $\det(\cdot)$ means the determinant of a $d\times d$ matrix, ∇Z, $\nabla Z'$ stand for the matrices of the first partial derivatives of $Z(z), Z'(z)$ with respect to the variable z, and $\wedge, \vee$ denote the minimum, maximum (resp.) between two real numbers.

Definition 2.59 (Green function spaces)**.** Let us denote by $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$ (or $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}(\overline{\Omega}\times(0,T],\mathbb{R}^n)$ when necessary), $k \geq 0$, $n \in \mathbb{N}$ and $0<\alpha<1$, the space of all continuous functions (or kernels) $\varphi(x,t,\xi)$ defined for x,ξ in $\Omega \subset \mathbb{R}^d$ and $0 < t \leq T$, with values in $\mathbb{R}^n$ (usually $n=1$ and $k \geq 0$) and such that the above infima (semi-norms) (2.77),..., (2.91) (of order k) are finite. Thus the maximum of the quantities (2.77),..., (2.91), denoted by $[\![\cdot]\!]_{k,\alpha} = [\![\cdot]\!]_{\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}}$, is the norm of the Banach space $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$. When $\alpha = 0$, we denote by $\mathcal{G}_k^0$ (or $\mathcal{G}_k^0(\overline{\Omega}\times(0,T],\mathbb{R}^n)$ when necessary), $k \geq 0$, and $n \in \mathbb{N}$, the space of all measurable functions (or kernels) $\varphi(x,t,\xi)$ defined for x,ξ in $\Omega \subset \mathbb{R}^d$ and $0<t\leq T$, with values in $\mathbb{R}^n$ (usually $n=1$ and $k\geq 0$) and such that the two infima (2.77) and (2.78) (of order k) are finite, with the norm $[\![\cdot]\!]_{k,0} = [\![\cdot]\!]_{\mathcal{G}_k^0}$. □

The Volterra equations (2.75) is solved in a Green function space $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$. We have

Theorem 2.60 (Green function)**.** *Under suitable conditions on the coefficients as discussed above, and in particular if the boundary coefficients b_i belongs to $C^{1+\alpha}(\partial\Omega)$, for any $i = 1, \dots, d$, then there exists the (strong) Green function $G(x,t,\xi)$ for the parabolic second order integro-differential operator $\partial_t + L - I$ in $\Omega \times (0,T]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega \times [0,T]$. Moreover $G = G_L + G_L \star Q$, where Q is the solution of the Volterra equation (2.75) in the Green function space $\mathcal{G}^0_{2-\gamma}$, given by (2.76) with $Q_0 = IG_L$ and the semi-norms $C(\nabla^\ell G_L \star Q, 4-\ell-\gamma)$, $K(\nabla^\ell G_L \star Q, 4-\ell-\gamma)$, $M_i(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $i = 1,2$, $N_i(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $i = 1,2,3$ and $R_1(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$ are finite, for $\ell = 0, 1$. Furthermore, if we assume Hölder continuous coefficients then $G(x,t,\xi)$ is also the classic Green function and solution of the Volterra equation (2.75) in the Green function space $\mathcal{G}^{\alpha,\frac{\alpha}{2}}_{2-\gamma}$. In this case, the semi-norms $M_3(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, $N_4(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha$, $R_2(\nabla^\ell G_L \star Q, 4-\ell-\gamma, \alpha)$, for $\ell = 0, 1$, and the semi-norms $C(\nabla^2 G_L \star Q, 2-\gamma)$, $K(\nabla^2 G_L \star Q, 2-\gamma)$, $M_2(\nabla G_L \star Q, 2-\gamma, 2\alpha)$, $M(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$, $N(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$, $N_i(\nabla G_L \star Q, 2-\gamma, 2\alpha)$, $i = 2,3$ and $R(\nabla^2 G_L \star Q, 2-\gamma, \alpha)$ are finite.* □

If $G(x,t,\xi)$ and $P(x,t,\xi)$ are the Green function and the Poisson kernel, respectively, then any smooth solution of the following (parabolic, differential) boundary value problem

$$\begin{cases} \partial_t u + Lu = Iu + f & \text{in } \Omega \times (0,T]\,, \\ u = \varphi & \text{on } \Omega \times \{0\}\,, \\ Bu = \psi & \text{on } \partial\Omega \times [0,T]\,, \end{cases}$$

is given by the expression

$$\begin{aligned} u(x,t) = &\int_0^t \mathrm{d}\tau \int_\Omega G(x,t-\tau,\xi) f(\xi,\tau)\mathrm{d}\xi + \\ &+ \int_\Omega G(x,t,\xi)\varphi(\xi)\mathrm{d}\xi + \\ &+ \int_0^t \mathrm{d}\tau \int_{\partial\Omega} P(x,t-\tau,\xi)\psi(\xi,\tau)\mathrm{d}\xi\,, \end{aligned}$$

and the *Chapman-Kolmogorov identity*

$$G(x,t+s,\xi) = \int_\Omega G(x,t,y)G(y,s,\xi)\mathrm{d}y\,,$$

for every x, ξ in Ω and t, s in $(0,T]$ is satisfied. In particular for $f = a_0$, $\varphi = 1$ and $\psi = b_0$ we obtain

$$\begin{aligned} 1 - \int_\Omega G(x,t,\xi)\mathrm{d}\xi \quad = \quad &\int_0^t \mathrm{d}\tau \int_\Omega G(x,t-\tau,\xi) a_0(\xi)\mathrm{d}\xi + \\ &+ \int_0^t \mathrm{d}\tau \int_{\partial\Omega} P(x,t-\tau,\xi) b_0(\xi)\mathrm{d}\xi\,. \end{aligned}$$

In particular, if $a_0 = 0$ and $b_0 = 0$ then

$$\int_\Omega G(x,t,\xi)\mathrm{d}\xi = 1 \qquad \forall\,(x,t) \in \Omega \times (0,T]\,,$$

which is one of the key property of a *transition density* function, used to describe *Markov processes*. The *weak maximum principle* implies that $G \geq 0$ and sometime the *strong maximum principle* yields the strictly positivity of the Green (and Poisson) functions.

All the above estimates are valid on $\Omega \times [0,T]$ for any $T > 0$. In an unbounded time interval we have the following

Theorem 2.61 (time-unbounded)**.** *Let $G(x,t,\xi)$ be the Green function for the parabolic second order integro-differential operator $\partial_t + L - I$ in $\Omega \times (0,\infty]$, with oblique boundary conditions given by first order differential operator B on $\partial\Omega \times [0,\infty]$ as given by* Theorem 2.60*. Then we have the following estimates: for every $\delta > 0$ there exist positive constants C_0, M_0 such that for any $t, t' \geq \delta$*

$$|\nabla^\ell G(x,t,\xi)| \leq C_0, \quad \ell = 0,1,2, \tag{2.93}$$

$$\begin{cases} |\nabla^\ell G(x,t,\xi) - \nabla^\ell G(x',t',\xi')| \leq M_0\big(|x-x'|^\alpha + \\ \qquad\qquad +|t-t'|^{\alpha/2} + |\xi-\xi'|^\alpha\big), \end{cases} \tag{2.94}$$

for any $\ell = 0,1$, x, ξ and x', ξ' in $\overline{\Omega}$. Moreover, if we assume Hölder continuous coefficients then for any $\delta > 0$ there exists a positive constant $c = c(\delta) > 0$ such that

$$G(x,t,\xi) \geq c, \qquad \forall (x,t,\xi) \in \overline{\Omega} \times (\delta,\infty] \times \Omega\,, \tag{2.95}$$

we also have the estimate

$$|\nabla G(x,t,\xi) - \nabla G(x,t',\xi)| \leq M_0\,|t-t'|^\alpha, \tag{2.96}$$

and we may let $\ell = 2$ in estimate of (2.94)*.* □

• *Remark* 2.62. Notice that from the technique used in proving of the above Theorem 2.61 we can estimate the constants C_0 and M_0 appearing in (2.93), (2.94) and (2.96) as t, t' become large, i.e., if we define

$$q(t) := \sup_x \int_\Omega G(x,t,y)\mathrm{d}y, \quad t \geq 1 \tag{2.97}$$

then we have for any $t' \geq t \geq T$ the estimates

$$C_0 \leq C_G\, q(T), \quad \text{and} \quad M_0 \leq M_G\, q(T), \quad T \geq 1 \tag{2.98}$$

where the constants C_G and M_G depend on the semi-norms $K_2(\nabla^\ell G, 2-\ell)$, $K_2(G,2)$, $M_1(\nabla^\ell G, 2-\ell, \alpha)$, $M_3(\nabla^\ell G, 2-\ell, \alpha)$ and $N_3(\nabla^\ell G, 2-\ell, \alpha)$, but all on the time interval $[0,1]$. This means that estimates for the Green function $G(x,t,\xi)$ on the Green spaces in the time interval $[0,\infty)$ are obtained from estimates on any bounded time interval $[0,\delta]$, with $\delta > 0$ plus a bound on the expression (2.97) of $q(t)$ as t becomes large. □

In all theses estimates applied to integro-differential (or Lévy-type) operators with a dominant second order differential part. Other situation can be found in Jacob [115, Vol II, Section 2.7, pp. 138–151], Kolokoltsov [134], Komatsu [135, 136], Mikulevicius and Pragarauskas [181, 182] and Mikulevicius and Rozovskii [183], among others.

2.13 Examples of Transition Functions

Green and transition functions are essentially the same objects, one is seen as the inverse of a functional operator (e.g., an integro-differential operator) and the other is the essence of a Markov-Feller processes.

Let us start with a couple of simple one-dimensional prototypes first in the whole real line and with boundary conditions in the real semi-line. First recall that given a locally compact separable complete metric space E, we define $C_0(E)$ as the Banach space of all continuous real functions on E vanishing at infinity, i.e., $f : E \to \mathbb{R}$, continuous and for any $\varepsilon > 0$ there exists a compact subset K of E such that $|f(x)| < \varepsilon$ for any x in $E \setminus K$. Note that we are using indistinctly $p(t, x, B)$ or $p(x, t, B)$ for the transition functions.

2.13.1 One-Dimensional

Example 2.1 (Wiener process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) := \frac{1}{\sqrt{2\pi t}} \Big\{ \int_B \exp\big[- \frac{(y-x)^2}{2t} \big] \mathrm{d}y \Big\},$$

for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the typical one-dimensional Brownian motion or Wiener process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$\begin{aligned} S(t)f(x) &:= \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x + \sqrt{t}z)\, \exp\big(- \frac{z^2}{2} \big) \mathrm{d}z, \end{aligned}$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator A is the differential operator A

$$\mathcal{D}(A) := \{ f \in C_0(\mathbb{R}) \cap C^2(\mathbb{R}) : f'' \in C_0(\mathbb{R}) \}, \quad Af := \frac{1}{2} f''.$$

The associated resolvent operator in $C_0(\mathbb{R})$ is given by

$$\begin{aligned} R(\lambda)f(x) &:= \int_{\mathbb{R}} f(y)\, r(\lambda, x, \mathrm{d}y) = \\ &= \frac{1}{\sqrt{2\lambda}} \int_{-\infty}^{+\infty} f(x + \frac{z}{\sqrt{2\lambda}})\, \exp\big(- |z| \big) \mathrm{d}z, \end{aligned}$$

where the resolvent kernel is

$$r(\lambda, x, B) := \frac{1}{\sqrt{2\lambda}} \int_B \exp\big(-\sqrt{2\lambda}|x-y|\big)\mathrm{d}y.$$

for every $\lambda > 0$, x in $\mathbb{R}$, and B in $\mathcal{B}$. A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$ and a realization with continuous paths can be constructed. □

Example 2.2 (Poisson process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t, x, B) := \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \mathbb{1}_B(x+k),$$

for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the typical one-dimensional Poisson process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} f(x+k),$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator is

$$\mathcal{D}(A) := C_0(\mathbb{R}), \qquad Af(x) := c[f(x+1) - f(x)], \quad \forall x \in \mathbb{R},$$

Notice that A is a nonlocal operator and that only a cad-lag realization of the above Poisson process can be constructed. We can generalize this example to a compound Poisson process $(P_t : t \geq 0)$, with parameters (c, μ), where $c > 0$ and μ is a probability distribution on $\mathbb{R}$. The probability transition function is

$$p(t, x, B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \mu^k(B), \qquad \mu^0 = \delta_0, \quad \text{and}$$
$$\mu^k(B) = (\mu^{k-1} \star \mu)(B) = \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for $k = 1, 2, \ldots$, for any $t > 0$, x in $\mathbb{R}$ and B in $\mathcal{B}$, where δ_0 is the Dirac measure at the origin. Since μ^k are all probability measures, the above series converges. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \\ = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}} f(x+y)\, \mu^k(\mathrm{d}y),$$

for every $t > 0$ and x in $\mathbb{R}$. Its infinitesimal generator is the bounded (integral) linear operator on $C_0(\mathbb{R})$, defined by

$$Af(x) := c \int_{\mathbb{R}} [f(x+y) - f(x)]\, \mu(\mathrm{d}y), \quad \forall x \in \mathbb{R}.$$

Again, only a cad-lag realization of the above Poisson process can be constructed. □

Example 2.3 (Cauchy process)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) := \frac{1}{\pi}\int_B \frac{t}{t^2+(y-x)^2}\mathrm{d}y,$$

for any $t>0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x+tz)\,\frac{1}{1+z^2}\,\mathrm{d}z,$$

for every $t>0$ and x in $\mathbb{R}$. Its infinitesimal generator is of the form

$$Af(x) := \frac{1}{\pi}\int_0^{\infty} \frac{f(x+y)+f(x-y)-2f(x)}{y^2}\mathrm{d}y, \quad \forall x\in\mathbb{R},$$

and the domain $\mathcal{D}(A)$ contains all twice-differentiable functions with compact support in $\mathbb{R}$. Notice that A is a nonlocal operator and that only a cad-lag realization of the above Cauchy process can be constructed. □

Example 2.4 (Wiener-Poisson)**.** On the state space $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t,x,B) := \mathrm{e}^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{k!}\frac{1}{\sqrt{2\pi t}}\Big\{\int_B \exp\big[-\frac{(x+k-y)^2}{2t}\big]\mathrm{d}y\Big\},$$

for any $t>0$, x in $\mathbb{R}$ and B in $\mathcal{B}$. This is the sum of independent Wiener and Poisson processes. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$\begin{aligned} S(t)f(x) &:= \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \\ &= \mathrm{e}^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^k}{\sqrt{2\pi}\,k!}\int_{-\infty}^{\infty} f(x+k+\sqrt{t}z)\,\exp\big(-\frac{z^2}{2}\big)\mathrm{d}z, \end{aligned}$$

for every $t>0$ and x in $\mathbb{R}$. Its infinitesimal generator A is the closure of the (closable) operator $\mathring{A}$

$$\begin{aligned} &\mathcal{D}(\mathring{A}) := C_0^0(\mathbb{R})\cap C^2(\mathbb{R}),\\ &\mathring{A}f(x) := \frac{1}{2}f''(x) + c[f(x+1)-f(x)], \quad \forall x\in\mathbb{R}, \end{aligned}$$

Only a cad-lag realization can be constructed as $(x+P_t+W_t : t\geq 0)$, where $(W_t : t\geq 0)$ is a standard Wiener process independent of the Poisson process $(P_t : t\geq 0)$. We can generalize this example to a $(W_t : t\geq 0)$ Wiener process (with drift b and covariance σ^2) and a $(P_t : t\geq 0)$ compound Poisson processes (with parameters (c,μ)), independent of each other. Thus b is a real constant,

$\sigma\,, c > 0$, and μ is a probability distribution on $\mathbb{R}$. Again, a cad-lag realization is given by $(X_t = x + W_t + P_t : t \geq 0)$ and the probability transition function is

$$p(t,x,B) = \int_B \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!}\, p_k(t, x-y)\mathrm{d}y,$$

$$p_0(t,x) = \Big\{ \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \exp\big[-\frac{(x+bt-y)^2}{2t}\big]\, \mu(\mathrm{d}y)\Big\},$$

$$p_k(t,x) = \int_{\mathbb{R}} p_k(t, x-y)\, \mu(\mathrm{d}y), \quad k = 1, 2, \ldots,$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. Again, since p_0 is a probability density and μ^k is a probability measure the above series converges. Notice that if μ^k denotes the k convolution as defined in Example 2.2, then p_k can be expressed as $p_w \star \mu^k$, where p_w is the probability density transition function of a Wiener process. The associated semigroup in $C_0(\mathbb{R})$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times$$
$$\times \int_{\mathbb{R}} \mu^k(\mathrm{d}y) \int_{-\infty}^{\infty} f(x+y+\sqrt{t}z)\, \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$. The infinitesimal generator is the closure of the (closable) integro-differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) := C_0^0(\mathbb{R}) \cap C^2(\mathbb{R}),$$
$$\mathring{A}f(x) := \frac{1}{2}\sigma f''(x) + bf'(x) + c\int_0^{\infty} [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

for every x in $\mathbb{R}$. Again, notice the nonlocal character of this unbounded operator. □

Example 2.5 (reflecting barrier)**.** On the state space $\mathbb{R}_0^+ := [0, \infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) :=$$
$$:= \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[-\frac{(y-x)^2}{2t}\big] + \exp\big[-\frac{(y+x)^2}{2t}\big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. This represents Brownian motion with reflecting barrier at $x = 0$ and the process itself can be constructed as $(X_t = |x + W_t| : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process in $\mathbb{R}$. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}_0^+} f(y)\, p(t,x,\mathrm{d}y) =$$
$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \check{f}(y)\, \exp\big[-\frac{(y-x)^2}{2t}\big]\mathrm{d}y,$$

where $\check{f}(y) := f(y)$ if $y \geq 0$ and $\check{f}(y) := f(-y)$ if $y \leq 0$, for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0\},$$
$$Af := \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$ and $X_t := |x + bt + W_t|$. □

The reflected Brownian motion of above Example 2.5 can also be constructed by means of local time as follows. First, for an given $x \geq 0$ we define τ_x the first exit time of the open region $(0, \infty)$, i.e., $\tau_x := \inf\{t \geq 0 : x + W_t \geq 0\}$. Next, we look at its running maximum, i.e., $M_t := \max\{x + W_s : \tau_x \leq s \leq t\}$, which except for a factor $1/2$ is called the *local time* of $(x + W_t : t \geq 0)$ at the origin. It can be proved that the process $(|x + W_t| : t \geq 0)$ has the same law as the process $(x + M_t - W_t : t \geq 0)$, which gives another realization of the reflected Brownian motion.

Example 2.6 (absorbing barrier)**.** On the state space $\mathbb{R}^+ := (0, \infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) :=$$
$$:= \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[-\frac{(y-x)^2}{2t} \big] - \exp\big[-\frac{(y+x)^2}{2t} \big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $x > 0$ and B in $\mathcal{B}$. This represents Brownian motion with absorbing barrier at 0, i.e., the Brownian motion particle *dies* at the first time when it hits the boundary $\{0\}$. The process itself can be constructed by stopping (or killing) the process $x + W_t$ at the first instant τ_x when it hits the boundary $\{0\}$, where W_t is a standard Wiener process in $\mathbb{R}$, i.e.,

$$\tau_x := \inf\{t > 0 : x + W_t = 0\},$$
$$X_t := x + W_t \quad t < \tau_x,\ x > 0,$$

thus $(X_t : t \geq 0)$ is the Brownian motion with initial value x at time $t = 0$ and absorbed (or otherwise said killed) at the origin. τ_x is the lifetime of the process X_t. Often we introduce an extra point (indicated by $\triangle$, or ∞ or ∂) to the state space $\mathbb{R}^+ := (0, \infty)$, called the *coffin* state, and defines X_t for all times by $X_t = \triangle$ for $t \geq \tau_x$. Its associated semigroup in $C_0(\mathbb{R}^+)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}^+} f(y)\, p(t, x, \mathrm{d}y) =$$
$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \hat{f}(y)\, \exp\big[-\frac{(y-x)^2}{2t} \big] \mathrm{d}y,$$

where $\hat{f}(y) := f(y)$ if $y \geq 0$ and $\hat{f}(y) := -f(-y)$ if $y \leq 0$, for every $t > 0$ and

$x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}^+) \cap C^2(\mathbb{R}^+) : f'' \in C_0(\mathbb{R}^+)\},$$
$$Af := \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$. Notice that $p(t, x, \mathbb{R}^+) < 1$ so that the process X_t dies, i.e., it does hit the boundary $x = 0$ in a finite time. The semigroup $S(t)$ may be extended to the space of continuous and bounded function in $[0, \infty)$, where $S(t)\mathbb{1} = 0$ for all $t > 0$. Moreover, the Banach space $C_0(\mathbb{R}^+)$ includes a vanishing boundary condition at infinity and also at $x = 0$, so that actually the condition $f(0) = 0$ in the definition of $\mathcal{D}(A)$ is redundant. Generally, we look at this as a process in the closure $[0, \infty)$ and we use the Banach space $C_0([0, \infty[)$ instead of $C_0(]0, \infty[)$. □

Sometimes we may use the complementary error function

$$\operatorname{Erfc}(x) := \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-v^2} \mathrm{d}v, \tag{2.99}$$

which satisfies $\operatorname{Erfc}(0) = 1$ and

$$\frac{2}{x + \sqrt{x^2 + 2}} \leq \sqrt{\pi}\, e^{x^2} \operatorname{Erfc}(x) \leq \frac{2}{x + \sqrt{x^2 + 1}}. \tag{2.100}$$

Indeed, by considering the functions

$$f(x) := \frac{1}{x + \sqrt{x^2 + 1}} - \mathrm{e}^{x^2} \int_x^\infty \mathrm{e}^{-y^2} \mathrm{d}y,$$
$$g(x) := \mathrm{e}^{x^2} \int_x^\infty \mathrm{e}^{-y^2} \mathrm{d}y - \frac{1}{x + \sqrt{x^2 + 2}},$$

which satisfy $f(0) = 1 - \frac{\pi}{2} > 0$ and $g(0) = \frac{\pi}{2} - \frac{1}{2} > 0$, we can estimate

$$\mathrm{e}^{x^2} \int_x^\infty \mathrm{e}^{-y^2} \mathrm{d}y \leq \frac{2}{x}\, \mathrm{e}^{x^2} \int_x^\infty 2y\, \mathrm{e}^{-y^2} \mathrm{d}y = \frac{1}{x},$$

so that $|f(x)| \leq \frac{2}{x}$ and $|g(x)| \leq \frac{2}{x}$. Calculations show that $f'(x) - 2xf(x) \leq 0$ and $g'(x) - 2xg(x) \leq 0$, for any $x \geq 0$, and the desired estimate (2.100) follows.

Example 2.7 (sticking barrier)**.** On the state space $\mathbb{R}_0^+ := [0, \infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) := \frac{1}{\sqrt{2\pi t}} \left(\int_B \Big\{ \exp\Big[-\frac{(y - x)^2}{2t} \Big] - \exp\Big[-\frac{(y + x)^2}{2t} \Big] \Big\} \mathrm{d}y \right) + \\ + \operatorname{Erfc}\Big(\frac{x}{\sqrt{2\pi t}} \Big)\, \mathbb{1}_B(0),$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. This represents Brownian motion with sticking barrier at $x = 0$, i.e., when the Brownian motion particle reaches $x = 0$ for the

first time, it sticks there forever. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f''(0) = 0\},$$
$$Af := \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$. Notice that comparing with the previous Example 2.6, now the state space $\mathbb{R}_0^+$ includes the barrier $x = 0$ and $p(t, x, \mathbb{R}_0^+) = 1$ for any $t > 0$ and x in $\mathbb{R}_0^+$. The semigroup takes the form

$$S(t)f(x) := \int_{\mathbb{R}_0^+} f(y)\, p(t, x, \mathrm{d}y) = \\ = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \hat{f}(y)\, \exp\big[- \frac{(y-x)^2}{2t}\big] \mathrm{d}y,$$

for every x in $\mathbb{R}_0^+$ and $t > 0$. Notice that the function $y \mapsto \hat{f}(y)$, where $\hat{f}(y) := f(y)$ if $y \geq 0$ and $\hat{f}(y) := 2f(0) - f(-y)$ if $y < 0$, is continuously differentiable whenever f is so. Thus, the function $x \mapsto S(t)f(x)$ can be defined as a smooth function, for every x in $\mathbb{R}$. The process itself can be constructed by stopping the process $x + W_t$ at the first instant τ_x when it hits the boundary $\{0\}$, where W_t is a standard Wiener process in $\mathbb{R}$, i.e.,

$$\tau := \inf\{t > 0 : x + W_t = 0\},$$
$$X_t := x + W_{t \wedge \tau_x} \quad t \geq 0,\ x > 0,$$

thus $(X_t : t \geq 0)$ is the Brownian motion with initial value x at time $t = 0$ and stopped at the origin. □

We may combine the reflecting barrier Example 2.5 and this sticking barrier to get a process where the domain of the infinitesimal generator $\mathcal{D}(A)$ has a boundary condition of the form $f'(0) - cf''(0) = 0$ with a positive constant c instead of just $f''(0) = 0$. This is called *sticky* barrier. Similarly, we may combine the reflecting barrier Example 2.5 with the absorbing barrier to get a process where the domain of the infinitesimal generator $\mathcal{D}(A)$ has a boundary condition of the form $f'(0) - cf(0) = 0$ with a positive constant c. This is called *elastic* barrier. The construction of a sticky Brownian motion or an elastic Brownian motion is more delicate, it starts with the reflecting Brownian motion and its local time at the barrier, the reader is referred to the classic book by Itô and McKean [113] for a complete analysis.

Example 2.8 (sticky Wiener)**.** On the state space $\mathbb{R}_0^+ := [0, +\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, y) := \frac{1}{\sqrt{2\pi t}} \Big(\exp\Big[- \frac{(y-x)^2}{2t}\Big] - \exp\Big[- \frac{(y+x)^2}{2t}\Big]\Big) + \\ + \delta(y)\, e^{\frac{t+2cx}{2c^2}} \operatorname{Erfc}\Big(\frac{t+cx}{c\sqrt{2t}}\Big) + \frac{1}{c}\, e^{\frac{t+2c(x+y)}{2c^2}} \operatorname{Erfc}\Big(\frac{t+c(x+y)}{c\sqrt{2t}}\Big),$$

for any $t > 0$, $x \geq 0$, $y \geq 0$. This represents a *slowly* reflecting Brownian motion on $[0, +\infty)$, i.e., when the Brownian motion particle reaches $x = 0$, it sticks there for some time. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = cf''(0)\},$$
$$Af := \frac{1}{2}f''.$$

We can visualize this process as a Brownian motion with a suitable time change, more specifically, starting with a standard Wiener process $x + W_t$ we have $X_t = x + W_{\ell_x(t)}$, where

$$\ell_x(t) = \inf\{s : A_s \leq t\}, \qquad A_t = \int_0^t \mathbb{1}_{\{x+W_s>0\}}\, \mathrm{d}s + c\, M_t,$$

with $M_t := \max\{x + W_s : \tau_x \leq s \leq t\}$ and $\tau_x := \inf\{t \geq 0 : x + W_t \geq 0\}$. □

Example 2.9 (elastic Wiener)**.** On the state space $\mathbb{R}_0^+ := [0, +\infty)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, y) := \frac{1}{\sqrt{2\pi t}}\Big(\exp\Big[-\frac{(y-x)^2}{2t}\Big] + \exp\Big[-\frac{(y+x)^2}{2t}\Big]\Big) - \\ -c\, \exp\Big(c(x+y) + \frac{c^2 t}{2}\Big)\, \mathrm{Erfc}\Big(\frac{x+y+ct}{\sqrt{2t}}\Big),$$

for any $t > 0$, $x \geq 0$, $y \geq 0$. This represents reflecting Brownian motion on $[0, +\infty)$ killed elastically at $x = 0$. The infinitesimal generator of its associated semigroup in $C_0(\mathbb{R}_0^+)$ is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = cf(0)\},$$
$$Af := \frac{1}{2}f''.$$

We can visualize this process as a reflected Brownian motion killed at a random time r, where $P(\{r > t\} \cap B) = \mathbb{E}\{\mathbb{1}_B \exp(-c\tau(t))\}$, B is any Borel set of $C(\mathbb{R}_0^+)$ and $\tau(t)$ is the local time of the Wiener process. □

Example 2.10 (doubly reflected Wiener)**.** We consider now a Brownian motion with state space $[0, a]$. where a is a positive real number. On the state space $[0, a]$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) := \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}}\Big(\int_B \Big\{\exp\big[-\frac{(2ka + y - x)^2}{2t}\big] + \\ + \exp\big[-\frac{(2ka + y + x)^2}{2t}\big]\Big\}\mathrm{d}y\Big),$$

for any $t > 0$, $0 \leq x \leq a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion with reflecting barrier at $x = 0$ and at $x = a$.

The process itself can be constructed as $(X_t = \varphi_a(x + W_t) : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process in $\mathbb{R}$ and φ_a is the function $x \mapsto \min\{(2a - x), x\}$ for x in $[0, 2a]$ and extended to $\mathbb{R}$ as a continuous periodic function with $2a$-period, so that φ_a maps $\mathbb{R}$ onto $[0, a]$. The associated semigroup in $C([0, a])$ is given by

$$S(t)f(x) := \int_0^a f(y)\, p(t, x, \mathrm{d}y) = \\ = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f[\varphi_a(y)] \, \exp\big[- \frac{(y-x)^2}{2t}\big] \mathrm{d}y,$$

for every $t > 0$ and $0 \leq x \leq a$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) := \{f \in C^2([0, a]) : f'(0) = 0,\ f'(a) = 0\},$$
$$Af := \frac{1}{2} f''.$$

A constant drift b can be added so that $Af = \frac{1}{2}f'' + bf'$ and $X_t := \varphi_a(x+bt+W_t)$. Notice that $p(t, x, [0, a]) = 1$ for any $t > 0$ and x in $[0, a]$. □

Example 2.11 (doubly absorbed Wiener)**.** We consider now a Brownian motion with state space $(0, a)$, where a is a positive real number. On the state space $(0, a)$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t, x, B) := \sum_{k \in \mathbb{Z}} \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[- \frac{(2ka + y - x)^2}{2t}\big] - \\ - \exp\big[- \frac{(2ka + y + x)^2}{2t}\big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $0 \leq x \leq a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion with absorbing barrier at $x = 0$ and at $x = a$, i.e., the Brownian motion particle *dies* at the first time when it hits the boundary $x = 0$ or $x = a$. The process itself can be constructed as $(X_t = x + W_{t \wedge \tau} : t \geq 0)$, where $(W_t : t \geq 0)$ is a standard Wiener process in $\mathbb{R}$ and τ is the first exit time from the open set $(0, a)$ for the process $x + W_t$, i.e.,

$$\tau := \inf\{t > 0 : x + W_t \leq 0 \ \text{ or } \ x + W_t \geq a\}, \quad t \geq 0,\ x > 0,$$

The associated semigroup in $C_0(]0, a[)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}_0^+} f(y)\, p(t, x, \mathrm{d}y) = \\ = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \hat{f}_a(y) \, \exp\big[- \frac{(y-x)^2}{2t}\big] \mathrm{d}y,$$

for every $t > 0$ and $0 \leq x \leq a$, where now $\hat{f}_a(y) = f[\varphi_a(y)]$ if $y \geq 0$ and $\hat{f}_a(y) =$

$-f[\varphi_a(y)]$ if $y \le 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(]0,a[) \cap C^2(]0,a[) : \\ : f'' \in C_0(]0,a[),\ f(0) = 0,\ f(a) = 0\},$$
$$Af := \frac{1}{2} f''.$$

Technically, the points 0 and a does not belong to the domain of definition of functions f in the Banach space $C_0(]0,a[)$, but we identify $C_0(]0,a[)$ with the subspace of $C([0,a]) = C_0([0,a])$ satisfying $f(0) = f(a) = 0$. Again notice that $p(t,x,]0,a[) < 1$ and that the semigroup $S(t)$ may be considered as defined on the Banach $C([0,a])$ where $S(t)\mathbb{1} = 0$ for all $t > 0$, so that the state of the process could be regarded as $[0,a]$. A constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$. and $X_t := x + bt \wedge \tau + W_{t\wedge\tau})$, where τ is now the first exit time from the open set $(0,a)$ for the process $(bt + W_t : t \ge 0)$. □

Some details on Brownian motion on a finite interval relative to Examples 2.10 and 2.11 can be found in Karatzas and Shreve [124, Section 2.8.c, pp. 97–100].

Example 2.12 (periodic Wiener)**.** We consider now a Brownian motion with state space $[0,a]$. where a is a positive real number. On the state space $[0,a]$ with its Borel σ-algebra $\mathcal{B}$ we consider

$$p(t,x,B) := \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2\pi t}} \Big(\int_B \Big\{ \exp\big[-\frac{(ka + y - x)^2}{2t} \big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $0 \le x \le a$ and B in $\mathcal{B}$, where $\mathbb{Z}$ is the set of all integer numbers. This represents Brownian motion on a circle (the interval $[0,a]$, with 0 and a identified). The process itself can be constructed as $(X_t = \psi_a(x + W_t) : t \ge 0)$, where $(W_t : t \ge 0)$ is a standard Wiener process in $\mathbb{R}$ and $\psi_a(x) = x - a[x/a]$ (where $[x]$ denotes the integral part of x), which maps $\mathbb{R}$ onto $[0,a]$. The associated semigroup in $C_\#(0,a) = \{f \in C([0,a]) : f(0) = f(a)\}$ is given by

$$S(t)f(x) := \int_0^a f(y)\, p(t,x,\mathrm{d}y) = \\ = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f[\psi_a(y)] \exp\big[-\frac{(y-x)^2}{2t} \big] \mathrm{d}y,$$

for every $t > 0$ and $0 \le x \le a$. The infinitesimal generator is a closed extension of the differential operator $Af := \frac{1}{2} f''$ with domain

$$\mathcal{D}(A) := \{f \in C_\#(0,a) \cap C^2([0,a]) :\ f''(0) = f''(a)\}.$$

Notice that even if ψ_a is not continuous, the composition $x \mapsto f[\psi_a(x)]$ is continuously differentiable for any continuous function f in the domain $\mathcal{D}(A)$. Also, a constant drift b can be added so that $Af = \frac{1}{2} f'' + bf'$ and $X_t := \psi_a(x + bt + W_t)$. □

More generally, we may consider a Sturm-Liouville problem in $[0,a]$ of the form

$$\begin{cases} \frac{1}{2}u'' + \lambda u = 0, \text{ in } (0,a), \\ \alpha_0 u(0) - \beta_0 u'(0) = \alpha_a u(a) + \beta_a u'(a) = 0, \end{cases} \tag{2.101}$$

where $\alpha_0, \beta_0, \alpha_a, \beta_a \geq 0$, $\alpha_0 + \beta_0 > 0$ and $\alpha_a + \beta_a > 0$. A periodic condition of the form $u(0) - u(a) = u'(0) - u'(a) = 0$ can also be used. There is a sequence of (positive) eigenvalues $(0 < \lambda_0 < \lambda_1 < \cdots)$, $\lambda_n \to \infty$ as $n \to \infty$, with its corresponding eigenfunctions $(u_0, u_1, \ldots)$, satisfying the boundary value problem (2.101) and form an orthonormal basis in $L^2(0,a)$. Certainly, u_n is a linear combination of $\sin(x\sqrt{2\lambda_n})$ and $\cos(x\sqrt{2\lambda_n})$ and if $a = \pi$ then when $\alpha_0\alpha_a = \beta_0\beta_a = 0$ and $\alpha_0\beta_a \neq \alpha_a\beta_0$ we have $2\sqrt{2\lambda_n} = 2n+1$ and when $\alpha_0\beta_a = \alpha_a\beta_0$ and $\alpha_0\alpha_a \neq 0$ or $\beta_0\beta_a \neq 0$ we have $\sqrt{2\lambda_n} = n$. In the case of periodic boundary conditions, if $a = 2\pi$ then $\sqrt{2\lambda_n} = n$. Define

$$p(t,x,B) := \sum_{n=0}^{\infty} \mathrm{e}^{-t\lambda_n} u_n(x) \int_B u_n(y)\mathrm{d}y, \tag{2.102}$$

for every x in $[0,a]$, $t > 0$ and B in $\mathcal{B}$. The maximum principle ensures that p in nonnegative, i.e, $0 \leq p(t,x,B) \leq 1$, for every t, x, B. Parseval equality yields

$$\int_0^a |p(t,x,B)|^2 \mathrm{d}x = \sum_{n=0}^{\infty} \mathrm{e}^{-t\lambda_n} |B|, \quad \forall x \in [0,a],\ t > 0,\ B \in \mathcal{B},$$

where $|B|$ denotes the Lebesgue measure of the Borel set B. Some more details are needed to discuss the convergence of the series (2.102), which is the eigenvalues and eigenfunctions expansion of the Green function or Green operator relative to the boundary value problem (2.101). From here, the associate semigroup and the stochastic process can be constructed. The interested reader may consult the pioneer paper Feller [80] related to parabolic differential equations and the associated semigroups.

Example 2.13 (reflecting Wiener-Poisson)**.** On the state space $\mathbb{R}_0^+ := [0,\infty)$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t,x,B) := \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!} \Big(\int_B \Big\{ \exp\big[-\frac{(x+k-y)^2}{2t} \big] + \\ + \exp\big[-\frac{(x+k+y)^2}{2t} \big] \Big\} \mathrm{d}y \Big),$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. This is a Wiener-Poisson process with reflecting barrier at $x = 0$ and the process itself can be constructed as $(X_t = |x + W_t + P_t| : t \geq 0)$, where $(W_t : t \geq 0)$ is a Wiener process independent of the Poisson process $(P_t : t \geq 0)$, both in in $\mathbb{R}$. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t,x,\mathrm{d}y) = \\ = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \int_{-\infty}^{\infty} \check{f}(x+k+\sqrt{t}z) \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

where $\check{f}(y) := f(y)$ if $y \geq 0$ and $\check{f}(y) := f(-y)$ if $y \leq 0$, for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0\},$$
$$Af(x) := \frac{1}{2} f''(x) + c[f(x+1) - f(x)], \quad \forall x \in \mathbb{R}_0^+.$$

Only a cad-lag realization can be constructed. We can generalize this example to a $(W_t : t \geq 0)$ Wiener process (with drift b and covariance σ) and a $(P_t : t \geq 0)$ compound Poisson processes (with parameters (c, μ)), independent of each other and with reflecting barrier at $x = 0$. The compound process is indeed a *subordinator*, i.e., increasing in t so that all jumps of the sum process $(x + W_t + P_t : t \geq 0)$ are inside the real semi-line $[0, \infty)$. Thus $c > 0$ and μ is a probability distribution on $(0, \infty)$. Again, a cad-lag realization is given by $(X_t = |x + W_t + P_t| : t \geq 0)$ and the probability transition function is

$$p(t, x, B) = \int_B \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!}\, p_k(t, x, y)\mathrm{d}y,$$
$$p_k(t, x, y) = \frac{1}{\sqrt{2\pi t}} \Big(\int_0^{\infty} \Big\{ \exp\big[-\frac{(x-y-z)^2}{2t} \big] + \\ + \exp\big[-\frac{(x+y-z)^2}{2t} \big] \Big\} \mu^k(\mathrm{d}z) \Big),$$
$$\mu^0 = \mu, \quad \mu^k(B) = \int_{\mathbb{R}\times\mathbb{R}} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for every $k = 1, 2, \ldots$, for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$. Here some work is necessary to ensure the proper convergence of the above series. Again notice that $\mu^k = \mu^{k-1} \star \mu$ is the k convolution of μ. Its associated semigroup in $C_0(\mathbb{R}_0^+)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}} f(y)\, p(t, x, \mathrm{d}y) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times \\ \times \int_0^{\infty} \mu^k(\mathrm{d}y) \int_{-\infty}^{\infty} \check{f}(x + y + \sqrt{t}z)\, \exp\big(-\frac{z^2}{2} \big) \mathrm{d}z,$$

for every $t > 0$ and $x \geq 0$. The infinitesimal generator is the integro-differential operator

$$\mathcal{D}(A) := \{f \in C_0(\mathbb{R}_0^+) \cap C^2(\mathbb{R}_0^+) : f'' \in C_0(\mathbb{R}_0^+),\ f'(0) = 0\},$$
$$Af(x) := \frac{1}{2}\sigma f''(x) + bf'(x) + c \int_0^{\infty} [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

for every x in $\mathbb{R}_0^+$. If the compound Poisson process P_t has the parameter-distribution μ in the whole space $\mathbb{R}$ then the sum process $x + W_t + P_t$ may have a jumps outside of the semi-line $[0, \infty)$. In this case, we may keep the expression

$(X_t = |x + W_t + P_t| : t \geq 0)$ and make appropriated modifications. For instance the semigroup takes the form

$$S(t)f(x) := \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{\sqrt{2\pi}\, k!} \times \\ \times \int_{\mathbb{R}} \mu^k(\mathrm{d}y) \int_{-\infty}^{\infty} \check{f}(x + y + \sqrt{t}z) \exp\big(-\frac{z^2}{2}\big)\mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}$ and $\check{f}$ as above, however, the boundary condition for the domain of the infinitesimal generator needs more work. It is clear that absorbing and sticking barriers can be considered for Wiener-Poisson processes by means of the expression with $\hat{f}$ and the stopping argument. □

As in Examples 2.6 and 2.7, we can discuss absorbing and sticking barriers for Wiener-Poisson processes by means of arguments similar to Example 2.13. This is on the space either $\mathbb{R}^+ = (0, \infty)$ or $\mathbb{R}_0^+ = [0, \infty)$ with its Borel σ-algebra $\mathcal{B}$ and for a given positive constant c, we consider

$$p(t, x, B) := \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{2\pi t\, k!}\, p_i(t, x + k, B), \tag{2.103}$$

for any $t > 0$, $x \geq 0$ and B in $\mathcal{B}$, where $p_i(t, x, B)$ is the transition function for absorbing barrier with $i = 1$ or for sticking barrier with $i = 2$, as in previous examples. Notice that

$$p_2(t, x, B) = p_1(t, x, B) + [1 - p_1(t, x, \mathbb{R})]\mathbb{1}_B(0).$$

In the case of the Wiener-Poisson process, the boundary condition for the absorbing barrier is clearly $f(0) = 0$. However, for the sticking barrier boundary condition is

$$\frac{1}{2} f''(0) + c[f(0 + 1) - f(0)] = 0,$$

i.e., the equation is satisfied up to the boundary.

Also the case of a Wiener-Poisson process with periodic conditions can be easier studied, e.g., a Wiener-Poisson process in $\mathbb{R}$ is combined with the operation modulo $[0, a]$, ψ_a, as in Example 2.12, which maps $\mathbb{R}$ into $[0, a]$.

Trying to extend the doubly reflected Wiener in an interval, Example 2.10, to a Wiener-Poisson process, we encounter a new difficulty, we may *jump outside the interval*. This forces us to make a decision on the jumps, e.g., a natural extension or reflection. This is a more delicate issue. For instance, if we want the reflection on an interval $[0, a]$, first we make a periodic condition on $[-a, a]$ and then we take the absolute value. However, if we want a natural extension, first we make a constant and continuous extension outside of the given interval $[0, a]$ and then we use the process in the whole line.

For instance, the reader may consult the books Mandl [164] for a comprehensive treatment of one-dimensional Markov processes. On the other hand, several examples (without jumps) can be found in Borodin and Salminen [30, Appendix 1, pp. 102-119]

2.13.2 Multi-Dimensional

In the whole space $\mathbb{R}^d$, $d \geq 2$ we have more difficulties. A central role is played by the *Gauss kernel* Γ_0 defined by

$$\Gamma_0(x,t,a) := \frac{1}{(2\pi t)^{d/2}\sqrt{\det a}} \exp\big(-\frac{x\cdot a^{-1}x}{2t}\big), \tag{2.104}$$

for every $t > 0$ and x in $\mathbb{R}^d$, where a is an invertible symmetric nonnegative $d \times d$-matrix, if we write by components $a = (a_{ij})$ then its inverse $a^{-1} = (a^{ij})$ and $x\cdot a^{-1}x = \sum_{ij} a^{ij}x_i x_j$. When a is the identity matrix $\mathbb{1}$ we write $\Gamma_0(x,t) := \Gamma_0(x,t,\mathbb{1})$. In analysis the constant $1/2$ is replaced by $1/4$ and called *heat kernel*. This function is infinitely differentiable in all its arguments and in particular, for any derivative ∂^ℓ, with $\ell = (\ell_1, \ell_2, \dots, \ell_d, \ell_t, \ell_a)$ and $|\ell| = \ell_1 + \ell_2 + \cdots + \ell_d + 2\ell_t$ we have

$$|\partial^\ell \Gamma_0(x,t,a)| \leq C t^{-(d+|\ell|)/2} \exp\big(-c\frac{|x|^2}{t}\big), \tag{2.105}$$

for every $t > 0$ and x in $\mathbb{R}^d$, for some positive constants $C = C(d,\ell,\delta)$ and $c = c(d,\ell,\delta)$, where the symmetric matrix a satisfies $\delta|\xi|^2 \leq \xi\cdot a\xi \leq |\xi|^2/\delta$ for any ξ in $\mathbb{R}^d$, for some $\delta > 0$. Also we have

$$\int_{\mathbb{R}^d} \partial^\ell \Gamma_0(x,t,a)\,\mathrm{d}x = \begin{cases} 1 & \text{if } \ell = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.106}$$

This $\Gamma_0(x,t,a)$ is the probability density transition function of a Wiener process in $\mathbb{R}^d$, with zero mean and co-variance a. The corresponding resolvent kernel is given by

$$\begin{aligned} R(\lambda,x) &:= \frac{1}{(2\pi t)^{-d/2}\sqrt{\det a}} \int_0^\infty \exp\big(-\frac{x\cdot a^{-1}x}{2t} - \lambda t\big)\mathrm{d}t = \\ &= \frac{2}{(2\pi t)^{-d/2}}\Big(\frac{2\lambda}{x\cdot a^{-1}x}\Big)^{(d/4-1/2)} \mathrm{K}_{d/2-1}\big(\sqrt{2\lambda x\cdot a^{-1}x}\big), \end{aligned}$$

for every $\lambda > 0$ and x in $\mathbb{R}^d$, where K_ν is the modified Bessel function of 2nd kind. In particular,

$$\mathrm{K}_{n-1/2}(z) = \sqrt{\frac{\pi}{2z}}\, z^n \Big(-\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}\Big)^n e^z, \quad n = 0, 1, \dots,$$

and so

$$R(\lambda,x) := \frac{1}{4\pi|x|}\exp(-\sqrt{2\lambda}|x|),$$

for $d = 3$ and $a = \mathbb{1}$, the identity matrix.

Example 2.14 (d-dimensional Wiener)**.** A Wiener process with vector mean b and co-variance matrix a has a transition probability function on the state space space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ defined by

$$p(t,x,B) := \int_B \Gamma_0(x-bt-y,t,a)\mathrm{d}y,$$

for every x in $\mathbb{R}^d$, $t>0$ and B in $\mathcal{B}$, where $\Gamma_0(x,t,a)$ is the Gauss kernel (2.104). Notice that if $(W_t : t \geq 0)$ is a standard Wiener process starting at the origin, i.e., with $W_0 = 0$, zero mean and co-variance matrix $\mathbb{1}$ or equivalently $p(t,x,\mathrm{d}y)\Gamma(x-\mathrm{d}y,t)$ as its transition probability function, then the process $X_t := x+bt+\sqrt{a}W_t$ is a realization of the above Wiener process starting at x. Also this can be constructed as the product of d independent one dimensional Brownian motions, i.e., the probability transition *density* function $\Gamma_0(x,t)$ is the product of d similar one dimensional expressions $\Gamma_0(x_i,t)$ as the one used in Example 2.1. The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}^d} f(y)\, p(t,x,\mathrm{d}y) =$$
$$= \int_{\mathbb{R}^d} f(x+bt+\sqrt{ta}\,z)\,\Gamma_0(z,1)\,\mathrm{d}z,$$

for every $t>0$ and x in $\mathbb{R}^d$. Its infinitesimal generator A is the closure of the (closable) differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) := C_0^0(\mathbb{R}^d)\cap C^2(\mathbb{R}^d),$$

$$\mathring{A}f := \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\,\partial_{ij}^2 f + \sum_{i=1}^{d} b_i\,\partial_i f,$$

where a_{ij} and b_i are the entries of the matrix a and the vector b. □

Example 2.15 (Ornstein-Uhlenbeck)**.** This is a modification of a Wiener process where a linear drift is added. Two matrices a and b describe the process X, namely,

$$X_t(x) = \mathrm{e}^{bt}x + \int_0^t \sqrt{a}\,\mathrm{e}^{b(t-s)}\,\mathrm{d}W_t,$$

where $(W_t : t \geq 0)$ is a standard Wiener process. The process $X_t(x)$ has a Gaussian distribution with mean $\mathrm{e}^{bt}x$ and covariance

$$q_t = \int_0^t \mathrm{e}^{bs}a\mathrm{e}^{b^*s}\mathrm{d}s, \quad t>0,$$

where b^* is the adjoint matrix. Thus, the transition probability function of the Ornstein-Uhlenbeck process $(X_t(x) : t \geq 0)$ on state space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ is given by

$$p(t,x,B) := \int_B \Gamma_0(\mathrm{e}^{bt}x-y,1,q_t^{-1})\,\mathrm{d}y,$$

for every x in $\mathbb{R}^d$, $t > 0$ and B in $\mathcal{B}$, where $\Gamma_0(x,t,a)$ is the Gauss kernel (2.104). The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}^d} f(y)\, p(t,x,\mathrm{d}y) = \\ = \int_{\mathbb{R}^d} f(\mathrm{e}^{bt}x + q_t\, z)\, \Gamma_0(z,1)\, \mathrm{d}z,$$

for every $t > 0$ and x in $\mathbb{R}^d$. Its infinitesimal generator is the closure of the (closable) differential operator $\mathring{A}$

$$\mathcal{D}(\mathring{A}) := C_0^0(\mathbb{R}^d) \cap C^2(\mathbb{R}^d),$$

$$\mathring{A}f := \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\, \partial_{ij}^2 f + \sum_{i,j=1}^{d} b_{ij} x_j\, \partial_i f,$$

where a_{ij} and b_{ij} are the entries of the matrices a and b. □

Example 2.16 (compound Poisson)**.** A compound poisson process with parameter $c > 0$ and μ, where μ is a distribution in $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$ has a transition probability function on the state space space $\mathbb{R}^d$ with its Borel σ-algebra $\mathcal{B}$ defined by

$$p(t,x,B) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}^d_*} \mathbb{1}_B(x+y)\, \mu^k(\mathrm{d}y),$$

$$\mu^0 = \mu, \quad \mu^k(B) = \int_{\mathbb{R}^d_* \times \mathbb{R}^d_*} \mathbb{1}_B(y+z)\, \mu^{k-1}(\mathrm{d}y)\, \mu(\mathrm{d}z),$$

for every $k = 1, 2, \ldots$, for any $t > 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$. The probability measures $\mu^k = \mu^{k-1} \star \mu$ are called the k-convolution of μ. Based on two sequences of independent identically distributes random variables with exponential and μ distribution, a canonical realization of the compound Poisson process can be constructed. The associated semigroup in $C_0(\mathbb{R}^d)$ is given by

$$S(t)f(x) := \int_{\mathbb{R}^d} f(y)\, p(t,x,\mathrm{d}y) = \\ = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(ct)^k}{k!} \int_{\mathbb{R}^d_*} f(x+y)\, \mu^k(\mathrm{d}y),$$

for every $t > 0$ and x in $\mathbb{R}^d$. Its infinitesimal generator is the integral operator

$$\mathcal{D}(A) := C_0(\mathbb{R}^d), \qquad Af(x) := c \int_{\mathbb{R}^d_*} [f(x+y) - f(x)]\, \mu(\mathrm{d}y),$$

which is clearly a nonlocal operator. If the distribution μ has support in an open semi-space $\mathbb{R}^d_+ := \mathbb{R}^{d-1} \times (0,\infty)$ then we may consider the compound Poisson process only in $\mathbb{R}^d_+$, which is called *subordinator* in the one dimensional case. □

It is clear that we may mix Examples 2.14 and 2.16 to produce a d-dimensional Wiener-Poisson process with probability density transition function defined by

$$\begin{cases} \Gamma(x,t) = \mathrm{e}^{-ct} \sum_{k=0}^{\infty} \frac{(c\,t)^k}{k!} \, (\Gamma_0 \star \mu^k)(x,t), \\ \mu^0 := \delta_0, \qquad \mu^k(B) := \int_{\mathbb{R}^d_* \times \mathbb{R}^d_*} \mathbb{1}_B(x+y) \, \mu(\mathrm{d}x) \, \mu^{k-1}(\mathrm{d}y), \\ (\Gamma_0 \star \mu^k)(x,t) := \int_{\mathbb{R}^d_*} \Gamma_0(x-y,t) \, \mu^k(\mathrm{d}y), \end{cases} \tag{2.107}$$

for every x in $\mathbb{R}^d$ and $t > 0$, where δ_0 is the Dirac measure at the origin. Since μ is a probability measure on $\mathbb{R}^d_*$, so is μ^k and the above series is clearly convergent. The infinitesimal generator takes the form

$$\begin{cases} Af(x) := \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \, \partial^2_{ij} f(x) + \sum_{i=1}^{d} b_i \, \partial_i f(x) + \\ \qquad\qquad + c \int_{\mathbb{R}^d_*} [f(x+y) - f(x)] \, \mu(\mathrm{d}y), \end{cases} \tag{2.108}$$

which is a second order integro-differential (non-local) operator.

Example 2.17 (Neumann)**.** This is a half-space normal reflecting barrier, i.e., on the semi-space $\bar{\mathbb{R}}^d_+ := \mathbb{R}^{d-1} \times [0,\infty)$, with the notation $x = (\tilde{x}, x_d)$, we consider the function

$$G^N_0(\tilde{x}, x_d, t, \xi_d) := \Gamma_0(\tilde{x}, x_d - \xi_d, t) + \Gamma_0(\tilde{x}, x_d + \xi_d, t),$$

for every $t > 0$, $x_d, \xi_d \geq 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. As in Example 2.5, we may define a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t,x,B) := \int_B G^N_0(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d) \, \mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. The arguments are the same, even the construction of the (standard) *normal reflected* Wiener process in a d-dimensional half-space is simple, since this is a $(d-1)$-dimensional Wiener process and an independent one-dimensional Wiener process with reflecting barrier at $x_d = 0$. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\mathcal{D}(A) := \{ f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+), \ \partial_d f(\tilde{x}, 0) = 0 \},$$
$$Af := \frac{1}{2} \Delta f,$$

where Δ is the usual Laplacian operator $\sum_{i=1}^d \partial_i^2$, here in the sense of Schwartz distribution derivative. Except for the $1/2$ factor, the local time correspond to the so-called *Poisson kernel* which is $P_0(t,x,\tilde{\xi}) := -2\,\Gamma_0(\tilde{x} - \tilde{\xi}, x_d, t)$, for any $t > 0$, $x = (\tilde{x}, x_d)$ in $\mathbb{R}^{d-1} \times (0,\infty)$ and $\tilde{\xi}$ in $\mathbb{R}^{d-1}$. □

Example 2.18 (Dirichlet)**.** This is a half-space normal reflecting barrier, i.e., on the semi-space $\mathbb{R}^d_+ := \mathbb{R}^{d-1} \times (0,\infty)$, with the notation $x = (\tilde{x}, x_d)$, we consider the function

$$G_0^D(\tilde{x}, x_d, t, \xi_d) := \Gamma_0(\tilde{x}, x_d - \xi_d, t) - \Gamma_0(\tilde{x}, x_d + \xi_d, t),$$

for every $t > 0, x_d, \xi_d > 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. As in Example 2.6, we may define a transition function on the state space $\mathbb{R}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) := \int_B G_0^D(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d)\, \mathrm{d}\xi,$$

for any $t > 0$, x in $\mathbb{R}^d_+$ and B in $\mathcal{B}$. The arguments are the same, even the construction of the (standard) *stopped* Wiener process in d-dimensional half-space is simple, since this is a $(d-1)$-dimensional Wiener process and an independent one dimensional Wiener process with absorbing barrier at $x_d = 0$. If the barrier $x_d = 0$ is of some interest, then we may proceed as in Example 2.7 and convert p into a probability transition function. To that effect, we notice that

$$G_0^D(\tilde{x}, x_d, t, \xi_d) = \frac{1}{\sqrt{2\pi t}} \Big\{ \exp\big[-\frac{(x_d - \xi_d)^2}{2t} \big] - \\ - \exp\big[-\frac{(x_d + \xi_d)^2}{2t} \big] \Big\} \Gamma_{d-1}(\tilde{x}, t),$$

where $\Gamma_{d-1}(\tilde{x}, t)$ has the same expression (2.104) with the identity matrix $a = \mathbb{1}$ in dimension $(d-1)$. Then we define

$$p(t, x, B) := \int_B \Big\{ \frac{1}{\sqrt{2\pi t}} \Big(\exp\big[-\frac{(x_d - \xi_d)^2}{2t} \big] - \\ - \exp\big[-\frac{(x_d + \xi_d)^2}{2t} \big] \Big) + \Big[1 - \frac{1}{\sqrt{2\pi t}} \times \\ \times \int_{-x_d}^{x_d} \exp\big(-\frac{z^2}{2t} \big) \mathrm{d}z \Big] \mathbb{1}_B(\tilde{\xi}, 0) \Big\} \Gamma_{d-1}(\tilde{x} - \tilde{\xi}, t) \mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. This yields the (standard) *sticking* Wiener process in a d-dimensional half-space. Expressions for the associated semigroup and its infinitesimal generator are obtained immediately from the one-dimensional case. □

A reflected Wiener process with zero mean and co-variance matrix a in a d-dimensional half-space presents more difficulties. After a rotation of coordinates, we can reduce the general case of a Wiener process with zero mean and covariance matrix a to the case where a is the identity matrix, but the boundary condition is a *oblique* reflection, i.e., instead of the condition $\partial_d f(\tilde{x}, 0) = 0$ on the domain of the infinitesimal generator $A = \Delta/2$ we have $b \cdot \nabla f(\tilde{x}, 0) = 0$, where ∇ is the gradient operator in the first d-dimensional variable, i.e., x, and $b = (b_1, \ldots, b_d)$ is a vector with $b_d > 0$. For the boundary value problem

$$\begin{cases} \partial_t u - \frac{1}{2}\Delta u(x,t) = 0, & \forall x \in \mathbb{R}^d_+,\ t > 0, \\ u(x, 0) = 0, & \forall x \in \mathbb{R}^d_+, \\ b \cdot \nabla u(x,t) = \psi(x,t), & \forall x \in \partial\mathbb{R}^d_+,\ t > 0, \end{cases} \tag{2.109}$$

where $\partial\mathbb{R}^d_+ := \mathbb{R}^{d-1} \times \{0\}$, we can calculate the Poisson kernel P^b_0 as

$$P^b_0(\tilde{x}, x_d, t) := \varphi_b(\tilde{x}, x_d, t)\, \Gamma_0(\tilde{x}, x_d, t), \tag{2.110}$$

and

$$\varphi_b(\tilde{x}, x_d, t) := -\frac{1}{|b|^2}\Big\{ b_d + \frac{|b|^2 x_d - b_d(b\cdot x)}{|b|\sqrt{t/2}} \times \\ \times \exp\Big[\frac{(b\cdot x)^2}{2t|b|^2}\Big] \int_{(b\cdot x)/|b|\sqrt{2t}}^{+\infty} \mathrm{e}^{-r^2} dr \Big\},$$

for any $t > 0$, $x = (\tilde{x}, x_d)$ in $\mathbb{R}^d_+$. Actually, we use $P^b_0(\tilde{x} - \tilde{\xi}, x_d, t)$ with $\tilde{\xi}$ in $\partial\mathbb{R}^d_+$.

Example 2.19 (oblique)**.** This is a half-space oblique reflecting barrier in the direction of the vector $b = (b_1, \dots, b_d)$ with $b_d > 0$. On the semi-space $\bar{\mathbb{R}}^d_+ := \mathbb{R}^{d-1} \times [0, \infty)$, with the notation $x = (\tilde{x}, x_d)$, we consider the function

$$G^b_0(\tilde{x}, x_d, t, \xi_d) := \Gamma_0(\tilde{x}, x_d - \xi_d, t) - \Gamma_0(\tilde{x}, x_d + \xi_d, t) - \\ -2\, b_d P^b_0(x, x_d + \xi_d, t),$$

for every $t > 0$, $x_d, \xi_d \geq 0$, and $\tilde{x}$ in $\mathbb{R}^{d-1}$. This yields a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) := \int_B G^b_0(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d) \mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. This is not a product of $(d-1)$ independent Brownian motions in $\mathbb{R}$ with an independent reflected Brownian motion in $[0, \infty)$, certainly, the function φ_b in (2.110) makes the coupling. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\mathcal{D}(A) := \{f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+),\ b \cdot \nabla f(\tilde{x}, 0) = 0\},$$
$$Af := \frac{1}{2}\Delta f,$$

but a realization of a d-dimensional (standard) Wiener process in $\bar{\mathbb{R}}^d_+$ with oblique reflection at the barrier $x_d = 0$ is obtained from general existence theorems based on the above transition probability function. □

The fact that an explicit expression can be found for the above transition function give specific estimates allowing the construct Green function for variable coefficients and integro-differential operators, the reader may consult the books Garroni and Menaldi [93, 94]. The case of a normal reflected Wiener-Poisson process can be treated as in the one dimensional case Example 2.13, however, the oblique reflection needs another method. Let us consider the case of an integro-differential operator of the form $A := \Delta/2 + I$, where

$$I\varphi(x) := c \int_{\mathbb{R}^d_+} [\varphi(x + y) - \varphi(x)]\, \mu(\mathrm{d}y), \quad \forall x \in \mathbb{R}^d_+, \tag{2.111}$$

where μ is now a probability measure in $\bar{\mathbb{R}}^d_+$ and $c > 0$. We define

$$G^b := G^b_0 + G^b_1 + \cdots + G^b_k + \cdots, \qquad G^b_k := G^b_0 \star IG^b_{k-1}, \tag{2.112}$$

where I is considered acting on the first d-dimensional variables, i.e., for any fixed $t > 0$, $\xi_d \geq 0$

$$IG^b_k(\tilde{x}, x, t, \xi_d) = c \int_{\mathbb{R}^d_+} [G^b_k(x + y, t, \xi_d) - G^b_k(x, t, \xi_d)]\, \mu(\mathrm{d}y),$$

for every $x = (\tilde{x}, x_d)$ in $\mathbb{R}^d_+$, for any $k \geq 0$, and the *kernel-convolution*

$$\begin{cases} (\varphi \star \psi)(\tilde{x}, x_d, t, \xi_d) := \\ \qquad = \displaystyle\int_0^t ds \int_{\mathbb{R}^d_+} \varphi(\tilde{x} - \tilde{y}, x_d, t - s, y_d)\, \psi(\tilde{y}, y_d, s, \xi_d)\mathrm{d}y, \end{cases} \tag{2.113}$$

for $\tilde{x}$ in $\mathbb{R}^{d-1}$, $x_d, \xi_d \geq 0$ and $t > 0$.

The semigroup property or Chapman-Kolmogorov identity for the transition function G^b_0, namely

$$G^b_0(\tilde{x}, x_d, t + s, \xi_d) = \int_{\mathbb{R}^d_+} G^b_0(\tilde{x} - \tilde{y}, x_d, t, y_d)\, G^b_0(\tilde{y}, y_d, s, \xi_d)\mathrm{d}y,$$

for every $\tilde{x}$ in $\mathbb{R}^{d-1}$, $x_d, \xi_d \geq 0$ and $s, t > 0$, and the explicit form of the function G^b_0 given in Example 2.19 yield the identity $G_k(x, t, \xi_d) = (t^k/k!)\, I^k G^b_0(x, t, \xi_d)$, where I^k is the k-iteration of the integral operator I.

Certainly, we use the technique of Section 2.12 to check in what sense the above series (2.112) converges. First we define the Green space $\mathcal{G}^0_k$ of continuous kernels $\varphi(x, t, \xi_d)$ for x in $\bar{\mathbb{R}}^d_+$, $t > 0$ and $\xi_d \geq 0$ such that

$$\begin{cases} |\varphi(x, t, \xi_d)| \leq C_0\, t^{-1+k-d/2}, \quad \forall x, t, \xi_d, \\ \displaystyle\int_{\mathbb{R}^d_+} \big[|\varphi(\tilde{y} - \tilde{\xi}, y_d, t, \xi_d)| + |\varphi(\tilde{x} - \tilde{y}, x_d, t, y_d)|\big]\mathrm{d}y \leq \\ \qquad\qquad \leq K_0\, t^{-1+k/2}, \quad \forall x, t, \xi, \end{cases} \tag{2.114}$$

for some constants C_0 and K_0, and the infimum of all such constants, denoted by $C(\varphi, k)$ and $K(\varphi, k)$, are semi-norms for $k > 0$.

It is easy to check that I maps the Green space $\mathcal{G}^0_k$ into itself,

$$C(I\varphi, k) \leq 2c\, C(\varphi, k) \quad \text{and} \quad K(I\varphi, k) \leq 2c\, K(\varphi, k), \tag{2.115}$$

for every φ, k, and that G^b_0 belongs to $\mathcal{G}^0_2$ in view of (2.105), which is valid for G^b_0 instead of Γ_0.

Therefore, G^b_k belongs to $\mathcal{G}^0_{2k+2}$ and

$$\begin{cases} C(\partial^\ell G^b_k, 2k + 2 - |\ell|) \leq \dfrac{(2c)^k}{k!}\, C(\partial^\ell G^b_0, 2 - |\ell|), \\ K(\partial^\ell G^b_k, 2k + 2 - |\ell|) \leq \dfrac{(2c)^k}{k!}\, K(\partial^\ell G^b_0, 2 - |\ell|), \end{cases} \tag{2.116}$$

for every $k = 1, 2, \dots$ and ∂^ℓ, with $\ell = (\ell_1, \ell_2, \dots, \ell_d, \ell_t, \ell_a)$ and $|\ell| = \ell_1 + \ell_2 + \cdots + \ell_d + 2\ell_t$. Because of the identity $G_k(x,t,\xi_d) = (t^k/k!)\, I^k G_0^b(x,t,\xi_d)$, the expression $G_k^b = G_0^b \star I G_{k-1}^b$ is not really used, not integration in the time variable is needed in this explicit case and the semi-norms (2.114) are meaningful even for $k \le 0$. Recall that $c > 0$ is the constant used in the definition of the integral operator I in (2.111).

At this point we have proved that the remainder of the series (2.112) defining $\partial^\ell G^b$, i.e., $\partial^\ell G_k^b + \partial^\ell G_{k+1}^b + \cdots$ converges in the Green space $\mathcal{G}^0_{2k+2-|\ell|}$, for $\ell \ge 0$, so that G^b is infinitely many time differential in all its arguments. Moreover, G^b satisfies the Volterra equation $G^b = G_0^b + G_0^b \star I G^b$, the Chapman-Kolmogorov identity, and

$$\int_{\mathbb{R}^d_+} G^b(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d)\, d\xi = \int_{\mathbb{R}^d_+} G_0^b(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d) \mathrm{d}\xi = 1,$$

since $I G_k^b$ has means zero for any $k \ge 1$.

To complete this explicit calculation, we denote by Δ_0 the Dirac measure at the origin to have

$$I^k \varphi(x) = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-1} \frac{(c\,t)^k}{k!} \int_{\mathbb{R}^d_+} \varphi(x+y)\, \bar{\mu}^k(\mathrm{d}y),$$
$$\mu^0 := \delta_0, \qquad \mu^k(B) := \int_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \mathbb{1}_B(x+y)\, \mu(\mathrm{d}x)\, \mu^{k-1}(\mathrm{d}y),$$

for every $k \ge 1$, which implies

$$G^b(x,t,\xi_d) = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-1} \frac{(c\,t)^k}{k!}\, \mu^k G_0^b(x,t,\xi_d),$$
$$\mu^k G_0^b(x,t,\xi_d) := \int_{\mathbb{R}^d_+} G_0^b(x+y,t,\xi_d)\, \mu^k(\mathrm{d}y),$$

and interchanging the order of the summation we obtain

$$\begin{cases} G^b(x,t,\xi_d) = \mathrm{e}^{-ct} \displaystyle\sum_{k=0}^{\infty} \frac{(c\,t)^k}{k!}\, \mu^k G_0^b(x,t,\xi_d), \\ \mu^0 := \delta_0, \quad \mu^k(B) := \displaystyle\int_{\mathbb{R}^d_+ \times \mathbb{R}^d_+} \mathbb{1}_B(x+y)\, \mu(\mathrm{d}x)\, \mu^{k-1}(\mathrm{d}y), \\ \mu^k G_0^b(x,t,\xi_d) := \displaystyle\int_{\mathbb{R}^d_+} G_0^b(x+y,t,\xi_d)\, \mu^k(\mathrm{d}y), \end{cases} \tag{2.117}$$

for every $k \ge 1$, for any x in $\bar{\mathbb{R}}^d_+$, $\xi_d \ge 0$ and $t > 0$. Since μ is a probability measure on $\bar{\mathbb{R}}^d_+$, so is μ^k and the above series is clearly convergent as the initial one given by (2.112). These arguments complement the one dimensional examples.

Example 2.20 (oblique Wiener-Poisson)**.** This is a half-space oblique reflecting barrier in the direction of the vector $b = (b_1, \dots, b_d)$ with $b_d > 0$, for a standard Wiener process in $\mathbb{R}^d$ and a compound Poisson process with parameters $c > 0$

and μ, where μ is a distribution on the open semi-space $\mathbb{R}^d_+ := \mathbb{R}^{d-1} \times (0, \infty)$, with the notation $x = (\tilde{x}, x_d)$. In the state space $\bar{\mathbb{R}}^d_+ := \mathbb{R}^{d-1} \times [0, \infty)$, the closed semi-space, we consider the function G^b defined by (2.112) or (2.117). This yields a transition probability function on the state space $\bar{\mathbb{R}}^d_+$ with its Borel σ-algebra $\mathcal{B}$

$$p(t, x, B) := \int_B G^b(\tilde{x} - \tilde{\xi}, x_d, t, \xi_d) \mathrm{d}\xi,$$

for any $t > 0$, x in $\bar{\mathbb{R}}^d_+$ and B in $\mathcal{B}$. Expressions for the associated semigroup and its infinitesimal generator can be obtained, e.g.,

$$\begin{aligned} &\mathcal{D}(A) := \{f \in C_0(\bar{\mathbb{R}}^d_+) : \Delta f \in C_0(\bar{\mathbb{R}}^d_+),\ b \cdot \nabla f(\tilde{x}, 0) = 0\}, \\ &Af := \frac{1}{2}\Delta f + If, \end{aligned}$$

where the integral operator I is given by (2.111). A realization of a d-dimensional (standard) Wiener-Poisson process in $\bar{\mathbb{R}}^d_+$ with parameter $c > 0$ and μ, and oblique reflection at the barrier $x_d = 0$ is obtained from general existence theorems based on the above transition probability function. □

It is possible to use an integral operator I of the form

$$\begin{cases} I\varphi(x) := \displaystyle\int_{\mathbb{R}^d_+} [\varphi(x + y) - \varphi(x) - y \cdot \nabla\varphi(x)]\, m(\mathrm{d}y), \\ \forall x \in \mathbb{R}^d_+, \quad \text{with} \quad \displaystyle\int_{\mathbb{R}^d_+} \frac{|y|^2}{1 + |y|}\, m(\mathrm{d}y) < \infty. \end{cases} \tag{2.118}$$

The definition (2.112) of G^b still valid but not (2.117). Because of the *constant coefficients* we can make explicit calculations and $G^b_k = G^b_0 \star IG^b_{k-1} = (t^k/k!)I^k G^b_0$ but we need to work harder to show the convergence of the series. For instance, if we assume

$$\int_{\mathbb{R}^d_+} \frac{|y|^{2-\alpha}}{1 + |y|}\, m(\mathrm{d}y) < \infty, \quad \alpha \in (0, 2],$$

then the integral operator I maps the Green space $\mathcal{G}^2_k$ (kernel φ satisfying condition (2.114) for $\partial_\ell\varphi$ of order $k - |\ell|$, with $|\ell| \le 2$) into the Green space $\mathcal{G}^0_{k+\alpha}$, with appropriate estimates, see previous Section 2.12 and the books Garroni and Menaldi [93, 94] for details.

On the other hand, the spectral theory of compact operators can be used to give an eigenvalues and eigenfunction expansion of the Green function or Green operator as in the Sturm-Liouville case.

Chapter 3

Stochastic Processes

A general stochastic process with values in a Polish space (separable complete metric space) E, usually locally compact (e.g., closed subset of $\mathbb{R}^d$ or a one-point compactification of an open), on a probability space $(\Omega, \mathcal{F}, P)$ is a family of random variables (measurable mappings, almost surely defined), denoted by either $X = X(t)$ or $X = X_t$, from Ω into E indexed by a set T (usually a subset of real numbers). We suppress the qualifier general when a suitable regularity on the paths is imposed, i.e., at least separable and measurable. From the phenomenological viewpoint, a stochastic process is identified by prescribed properties given on the family of finite-dimensional distributions. A priori, a stochastic process can be viewed as a random variable X with values in the product topological space E^T, which is not a Polish space if E and T are uncountable. Mathematically, we add some regularity conditions (such as continuity) on the paths of the stochastic process ($X(\cdot, \omega)$, for each ω), so that we can consider a much smaller subset of E^T, with a structure of Polish space (non-locally compact), where the process actually lives. Usually, these Polish spaces are either $C([0, \infty), E)$, the space of E-valued continuous functions, or $D([0, \infty), E)$, the space of cad-lag functions as described in section 1.12. Actually, we use always the cad-lag regular form of a stochastic process, that is, given a stochastic process $\{Y(t) : t \in T\}$ with an uncountable (index) subset $T \subset [0, \infty)$, the *cad-lag regularisation* (usually this is also a version of $Y(t)$) is the stochastic process X given by $X(t, \omega) = \lim_{s \in \mathbb{Q}, s \to t} Y(t, \omega)$ and defined for any t in T, where the limits from the right and from the left (on the rational or dyadic-rational numbers $\mathbb{Q}$) exist finitely. Therefore, we say that the stochastic process Y is *regularisable* if the cad-lag regularisation is defined for any t in T. It can be proved that a process is regularisable if and only if the process and its (rational) up-crossings are locally bounded, moreover, the set of regularisable stochastic processes is measurable, see Rogers and Williams [214, Chapter 2, Section 5, pp. 163–166].

We are going to deal only with stochastic processes admitting a cad-lag version, unless otherwise stated. Thus for the initial (nominal or reference) stochastic process, we take a canonical realization, i.e., a complete probability space $(\Omega, \mathcal{F}, P)$ and a random variable X with values in the Polish space

$D([0,\infty),E)$. Usually, we may want to set $\Omega = D([0,\infty),E)$, $\mathcal{F} = \mathcal{B}(\Omega)$, $X(t) = \omega(t)$, $X_-(t) = \omega(t-)$, and use the (complete) σ-algebras $\{\mathcal{F}(t),\, t \in T\}$ and $\{\mathcal{F}_-(t),\, t \in T\}$ generated by the family of E-valued random variables $\{X(s),\, s \le t,\, s \in T\}$ and $\{X_-(s),\, s \le t,\, s \in T\}$, respectively. Most of the times, $E \subset \mathbb{R}^d$ and P is the (unique) Wiener measure on $D([0,\infty),\mathbb{R}^d)$ with support in $C([0,\infty),\mathbb{R}^d)$. Other processes Y, Z, ... are regarded as random variables with values in the (non-locally compact) Polish space $D([0,\infty),E)$ or with values in $C([0,\infty),E)$ if possible. Sometimes, these processes are regarded also as a maps from $\Omega \times T$ into E. For instance, an increasing process Y will have also the property (on the paths) that $t \mapsto Y(t,\omega)$ is increasing, for every ω in $\Omega \setminus N$, with $P(N) = 0$. Thus, unless otherwise stated, a *E-valued stochastic process* is a measurable function from a complete probability space $(\Omega,\mathcal{F},P)$ into the canonical space $D([0,\infty),E)$, i.e., a $D([0,\infty),E)$-valued random variable, which imposes the cad-lag regularity on paths. In any case, it is implicitly assumed that we have choose a probability space $(\Omega,\mathcal{F},P)$ where regular conditional probability exists.

Our main interest is (strong) Markov processes admitting a cad-lag realization and having a (continuous) transition function. These stochastic processes are identified by (1) a prescribed (continuous) transition function, or (2) a given (continuous) Markov-Feller semigroup, which in turn can be obtained by its infinitesimal generator. With the previous introduction, the purpose of this chapter is to discuss (in some detail) certain topics in *stochastic analysis* as a tool to describe (or generate) Markov processes. Certainly, it is implicity assume a minimum understanding of probability, e.g., see Bremaud [32, Appendix A1, pp. 255–295].

3.1 Discrete Time

To motivate some delicate points in the theory of continuous time processes we discuss first *sequences of random variables*, i.e., random processes in discrete time. First, a filtered space is a complete probability space $(\Omega,\mathcal{F},P)$ and an increasing sequence (so-called *filtration*) of sub σ-algebras $(\mathcal{F}_n : n = 0,1,\ldots)$, $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, for all $n = 1,2,\ldots$, such that $\mathcal{F}_0$ contain all null sets. A stochastic sequence (or process) $(X_n : n = 0,1,\ldots)$ is a sequence of $\mathbb{R}$-valued (or $\mathbb{R}^d$-valued) random variables, 'identified' almost surely (i.e., P-equivalence class). Its associated natural filtration is the sequence $(\mathcal{F}_n : n = 0,1,\ldots)$ of sub σ-algebras generated by $\{X_0,X_1,\ldots,X_n\}$ and augmented with all null sets, i.e., $\sigma[X_0,X_1,\ldots,X_n]$ and all null sets. Given a filtered space, a stochastic sequence (or process) $(X_n : n = 0,1,\ldots)$ is called *adapted* if every random variable X_n is $\mathcal{F}_n$-measurable. Also, it is called *predictable* if every random variable X_n is $\mathcal{F}_{n-1}$-measurable, for any $n = 1,2,\ldots$, here X_0 is ignored or taken equal to zero. A stopping time η is a maps (identified almost surely) from Ω into the set $\{0,1,\ldots,\infty\}$ such that $\{\eta \le n\}$ (or equivalently $\{\eta = n\}$) belongs to $\mathcal{F}_n$ for any $n \ge 0$, where $\mathcal{F}_\infty := \mathcal{F}$. For an given stopping time, the σ-algebra $\mathcal{F}_\eta$ is defined as the collection of all subsets A in $\mathcal{F}$ such that $A \cap \{\eta \le n\}$ (or equivalently

$A \cap \{\eta = n\}$) belongs to $\mathcal{F}_n$, for any $n \geq 0$. Note that a typical stopping time is the *hitting time* (or entry time) of a Borel subset B of $\mathcal{B}(\mathbb{R})$ (or $\mathcal{B}(\mathbb{R}^d)$) for a stochastic sequence $(X_n : n = 0, 1, \ldots)$, i.e., $\eta := \inf\{n \geq 0 : X_n \in B\}$, where $\eta = \infty$ if X_n does not belong to B for any $n \geq 0$.

Similar to Kolmogorov's extension theorem (see Theorem 1.30 in Chapter 1) we can mention the following construction of the direct product of probability spaces (e.g., Halmos [104, Section VII.38, Theorem B, pp. 157–158]), namely, there exists a unique probability measure P on the (countable) product space $\Omega = \prod_n \Omega_n$ with the product σ-algebra $\mathcal{F}$ (generated by the collection of cylindrical sets $C_n = \prod_{k=1}^n F_k \times \prod_{k=n+1}^\infty \Omega_k$, with F_k in $\mathcal{F}_k$,) such that $P(C_n) = \prod_{k=1}^n P_k(F_k)$ for every cylindrical set. Note that the countable assumption is really not an issue, it can be easily dropped.

A direct consequence of the above result is the construction of sequences of independent and identically distributed $\mathbb{R}^d$-valued random variables, i.e., given a distribution μ on $\mathbb{R}^d$ the exists a stochastic sequence $(Z_n : n = 0, 1, \ldots)$ on a complete probability space $(\Omega, \mathcal{F}, P)$ such that

$$(1) \qquad P(Z_k \in B) = \mu(B), \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

$$(2) \qquad P(Z_k \in B_k,\ \forall k = 1, \ldots, n) = \prod_{k=1}^n P(Z_k \in B_k),$$

for every B_k in $\mathcal{B}(\mathbb{R}^d)$ and any $n \geq 1$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra in $\mathbb{R}^d$. Thus, the series of partial sum $X_0 := 0$, $X_n := \sum_{k=1}^n Z_k$ is called a *random walk* in $\mathbb{R}^d$ or a d-dimensional random walk with incremental distribution μ.

The reader is also referred to Tulcea's theorem (e.g., Neveu [188, Section V.1, pp. 153–159], Shiryayev [227, Section II.9, Theorem 2, pp. 243–250]), which is specially designed for construction of Markov chains (processes) from transition functions. To present this result on *product probability*, we need some notation. First, a *transition probability* between two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ is a function $Q : \Omega \times \mathcal{F}' \to [0, 1]$, $Q(\omega, F')$, which is measurable in ω and a probability in F'. Note two particular cases, (1) $Q(\omega, F') = P(F')$ a fixed probability on $(\Omega', \mathcal{F}')$ for every ω in Ω, and (2) $Q(\omega, F') = \mathbb{1}_{\{q(\omega) \in F'\}}$ where $q : \Omega \to \Omega'$ is a measurable function.

For $(\Omega_i, \mathcal{F}_i)$ a sequence of measurable spaces, the product σ-algebra $\mathcal{F} = \prod_{i=1}^\infty \mathcal{F}_i$ on the product space $\Omega = \prod_{i=1}^\infty \Omega_i$ is generated by the cylindrical sets

$$C_n := \prod_{i=1}^n F_i \times \prod_{i=n+1}^\infty \Omega_i, \quad \text{with} \quad F_i \in \mathcal{F}_i,\ \forall i, \quad n = 1, 2, \ldots \tag{3.1}$$

For a fixed n, denote by $\mathcal{F}^n$ a sub σ-algebra of $\mathcal{F}$ generated by the n-cylindrical sets as above. It is clear that $\mathcal{F}^n$ can be identified with the σ-algebra $\prod_{i=1}^n \mathcal{F}_i$ of finite product space $\prod_{i=1}^n \Omega_i$, and that $\mathcal{F}$ is generated by the algebra $\cup_n \mathcal{F}^n$.

Let P_1 be a probability on $(\Omega_1, \mathcal{F}_1)$ and Q_k be a transition probability from finite product space $(\prod_{i=1}^{k-1} \Omega_i, \prod_{i=1}^{k-1} \mathcal{F}_i)$ into $(\Omega_k, \mathcal{F}_k)$, for $k \geq 2$. We desire to construct a probability P on the infinite product space $(\Omega, \mathcal{F})$ such that

$$P(C_n) = \int_{F_1} P_1(\mathrm{d}\omega_1) \int_{F_2} Q_2(\omega_1, \mathrm{d}\omega_2) \ldots \int_{F_n} Q_n(\omega_1, \ldots, \omega_{n-1}, \mathrm{d}\omega_n),$$

for any cylindrical set C_n as in (3.1). Note that if P_n denotes the restriction of P to $\prod_{i=1}^n \mathcal{F}_i$ (i.e., the finite-dimensional distributions of P) then the right-hand term prescribes a particular form for P_n, where a disintegration (by means of the transition probability Q_n) is assumed a priori. Comparing with Kolmogorov's extension theorem, we assume that the finite-dimensional distributions enjoy a disintegration condition, instead of a topological assumption in the spaces Ω_i.

Now, for a fixed n, consider the following expression constructed backward by induction:

$$
\begin{aligned}
&P(\omega_1,\dots,\omega_n;F) = \mathbb{1}_{F^n}(\omega_1,\dots,\omega_n), \quad F = F^n \times \prod_{i=n+1}^{\infty} \Omega_i,\ F^n \in \prod_{i=1}^{n} \mathcal{F}_i,\\
&P(\omega_1,\dots,\omega_{k-1};F) = \int_{\Omega_k} P(\omega_1,\dots,\omega_{k-1},\omega_k;F)\, Q_k(\omega_1,\dots,\omega_{k-1},\mathrm{d}\omega_k),\\
&P(\omega_1;F) = \int_{\Omega_2} P(\omega_1,\omega_2;F)\, Q_2(\omega_1,\mathrm{d}\omega_2),\\
&P(F) = \int_{\Omega_1} P(\omega_1;F)\, P_1(\mathrm{d}\omega_1).
\end{aligned}
$$

A Fubini-Tonelli type theorem ensures that each step of the above construction is possible and that $P(\omega_1,\dots,\omega_k;F)$ is a transition probability from the (finite) product space $(\prod_{i=1}^k \Omega_i, \prod_{i=1}^k \mathcal{F}_i)$ into $(\Omega,\mathcal{F}^n)$, for any $k = n,\dots,1$; and that $P(F)$ is a probability on $(\Omega,\mathcal{F}^n)$. It is also clear that for cylindrical sets as (3.1) we have

$$
\begin{aligned}
&P(C_n) = \int_{F_1} P_1(\mathrm{d}\omega_1) \int_{F_2} Q_2(\omega_1,\mathrm{d}\omega_2) \dots \int_{F_n} Q_n(\omega_1,\dots,\omega_{n-1},\mathrm{d}\omega_n),\\
&P(\omega_1,\dots,\omega_{k-1};F) = \Big(\prod_{i=1}^{k-1} \mathbb{1}_{F_i}(\omega_i)\Big) \int_{F_k} Q_k(\omega_1,\dots,\omega_{k-1},\mathrm{d}\omega_k) \times\\
&\quad \times \int_{F_{k+1}} Q_{k+1}(\omega_1,\dots,\omega_{k-1},\omega_k,\mathrm{d}\omega_{k+1}) \dots \int_{F_n} Q_n(\omega_1,\dots,\omega_{n-1},\mathrm{d}\omega_n),\\
&P(\omega_1,\dots,\omega_n;C_n) = \prod_{i=1}^{n} \mathbb{1}_{F_i}(\omega_i),
\end{aligned}
$$

and therefore, $P(\omega_1,\dots,\omega_n;F) = P(\omega_1,\dots,\omega_{n-1};F)$ for any F in $\mathcal{F}^{n-1}$. This last property allows us to consider $n = 1,2,\dots$ and to extend (uniquely) the definition of $P(\omega_1,\dots,\omega_n;F)$ to F in the algebra $\cup_n \mathcal{F}^n$.

Theorem 3.1 (Tulcea)**.** *Under the above notation, the function $P_n(\omega,F) = P(\omega_1,\dots,\omega_n;F)$, with $\omega = (\omega_1,\dots,\omega_n,\dots)$, is a transition probability from $(\Omega,\mathcal{F}^n)$ into $(\Omega,\mathcal{F})$. Moreover $(\Omega,\mathcal{F},P)$ is a probability space on which P_n provides a regular conditional probability for $\mathcal{F}^n$.* □

Proof. Only a brief idea is given. The central point is show the σ-additivity of P_n on the algebra $\cup_n \mathcal{F}^n$ with $P_0 = P$, and then to use Caratheodory extension to have a probability on $\mathcal{F}$. To this purpose, suppose that there exists a decreasing sequence $\{A_k\}$ in $\cup_n \mathcal{F}^n$ such that $\cap_k A_k = \emptyset$ with $\lim_k P(A_k) \neq 0$. Then, the above construction of the P_1 show that there exists a ω_1^* such that

$\lim_k P(\omega_1^*; A_k) \neq 0$, and by induction, we can construct a sequence $\omega^* = (\omega_1^*, \dots, \omega_n^*, \dots)$ such that $\lim_k P(\omega_1^*, \dots, \omega_n^*; A_k) \neq 0$. Since A_k belongs to some $\mathcal{F}^m$ with $m = m(k)$, from the construction of P we obtain $P(\omega_1^*, \dots, \omega_n^*; A_k) = \mathbb{1}_{A_k}(\omega^*)$ if $n \geq m(k)$. Hence ω belongs to A_k for every k, which is a contradiction. □

First let us discuss Kolmogorov's extension theorem (see Theorem 1.30 in Chapter 1) in a general product space $\Omega = \prod_{t\in T} \Omega_t$, $\mathcal{F} = \prod_{t\in T} \mathcal{F}_t$. We assume given a family of finite-dimensional distributions $\{P_s : s \in T^n,\ n = 1, 2, \dots\}$ on $\Omega^s = \Omega_{t_1} \times \cdots \times \Omega_{t_n}$, with $s = (t_1, \dots, t_n)$ which satisfies the consistency conditions, namely

(a) if $s = (s_{i_1}, \dots, s_{i_n})$ is a permutation of $t = (t_1, \dots, t_n)$ then for any B_i in $\mathcal{F}_{t_i}$, $i = 1, \dots, n$, we have $P_t(B_1 \times \cdots \times B_n) = P_s(B_{i_1} \times \cdots \times B_{i_n})$,

(b) if $t = (t_1, \dots, t_n, t_{n+1})$ and $s = (t_1, \dots, t_n)$ and B in $\mathcal{F}^s = \mathcal{F}_{t_1} \times \cdots \times \mathcal{F}_{t_n}$ then $P_t(B \times \Omega_{t_{n+1}}) = P_s(B)$.

If a total order is given on the index set T, it is enough to have the finite-dimensional distributions defined only for $(s_1, s_2, \dots, s_n)$ such that $s_1 < s_2 < \cdots < s_n$ and to satisfy only a consistency conditions of the form

(b') if $t = (t_1, \dots, t_n)$ and $s = (s_1, \dots, s_m)$ with $t_1 < \cdots < t_n < r < s_1 < \cdots < s_m$ and $A \times B$ in $\mathcal{F}^t \times \mathcal{F}^s$ then $P_{(t,r,s)}(A \times \Omega_r \times B) = P_{(t,s)}(A \times B)$, for any $n, m = 0, 1, \dots$.

Consistency along is not sufficient to ensure the existence of a probability P defined on $(\Omega, \mathcal{F})$ such that P_s be the restriction (or image trough the projection) of P over $(\Omega^s, \mathcal{F}^s)$. Some sort of topology is necessary on Ω_t so that P_s results inner regular (e.g., see Doob [60, pp. 403, 777], Neveu [188, Section III.3, pp. 74–81]), for instance, if Ω_t is a Lusin space (i.e., Ω_t is homeomorphic to a Borel subset of a compact metrizable space) and $\mathcal{F}_t = \mathcal{B}(\Omega_t)$ its Borel σ-algebra then every probability measure is inner regular. Under these conditions, the construction of the measure P is possible.

It is interesting to note that there is almost no difficulty to extend Tulcea's construction to a general product space with an index non necessarily countable. Indeed, we assume that P_s, with $s = (t_1, \dots, t_n)$, has the form

$$P_s(C_n) = \int_{F_1} P_{t_1}(\mathrm{d}\omega_1) \int_{F_2} Q_{t_1,t_2}(\omega_1, \mathrm{d}\omega_2) \dots \int_{F_n} Q_{t_1,\dots,t_n}(\omega_1, \dots, \omega_{n-1}, \mathrm{d}\omega_n),$$

for some family of transition probabilities $\{Q_s : s = (s', t),\ s' \in T^{n-1},\ n \geq 2,\ t \in T\}$ from $(\Omega^{s'}, \mathcal{F}^{s'})$ into $(\Omega_t, \mathcal{F}_t)$, and any cylindrical set $C_n = \prod_{t\in T} F_t$ with $F_t = \Omega_t$ if $t \neq t_i$ for every i, and $F_{t_i} \in \mathcal{F}_{t_i}$. Hence, we can construct a family of consistent probability on any countable product. Since only a countable number of finite-dimensional is involved in proving the σ-additivity, we have a probability in general product space Ω. Thus, the disintegration of the finite-dimensional distributions in term of the transition probabilities $\{Q_s : s \in T^n,\ n \geq 2\}$ replace the extra condition on inner regular measures. Moreover, Tulcea's construction

yields an expression for a regular conditional distribution on any countable subset of indexes.

A very useful and well know result is the following

Lemma 3.2 (Borel-Cantelli)**.** *Let $(F_1, F_2, \ldots)$ a sequence of measurable sets in a probability space $\Omega, \mathcal{F}, P$). (1) If $\sum_n P(F_n) < \infty$ then $P(\cap_n \cup_{k\geq n} F_k) = 0$. (2) If $(F_1, F_2, \ldots)$ is independent and $\sum_n P(F_n) = \infty$ then $P(\cap_n \cup_{k\geq n} F_k) = 1$.* □

It may be worthwhile to recall that *independence* is stable under weak convergence, i.e., if a sequence $(\xi_1, \xi_2, \ldots)$ of $\mathbb{R}^d$-valued random variables converges weakly (i.e., $\mathbb{E}\{f(\xi_n)\} \to \mathbb{E}\{f(\xi)\}$ for any bounded continuous function) to a random variable ξ then the coordinates of ξ are independent if the coordinates of ξ_n are so. On the other hand, for any sequence $(\mathcal{F}_1, \mathcal{F}_2, \ldots)$ of σ-algebras the *tail* or terminal σ-algebra is defined as $\mathcal{F}_{tail} := \cap_n \vee_{k\geq n} \mathcal{F}_k$, where $\vee_{k\geq n}\mathcal{F}_k$ is the smaller σ-algebra containing all σ-algebras $\{\mathcal{F}_k : k \geq n\}$. An important fact related to the independence property is the so-called Kolmogorov's zero-one law, which states that any tail set (that is measurable with respect to a tail σ-algebra) has probability 0 or 1.

Another typical application of Borel-Cantelli lemma is to deduce almost surely convergence from convergence in probability, i.e., if a sequence $\{x_n\}$ converges in probability to x (i.e., $P\{|x_n - x| \geq \varepsilon\} \to 0$ for every $\varepsilon > 0$) with a stronger rate, namely, the series $\sum_n P\{|x_n - x| \geq \varepsilon\} < \infty$, then $x_n \to x$ almost surely.

A key tool to study sequences of random variables is the martingale concept.

Definition 3.3 (discrete martingale)**.** A stochastic sequence $(X_n : n = 0, 1, \ldots)$ is called a *martingale* relative to a filtration $(\mathcal{F}_n : n = 0, 1, \ldots)$ if

$$\mathbb{E}\{|X_n|\} < \infty, \ \forall n, \ \text{ and } \ \mathbb{E}\{X_n \,|\, \mathcal{F}_{n-1}\} = X_{n-1}, \ \text{ a.s., } n \geq 1.$$

A *super or sub martingale* is defined similarly, replacing the equal sign $=$ by the $\leq$ or $\geq$ signs, respectively. □

Note that X_n turns out to be $\mathcal{F}_n$-measurable and it is determined almost surely, actually we take X_n as a $\mathcal{F}_n$-measurable function defined everywhere. If only the complete probability space $(\Omega, \mathcal{F}, P)$ is given, then the filtration $(\mathcal{F}_n : n = 0, 1, \ldots)$ is naturally generated by the stochastic sequence $(X_n : n = 0, 1, \ldots)$, i.e., $\mathcal{F}_n$ is the smallest sub σ-algebra of $\mathcal{F}$ containing all null sets and rendering measurable the random variables $\{X_0, X_1, \ldots, X_n\}$. A super-martingale *decreases on average* while a sub-martingale *increases on average*. Since X_0 is integrable, we may focus our attention on sequences with $X_0 = 0$. A typical example of martingale is a real valued random walk or $\mathbb{R}^d$-valued random walk since (super-/sub-) martingales can be defined by coordinates when dealing with $\mathbb{R}^d$-valued random variables. Also, if φ is a convex and increasing real-valued function such that $\mathbb{E}\{\varphi(X_n)\} < \infty$ for some sub-martingale $(X_n : n = 0, 1, \ldots)$ then the stochastic sequence $(\varphi(X_n) : n = 0, 1, \ldots)$ is also a sub-martingale.

In most cases, the filtration $\mathcal{F}_n$ is generated by another sequence of random variables $\{Y_0, Y_1, \ldots\}$, i.e., $\mathcal{F}_n = \sigma[Y_0, \ldots, Y_n]$, which is regarded as the history. In this case, $X_n = h_n(Y_0, \ldots, Y_n)$ for some Borel function $h_n \colon \mathbb{R}^{n+1} \to \mathbb{R}$, e.g., see Karr [127].

Many important results are found in the study of martingales, related to estimates and representation, we will mention only some of them. For Doob's upcrossing estimate, denote by $U_N(X, [a, b])$ the number of up-crossings of $[a, b]$ by time N for a fixed ω, i.e., the largest k such that $0 \le s_1 < t_1 < \cdots < s_k < t_k \le N$, $X_{s_i} < a$ and $X_{t_i} > b$, for any $i = 1, 2, \ldots k$. Then for any super-martingale the estimate

$$(b - a)\, \mathbb{E}\{U_N(X, [a, b])\} \le \mathbb{E}\{(X_N - a)^-\} \tag{3.2}$$

holds. Note that the number of steps does not appear directly on the right-hand side, only the final variable X_N is relevant. To show this key estimate, by induction, we define $C_1 := \mathbb{1}_{X_0 < a}$, i.e., $C_1 = 1$ if $X_0 < a$ and $C_1 = 0$ otherwise, and for $n \ge 2$,

$$C_n := \mathbb{1}_{C_{n-1}=1}\, \mathbb{1}_{X_{n-1} \le b} + \mathbb{1}_{C_{n-1}=0}\, \mathbb{1}_{X_{n-1} < a}$$

to construct a bounded nonnegative super-martingale $Y_n := \sum_{k=1}^{n} C_k (X_k - X_{k-1})$. Clearly, the sequence $(C_n : n = 1, 2, \ldots)$ is predictable. Based on the inequality

$$Y_N \ge (b - a)\, U_N(X, [a, b]) - [X_N - a]^-,$$

for each ω, the estimate (3.2) follows.

The Doob's super-martingale convergence states that for a super martingale $(X_n : n = 0, 1, \ldots)$ bounded in L^1, i.e., $\sup_n |X_n| < \infty$ the limits $X_\infty := \lim_n X_n$ exists almost surely. The convergence is in L^1 if and only if the sequence $(X_n : n = 0, 1, \ldots)$ is uniformly integrable, and in this case we have $\mathbb{E}\{X_\infty \,|\, \mathcal{F}_n\} \le X_n$, almost surely, with the equality for a martingale. To prove this convergence, we express the set Ω_0 of all ω such that the limit $\lim_n X_n(\omega)$ does not exist in the extended real number $[-\infty, +\infty]$ as a countable union of subsets $\Omega_{a,b}$ where $\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)$, for any rational numbers $a < b$. By means of the upcrossing estimate (3.2) we deduce

$$\Omega_{a,b} \subseteq \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega : U_n(X, [a, b]) > m\},$$
$$P\Big(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega : U_n(X, [a, b]) > m\}\Big) = 0,$$

which yields $P(\Omega_0) = 0$. Thus the limit exists in $[-\infty, +\infty]$ and by Fatou's Lemma, it is finite almost surely.

If $p > 1$ and $(X_n : n = 0, 1, \ldots)$ is a nonnegative sub-martingale bounded in L^p then Doob's L^p inequality reads as follows

$$\|\sup_n X_n\|_p \le p' \sup_n \|X_n\|_p, \qquad \text{with } 1/p + 1/p' = 1, \tag{3.3}$$

where $\|\cdot\|_p$ denotes the in $L^p := L^p(\Omega, \mathcal{F}, P)$. Note that $(p')^p \le 4$ for every $p \ge 2$. Indeed, if the set Ω_n^r of all ω where $\sup_{k\le n} X_k \ge r$ is expressed as the disjoint union

$$\Omega_n^r = \bigcup_{k=0}^{n} \Omega_{n,0}^r \quad \text{with}$$
$$\Omega_{n,k}^r := \{X_0 < r\} \cap \{X_1 < r\} \cap \cdots \{X_{k-1} < r\} \cap \{X_k \ge r\},$$

and $\Omega_{n,0}^r := \{X_0 \ge r\}$, then we have $X_k \ge r$ on $\Omega_{n,k}^r$, which yields the Doob's maximal inequality

$$r\, P\big(\sup_n X_n \ge r\big) \le \mathbb{E}\{X_n \mathbb{1}_{\sup_n X_n \ge r}\} \le \mathbb{E}\{X_n\}.$$

Using the above estimate for $|X_n|^p$ instead of X_n, we obtain the p-estimate (3.3).

The Doob's decomposition gives a clean insight into martingale properties. Let $(X_n : n = 0, 1, \ldots)$ be a stochastic sequence of random variables in L^1, and denote by $(\mathcal{F}_n : n = 0, 1, \ldots)$ its natural filtration, i.e., $\mathcal{F}_n := \sigma[X_0, X_1, \ldots, X_n]$. Then there exists a martingale $(M_n : n = 0, 1, \ldots)$ relative to $(\mathcal{F}_n : n = 0, 1, \ldots)$ and a predictable sequence $(A_n : n = 0, 1, \ldots)$ with respect to $(\mathcal{F}_n : n = 0, 1, \ldots)$ such that

$$X_n = X_0 + M_n + A_n, \ \forall n, \quad \text{and} \quad M_0 = A_0 = 0. \tag{3.4}$$

This decomposition is unique almost surely and the stochastic sequence $(X_n : n = 0, 1, \ldots)$ is a sub-martingale if and only if the stochastic sequence $(A_n : n = 0, 1, \ldots)$ is monotone increasing, i.e., $A_{n-1} \le A_n$ almost surely for any n. Indeed, define the stochastic sequences $(A_n : n = 1, \ldots)$ by

$$A_n := \sum_{k=1}^{n} \mathbb{E}\{X_k - X_{k-1} \,|\, \mathcal{F}_{k-1}\}, \quad \text{with} \ \ \mathcal{F}_k := \sigma[X_0, X_1, \ldots, X_k]$$

and $(M_n : n = 1, \ldots)$ with $M_n := X_n - X_0 - A_n$ to obtain the decomposition (3.4). This implies that the only *deterministic* martingale is a constant.

Given a martingale $M = (M_n : n = 0, 1, \ldots)$ with each M_n in L^2 and $M_0 = 0$, we may use the above decomposition to express the sub-martingale $M^2 = (M_n^2 : n = 0, 1, \ldots)$ as $M^2 = N + A$, where $N = (N_n : n = 0, 1, \ldots)$ is a martingale and $A = (A_n : n = 0, 1, \ldots)$ is a predictable increasing sequence, both N and A null at $n = 0$. The stochastic sequence A is written as $\langle M \rangle$ and called the *angle-brackets* sequence of M. Note that

$$\mathbb{E}\{M_n^2 - M_{n-1}^2 \,|\, \mathcal{F}_{n-1}\} = \mathbb{E}\{(M_n - M_{n-1})^2 \,|\, \mathcal{F}_{n-1}\} = A_n - A_{n-1},$$

for every $n \ge 1$. Similarly, define the stochastic sequence (of quadratic variation)

$$[M]_n := \sum_{k=1}^{n} (M_k - M_{k-1})^2, \quad \forall n \ge 1,$$

and $[M]_0 = 0$. Then the stochastic sequence $V = (V_n : n = 1, 2, \ldots)$,

$$V_n := M_n^2 - [M]_n = \sum_{k=1}^{n} 2M_{k-1}\, M_k$$

is a martingale. Note that $[M]$ is an adapted sequence while $\langle M \rangle$ is predictable, so the strength of the Doob's decomposition. It is clear that

$$\mathbb{E}\{|M_n|^2\} = \mathbb{E}\{\langle M \rangle_n\} = \mathbb{E}\{[M]_n\}, \quad \forall n \geq 1,$$

which combined with the p-estimate (3.3), $p = 2$, yields

$$\mathbb{E}\{\sup_{k \leq n} |M_k|^2\} \leq 4 \sup_{k \leq n} \mathbb{E}\{\langle M \rangle_k\}, \quad \forall n \geq 1.$$

Actually, this generalize into the following Davis-Burkhölder-Gundy inequality

$$c_p\, \mathbb{E}\{([M]_n)^{p/2}\} \leq \mathbb{E}\{\sup_{k \leq n} |M_k|^p\} \leq C_p\, \mathbb{E}\{([M]_n)^{p/2}\}, \tag{3.5}$$

valid for any $n \geq 1$ and $p > 0$ and some constants $C_p > c_p > 0$ independent of the martingale $(M_n : n = 0, 1, \ldots)$. Even for $p = 1$, we may use $C_1 = 3$ in the right-hand side of (3.5). Moreover, the L^2-martingale $(M_n : n = 0, 1, \ldots)$ may be only a *local martingale* (i.e., there exists a sequence of stopping times $\eta = (\eta_k : k = 0, 1, \ldots)$ such that $M^{\eta,k} := (M_n^{\eta,k} : n = 0, 1, \ldots)$, defined by $M_n^{\eta,k}(\omega) := M_{n \wedge \eta_k(\omega)}(\omega)$, is a martingale for any $k \geq 0$ and $\eta_k \to \infty$ almost surely), the time n may be replaced by a stopping time η (or ∞), the angle-brackets $\langle M \rangle$ can be used in lieu of $[M]$, and the above inequality holds true. All these facts play an important role in the continuous time case.

Let $X = (X_n : n = 0, 1, \ldots)$ be a sub-martingale with respect to $(\mathcal{F}_n : n = 0, 1, \ldots)$ and uniformly integrable, i.e., for every ε there exists a sufficiently large $r > 0$ such that $P(|X_n| \geq r) \leq \varepsilon$ for any $n \geq 0$. Denote by $A := (A_n : n = 0, 1, \ldots)$ and $M := (M_n : n = 0, 1, \ldots)$ the predictable and martingale sequences given in the decomposition (3.4), $X_n = X_0 + M_n + A_n$, for all $n \geq 0$. Since X is a sub-martingale, the predictable sequence A is monotone increasing. The Doob's optional sampling theorem implies that the martingale M is uniformly integrable, moreover $A_\infty := \lim_n A_n$ is integrable and the families of random variable $\{X_\eta : \eta \text{ is a stopping}\}$ and $\{M_\eta : \eta \text{ is a stopping}\}$ are uniformly integrable. Furthermore, for any two stopping times $\eta \leq \theta$ we have

$$\mathbb{E}\{M_\theta \,|\, \mathcal{F}_\eta\} = M_\eta, \quad \text{a.s.} \quad \text{and} \quad \mathbb{E}\{X_\theta \,|\, \mathcal{F}_\eta\} \geq X_\eta, \quad \text{a.s.} \tag{3.6}$$

We skip the proof (easily found in the references below) of this fundamental results. Key elements are the convergence and integrability of the limit $M_\infty := \lim_n M_n$ (almost surely defined), which allow to represent M_n as $\mathbb{E}\{M_\infty \,|\, \mathcal{F}_n\}$. Thus, specific properties of the conditional expectation yield the result.

For instance, the reader is referred to the books Bremaud [32], Chung [43], Dellacherie and Meyer [58, Chapters I–IV], Doob [59, 61], Karlin and Taylor [125, 126], Nelson [187], Neveu [189], Williams [253], among others.

3.2 Filtered Spaces

We have seen that the study of stochastic processes leads to the discussion on probability measures in separable complete metric spaces, i.e., in Polish spaces (recall that a countably product of Polish spaces is a Polish space with the product topology and that any Borel set of a Polish space is a continuous image of the product Polish space $\mathbb{N}^\infty$, sequences of nonnegative integer numbers). Natural models for stochastic dynamical systems are the Markov processes. Since past, present and future information are represented by sets of events, a systematic analysis of families of σ-algebras is necessary, see Section 1.9. Thus, the starting point is a (complete) probability space with either a filtration or a process with its natural filtration.

Remark that all properties concerning a Filtered space are not preserved accross any possible version of a given process, e.g., if X and Y are versions of the same process (namely, $P\{X(t) \neq Y(t)\} = 0$ for every $t \geq 0$) then X may be adapted to the filtration but not Y. In this way, for a given process X we can take any version, but as soon as a filtration is involved, the given version kept. However, changing the processes in an evanescent set is fine, since the filtration is complete.

Definition 3.4. Given a *filtered space* $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$, this is a (usually complete) probability space $(\Omega, \mathcal{F}, P)$ and a (standard) filtration $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$, i.e., $\mathcal{F}(0)$ contains all P-negligible (or P-null) sets in $\mathcal{F}$ (complete) and $\mathcal{F}(t) = \bigcap_{\varepsilon>0} \mathcal{F}(t+\varepsilon)$ (right-continuous). Sometimes, this is also called *stochastic basis*, see Jacod and Shiryaev [117, Chapter 1].
(a) The σ-algebra $\mathcal{M}$ of *progressively measurable* sets is composed by all subsets A of $\Omega \times [0,\infty)$ such that $A \cap (\Omega \times [0,t])$ belongs to $\mathcal{F}(t) \times \mathcal{B}([0,t])$ for every $t \geq 0$.
(b) The σ-algebra $\mathcal{O}$ of *optional* or *well-measurable* sets is generated by sets of the form $F_0 \times \{0\}$ and $F \times [s,t)$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$.
(c) The σ-algebra $\mathcal{P}$ of *predictable* sets is generated by sets of the form $F_0 \times \{0\}$ and $F \times (s,t]$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$. □

Note that sometimes the variables t and ω may be exchanged so that the σ-algebras $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ are regarded as defined on $[0,\infty)\times\Omega$ instead of $\Omega\times[0,\infty)$. As long as no confusion arrives, we will ignore this fact.

It may be convenient to use the notation $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$ for a filtration and assume that $\mathcal{F}$ is the minimal σ-algebra containing all sets belonging to $\mathcal{F}(t)$ for some $t \geq 0$, so that $(\Omega, \mathbb{F}, P)$ denotes a stochastic basis. If a given filtration $(\mathcal{F}_0(t) : t \geq 0)$ does not satisfy the usual conditions of completeness and right-continuity then its *usual augmentation* $(\mathcal{F}(t) : t \geq 0)$ is defined as the minimal filtration satisfying $\mathcal{F}_0(t) \subset \mathcal{F}(t)$ for all t, plus the usual conditions. The σ-algebra $\mathcal{F}(t)$ can be constructed in two steps, first $\mathcal{F}_0(t)$ is completed with all null sets to a new σ-algebra $\bar{\mathcal{F}}_0(t)$ and second $\bar{\mathcal{F}}_0(t)$ is made right-continuous by defining $\mathcal{F}(t) = \bigcap_{s>t} \bar{\mathcal{F}}_0(s)$. Thus, for any F in $\mathcal{F}(t)$ there exists a F_0 in $\mathcal{F}_0(t+) = \bigcap_{s>t} \mathcal{F}_0(s)$ such that $F \Delta F_0 = (F \setminus F_0) \cup (F_0 \setminus F)$ is a null set.

Completing the family of increasing σ-algebras to become a filtration (satisfying the usual condition) is a routing task, however this is an important issue when dealing with the *strong Markov property* as discussed later on. The passage from $\mathcal{F}_0(t)$ to $\mathcal{F}_0(t+)$ is a very technical matter that we have to deal when the realization or simulation of a processes is studied. For instance, it can be easily proved that the completion of history of a Lévy process (or of a *right-constant* process) is actually right-continuous, e.g., see Bremaud [32, Appendix A2.3, pp. 303–311], Davis [56, Appendix A2, Theorem A2.2, pp. 259–261], Protter [206, Section 1.4, Theorem 31, pp. 22–23].

Note that the three σ-algebras defined in term of the filtration $(\mathcal{F}(t) : t \geq 0)$ are all on $\Omega \times [0, \infty)$ and not on Ω alone. We have $\mathcal{P} \subset \mathcal{O} \subset \mathcal{M}$ and it can be proved that the predictable class $\mathcal{P}$ (optional class $\mathcal{O}$, resp.) is the minimal σ-algebra for which adapted left-continuous (right-continuous, resp.) processes are measurable as function from $\Omega \times [0, \infty)$ into the state space (e.g., $E \subset \mathbb{R}^d$).

Recall that a random variable τ with values in $[0, \infty]$ is called a *stopping time* (or optional time) if sets of the form $\{\omega : \tau(\omega) \leq t\}$ are measurable with respect to $\mathcal{F}(t)$ for every $t \geq 0$. This is equivalent to imposing that the stochastic interval $[\![\tau, \infty[\![$ is optional, see Definition 1.8 in Chapter 1. Thus, $\mathcal{P}$ ($\mathcal{O}$, resp.) are generated by stochastic interval of the form $[\![0, \tau]\!]$ ($[\![0, \tau[\![$, resp.) where τ is any stopping time. Filtration satisfying the usual condition (right-continuity and completeness) are necessary to be able to identify a stopping time with it equivalent class, as explained below.

Assume that a right-continuous filtration $\mathbb{F} = \{F(t) : t \geq 0\}$ is given. If $\mathcal{O}$ is an open set of $\mathbb{R}^d$ and X is a cad-lag $\mathbb{F}$-adapted process with values in $\mathbb{R}^d$, then the *hitting time* $\tau_{\mathcal{O}}$ of an open set $\mathcal{O}$ is a stopping time, where

$$\tau_{\mathcal{O}} := \inf\big\{t > 0 \,:\, X(t) \in \mathcal{O}\big\},$$

and $\tau_{\mathcal{O}} = +\infty$ if $X(t) \in \mathbb{R}^d \setminus \mathcal{O}$ for every $t \geq 0$. Indeed the relation $\tau_{\mathcal{O}}(\omega) < t$ if and only if $X(s, \omega) \in \mathcal{O}$ for some rational number in $[0, t)$ shows that the event $\{\tau_{\mathcal{O}} < t\}$ belongs to $\mathcal{F}(t)$ and so $\{\tau_{\mathcal{O}} \leq t\}$ is in $\mathcal{F}(t+)$. Similarly, if C is a closed set of $\mathbb{R}^d$ then the *contact time* $\tilde{\tau}_{C}$ of a closed set C is also an stopping time, where

$$\tilde{\tau}_{C} := \inf\big\{t \geq 0 \,:\, \text{either } X(t) \in C \text{ or } X(t-) \in C\big\},$$

with $X(0-) = X(0)$. Indeed, use the fact that $\tilde{\tau}_{C}(\omega) \leq t$ if and only if the infimum over all rational numbers s in $[0, t]$ of the distance from $X(s, \omega)$ to C is zero. However, if K is a compact set of $\mathbb{R}^d$ then to check that the *entry time* $\bar{\tau}_{K}$ of a compact set K is also an stopping time, where

$$\bar{\tau}_{K} := \inf\big\{t \geq 0 \,:\, X(t) \in K\big\},$$

is far more delicate. The argument uses ordinal numbers and involves the assumption of P-completion $\bar{\mathbb{F}}$ for the filtration $\mathbb{F}$, i.e., $\bar{\mathbb{F}}$ is the minimal right-continuous filtration such that $\mathcal{F}(0)$ contains all P-nulls sets in $\bar{\mathcal{F}}$, the P-completion of $\mathcal{F}$. In this case, for any $\bar{\mathbb{F}}$-stopping time T there exists a $\mathbb{F}$-stopping time S such that $P\{T = S\} = 1$ and $\bar{F}(T)$ is the smaller σ-algebra containing

$\mathcal{F}(S+)$ all P-null sets. Furthermore, by means of the so-called *analytic sets* (i.e., continuous or Borel images of Borel sets in a Polish space), a deeper result shows that the hitting time of any Borel set is indeed a stopping time, e.g., see Doob [60, pp. 419–423] or Rogers and Williams [214, Sections II.73–76].

A process X with values in $E \subset \mathbb{R}^d$ is progressively measurable (resp. optional or predictable) if the map $(t,\omega) \mapsto X(t,\omega)$ or equivalently $(t,\omega) \mapsto \big(t, X(t,\omega)\big)$ is measurable with respect to $\mathcal{M}$ (resp. $\mathcal{O}$ or $\mathcal{P}$). In particular, if $(t,x) \mapsto a(t,x)$ is a Borel function and X is progressively measurable (respectively, optional or predictable) then so is the map $(t,\omega) \mapsto a\big(t, X(t,\omega)\big)$. On the other hand, a stopping time τ is called *predictable* if there exists an *announcing sequence* of stopping times $\{\tau_n : n = 1, 2, \dots\}$, i.e., τ_n increases to τ and $P(\tau_n < \tau \text{ if } \tau > 0) = 1$; sometimes the condition $P(\tau > 0) = 1$ is also requested. It is not hard to show that τ is a predictable time if and only if the stochastic interval $[\![\tau, \infty[\![$ is predictable. Note that $\tau + \varepsilon$ is a predictable (stopping) time for any stopping time τ and any constant $\varepsilon > 0$. Moreover, if τ and θ are predictable times then all stochastic intervals that have τ, θ, 0 or ∞ as endpoints are predictable sets. There are many interesting measurability question on these points, e.g., see Bichteler [25, Section 3.5]

Based on the alternative way of generating the σ-algebras $\mathcal{O}$ and $\mathcal{P}$ (as mentioned above), we deduce that a right-continuous (resp. left-continuous) progressively measurable process is optional (resp. predictable). When working with cad-lag (continuous) processes, the difference between the progressively measurable and optional (predictable) processes have almost no importance. Recall that given a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ a stochastic process X is called *adapted* if the random variable $\omega \mapsto X(t,\omega)$ is $\mathcal{F}(t)$-measurable for any $t \geq 0$. Thus, any adapted cad-lag process is progressively measurable.

The concept of a *predictable* (also called *previsible*) or *optional* process implies that of *adapted process* (to a given filtration) in a way suggested by the name. Denote by $\mathcal{P}_{\mathcal{R}}$ the family of subsets of $\Omega \times [0,\infty)$ containing all sets of the form $F_0 \times \{0\}$ and $F \times (s,t]$, where $F_0 \in \mathcal{F}(0)$ and $F \in \mathcal{F}(s)$ for $s < t$ in $(0,\infty)$, is called the *class of predictable rectangles*. Sometimes, the sets $F_0 \times \{0\}$ need special consideration and some authors prefer to remove these type of sets from the definition of the σ-algebra $\mathcal{P}$. As it was defined, the σ-algebra $\mathcal{P}$ of subsets of $[0,\infty) \times \Omega$ generated by all predictable rectangles is called the *predictable* σ-algebra associated with the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Another equivalent way of generating the predictable σ-algebra is to define $\mathcal{P}$ as generated by the stochastic intervals of the form $[\![0, \tau]\!]$, for stopping times τ with respect to the given filtration $\{\mathcal{F}(t) : t \geq 0\}$.

Sometimes, the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is *quasi-left continuous*, namely, $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for any *predictable* stopping time τ. This is the case of a (continuous) Markov-Feller process (see Rogers and Williams [214, Chapter 6, Theorem 18.2, pp. 346–347]).

It should be clear by now that filtered spaces are a fundamental feature of the theory of stochastic processes and definitions of our central object, Markov processes, will involve a filtration. Heuristically speaking, the σ-algebra $\mathcal{F}(t)$ is the collection of event that may occur before or at the time t (i.e., the set

of possible *pasts* up to time t). In what follows, unless otherwise stated, we assume that filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ is given, sometime denoted by $(\Omega, \mathbb{F}, P)$ and also called stochastic basis, with $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$, $\mathcal{F} = \sigma\{\mathcal{F}(t) : t \geq 0\}$, and the three σ-algebras $\mathcal{P} \subset \mathcal{O} \subset \mathcal{M}$, called predictable, optional and progressively measurable, respectively, as in Definition 3.4. Moreover, if Ω is also a Polish space then $\mathcal{F}$ contains the Borel σ-algebra $\mathcal{B}(\Omega)$, actually, $\mathcal{F} = \mathcal{B}(\Omega)$ in most of the cases.

On a given measurable space $(\Omega, \mathcal{F})$ we may define its *universal completion* $\mathcal{F}^0 = \bigcap_P \mathcal{F}^P$, where $\mathcal{F}^P$ is the completion of $\mathcal{F}$ relative to P, and the intersection is over all probability measures P on $\mathcal{F}$. This is to say that $A \in \mathcal{F}^0$ if and only if for every P there exist B and N in $\mathcal{F}$ such that $B \setminus N \subset A \subset B \cup N$ and $P(N) = 0$ (since B and N may depend on P, clearly this does not necessarily imply that $P(N) = 0$ for every P). Thus, a universally complete measurable space satisfies $\mathcal{F} = \mathcal{F}^0$. The concept of universally measurable is particularly interesting when dealing with measures in a Polish space Ω, where $\mathcal{F} = \mathcal{B}(\Omega)$ is its Borel σ-algebra, and then any subset of Ω belonging to $\mathcal{F}^0$ is called universally measurable. In this context, it is clear that a Borel set is universally measurable. Also, it can be proved that any analytic set is universally measurable, and on any uncountable Polish space there exists a analytic set (with not analytic complement) which is not a Borel set, e.g., see Dudley [62, Section 13.2].

A detailed discussion on the strong Markov property involves a *measured filtration* $(\Omega, \mathbb{F}, \mathbb{P})$, i.e., besides the filtration $(\Omega, \mathbb{F})$, we have a class of probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F})$, with $\mathcal{F} = \mathcal{F}(\infty)$. Then, the $\mathbb{P}$-universal completion $\mathbb{F}^0$ (i.e., the family $\mathbb{P}$ is implicitly understood in the notation) of the filtration $\mathbb{F}$ is defined by adding all null sets, i.e.,

$$\mathcal{F}^0(t) = \bigcap_{\varepsilon > 0} \bigcap_{P \in \mathbb{P}} \sigma\big(\mathcal{F}(t+\varepsilon), \mathcal{N}^P\big), \quad \forall t \geq 0,$$

where $\mathcal{N}^P$ denotes the σ-algebra of $(P, \mathcal{F})$-null sets (i.e., all the subsets of some set N in $\mathcal{F}$ with $P(N) = 0$). Hence the new filtration $\mathbb{F}^0$ is right-continuous and universally complete (i.e., the universal completion of $\mathcal{F}^0(t)$ is again $\mathcal{F}^0(t)$, for every t) by construction but not necessarily complete with respect to a particular probability P in the class $\mathbb{P}$. Furthermore, sometimes not all $(P, \mathcal{F})$-null sets are necessary, and completion arguments are reviewed. Countable unions of sets N with $P(N) = 0$ and the property of being $\mathcal{F}(t)$-measurable (for some finite t) are called *P-nearly empty* sets. Thus, a set N in $\mathcal{F} = \mathcal{F}(\infty)$ with $P(N) = 0$ which is not in $\mathcal{F}(t)$, for every t finite, may not be nearly empty. Then, a measured filtration is called *$\mathbb{P}$-regular* if $\mathcal{F}(t) = \mathcal{F}^{\mathbb{P}}(t)$ for every $t \geq 0$, where $\mathcal{F}^{\mathbb{P}}(t) = \bigcap_{P \in \mathbb{P}} \mathcal{F}^P(t)$, with $\mathcal{F}^P(t)$ the σ-algebra of all subsets A of Ω such that for every P in the class $\mathbb{P}$, the symmetric difference $(A \setminus A_P) \cup (A_P \setminus A)$ is P-nearly empty for some A_P in $\mathcal{F}(t)$. Note that $\mathcal{F}^P(t)$ contains the completion of $\mathcal{F}(t)$ relative to the restriction of P to $\mathcal{F}(t)$ (so it is universally complete), but it could be smaller than $\sigma\big(\mathcal{F}(t), \mathcal{N}^P\big)$. Moreover, filtration $\{\mathcal{F}^{\mathbb{P}}(t+) : t \geq 0\}$ is also $\mathbb{P}$-regular, and called the *$\mathbb{P}$-natural enlargement* of $\mathbb{F}$. Essentially, as long as we work with a right-continuous regular filtration, the technical points about

measurability are resolved, this is usually referred to as the natural conditions. For instance, see the book Bichteler [25, Section 1.3] for a comprehensive study.

When a process is viewed as a function of two-variables, $(t,\omega) \mapsto X(t,\omega)$, properties like continuity or monotonicity refers to the path (i.e., to the function $t \mapsto X(t,\omega)$ for a fixed ω, which are global property on the variable t only), while properties like integrable or bounded may refer to either one of the variables or even to both variables. However, the qualifier integrable usually refers to the random variable $\omega \mapsto X(t,\omega)$, for a fixed $t \geq 0$. As discussed later, when a filtration is given, the term *locally* will apply to any property of a process involving both variables with a precise meaning, e.g., a process X is locally bounded if there is a (increasing) sequence $(\tau_n : n \geq 1)$ of stopping times satisfying $P(\tau_n \uparrow \infty) = 1$ such that the stopped process $X_n(t) := X(t \wedge \tau_n)$ is bounded. In all statements and procedures, processes are considered equals if their paths differ in a set of measure zero (i.e., they are indistinguishable one of each other), but sometimes we may select a particular element of its equivalent class to perform a specific construction. All these terminologies become clear from the context of the discussion.

3.3 Bounded Variation

Let us consider real-valued (or $\mathbb{R}^d$-valued) processes $(X(t) : t \geq 0)$ in a probability space we $(\Omega, \mathcal{F}, P)$. If a filtration $\mathbb{F} = (\mathcal{F}(t) : t \geq 0)$ is also given then the term *adapted* (to $\mathbb{F}$) is implicitly assumed (although sometimes is explicitly mentioned) and the qualifier *raw* is to be used to explicitly assume not necessarily adapted.

3.3.1 Pathwise Properties

An *(monotone) increasing* [*(monotone) decreasing*, resp.] stochastic process $(X(t) : t \geq 0)$ is such that the function $t \mapsto X(t,\omega)$ is increasing [decreasing, resp.] for every ω, except perhaps in a null set. Because an increasing function has left and right-hand limits at each points, it is convenient to normalize the process to be right-continuous. Thus an increasing process is a random variable X (almost surely defined) with values in the sample space $D([0,\infty),\mathbb{R})$ such that $X(t) \geq X(s)$ for every $t \geq s$. Also, vector valued process (i.e., in $D([0,\infty),\mathbb{R}^d)$) can be considered.

Similarly, a stochastic process $X = (X(t) : t \geq 0)$ is said to be of *locally bounded variation* in $[0,\infty)$ (or simplifying, of *finite variation*) if it is a random variable X with values in the sample space $D([0,\infty),\mathbb{R})$ and its variation process $\{\mathrm{var}(X,[0,t]) : t \geq 0\}$ is finite,

$$\mathrm{var}(X,[0,t]) := \sup\{\sum_{i=1}^{n} |X(t_i) - X(t_{i-1})| \,:\, 0 = t_0 < t_1 < \cdots < t_n = t\},$$

where the supremum is taken over all partitions of the interval $[0,t]$, var is referred to as the *variation operator*. Clearly, as long as the process is cad-lag,

we may only consider some countable family of partitions, e.g., partitions with $t_i = i2^{-n}$ for $i = 0, 1, \ldots, 2^n$. It can be defined the positive $\{\text{var}^+(X,[0,t]) : t \geq 0\}$ and the negative $\{\text{var}^-(X,[0,t]) : t \geq 0\}$ variation processes exchanging the absolute value $|\cdot|$ with the positive part $[\cdot]^+$ and the negative part $[\cdot]^-$ of a real number in the above definition. Note that because X is cad-lag, the supremum can be taken over partitions with end points $t_1 < \cdots < t_{n-1}$ in a countable dense set so that the functions $X \mapsto \text{var}(X,\cdot)$, $X \mapsto \text{var}^+(X,\cdot)$ and $X \mapsto \text{var}^-(X,\cdot)$ are measurable from $D([0,\infty),\mathbb{R})$ or $C([0,\infty),\mathbb{R})$ into itself.

It can be checked that, e.g., Gordon [103, Chapters 4 and 6],

$$\begin{aligned}&\text{var}(X,[0,t]) = \text{var}^+(X,[0,t]) + \text{var}^-(X,[0,t]) \quad \text{and}\\ &X_t - X_0 = \text{var}^+(X,[0,t]) - \text{var}^-(X,[0,t]) \quad \forall t \in [0,\infty),\end{aligned}$$

and that the three variation processes

$$\{\text{var}(X,[0,t]) : t \geq 0\}, \quad \{\text{var}^+(X,[0,t]) : t \geq 0\}, \quad \{\text{var}^-(X,[0,t]) : t \geq 0\}$$

are (monotone) increasing (and cad-lag); and they are adapted, optional or predictable if the initial process X is so. Thus we can look at a locally bounded variation process X as two random variables $\text{var}^+(X,[0,t])$ and $\text{var}^-(X,[0,t])$ with values in the sample space $D = D([0,\infty),\mathbb{R})$, i.e., a probability measure P on D with the Borel σ-algebra $\mathcal{B}(D)$ and two increasing and measurable maps $\text{var}^+(X,[0,t])$ and $\text{var}^-(X,[0,t])$ from D into itself. Note that $\text{var}^+(X,[0,t])$ and $\text{var}^-(X,[0,t])$ are minimal in the sense that if X is of bounded variation and $X = Y - Z$ with each Y and Z monotone increasing then $\text{var}^+(X,[0,t]) - \text{var}^+(X,[0,s]) \leq Y_t - Y_s$ and $\text{var}^-(X,[0,t]) - \text{var}^-(X,[0,s]) \leq Z_t - Z_s$, for every $t \geq s$. This is the so-called Jordan decomposition. On the other hand, given a (cad-lag) bounded variation process X, its *continuous part* is defined as $X^c(t) := X(t) - X^{jp}(t)$, where the *jump part* is defined by $X^{jp}(t) := \sum_{0<s\leq t} \delta X(s)$, with $\delta X(s) = X(s) - X(s-)$. It is clear that, by rearranging the jumps, we can rewrite the jumps part as $X^{jp}(t) = \sum_n [X(\tau_n) - X(\tau_n-)]\mathbb{1}_{\tau_n \leq t}$, where the series converges absolutely for any t and the random times τ_n are stopping times if the process X^{jp} is adapted, see Sato [220, Lemma 21.8, Chapter 4, pp. 138–140]. Moreover, since the continuous part $X^c(t)$ is still of bounded variation, it is differentiable almost everywhere and we have $X^c(t) = X^{ac}(t) + X^{sc}(t)$, where $X^{ac}(t) := \int_0^t \dot{X}^c(s)\mathrm{d}s$ is called the *absolutely continuous part* and $X^{sc}(t)$ is the *singular continuous part*. Thus, any bounded variation process X can be written as a unique sum $X^{ac} + X^{sc} + X^{jp}$ called Lebesgue decomposition.

On the other hand, for any cad-lag process X and any $\varepsilon > 0$ we can define the ε-jumps process as

$$X^{\varepsilon j}(t) := \sum_{0<s\leq t} \delta X(t) \mathbb{1}_{|\delta X(t)| \geq \varepsilon}, \quad \forall t > 0,$$

and the ε-almost continuous process $X^{\varepsilon c}(t) := X(t) - X^{\varepsilon j}(t)$. However, the continuous part, i.e., $\lim_{\varepsilon \to 0} X^{\varepsilon c}(t)$ may not be defined in general. Certainly, this would be $X^c(t)$ when X has locally bounded variation. However, the above

limit will make sense in the L^2 topology for *square integrable local martingales* and defined for every process which is a local martingale, as discussed later.

A monotone increasing or a locally bounded variation process X induces a Borel (positive or signed) measure on $[0,\infty)$ by setting

$$\mu(\{0\}) = X(0,\omega) \quad \text{and} \quad \mu(]a,b]) = X(b,\omega) - X(a,\omega), \quad 0 < a < b,$$

for each sample path $X(\cdot,\omega)$, which is referred to as the Lebesgue-Stieltjes measure. Conversely, if a (Borel) locally finite signed measure μ on $[0,\infty)$ is defined for ω which is *weakly measurable* i.e. $\omega \mapsto \mu(A,\omega)$ is measurable for each Borel subset A of $[0,\infty)$ then we can define process of bounded variation as $X(t,\omega) = \mu([0,t],\omega)$, for any $t \geq 0$. Based on Fubini's theorem, it can be proved that given two processes X and Y of locally bounded variation (cad-lag) we have the integration-by-part formula

$$X(b)Y(b) - X(a)Y(a) = \int_{(a,b]} X(t-)\mathrm{d}Y(t) + \\ + \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\leq b} \delta X(t)\,\delta Y(t).$$

When the integrands $Y(t-)$ and $X(t-)$ are left-continuous and the integrator $X(t)$ and $Y(t)$ are right-continuous as above, the integral can be regarded in the Riemann-Stieltjes sense, where $X(t-)$ is the left-hand limit at t. Also, the last two terms may be grouped and considered as an integral in the sense of Lebesgue-Stieltjes, i.e.,

$$\int_{(a,b]} Y(t)\mathrm{d}X(t) = \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\leq b} \delta X(t)\,\delta Y(t),$$

where $\delta X(t) = X(t) - X(t-)$ is the jump at t. Moreover, given a locally bounded variation (cad-lag) process Y, the equation

$$X(t) = 1 + \int_{(0,t]} X(t-)\mathrm{d}Y(t), \ \forall t \geq 0,$$

has a unique solution X, in the class of locally bounded variation process (cad-lag), which is explicitly given by the formula

$$X(t) = \exp\left[Y^c(t) - Y^c(0)\right] \prod_{0<s\leq t} (1 + \delta Y(s)),$$

where $Y^c(t) = Y(t) - \sum_{0<s\leq t} \delta Y(s)$ is the continuous part of the process $(Y(t) : t \geq 0)$, the (infinite) product is the exponential of the absolutely convergence series $\sum_{0<s\leq t} \ln(1 + \delta Y(s))$, and clearly, if $Y(\tau) - Y(\tau-) = -1$ for some $\tau > 0$ then $X(t) = 0$ for any $t \geq \tau$, see Shiryayev [227, pp. 204–208], Doob [61, pp. 160–166], Chung and Williams [45, pp. 4–6].

An *elementary process* (or piecewise constant over stochastic intervals) is a stochastic process of the form

$$Y(t) = \sum_{i=1}^{n} Y_{i-1} \mathbb{1}_{(\tau_{i-1},\tau_i]}(t), \qquad t \geq 0, \tag{3.7}$$

where $0 = \tau_0 \le \tau_1 \le \ldots \le \tau_n$ are stopping times and Y_{i-1} is a $\mathcal{F}(\tau_{i-1})$ measurable random variable for any i. It is called *bounded* if all Y_{i-1} are bounded random variables. The set of (bounded) elementary processes form a subalgebra (i.e., closed by natural addition and multiplication) of predictable sets. Note that processes of the form (3.7) are left-continuous with right-hand limits, i.e., the (right-hand limit) process $Y(t+)$ is cad-lag, and satisfies $Y(0) = 0$.

If Y is an elementary process and X is a locally bounded variation process (cad-lag) then we may take Y as an *integrand* and X as an *integrator* to construct the integral process for $t \ge 0$ by

$$Z(t) = \int_{(0,t]} Y(s)\mathrm{d}X(s) = \sum_{i=1}^{n} Y_{i-1}[X(t \wedge \tau_i) - X(t \wedge \tau_{i-1})]. \tag{3.8}$$

This integral extends to all predictable processes Y in either Riemann-Stieltjes or Lebesgue-Stieltjes sense. In particular, the above integral makes sense for any bounded adapted cag-lad (left-continuous with right-hand limits) Y. Actually, if the Lebesgue-Stieltjes integral is used then $\mathrm{d}X(t)$ means $\mathrm{d}\mu_X$ (the signed measure induced by the cad-lag bounded variation process X) integration over the interval $(0,t]$, but if the Riemann-Stieltjes integral is used then we mean

$$\int_{(0,t]} Y(s)\mathrm{d}X(s) = \lim_{\alpha \to 0^+,\, \beta \to t^+} \int_{\alpha}^{\beta} Y(s)\mathrm{d}V(s) = \int_{]0,t]} Y(s)\mu_X(\mathrm{d}s).$$

Clearly, the Lebesgue-Stieltjes integral makes also sense when the integrand Y is not necessarily cag-lad. Moreover, from the integration by part formula we deduce the following property. If a function f is continuously differentiable from $\mathbb{R}^d$ into $\mathbb{R}$ and $X = (X_1, \ldots, X_d)$ is a stochastic process with values in $\mathbb{R}^d$ where each components X_i is a process of locally bounded variation then we have

$$\begin{cases} f(X(t)) - f(X(0)) = \displaystyle\int_{(0,t]} \nabla f(X(s-))\mathrm{d}X(s) + \\ \qquad + \displaystyle\sum_{0<s\le t} \big[f(X(s)) - f(X(s-) - \nabla f(X(s-))\,\delta X(s))\big], \end{cases} \tag{3.9}$$

If the initial process X is adapted then the integral (3.8) defines an adapted cad-lag stochastic process $(Z(t) : t \ge 0)$ of locally bounded variation, with

$$\mathrm{var}(Z, [0,t]) = \int_{(0,t]} |Y(s)|\,\mathrm{var}(X, \mathrm{d}s).$$

Actually, as long as the above integral is finite (in Lebesgue sense) with a predictable processes $Y(s)$ and locally bounded variation process $X(s)$, the integral (3.8) defines a process $Z(t)$ with locally bounded variation.

Sometimes it is necessary to make a *time change* in Stieltjes integrals. For a given increasing cad-lag process A with values in $[0, \infty]$ consider

$$A^{-1}(s) := \inf\{t \ge 0 : A(t) > s\}, \quad \forall s \ge 0, \tag{3.10}$$

with $A^{-1}(s) = 0$ for $s \le \sup\{t \ge 0 : A(t) = 0\}$, and $A^{-1}(s) := +\infty$ if $A(t) \le s$ for all $t \ge 0$. This define a cad-lag process $(A^{-1}(s) : s \ge 0)$ with the properties

$$A^{-1}(s-) = \inf\{t \ge 0 : A(t) \ge s\}, \quad \forall s \ge 0,$$
$$A[A^{-1}(s)] \ge s, \quad \forall s \ge 0,$$
$$A(t) = \inf\{s \ge 0 : A^{-1}(s) > t\}, \quad \forall t \ge 0.$$

If A is continuous then A^{-1} may not be continuous (when A is not strictly increasing). The following change of variables formula can be obtained. For any nonnegative Borel measurable function f on $[0, \infty)$ we have

$$\int_{[0,\infty)} f(t)\mathrm{d}A(t) = \int_0^\infty f(A^{-1}(s))\, \mathbb{1}_{A^{-1}(s)<\infty}\mathrm{d}s,$$
$$\int_{[u(a),u(b)]} f(t)\mathrm{d}A(t) = \int_{[a,b]} f(u(t))\mathrm{d}A(u(t)),$$

for any continuous non-decreasing process u on the bounded interval $[a, b]$.

A typical example is a real-valued Poisson process $X(t)$ with parameter $c > 0$, which is a process of bounded variation and a jumps process of the form $X(t) = \sum_n \mathbb{1}_{t \ge \theta_n}$, where $\theta_n := \tau_1 + \cdots + \tau_n$ and $(\tau_1, \tau_2, \ldots)$ is a sequence of independent exponentially distributed (with parameter c) random variables. If Y is a (cad-lag) bounded adapted process with respect to X then the following (cad-lag) adapted processes are defined by the Riemann-Stieltjes integrals

$$M(t) := \int_{]0,t]} Y(s-)\mathrm{d}X(s) - c\int_0^t Y(s)\mathrm{d}s, \quad \forall t \ge 0,$$
$$N(t) := M^2(t) - c\int_0^t Y^2(s)\mathrm{d}s, \quad \forall t \ge 0,$$
$$E(t) := \exp\Big\{\int_{]0,t]} Y(s-)\mathrm{d}X(s) + c\int_0^t [1 - \mathrm{e}^{Y(s)}]\mathrm{d}s\Big\}, \ \forall t \ge 0.$$

Taking the left-hand limit $Y(s-)$ in the integral with respect to X is essential to make the Riemann-Stieltjes integral meaningful. It will be seen later that these three processes M, N, E are martingales and the above integral will be called stochastic integral when the process Y is predictable with respect to X.

Note that all arguments made above for locally bounded variation process are of a pathwise character, without any assumption of integrability in Ω.

3.3.2 Integrable Finite Variation

No specific difference was made in the previous pathwise discussion regarding path with bounded variation within any bounded time-interval and within the half (or whole) real line, i.e., bounded variation paths (without any other qualitication) refers to any bounded time-interval, and so the limit $A(+\infty)$ for a monotone paths could be infinite. Moreover, no condition on integrability (with

respect to the probability measure) was assumed, and as seen later, this integrability condition (even locally) is realted to the concept of martingales.

Now, we mention that an important role is played by the so-called *integrable increasing processes* in $[0,\infty)$, i.e., processes A with (monotone) increasing path such that

$$\mathbb{E}\{\sup_t A(t)\} = \mathbb{E}\{\lim_{t\to\infty} A(t)\} = \mathbb{E}\{A(\infty)\} < \infty,$$

and processes with *integrable bounded variation* or *integrable finite variation* on $[0,\infty)$, i.e., processes A where the variation process $\{\mathrm{var}(A,[0,t]) : t \geq 0\}$ satisfies

$$\mathbb{E}\{\sup_t \mathrm{var}(A,[0,t])\} = \mathbb{E}\{\mathrm{var}(A,[0,\infty[)\} < \infty,$$

or equivalently, $A = A^+ - A^-$ where A^+ and A^- are integrable increasing processes in $[0,\infty)$. These two concepts are localized as soon as a filtration is given, e.g., if there exists a (increasing) sequence of stopping times $(\tau_n : n \geq 1)$ satisfying $P(\lim_n \tau_n = \infty) = 1$ such that the stopped process $A_n(t) := A(t \wedge \tau_n)$ is an integrable increasing process in $[0,\infty)$ for any n then A is a *locally integrable increasing process* in $[0,\infty)$. Note that processes with *locally integrable bounded variation* or *locally integrable finite variation* on $[0,\infty)$, could be misinterpreted as processes such that their variations $\{\mathrm{var}(A,[0,t]) : t \geq 0\}$ satisfy $\mathbb{E}\{\mathrm{var}(A,[0,t])\} < \infty$, for any $t > 0$. It is worth to remark that any predictable process of bounded (or finite) variation (i.e., its variation process is finite) is indeed of locally integrable finite variation, e.g., see Jacod and Shiryaev [117, Lemma I.3.10]. Moreover, as mentioned early, the qualifiers increasing or bounded (finite) variation implicitly include a cad-lag assumption, also, the qualifier locally implicitly includes an adapted condition. In the rare situation where an adapted assumption is not used, the tern *raw* will be explicitly used.

A simple application of the change of time (3.10), i.e., the following expression for a cad-lag increasing process A,

$$\mathbb{E}\Big\{\int_0^T X(t)\mathrm{d}A(t)\Big\} = \int_0^\infty \mathbb{E}\big\{X\big(A^{-1}(s)\big)\mathbb{1}_{A^{-1}(s)<\infty}\big\}\mathrm{d}s,$$

proves that for any two nonnegative measurable processes (non necessarily adapted) X and Y satisfying $\mathbb{E}\big\{X(\tau)\mathbb{1}_{\tau<\infty}\big\} = \mathbb{E}\big\{Y(\tau)\mathbb{1}_{\tau<\infty}\big\}$, for every stopping time τ, we have

$$\mathbb{E}\Big\{\int_0^r X(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^r Y(t)\mathrm{d}A(t)\Big\}, \quad \forall r \in (0,\infty].$$

Now, if $\mathcal{F}_\tau$ denotes the σ-algebra associated with a stopping time (see Definition 1.8) then the condition

$$\mathbb{E}\big\{X(\tau)\mathbb{1}_{\tau<\infty}\big|\mathcal{F}(\tau)\big\} = \mathbb{E}\big\{Y(\tau)\mathbb{1}_{\tau<\infty}\big|\mathcal{F}(\tau)\big\}, \quad \text{a.s.}$$

implies

$$\mathbb{E}\Big\{\int_\tau^\infty X(t)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\} = \mathbb{E}\Big\{\int_\tau^\infty Y(t)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\}, \tag{3.11}$$

almost surely, which is used later with martingale.

On the other hand, we can verify that if A and B are two cad-lag increasing processes (non necessarily adapted) such that

$$\mathbb{E}\big\{A(t) - A(s)\big|\mathcal{F}(s)\big\} = \mathbb{E}\big\{B(t) - B(s)\big|\mathcal{F}(s)\big\}, \quad \text{a.s.},$$

for every (extended) real numbers $0 \le s \le t \le \infty$, then we have

$$\mathbb{E}\Big\{\int_0^T X(t-)\mathrm{d}A(t)\big|\mathcal{F}(\tau)\Big\} = \mathbb{E}\Big\{\int_0^T X(t-)\mathrm{d}B(t)\big|\mathcal{F}(\tau)\Big\}, \tag{3.12}$$

for every $T \ge 0$ and for every nonnegative cad-lag adapted process X.

Let us go back to the relation of locally bounded variation process X with a Borel (positive or signed) measure on $[0,\infty)$

$$\mu(\{0\}) = X(0,\omega), \qquad \mu(]a,b]) = X(b,\omega) - X(a,\omega), \quad 0 < a < b$$

and abandon the pathwise analysis. Similar to the null sets in Ω, a key role is played by *evanescent* sets in $[0,\infty)\times\Omega$, which are defined as all sets N in the product σ-algebra $\mathcal{B}([0,\infty))\times\mathcal{F}$ such that $P(\{\cup_t N_t\}) = 0$, where N_t is the t section $\{\omega : (\omega,t) \in N\}$ of N. For a given process A of *integrable bounded variation*, i.e., such that

$$\mathbb{E}\{\sup_t \mathrm{var}(A,[0,t]\} < \infty,$$

we may define (bounded) signed measure μ_A (this time) on $[0,\infty)\times\Omega$ by the formula

$$\mu_A(]a,b]\times F) = \mathbb{E}\Big\{\mathbb{1}_F \int_{]a,b]} \mathrm{d}A(t)\Big\}, \quad \forall b > a \ge 0,\ F \in \mathcal{F}. \tag{3.13}$$

Since progressively, optional or predictable measurable sets are naturally identified except an evanescent set, the measure μ_A correctly represents a process A with integrable bounded variation. Conversely, a (bounded) signed measure μ on $[0,\infty)\times\Omega$ corresponds to some process A if and only if μ is a so-called signed P-measure, namely, if for any set N with vanishing sections (i.e., satisfying $P\{\omega : (\omega,t) \in N\} = 0$ for every t) we have $\mu(N) = 0$. A typical case is the point processes, i.e.,

$$A(t) := \sum_n a_n \mathbb{1}_{\tau_n \ge t},$$

where $\tau_{n-1} \le \tau_n$ and $\tau_{n-1} < \tau_n$ if $\tau_n < \infty$ is a sequence of stopping times and a_n is $\mathcal{F}(\tau_n)$-measurable random variable with values in $\mathbb{R}_* := \mathbb{R}\setminus\{0\}$, for every n. Then, for each fixed ω the function $t \to A(t,\omega)$ is piecewise constant, but

even if all the random variable a_n are bounded, the variation of the process A may not be integrable. The measure μ_A takes the form

$$\mu_A(X) = \sum_n E\{a_n\, X(\tau_n)\} = \mathbb{E}\Big\{ \int_{[0,\infty)} \int_{\mathbb{R}_*} a\, X(t,\omega)\, \nu_A(\mathrm{d}t, \mathrm{d}a, \omega)\Big\},$$

$$\nu_A(B,\omega) := \#\{n : (\tau_n(\omega), a_n(\omega)) \in B\},$$

for every B in $\mathcal{B}([0,\infty)\times\mathbb{R}_*)$, where $\#$ denotes the number of elements in a set and X is any bounded measurable process, in particular of the form $X(t,\omega) = \mathbb{1}_{]a,b]}(t)\, \mathbb{1}_F(\omega)$, for some set F in $\mathcal{F}$. It may seem more complicate to use the random measure ν_A defined on $[0,\infty)\times\mathbb{R}_*$, but indeed this is characteristic to jumps processes. The reader is referred to the discussions in the books by Dellacherie and Meyer [58, Section VI.2, pp. 113–164], Jacod and Shiryaev [117, Section 1.3, pp. 27–32], Rogers and Williams [214, Sections VI.19–21, pp. 347–352], and Elliott [73], Protter [206], among others, to complement the above remarks and following theorem–definition

Definition 3.5 (compensator)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a given filtered space. For any bounded (or integrable) measurable process X there exists a unique predictable process ${}^pX$, called *predictable projection*, such that for any predictable stopping time τ we have $\mathbb{E}\{{}^pX\, \mathbb{1}_{\tau<\infty}\} = \mathbb{E}\{X\, \mathbb{1}_{\tau<\infty}\}$. It is proved that a process A with integrable bounded variation is predictable if and only if $\mu_A(X) = \mu_A({}^pX)$ for any bounded measurable process X, see (3.13). Now, given a process A with integrable bounded variation with a corresponding signed P-measure μ_A on $[0,\infty)\times\Omega$, the *dual predictable projection* of μ_A is defined by duality as follows

$$\mu_A^p(X) = \mathbb{E}\Big\{ \int_{[0,\infty)} {}^pX(t,\omega)\mathrm{d}A(t,\omega)\Big\},$$

for any bounded measurable process X. Since μ_A^p is a signed P-measure which commutes with the predictable projection, its corresponding process with integrable bounded variation, denoted by A^p, is predictable and satisfies

$$\mathbb{E}\{\int_{[0,\infty)} X(t,\omega)\mathrm{d}A^p(t,\omega) = \mathbb{E}\{\int_{[0,\infty)} {}^pX(t,\omega)\mathrm{d}A(t,\omega),$$

for any bounded measurable process X, and called the *compensator* of A. □

Similarly to above, we may define the *optional projection*, and *dual optional projection*, with the notations ${}^oX$, μ_A^o and A^o. Clearly, the above statements can be localized, i.e., the process X can only be assumed locally bounded or locally integrable, and the process A can only be supposed with locally integrable finite variation.

It will be stated later that the dual predictable projection μ_A^p corresponding to a signed P-measure μ_A of an adapted process A with integrable bounded variation is actually characterized by the fact that the (Stieltjes integral) process

$$\int_{[0,t]} X(t-,\omega)\mathrm{d}A(t,\omega) - \int_{[0,t]} X(t-,\omega)\mathrm{d}A^p(t,\omega). \quad t \geq 0$$

is a martingale for any bounded adapted process X. It is clear that $t \mapsto X(t-)$ is a predictable process and that in the above martingale condition it suffices to take processes of the form $X(t) := \mathbb{1}_{t \le \tau}$ for some stopping time τ, i.e., the process $t \mapsto A(t \wedge \tau) - A^p(t \wedge \tau)$ is a martingale.

Related with the compensator definition is the (unique) decomposition of any positive increasing adapted right-continuous process A into the sum of a continuous increasing adapted process ${}^cA$ with ${}^cA(0) = 0$ and a right-continuous increasing adapted process ${}^jA$ which can be expressed as follows:

$$ {}^jA(t) = \sum_n a_n \, \mathbb{1}_{t \ge \tau_n}, $$

where $\{\tau_n\}$ is a sequence of stopping times with bounded disjoint graphs and a_n is a bounded positive $\mathcal{F}(\tau_n)$-measurable function for every n. The proof of this fact is rather simple, first define inductively $\tau^{i,0} = 0$ and

$$ \tau^{i,j} = \inf\{t > \tau^{i,j} : A(t+) - A(t-) \ge 1/i\}, $$

and then $\tau_k^{i,j} := \tau^{i,j}$ if $A(t+) - A(t-) \le k+1$ and $\tau^{i,j} \le k$, and $\tau_k^{i,j} := \infty$ otherwise. Clearly $\{\tau_k^{i,j}\}$ is countable and can be rewritten as $\{\tau_n' : n = 1, 2, \ldots\}$, which is a sequence of stopping times with bounded graphs. Again, defining $\tau_n := \tau_n'$ if $\tau_i \ne \tau_n$ for every $i = 1, \ldots, n$ and $\tau_n := \infty$ otherwise, we get the desired sequence, with $a_n := A(\tau_n+) - A(\tau_n-)$.

Similarly, if A is as above and $\varphi : [0, \infty) \to [0, \infty)$ is a continuously differentiable function and for a given $r \ge 0$ we set

$$ \tau_r = \inf\{t \ge 0 : A(t) \ge r\} \quad \text{and} \quad \theta_r = \inf\{t \ge 0 : A(t) > r\}, $$

which are both stopping times (as seen later, τ_r is predictable), then for every bounded measurable process X we have

$$ \int_0^\infty X(s) \mathrm{d}\varphi(A(s)) = \int_0^\infty X(\tau_r) \, \varphi'(r) \, \mathbb{1}_{\tau_t < \infty} \mathrm{d}r = $$
$$ = \int_0^\infty X(\theta_r) \, \varphi'(r) \, \mathbb{1}_{\theta_t < \infty} \mathrm{d}r. $$

Details on the proof of these results can be found in Bichteler [25, Section 2.4, pp. 69–71].

As mentioned above, another measure associated with a process X is the so-called *jumps measure*, which is a random measure on $[0, \infty) \times \mathbb{R}_*$, with $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$ with integer values and defined, for each ω, by

$$ \nu(]a, b] \times B) := \#\{t : a < t \le b, \; X(t) - X(t-) \in B\}, $$

for every $b > a \ge 0$ and B in $\mathcal{B}(\mathbb{R}_*)$, i.e., ν is the number of jumps of the process X in the time interval $]a, b]$ which belongs to the set B. Typically, if $X = P$ is a Poisson measure process then the compensator of ν is indeed the (deterministic) Lévy measure m. Clearly, the above integer-valued random

measure is defined even if the process X is not of locally bounded variation, only the cad-lag property is used.

On the other hand, we may define the quadratic variation of X over a partition $\pi = (t_0 < t_1 < \cdots < t_n)$ is given by

$$\mathrm{var}_2(X, \pi) := \sum_{i=1}^{n} |X(t_i) - X(t_{i-1})|^2,$$

and then quadratic variation operator as

$$\begin{cases} \mathrm{var}_2(X, [0,t]) := \lim_{r \to 0} \mathrm{var}_2(X, [0,t], r), \\ \mathrm{var}_2(X, [0,t], r) := \sup\{\mathrm{var}_2(X, \pi)\}, \quad \forall t > 0, \end{cases} \tag{3.14}$$

where the supremum is taken over all partitions $\pi_t = (0 = t_0 < \cdots < t_n = t)$ of the interval $[0,t]$, with $t_i - t_{i-1} \le r$. It is easy to imagine a process with only jumps such that $\mathrm{var}_2(X, [0,t]) < \infty$ but $\mathrm{var}(X, [0,t]) = \infty$ for any $t > 0$, i.e., the sum of *small* jumps at the origin is infinite but the sum of the square converges. Moreover, if the process X is continuous with bounded variation then the estimate $\mathrm{var}_2(X, \pi) \le w(X, \pi)\, \mathrm{var}(X, \pi)$ shows that necessarily $\mathrm{var}_2(X, \pi) \to 0$ as the mesh of the partition $\delta(\pi) := \max_i\{t_i - t_{i-1}\}$ vanishes, where $w(X, \pi)$ is the modulus of continuity of X on π, namely,

$$w(X, \pi) = \sup_i \sup \big\{ |X(t) - X(s)| : t, s \in [t_{i-1}, t_i] \big\}.$$

However, we may construct a continuous process X with unbounded variation and with the above vanishing quadratic variation property. Furthermore, for a process X with vanishing quadratic variation we can setup and define the Riemann-Stieltjes integral to show that

$$\int_a^b [X(t)]^m \, \mathrm{d}X(t) = \frac{1}{m+1} \Big[[X(b)]^{m+1} - [X(a)]^{m+1} \Big],$$

for every $b > a \ge 0$. For instance, if $X = W$ is a Wiener process then $\mathbb{E}\{\mathrm{var}_2(W, \pi)\} = t_n - t_0$ and the relevance of the quadratic variation is clear when a pathwise analysis is not available. As seen later, this is a typical behavior for martingale processes. We refer the interested reader to Doob [61, Chapters X–XI, pp. 157–204] for a neat analytic approach.

The technique to treat cad-lag processes is essentially as follows. On one hand, the pathwise study is efficient for cad-lag process with local bounded variation paths. This includes (1) continuous process with local bounded variation path and (2) jump processes X with jumps of local bounded variation, i.e., $\sum_{s \le t} |\delta X(s)| < \infty$ for every $t > 0$. For every cad-lag jump process X, there is only a finite number of jumps that are larger than any deterministic constant r, so the number of larger jumps is finite, i.e., $\sum_{s \le t} \mathbb{1}_{\{|\delta X(s)| \ge r\}} < \infty$. Thus,

$$X(t) = \lim_{r \to 0} \sum_{s \le t} \mathbb{1}_{\{|\delta X(s)| \ge r\}} \delta X(s), \quad \forall t > 0,$$

but the series may not converge absolutely. In particular, we look at jumps processes with jumps satisfying $\sum_{s\le t}|\delta X(s)|^2<\infty$ for every $t>0$, which does not necessarily have local bounded variation paths. For these processes, the pathwise arguments are not longer valid. By imposing a local integrability (with respect to the path, i.e., $\mathbb{E}\{\sum_{s\le\tau_k}|\delta X(s)|^2\}<\infty$), for some increasing sequence of stopping time $\{\tau_1,\tau_2,\dots\}$ with $\tau_k\to\infty$, the compensator/martingale theory can be used. This is part of the stochastic integral theory, where jump processes are better viewed as random measure. Continuous martingale processes with no local bounded variation paths are also studied with non-pathwise technique. A more pure analytic point of view is the use of the so-called *orthogonal random measure*, see Definition 3.33.

Some arguments use a enumeration of the jumps, certainly, they are denumerable but to have them in a ordered way, we need to use ordinal numbers. An intuitive feeling is the fact that we can count through countable ordinals (where each nonempty subset has a first element) as follows:

$$1,2,3,\dots,\infty,\infty+1,\infty+2,\dots,2\infty,2\infty+1,\dots$$
$$3\infty,3\infty+1,\dots,\infty^2,\dots,\infty^3,\dots,\infty^\infty,\infty^\infty+1,\dots$$

where ∞ is the first infinite ordinal. Each countable ordinal is either a successor $\alpha+1$ of some countable ordinal α or a limit ordinal $\beta=\sup\{\alpha:\alpha<\beta\}$, which is the supremum of ordinals less than it. For instance, to count the jumps of a cad-lag process X, first we set $\tau_0=0$, $a_0=X(0)$ and given an ordinal i with successor $i+1$ we define

$$\tau_{i+1}=\inf\big\{t\ge\tau_i:X(t)\neq X(t-)\big\},\qquad a_{i+1}=X(\tau_{i+1})-X(\tau_{i+1}-),$$

while, given a limit ordinal i we define

$$\tau_i=\sup_{j<i}\tau_j,\qquad a_i=X(\tau_i)-X(\tau_i-).$$

Thus, for each countable ordinal i we have defined τ_i and a_i such that $i\le j$ implies $\tau_i\le\tau_j$, where τ_i may be infinite for some ω. Because there is a countable number of jumps, we have $\sup_i\tau_i(\omega)=\infty$ and so, for every t and ω there is a first (necessary countable) ordinal such that for $\kappa=\kappa(t,\omega)$ we have $\tau_i\wedge t=\tau_{i+1}\wedge t$, for every $i\ge\kappa$. This means that all the jumps of X within $[0,t]$ are listed with τ_i and a_i for $i\le\kappa$, the problem is that the possible values of the $\kappa(t,\omega)$ is uncountable (in much the same way that the number of finite ordinals is infinite), so that τ_i may not be an stopping time for some limit ordinal i. However, τ_i is almost surely equal to a stopping time. Indeed, set

$$c_i=\mathbb{E}\{\exp(-\tau_i)\},\qquad c=\inf_i c_i,$$

where the infimum is taken over all countable ordinals. Thus, there exists a sequence of (countable) ordinals $\{\iota(n):n\ge1\}$ independent of ω such that $c_{\iota(n)}\to c$ as $n\to\infty$. If $\iota(\infty)$ is the countable ordinal $\lim_n\iota(n)$ we have $c_{\iota(\infty)}=c$ and the stopping time $\sup\big\{\tau_{\iota(n)}:n\big\}$ is equal to $\tau_{\iota(\infty)}$ almost surely. Hence,

each τ_i in the above construction can be chosen to be a stopping time. On the other hand, to construct a (purely) jump process from its jumps, we need some assumptions on the cad-lag process, e.g., locally bounded variation. Indeed, by induction procedure, we may define the sum or series, starting from $S_0 = a_0$, we set $S_{i+1} = a_{i+1} + S_i$ if i has a successor and $S_i = a_i + \sum_{j<i} S_j$ (which converges absolutely) if i is a limit ordinal. Hence, the process $S_{\kappa(t,\omega)}(\omega)$ or equivalently $\sum_{i\le\kappa(t,\omega)} a_i(\omega)$ is optional. This same argument applies to semi-martingales, as seen in the next section.

3.4 Martingales

Related to the Markov processes with values in $\mathbb{R}^d$, is the concept of (vector) *martingales*. Moreover, the martingale property can be extended to processes with values in Hilbert, Banach or co-nuclear (the strong dual of a nuclear space) spaces, e.g., see Kallianpur and Xiong [123, Chapter 3, pp. 85–126] and Métivier [178].

Definition 3.6 (general martingale)**.** A (general) martingale with states in $E \subset \mathbb{R}^d$ is a (complete) probability measure P on $(\Omega, \mathcal{F})$, together with a measurable mapping M (P-equivalence class) from $(\Omega, \mathcal{F})$ into $(E^T, \mathcal{B}^T(E))$ and an increasing family of completed σ-algebras $(\mathcal{F}_t : t \in T)$ on $(\Omega, \mathcal{F})$ satisfying the *martingale property*

$$\mathbb{E}\{|M_t|\} < \infty,\ \forall t, \qquad \mathbb{E}\{M_t \,|\, \mathcal{F}_s\} = M_s, \quad \text{a.s.} \quad \forall t > s,$$

where M_t is the t-component of M. If the family of σ-algebras $(\mathcal{F}_t : t \in T)$ is not mentioned, then it is assumed $(\mathcal{F}_t : t \in T)$ is the history $(\mathcal{H}_t : t \in T)$ of the process $(M_t : t \in T)$, i.e., $\mathcal{H}_t$ is generated by the random variables $\{M_s : s \le t\}$ and the null sets. Moreover, we say that the martingale is *cad-lag* if $(\mathcal{F}_t : t \in T)$ is a filtration satisfying the usual conditions and except on a set of P-probability zero, the paths of $(M_t : t \in T)$ are cad-lag. The martingale is *continuous* if their paths are continuous. Furthermore, if $d = 1$, i.e., with values in $\mathbb{R}$, we may define also *super-* or *sub-*martingale by replacing the equal sign by either $\le$ or $\ge$ in the above condition. □

In most of the cases considered here, the index T is a bounded real interval or $[0, \infty)$, and the probability P is fixed, so that a (good) particular member of the P-equivalence class is used and called (super- or sub-)martingale. As usually, the conditional expectation operator identifies an equivalence class of processes satisfying the above condition and so another condition on the sample path is needed to make the above martingale condition workable in continuous time, e.g., a minimal condition would be a separable martingale and a more reasonable condition is right-continuity in probability. It is clear that if $(M_t : t \ge 0)$ is a cad-lag martingale relative to (or with respect to) $(\mathcal{F}_t : t \ge 0)$ then it is also a cad-lag martingale relative to its canonical (or natural) filtration $(\mathcal{H}_t : t \ge 0)$, the history of the process, see Definitions 1.26 and 1.27 on Markov processes in

Chapter 1. Certainly, if $(M_t : t \geq 0)$ is a super-martingale then $(-M_t : t \geq 0)$ is a sub-martingale.

When the filtration is the history of the process, the second condition in the martingale property of the above Definition 3.6 can be rephrased as follows

$$\mathbb{E}\Big\{M(t)\prod_{i=1}^{n} h_i\big(M(s_i)\big)\Big\} = \mathbb{E}\Big\{M(s)\prod_{i=1}^{n} h_i\big(M(s_i)\big)\Big\} \tag{3.15}$$

for any integer n, for every $0 \leq s_1 < s_2 \leq \cdots \leq s_n \leq s < t$, any (real-valued) Borel and bounded (in $\mathbb{R}^d$) functions h_i, $i = 1, \ldots, n$. Moreover, if the process $(M_t : t \geq 0)$ is right-continuous in L^1, i.e.,

$$\lim_{t \downarrow s} \mathbb{E}\{|M_t - M_s|\} = 0, \quad \forall s \geq 0,$$

then, applying the martingale property (3.15) to continuous functions h_i and $s_i + \varepsilon_i$, $s + \varepsilon$, with $0 < \varepsilon_i \leq \varepsilon$, we deduce another expression of the martingale property, namely,

$$\mathbb{E}\Big\{M(t)\prod_{i=1}^{n} h_i\big(M(s_i + 0)\big)\Big\} = \mathbb{E}\Big\{M(s)\prod_{i=1}^{n} h_i\big(M(s_i + 0)\big)\Big\} \tag{3.16}$$

for any integer n, for every $0 \leq s_1 < s_2 \leq \cdots \leq s_n \leq s < t$, any (real-valued) continuous and bounded (in $\mathbb{R}^d$) functions h_i, $i = 1, \ldots, n$. Note that relation (3.16) represents the second condition in the martingale property of Definition 3.6, where $(\mathcal{F}_t : t \geq 0)$ is the smallest filtration satisfying the usual conditions which makes the process $(M_t : t \geq 0)$ adapted. This proves that if $(M_t : t \geq 0)$ is a right-continuous (actually it suffices that it be right-continuous in probability) martingale with respect to $(\mathcal{F}_t : t \geq 0)$ then $(M_t : t \geq 0)$ is also a martingale relative to the (possible larger) right-continuous filtration $(\bar{\mathcal{F}}_t : t \geq 0)$, with $\bar{\mathcal{F}}_t = \cap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$. Clearly, if $(M_t : t \geq 0)$ and $(N_t : t \geq 0)$ are two sub-martingales (or super-martingales, respectively) relative to the same filtration $(\mathcal{F}_t : t \geq 0)$ then the new process $(M_t \vee N_t : t \geq 0)$ (or $(M_t \wedge N_t : t \geq 0)$, respectively) is also a sub-martingale (or super-martingale, respectively).

It is clear that the martingale condition does not distinguish modifications of the process, so it may be possible to have a (general) martingale which paths are not necessarily cad-lag, with a filtration which is not necessarily right-continuous (or completed). Thus, the assumption of a setup with a filtration satisfying the usual conditions is not at all granted, completion with null sets is a rather technical condition, but the right-continuity is essential to the well behavior and mathematically workable study of (sub-/super-) martingale processes. This is illustrated by the Doob's regularization result, which uses the following concept.

Let D be a countable dense set in $\mathbb{R}$ (e.g., the rational numbers) and x be a function from D into $\mathbb{R}$. The function x is called *regularisable* if the right-hand and left-hand limits exist finitely within D for every real value, i.e., for every t in $\mathbb{R}$ there exist real values $x(t+)$ and $x(t-)$ such that for every $\varepsilon > 0$ there is a $\delta > 0$ (possible depending on ε, $x(\cdot)$ and t) such that $0 < s - t < \delta$ implies $|x(s) - x(t+)| < \varepsilon$ and $0 < t - s < \delta$ implies $|x(s) - x(t-)| < \varepsilon$. Clearly, when

the initial function x is defined in an interval I of $\mathbb{R}$, first the function x is restricted to the set $I \cap D$ and then the above definition is applied for t in the interval I. In most cases, the countable dense set D can be arbitrary chosen or easily understood from the context. Usually $I = [0, \infty)$ and so $x(0-)$ is either not defined or set equal to 0 for the sake of completeness. If a function x is regularisable (within D) then x_+ and x_- (right and left limits) denote the new functions obtained as the pointwise limits.

An interesting point is that a function $x : D \cap [0, \infty) \to \mathbb{R}$ is regularisable if and only if for every integer N and any compact interval $[a, b]$ with $a < b$ in D, the following expressions are finite,

$$\|x\|_{\infty, D\cap[a,b]} := \sup\big\{|x(s)| \,:\, s \in D \cap [a, b]\big\},$$
$$U_N(x, [a, b]) := \sup\big\{k \,:\, 0 \le s_1 < r_1 < s_2 < \ldots < s_k < r_k \le N, \\ x(s_i) < a,\, x(r_i) > b,\, s_i, r_i \in D,\, \forall i\big\}.$$

As in the discrete case, $U_N(x, [a, b])$ is called the upcrossings of the interval $[a, b]$ by time N.

Theorem 3.7. *Let $M = \{M_t : t \ge 0\}$ be a real-valued family of random variables in a probability space $(\Omega, \mathcal{F}, P)$ which satisfies the (super- or sub-) martingale property relative to an increasing family of σ-algebras $\{\mathcal{F}(t) : t \ge 0\}$, i.e.,*

$$\mathbb{E}\{|M(t)|\} < \infty,\ \forall t, \qquad \mathbb{E}\{M(t) \mid \mathcal{F}_s\} = M_s,\quad a.s.\ \ \forall t > s \ge 0,$$

with $=$ replaced by $\le$ or $\ge$ when (super- or sub-) is used. Then M is regularisable except in a set of probability zero, and the processes $M_+ = \{M(t+) : t \ge 0\}$ and $M_- = \{M(t-) : t \ge 0\}$ are cad-lag (super- or sub) martingales relative to $\{\mathcal{F}(t+) : t \ge 0\}$ and $\{\mathcal{F}(t-) : t \ge 0\}$, respectively. Moreover, if the function $t \mapsto \mathbb{E}\{M(t)\}$ is right-continuous (resp., left-continuous) then M_+ (resp., M_-) is a version of M. □

For a complete detail on the proof see, e.g., Dellacherie and Meyer [58, Section VI.1] or Rogers and Williams [214, Section II.5, Subsections 65–67, pp. 169–174]. Clearly, the above results include the following statement. Let $\{\mathcal{F}(t) : t \ge 0\}$ be a right-continuous and complete the filtration, and assume that the function $t \mapsto \mathbb{E}\{M(t)\}$ is right-continuous (e.g., this mean right-continuity holds if M is a martingale). Then M_+ is a version of M, which is a cad-lag (super- or sub-) martingale relative to the filtration $\{\mathcal{F}(t) : t \ge 0\}$. Moreover, if M is separable then so is M_+ and therefore M_+ is indistinguishable from M, i.e., M itself is a cad-lag (super- or sub-) martingale.

An integrable process with independent increments and zero mean is not always a typical example of martingale, some regularity on the path is needed. For instance, if $\{w(t) : t \ge 0\}$ is a standard Wiener process in $\mathbb{R}^d$ then it is also a continuous martingale, and if $\{p(t) : t \ge 0\}$ is a standard Poisson process then $(M_t : t \ge 0)$, with $M_t = p(t) - \mathbb{E}\{p(t)\}$, is a cad-lag martingale. In general, we will see that if $(X_t : t \ge 0)$ is a cad-lag Markov process with infinitesimal

generator A (see Section 2.8 on Markov-Feller semigroups in Chapter 2) then the stochastic process

$$M_t = \varphi(X_t) - \int_0^t A\varphi(X_s)\mathrm{d}s,\ \forall t \geq 0$$

is a cad-lag martingale, for any (smooth) function φ in the domain $\mathcal{D}(A)$ of the infinitesimal generator A. In fact, as seen later, this is a characterization of the Markov processes in terms of the so-called *martingale problem*. On the other hand, the concept of martingale is a sort of complementary definition with respect to bounded variation processes, in the sense that the only continuous martingale of bounded variation is the trivial or constant process.

To study martingales we begin with either (1) a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$ satisfying the usual conditions and we look at cad-lag stochastic processes $(X_t : t \geq 0)$ as random variables with valued in the canonical space $D([0,\infty), \mathbb{R}^d)$ or (2) a canonical space $D = D([0,\infty), \mathbb{R}^d)$, with its Borel σ-algebra $\mathcal{B}$, the canonical process $(X_t := \omega(t),\ t \geq 0)$ and it associated the filtration $(\mathcal{F}_t : t \geq 0)$ and we look for probability measures on D. Thus, a cad-lag martingale is viewed as a random variable with values in the canonical space D, identified with its equivalence class, namely, all processes which are indistinguishable (or equivalent) of it, and as long as we use *cad-lag* (or *separable*) processes this agree with the notion of version (or modification), see Section 1.6 in Chapter 1.

We rephrase the above martingale concept

Definition 3.8 (martingale)**.** A martingale (process) relative to a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ is a random variable M (P-equivalence class) with values into the canonical space $D([0,\infty), \mathbb{R}^d)$ satisfying the *martingale property*

$$\mathbb{E}\{|M(t)|\} < \infty,\ \forall t, \qquad \mathbb{E}\{M(t) \mid \mathcal{F}(s)\} = M(s),\ \text{ a.s. } \forall t > s,$$

where $M(t) := M(\omega)(t)$. If the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is not mentioned, then it is assumed that $\{\mathcal{F}(t) : t \geq 0\}$ is the smallest filtration satisfying the usual condition, which renders the process $\{M(t) : t \geq 0\}$ adapted. Moreover, the martingale is called *continuous* if M take values into the canonical space $C([0,\infty), \mathbb{R}^d)$ almost surely, and it is called *uniformly integrable* if the family of random variables $\{M(t),\ t \geq 0\}$ is uniformly integrable, i.e., for any $\varepsilon > 0$ there is a $r > 0$ sufficiently large such that $P\{|M(t)| \geq r\} \leq \varepsilon$, for any t in $[0,\infty)$. When $d = 1$, i.e., with values in $\mathbb{R}$, we may define also *super-* or *sub-*martingale by replacing the equal sign by either $\leq$ or $\geq$ in the above condition. Sometimes, martingales are considered in a bounded time interval instead of the semi-line $[0,\infty)$. □

First, note the role of uniformly integrability by mentioning Doob's martingale convergence and optional-sampling results

Theorem 3.9. *If M is martingale bounded in L^1, i.e., $\sup_t \mathbb{E}\{|M(t)|\} < \infty$, the limit $M(\infty) := \lim_{t\to\infty} M(t)$ exists almost surely and the convergence of $M(t)$ to $M(\infty)$ is in L^1 if and only if the martingale is uniformly integrable. On the*

other hand, if M is an uniformly integrable martingale then (a) *the family of $\mathbb{R}^d$-valued random variable $\{M(\tau) : \tau$ is a stopping time$\}$ is uniformly integrable, and* (b) *for any stopping times $\tau \le \theta$ the equality $\mathbb{E}\{M(\theta)\,|\,\mathcal{F}(\tau)\} = M(\tau)$ holds almost surely.* □

As in the discrete case, the proof is mainly based on the Doob's upcrossing estimate. A (super-/sub-) martingale M satisfying the property (a) of the above theorem is called of *class (D)* (Dirichlet class). Note that an uniformly integrable super(or sub)-martingale need not to be of class (D). However, for any nonnegative sub-martingale X we have

$$r\,P\big(\sup_{s\le t} X(s) \ge r\big) \le \mathbb{E}\{X(t)\mathbb{1}_{\sup_{s\le t} X(s)\ge r}\} \le \mathbb{E}\{X(t)\}, \tag{3.17}$$

and therefore

$$\|\sup_{s\le t} X(s)\|_p \le p'\,\|X(t)\|_p, \quad \text{with } 1/p + 1/p' = 1, \tag{3.18}$$

actually, valid even if t is replaced by a stopping time τ. Here $\|\cdot\|_p$ denotes the norm in $L^p(\Omega, P, \mathcal{F})$.

Note that (3.11) implies that for any positive cad-lag martingale M, which is written as $M(t) = \mathbb{E}\{M(\infty)|\mathcal{F}(t)\}$ if M is uniformly integrable, and any local integrable increasing process A with $A(0) = 0$, we have

$$\mathbb{E}\big\{X(t)A(t)\big\} = \mathbb{E}\Big\{\int_0^t Y(s)\mathrm{d}A(s)\Big\}, \quad \forall t \in (0,\infty),$$

and even for $t = \infty$ if M is uniformly integrable.

Now, based on (3.12), an local integrable increasing process A with $A(0) = 0$ is called *natural* if

$$\mathbb{E}\Big\{\int_0^T M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^T M(t-)\mathrm{d}A(t)\Big\}, \quad \forall T \in \mathbb{R}, \tag{3.19}$$

for every nonnegative, bounded and cad-lag continuous martingale M. Since the process $Y(t) = Y(t)\mathbb{1}_{t<\tau} + Y(\tau)\mathbb{1}_{s\ge\tau}$ is a martingale for any stopping time τ, we deduce that (3.19) is equivalent to either

$$\mathbb{E}\Big\{\int_0^\infty M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^\infty M(t-)\mathrm{d}A(t)\Big\}, \tag{3.20}$$

or

$$\mathbb{E}\Big\{\int_0^\tau M(t)\mathrm{d}A(t)\Big\} = \mathbb{E}\Big\{\int_0^\tau M(t-)\mathrm{d}A(t)\Big\}, \quad \forall \text{ stopping time } \tau,$$

and that the increasing process $B(t) = B(t)\mathbb{1}_{t<\tau} + B(\tau)\mathbb{1}_{s\ge\tau}$ is also natural. Finally, if A is an integrable increasing natural process then (3.20) holds for any uniformly integrable cad-lag martingale M.

With all these properties in place, we can check that if X is a cad-lag sub martingale and A and B are two cad-lag increasing natural processes such that

$A(0) = B(0) = 0$ and $X - A$ and $X - B$ are (cad-lag) martingale then $A(t) = B(t)$, almost surely, for every $t \geq 0$.

As we can check later, this notion of (cad-lag) increasing natural process agrees with the more general condition of (cad-lag) increasing predictable process. For further details, the reader may want to take a look at certain points in the book by Meyer [179].

The following decomposition is extremely useful to extend the previous result to sub-martingales.

Theorem 3.10 (Doob-Meyer)**.** *If X is a (continuous) sub-martingale of class (D) then there exists a uniformly integrable martingale M and an integrable predictable (continuous) monotone increasing process A, both null at time zero such that $X = X(0) + M + A$. Moreover, this decomposition is unique.* □

For instance, a comprehensive proof of this fundamental results can be found Rogers and Williams [214, Section VI.6, pp. 367–382]. In particular, if X is an adapted (cad-lag) increasing process satisfying $\mathbb{E}\{\sup_t |X(t)|\} < \infty$ then X is a sub-martingale of class (D) and the above decomposition yields the predictable *compensator* as in Definition 3.5. Certainly, this can be extended to integrable bounded variation processes, by using the positive and negative variation.

Therefore, the previous convergence Theorem 3.9 can be extended to super-/sub-martingales of class (D) and the process $A = A_X$ is called the (predictable) *compensator* of the sub-martingale M. Note that μ_A on $[0, \infty) \times \Omega$ associated with the increasing process A, as defined by (3.13), satisfies

$$\mu_A(\rrbracket \tau, \theta \rrbracket) = \mathbb{E}\{A_\theta - A_\tau\} = \mathbb{E}\{X_\theta - X_\tau\},$$

for any stopping times $\tau \leq \theta$ and where the the stochastic interval $\rrbracket \tau, \theta \rrbracket$ is viewed as the subset $\{(\omega, t) : \tau(\omega) < t \leq \theta(\omega)\}$ of $[0, \infty) \times \Omega$. This is one of the key elements used in the definition of the process A, i.e., the fact that for any given sub-martingale X of class (D) we can construct a unique bounded (positive) measure on $[0, \infty) \times \Omega$ defined by $\mu_A(]\tau, \theta]) = \mathbb{E}\{X_\theta - X_\tau\}$. Actually, it is also established in the Doléans' proof of the above decomposition, that for a quasi-leftcontinuous (or regular) sub-martingale X (i.e., $\mathbb{E}\{X(\tau)\} = \mathbb{E}\{X(\tau-)\}$ for any predictable stopping time τ or equivalently A is continuous) the predictable (or conditional) variation for a partition $\pi = (t_0 < t_1 < \cdots < t_n)$,

$$^{\mathrm{p}}\mathrm{var}(X, \pi) := \sum_{i=1}^{n} \mathbb{E}\{X(t_i) - X(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\}, \tag{3.21}$$

which is equal to $^{\mathrm{p}}\mathrm{var}(A, \pi)$, has the following property: for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbb{E}\{|^{\mathrm{p}}\mathrm{var}(X, \pi_t) - A(t)|\} \leq \varepsilon$ for any partition $\pi = \pi_t$ with $t_n = t$ and $t_i - t_{i-1} \leq \delta$, for every $i = 1, 2, \ldots, n$.

Let us denote by $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ the space of square-integrable martingales M null at time zero, i.e., besides the martingale conditions in Definition 3.8 we impose $M(0) = 0$ and $\sup_{t \geq 0} \mathbb{E}\{|M(t)|^2\} < \infty$. A square-integrable martingale M is uniformly integrable and the convergence theorem applies to

produce a $\mathcal{F}(\infty)$-measurable random variable $M_\infty = M(\infty)$ with values in $\mathbb{R}$ (or $\mathbb{R}^d$) and square-integrable such that $M(t) = \mathbb{E}\{M(\infty)\,|\,\mathcal{F}(t)\}$. Hence, the space $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ can be identified with the closed subspace of the Hilbert space $L^2(\Omega, P, \mathcal{F}_\infty)$, $\mathcal{F}_\infty = \mathcal{F}(\infty)$, satisfying $\mathbb{E}\{M(\infty)\,|\,\mathcal{F}(0)\} = 0$. Note that if M^* denotes the sup-process defined by $M^*(t) := \sup_{s\leq t} |M(s)|$ and its limit $M^*(\infty) := \sup_{t\geq 0} |M(t)|$ then we have

$$\mathbb{E}\{|M^*(\infty)|^2\} \leq 4 \sup_{t\geq 0} \mathbb{E}\{|M(t)|\} = 4\,\mathbb{E}\{|M(\infty)|^2\},$$

after using Doob's estimate (3.18). Therefore, we can treat $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ as a Banach space with the norm $\|M^*(\infty)\|_p$, with $p = 2$, for any element M, without changing the topology. Moreover, the space of *continuous* square-integrable martingale processes, denoted by $M^2_c(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ is a closed subspace of the Hilbert space $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$. Thus, we may consider its orthogonal complement referred to as *purely discontinuous* square-integrable martingale processes null at time zero and denoted by $M^2_d(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, of all square-integrable martingale processes Y null at time zero satisfying $\mathbb{E}\{M(\infty)\,Y(\infty)\} = 0$ for all elements M in $M^2_c(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, actually, M and Y are what is called *strongly orthogonal*, i.e., $(M(t)\,Y(t) : t \geq 0)$ is an uniformly integrable martingale. The concept of strongly orthogonal is actually stronger than the concept of *orthogonal* in M^2 and weaker than imposing $M(t) - M(s)$ and $Y(t) - Y(s)$ independent of $\mathcal{F}(s)$ for every $t > s$.

Let M be a (continuous) square-integrable martingale process null at time zero, in a given filtered space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$. Based on the above argument M^2 is a sub-martingale of class (D) and Doob-Meyer decomposition Theorem 3.10 applies to get a unique predictable (continuous) increasing process $\langle M\rangle$, referred to as the *predictable quadratic variation* process. Thus, for a given element M in $M^2(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$, we have a unique pair M_c in $M^2_c(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ and M_d in $M^2_d(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ such that $M = M_c + M_d$. Applying Doob-Meyer decomposition to the sub-martingales M_c and M_d we may define (uniquely) the so-called *quadratic variation* (or optional quadratic variation) process by the formula

$$[M](t) := \langle M_c\rangle(t) + \sum_{s\leq t} (M_d(s) - M_d(s-))^2, \quad \forall t > 0. \tag{3.22}$$

Note that $[M_c] = \langle M_c\rangle$ and $M_d(t) - M_d(t-) = M(t) - M(t-)$, for any $t > 0$, but that $\langle M\rangle$ is the predictable (dual) projection of the increasing process $[M]$, as defined in Section 3.3. We re-state these facts for a further reference

Theorem 3.11 (quadratic variations)**.** *Let M be a (continuous) square-integrable martingale process null at time zero, in a given filtered space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$. Then* (1) *there exists a unique predictable (continuous) integrable monotone increasing process $\langle M\rangle$ null at time zero such that $M^2 - \langle M\rangle$ is a (continuous) uniformly integrable martingale, and* (2) *there exists a unique optional (continuous) integrable monotone increasing process $[M]$ null at time zero such that $[M](t) - [M](t-) = (M(t) - M(t-))^2$, for any $t > 0$, and $M^2 - [M]$ is a*

(continuous) uniformly integrable martingale. Moreover $M = 0$ if and only if either $[M] = 0$ or $\langle M \rangle = 0$. □

Also, the optional quadratic variation can be defined by means of the stochastic integral and for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\mathbb{E}\Big\{ \sup_{0<t\leq 1/\varepsilon} |\mathrm{var}_2(M, \pi_t) - [M](t)| \Big\} \leq \varepsilon,$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \leq \delta$, for every $i = 1, 2, \ldots, n$, where $\mathrm{var}_2(M, \pi_t)$ is the optional quadratic variation operator

$$\mathrm{var}_2(M, \pi_t) := \sum_{i=1}^{n} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^2\}, \tag{3.23}$$

and t_i in the partition may be stopping times. This previous limit could be used as definition of $[M]$, and then we may define the predictable variation process $\langle M \rangle$ as the compensator of the optional quadratic variation $[M]$.

If the filtration can be chosen quasi-left continuous (i.e., satisfying $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable stopping time τ) or equivalently if the predictable variation process $\langle M \rangle$ is continuous then for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t) - \langle M \rangle(t)|\} \leq \varepsilon$ for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \leq \delta$, for every $i = 1, 2, \ldots, n$, where $\mathrm{var}_2(M, \pi_t)$ is the optional quadratic variation operator and ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t)$ is the predictable quadratic variation operator defined by (3.21), i.e., ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi_t)$ converges in L^1 to $\langle M \rangle$ as the mesh goes to zero.

These are key results in the study of martingales and foundation of the stochastic integrals for continuous martingales. To understand the convergence in the L^1-norm of the predictable quadratic variation as defined in Theorem (3.11), first we realize that the predictable quadratic variation operator on M is equal to the predictable variation operator on $\langle M \rangle$, i.e., ${}^{\mathrm{p}}\mathrm{var}_2(M, \pi) = {}^{\mathrm{p}}\mathrm{var}(\langle M \rangle, \pi)$, as defined by (3.21). Setting $A^k(s) := \min\{\langle M \rangle(s), k\}$, for $s \geq 0$, for a given partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ we consider the (finite) sequence of bounded random variables

$$x_i := \mathbb{E}\{A^k(t_i) - A^k(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\} - [A^k(t_i) - A^k(t_{i-1})],$$

for $i = 1, 2, \ldots, t_n$, which are orthogonal in $L^2(\Omega, \mathcal{F}, P)$. Based on the elementary bound $(a - b)^2 \leq 2a^2 + 2b^2$ and Jensen's inequality we obtain

$$\mathbb{E}\{x_i^2\} \leq 4\,\mathbb{E}\{[A^k(t_i) - A^k(t_{i-1})]^2\},$$

which yields

$$\mathbb{E}\{[{}^{\mathrm{p}}\mathrm{var}(A^k, \pi_t) - A^k(t)]^2\} = \sum_{i=1}^{n} \mathbb{E}\{x_i^2\} \leq$$
$$\leq 4 \sum_{i=1}^{n} \mathbb{E}\{[A^k(t_i) - A^k(t_{i-1})]^2\} \leq 4\,\mathbb{E}\{\rho(A^k, [0, t], \delta)\, A^k(t)\},$$

where $\rho(A^k, [0,t], \delta)$ is the modulus of continuity of A^k, i.e.,

$$\rho(A^k, [0,T], \delta) := \sup\{A^k(t) - A^k(s) \,:\, 0 \le s < t \le T\},$$

and $0 < t_i - t_{i-1} \le \delta$, $i \ge 1$. Since

$$\begin{aligned}\mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t) - \langle M\rangle(t)|\} \le \\ \le \mathbb{E}\{|{}^{\mathrm{p}}\mathrm{var}(A^k,\pi_t) - A^k(t)|\} + \mathbb{E}\{\langle M\rangle(t) - A^k(t)\},\end{aligned}$$

we obtain the L^1 convergence stated in quadratic variation theorem above, when we assume that $\langle M\rangle$ is continuous, which is equivalent to the quasi-left continuity of the filtration.

Moreover, if M is a continuous martingale with $M(t) = 0$ then

$${}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t) \le {}^{\mathrm{p}}\mathrm{var}(M,\pi_t)\,\rho(M,[0,t],\delta),$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \le \delta$, $i = 1, 2, \ldots$. Thus, the predictable square variation process $\langle M\rangle$ vanishes, when M has almost surely path of bounded variation, i.e., M is the null process.

Similarly, the predictable ℓ-variation ($\ell > 2$) of any ℓ-integrable martingale M null at time zero is defined by

$${}^{\mathrm{p}}\mathrm{var}_\ell(M,\pi) := \sum_{i=1}^{n} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^\ell \,|\, \mathcal{F}(t_{i-1})\}, \tag{3.24}$$

with $\pi = (t_0 < t_1 < \cdots < t_n)$. Therefore, the inequality

$${}^{\mathrm{p}}\mathrm{var}_\ell(M,\pi_t) \le {}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t)\,[\rho(\langle M\rangle,[0,t],\delta)]^{\ell/2-1},$$

for any partition $\pi_t = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $0 < t_i - t_{i-1} \le \delta$, $i = 1, 2, \ldots$ and $\ell > 2$, proves that ${}^{\mathrm{p}}\mathrm{var}_\ell(M,\pi_t) \to 0$ almost surely as the mesh of the partition δ vanishes. These two facts about the convergence of the predictable quadratic variation and ℓ-variation ($\ell > 2$) are essential to define the stochastic integral.

With all this in mind, for any two square-integrable martingale process null at time zero M and N we define the predictable and optional quadratic covariation processes by

$$\begin{cases} \langle M, N\rangle := \big(\langle M+N\rangle - \langle M-N\rangle\big)/4, \\ [M, N] := \big([M+N] - [M-N]\big)/4, \end{cases} \tag{3.25}$$

which are processes of integrable bounded variations.

Most of proofs and comments given in this section are standard and can be found in several classic references, e.g., the reader may check the books by Dellacherie and Meyer [58, Chapters V–VIII], Jacod and Shiryaev [117], Karatzas and Shreve [124], Neveu [189], Revuz and Yor [212], among others.

Given a cad-lag integrable process X we can associate the so-called Föllmer *finitely additive* measure on the ring $\mathcal{R}$ generated by all predictable rectangles by the expression

$$\begin{cases} \lambda_X(]s,t] \times F) := \mathbb{E}\{(X_t - X_s)\, \mathbb{1}_F\}, \quad \forall t > s \geq 0,\ F \in \mathcal{F}_s, \\ \lambda_X(\{0\} \times F) := 0 \quad \forall F \in \mathcal{F}_0. \end{cases} \tag{3.26}$$

The *variation* of λ_X for any A in $\mathcal{R}$ is defined by

$$|\lambda_X|(A) := \sup\Big\{ \sum_{i=1}^{n} |\lambda_X(A_i)| \Big\}, \tag{3.27}$$

where the supremum is taken over all finite partition of A, i.e., $A = \cup_{i=1}^{n} A_i$, with A_i in $\mathcal{R}$ and $A_i \cap A_j = \emptyset$ if $i \neq j$. Replacing the absolute value $|\cdot|$ in the above definition (3.27) with the positive or negative part, the positive or negative variation λ_X^+ or λ_X^- is also defined. It satisfies $\lambda_X^+ + \lambda_X^- = |\lambda_X|$ and $\lambda_X^+ - \lambda_X^- = \lambda_X$. It is easily seen that the three variations $|\lambda_X|$, λ_X^+ and λ_X^- are finitely additive measure.

The process X is called a *quasi-martingale* if its variation $|\lambda_X|([0,t] \times \Omega)$ given by (3.27) is finite for every $t > 0$. An interesting point is that the sum of a bounded martingale and an adapted (cad-lag) integrable process or the difference of two non-negative super (or sub) martingale is a quasi-martingale, and indeed (see Protter [206, Theorem II.3.8, pp. 96-97]) any quasi-martingale admits a decomposition as a difference of two positive right continuous super (or sub) martingales. The notion of quasi-martingale can be easily generalized to the multi-dimensional case (even with values in a Banach space). It can be proved that X is a quasi-martingale if and only if

$$\sup\Big\{ \sum_{i=1}^{n} \big|\mathbb{E}\{X(t_i) - X(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\}\big| \Big\} < \infty,$$

where the supremum is taken over all partition of the form $0 = t_0 < t_i < \cdots < t_n = t$, any $n \geq 1$.

If the initial process X has integrable bounded variation then λ_X can be extended to a σ-additive measure on the predictable σ-algebra $\mathcal{P}$. Conversely, the finitely additive measure λ_X on $\mathcal{R}$ can be extended to σ-additive measure on $\mathcal{P}$ if X is a quasi-martingale of the class (D), i.e., such that the family of random variables $\{X(\tau) : \tau \text{ is a stopping time}\}$ is uniformly integrable. The interested reader may consult the book by Métivier [178].

3.5 Semi-Martingales

Starting from a (super-/sub-) martingale $(M(t) : t \geq 0)$ relative to a filtration $(\mathcal{F}(t) : t \geq 0)$ and a stopping time τ, we may stop M at time τ and preserve the martingale property, i.e., define a new (super-/sub-) martingale $(M(t \wedge \tau) : t \geq 0)$ relative to the (stopped) filtration $(\mathcal{F}(t \wedge \tau) : t \geq 0)$. Thus, the martingale

property is stable under the above stopping time operation and give rise to the following concept.

Definition 3.12 (localization)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a given filtered space. The term *locally* or *local* is applied to a property relative to a stochastic processes $\{X(t) : t \geq 0\}$ with the understanding that there exists a sequence of stopping times τ_n, with $\tau_n \to \infty$, such that the stopped process $\{X(t \wedge \tau_n) : t \geq 0\}$ satisfies the required property for any n, e.g., we say that $\{X(t) : t \geq 0\}$ is a local martingale or locally integrable or locally bounded if for any n the process $\{X(t \wedge \tau_n) : t \geq 0\}$ is respectively a martingale or integrable or bounded. The sequence $\{\tau_n : n = 1, 2, \ldots\}$ is called a *reducing sequence* for the process $\{X(t) : t \geq 0\}$. □

In some cases, it may be some ambiguity regarding the above definition, e.g., when we refer to a *local* uniform integrable martingale or to a uniform integrable *local* martingale, fortunately, we can prove that all cases used here are exactly the same. One of the reasons for using the above localization is the following construction: if $(\tau_n : n \geq 1)$ is a reducing sequence of stopping times for a local martingale X defined on $[0, \tau), \tau_n \to \tau$ then define $\tau_0 := 0$ and

$$\gamma(t) := \begin{cases} t - k + 1 & \text{if } \tau_{k-1} + k - 1 \leq t < \tau_k + k - 1, \\ \tau_k & \text{if } \tau_k + k - 1 \leq t < \tau_k + k, \end{cases}$$

which yields $\gamma(t) = t$ on $[0, \tau_1]$, $\gamma(t) = \tau_1$ on $[\tau_1, \tau_1 + 1]$, $\gamma(t) = t - 1$ on $[\tau_1 + 1, \tau_2 + 1]$, $\gamma(t) = \tau_2$ on $[\tau_2 + 1, \tau_2 + 2]$, $\gamma(t) = t - 2$ on $[\tau_2 + 2, \tau_3 + 2]$, $\gamma(t) = \tau_3$ on $[\tau_3 + 2, \tau_3 + 3]$, etc. Strictly speaking γ compresses the time interval $[0, \infty)$ onto $[0, \tau)$ and $X^\gamma := (X(\gamma(t)) : t \geq 0)$ is well defined, actually X^γ is a martingale relative to $(\mathcal{F}(\gamma(t)) : t \geq 0)$. This construction is a key element to extend previous results on (super-/sub-) martingales to *local* (super-/sub-) martingales, where integrability is no more an issue. Actually, by means of the Doob's optional sampling theorem if (super-/sub-) martingale X relative to the filtration $(\mathcal{F}(t) : t \geq 0)$ then $X^\gamma := (X(\gamma(t)) : t \geq 0)$ is a (super-/sub-) martingale relative to the filtration $(\mathcal{F}(\gamma(t)) : t \geq 0)$ for any family of stopping times such that $P(\gamma(s) \leq \gamma(t) < \infty) = 1$ for any $0 \leq s \leq t$, for instance, see the books by Ikeda and Watanabe [110, pp. 32–34] or Durrett [67, pp. 38–42] for more details and comments.

A very important point in the localization principle is the fact that when a property is localized, we are not given away only the integrability $\mathbb{E}\{|X(t)|\} < \infty$, for any $t \geq 0$, more is included. For instance, there are continuous non-negative super-martingales which are local martingales but not martingales, a typical example is a $M(t) := 1/|W(t)|$, where W is a Wiener process in $\mathbb{R}^3$ with $|W(0)| = 1$. Indeed, by means of the (Gaussian) density of W we may show that $\mathbb{E}\{|M(t)|\} < \infty$ and because $1/|x|$ is a fundamental solution for the Laplace equation we can complete the argument. Note that $\tau_n := \inf\{t \geq 0 : |M(t)| > n\}$ is a reducing sequence for M but the family of random variables $\{M(t \wedge \tau_n) : n = 1, 2, \ldots\}$ is not uniformly integrable. On the other hand, a

local (super-/sub-) martingale X satisfying

$$\mathbb{E}\{\sup_{s\leq t}|X(s)|\}<\infty,\quad \forall t,$$

is indeed a (super-/sub-) martingale, note the sup inside the mathematical expectation.

For any local sub-martingale we may define a reducing sequence as follows $\tau_n := \inf\{t\in[0,n] : |X(t)|\geq n\}$. Thus, a local sub-martingale is locally of class (D) and Theorem 3.10 applies to the stopped process. Thus the uniqueness yields the following local version of Doob-Meyer decomposition: A local sub-martingale X can be expressed as $X = X(0) + M + A$, where M is a local martingale and A is a predictable locally integrable monotone increasing process, both null at time zero. The case where the (local) predictable compensator A is continuous is very import. As mentioned above, these are quasi-left continuous processes, which are characterized by the condition either $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ or $P\{X(\tau) = X(\tau-)\} = 1$ valid for any predictable stopping time τ.

Note that not all local martingales are locally square-integrable martingale. For instance a local martingale X with locally square-integrable jump process $\delta X = (X(t) - X(t-) : t > 0)$ is actually a locally square-integrable martingale, so that continuous local martingales are locally square-integrable martingale. Hence, for a given local martingale M the predictable quadratic variation process $\langle M_c\rangle$ is defined as the unique predictable locally integrable monotone increasing process null at time zero such that $M_c^2 - \langle M_c\rangle$ is a (continuous) local martingale. Next, the (optional) quadratic variation process $[M]$ is defined as

$$[M](t) := \langle M_c\rangle(t) + \sum_{s\leq t}[M(s) - M(s-)]^2,\quad \forall t\geq 0, \tag{3.28}$$

where the second term in the right-hand side is an optional monotone increasing process null at time zero, not necessarily locally integrable (in sense of the localization in Ω defined above). An important point here is the fact that the square of the jumps are locally integrable, i.e., the process $\sqrt{[M]}$ is locally integrable and therefore

$$\sum_{s\leq t}[M(s) - M(s-)]^2<\infty,\quad \forall t>0, \tag{3.29}$$

almost surely. This follows from the use the compensator of Definition 3.5 and two facts: (1) for any cad-lag process there is only a finite number of jumps greater than a positive constant, i.e., $|M(s) - M(s-)| > \varepsilon$, almost surely, and (2) any local martingale with jumps bounded by a constant is locally square-integrable.

On the other hand, given a local martingale M and a real number κ there exists two local martingales V_κ and N_κ such that $M = V_\kappa + N_\kappa$, where V is a locally bounded (or finite) variation process and the jumps of N are bounded by κ. Thus, a local martingale is the sum of a local square-integrable martingale and a locally finite variation process.

It is also clear that we can write $M = M_c + M_d$, where M_c is the continuous local martingale part and M_d is the so-called purely discontinuous local martingale part, so that $M(t) - M(t-) = M_d(t) - M_d(t-)$ for any $t > 0$. Beside the defining fact that any square-integrable purely discontinuous martingale is orthogonal to any square-integrable integrable continuous martingale, we may define purely discontinuous martingales as locally uniform L^2-limits of local martingales with local finite variation, i.e., M_d is purely discontinuous if and only if there exists a sequence $\{X_n : n \geq 1\}$ of locally integrable finite variation processes of the form

$$X_n(t) = M_d(0) + A_n(t) - A_n^p(t), \qquad A_n(t) = \sum_{0<s\leq t} \big(X_n(s) - X_n(s-)\big),$$

where A_n^p is the compensator of A_n as in Definition 3.5, such that

$$\mathbb{E}\big\{ \sup_{0\leq t\leq T} |X_n(t) - M_d(t)|^2 \big\} \to 0,$$

for any constant $T > 0$, e.g., see Kallenberg [121, Theorem 26.14, pp. 527–529].

On the other hand, if the local martingale M is also locally square-integrable then the predictable quadratic variation process $\langle M \rangle$ is defined as the unique predictable locally integrable monotone increasing process null at time zero such that $M^2 - \langle M \rangle$ is a local martingale. In this case $\langle M \rangle$ is the predictable compensator of $[M]$. Hence, via the predictable compensator we may define the angle-bracket $\langle M \rangle$ when M is only a local martingale, but this is not actually used. An interesting case is when the predictable compensator process $\langle M \rangle$ is continuous, and therefore $[M] = \langle M \rangle$, which is the case when the initial local martingale is a quasi-left continuous process. Finally, the optional and predictable quadratic variation processes are defined by coordinates for local martingale with values in $\mathbb{R}^d$ and even the *co-variation* processes $\langle M, N \rangle$ and $[M, N]$ are defined by orthogonality as in (3.25) for any two local martingales M and N. For instance we refer to Rogers and Williams [214, Theorem 37.8, Section VI.7, pp. 389–391]) where it is proved that $[M, N]$ defined as above (for two local martingales M and N) is the unique optimal process such that $MN - [M, N]$ is a local martingale where the jumps satisfy $\delta[M, N] = \delta M\, \delta N$.

It is of particular important to estimate the moments of a martingale in term of its quadratic variation. For instance, if M is a square-integrable martingale with $M(0) = 0$ then $\mathbb{E}\{|M(t)|^2\} = E\{[M](t)\} = \mathbb{E}\{\langle M \rangle(t)\}$. If M is only locally square-integrable martingale then

$$\mathbb{E}\{|M(t)|^2\} \leq E\{[M](t)\} = \mathbb{E}\{\langle M \rangle(t)\}.$$

In any case, by means of the Doob's maximal inequality (3.18), we deduce

$$\mathbb{E}\big\{ \sup_{0\leq t\leq T} |M(t)|^2 \big\} \leq 4\, \mathbb{E}\{\langle M \rangle(T)\},$$

for any positive constant T, even a stopping time. This can be generalized to the following estimate: for any constant p in $(0, 2]$ there exists a constant C_p

depending only on p (in particular, $C_2 = 4$ and $C_1 = 3$) such that for any local martingale M with $M(0) = 0$ and predictable quadratic variation $\langle M\rangle$ we have the estimate

$$\mathbb{E}\{\sup_{0\le t\le T}|M(t)|^p\} \le C_p\,\mathbb{E}\{\big(\langle M\rangle(T)\big)^{p/2}\}, \tag{3.30}$$

for every stopping time T. If $\langle M\rangle$ is continuous (i.e., M is quasi-left continuous), we can proceed as follows. For a given $r > 0$ and a local martingale M we set $\tau_r := \inf\{t \ge 0 : \langle M\rangle(t) \ge r^2\}$, with $\tau_r = 0$ if $\langle M\rangle(t) < r^2$ for every $t \ge 0$. Since $\langle M\rangle$ is continuous we have $\langle M\rangle(\tau_r) \le r^2$ and $(M(t\wedge\tau_r) : t \ge 0)$ is a bounded martingale. Thus, for any $c > 0$ we have

$$\begin{aligned} P(\sup_{t\le T\wedge\tau_r} M^2(t) > c^2) &\le \frac{1}{c^2}\mathbb{E}\{M^2(T\wedge\tau_r)\} = \\ &= \frac{1}{c^2}\mathbb{E}\{\langle M\rangle(T\wedge\tau_r)\} \le \frac{1}{c^2}\mathbb{E}\{r^2\wedge\langle M\rangle(T)\}. \end{aligned}$$

Hence, for $r = c$ we obtain

$$\begin{aligned} P(\sup_{t\le T} M^2(t) > c^2) &\le P(\tau_c < T) + P(\sup_{t\le T\wedge\tau_c} M^2(t) > c^2) \le \\ &\le P(\langle M\rangle(t) > c^2) + \frac{1}{c^2}\mathbb{E}\{c^2\wedge\langle M\rangle(T)\}. \end{aligned}$$

Now, setting $c = r^{1/p}$, integrating in r and using Fubini's theorem we deduce

$$\begin{aligned} \mathbb{E}\{\sup_{t\le T}|M(t)|^p\} &= \int_0^\infty P(\sup_{t\le T} M^2(t) > r^{2/p})\,dr \le \\ &\le \int_0^\infty \big[P(\sup_{t\le T}\langle M\rangle(t) r^{2/p}) + \\ &\quad + \frac{1}{r^{2/p}}\mathbb{E}\{r^{2/p}\wedge\langle M\rangle(T)\}\big]\,dr = \frac{4-p}{2-p}\,\mathbb{E}\big\{\big[\langle M\rangle(T)\big]^{p/2}\big\}, \end{aligned}$$

so that we can take $C_p = (4-p)/(2-p)$, for $0 < p < 2$. If $\langle M\rangle$ is not continuous, then it takes longer to establish the initial bound in c and r, but the estimate (3.30) follows. This involves Lenglart–Robolledo inequality, see Liptser and Shiryayev [158, Section 1.2, pp. 66–68].

A very useful estimate is the so-called Davis-Burkhölder-Gundy inequality for local martingales vanishing at the initial time, namely

$$c_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\} \le \mathbb{E}\{\sup_{t\le T}|M(t)|^p\} \le C_p\,\mathbb{E}\{\big([M](T)\big)^{p/2}\}, \tag{3.31}$$

valid for any $T \ge 0$ and $p \ge 1$ and some universal constants $C_p > c_p > 0$ independent of the filtered space, T and the local martingale M. In particular, we can take $C_1 = C_2 = 4$ and $c_1 = 1/6$. Moreover, a stopping time τ can be used in lieu of the time T and the above inequality holds true.

Note that when the martingale M is continuous the optional quadratic variation $[M]$ may be replaced with the predictable quadratic variation angle-brackets $\langle M\rangle$. Furthermore, the p-moment estimate (3.30) and (3.31) hold for any $p > 0$

as long as M is a continuous martingale. All these facts play an important role in the continuous time case. By means of this inequality we show that any local martingale M such that $\mathbb{E}\{|M(0)| + \left(\sup_{t>0}[M](t)\right)^{1/2}\} < \infty$ is indeed a uniformly integrable martingale. For instance, we refer to Kallenberg [121, Theorem 26.12, pp. 524–526], Liptser and Shiryayev [158, Sections 1.5–1.6, pp. 70–84] or Dellacherie and Meyer [58, Sections VII.3.90–94, pp. 303–306], for a proof of the above Davis-Burkhölder-Gundy inequality for (non-necessary continuous) local martingale and $p \geq 1$, and to Revuz and Yor [212, Section IV.4, pp. 160–171] for continuous local martingales.

Now, combining bounded variation processes with martingales processes and localization arguments, we are led to the following definition.

Definition 3.13 (semi-martingale)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a given filtered space. A *semi-martingale* is a random variable X (P-equivalence class) with values into the canonical space $D([0,\infty), \mathbb{R}^d)$ which can be expressed as $X = X(0) + A^+ - A^- + M$, where $X(0)$ is a $\mathbb{R}^d$-valued $\mathcal{F}(0)$-measurable random variable, A^+, A^-, are adapted monotone increasing locally integrable processes and M is a local martingale, satisfying $A^+(0) = A^-(0) = M(0) = 0$. Thus, $A := A^+ - A^-$ is a process with locally integrable bounded variation paths. □

Based on the uniqueness of Doob-Meyer decomposition, a local martingale null at time zero with locally bounded variation is identically zero if it is predictable (in particular if it is continuous or deterministic). Since there are non-constant martingales with locally bounded variation paths (e.g., purely discontinuous local martingales), the decomposition in the definition of semi-martingale is not necessarily unique. Usually, the above definition of semi-martingale is known as *special semi-martingale*, but this is sufficiently general for our study. These (special) semi-martingales include a natural condition of local integrability (local first moment) on the bounded variation part (the adapted process A). The equivalent of this local integrability property, applied to the martingale part (the process M), is actually a necessary condition for martingale. Unless explicitly mentioned, we drop the adjective *special* in using of the name semi-martingale but we may call *general* or *non-special* semi-martingale when the process A in the above definition may not be locally integrable. Note that the only reason why the process A may not be integrable is because of the *large jumps*. It is clear then that a (special) semi-martingale is the difference of two local sub-martingales. Moreover, a local sub-martingale zero at the origin can be written in a unique manner as the sum of a local martingale and an increasing predictable process, both zero at the origin. Thus, the concept of special semi-martingales is equivalent to that of quasi-martingales, e.g. see Kallenberg [121], Protter [206].

Theorem 3.14. *Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a filtered space. Then every semi-martingale $X = (X(t) : t \geq 0)$ admits the unique canonical decomposition $X = X(0) + A + M$, where A is a predictable process with locally integrable variation and M is a local martingale, both satisfying $A(0) = M(0) = 0$. Moreover,*

the quadratic variation $[M]$ *defined by* (3.28) *is the unique optional monotone increasing process such that* $M^2-[M]$ *is a local martingale and the jumps* $\delta[M] = \delta M\,\delta M$, *where* $\delta M(t) := M(t) - M(t-)$. *Furthermore, the processes* $\sqrt{[M]}$ *(by coordinates) and* $\sup\{|X(s) - X(0)| : 0 \le s \le t\}$ *are locally integrable. If the semi-martingale* X *is quasi-left continuous (i.e., either* $P\{X(\tau-) = X(\tau)\} = 1$ *or* $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ *for every predictable stopping time* τ*), then the process* A *in the semi-martingale decomposition is continuous.* □

Note that the local martingale appearing in the above expression has a unique representation $M = M_c + M_d$, where M_c (respectively M_d) is the continuous (purely discontinuous) part. Also, if M is a local martingale with $M(0) = 0$ and $[M]$ denotes its (optional) quadratic variation (or characteristic) then for any $t > 0$ and any sequence of partitions $(\pi_k : k = 1, 2, \ldots)$, with π_k of the form $(0 = t_0 < t_1 < \cdots < t_n = t)$ and the mesh (or norm) of π_k going to zero we have $\mathrm{var}_2(M, \pi_k) \to [M](t)$ in probability as $k \to 0$, see Liptser and Shiryayev [158, Theorem 1.4, pp. 55–59].

Semi-martingales are stable under several operations, for instance under stopping times operations and localization, see Jacod and Shiryaev [117, Theorem I.4.24, pp. 44-45].

Observe that a process X with independent increments (i.e., which satisfies for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the random variables $\{X(t_0),\, X(t_1) - X(t_0), \ldots,\, X(t_n) - X(t_{n-1})\}$ are independent) is not necessarily a semi-martingale, e.g., deterministic cad-lag process null at time zero is a process with independent increments, but it is not a general semi-martingale (not necessarily special!) unless it has finite variation. It is clear that $\mathbb{R}^d$-valued processes with independent increments are completely described by their characteristic functions, namely

$$\hat{X}(\lambda, t) := \mathbb{E}\{\exp(\mathtt{i}\,\lambda \cdot X(t))\}, \quad \forall t \ge 0,\ \lambda \in \mathbb{R}^d,$$

which is a complex-valued cad-lag function. It can be proved that a process X with independent increments is a general semi-martingale if and only if the (deterministic) function $t \mapsto \hat{X}(\lambda, t)$ has locally bounded variation for any λ in $\mathbb{R}^d$. Moreover, a process X with independent increments has the form $X = Y + A$, where Y is a general semi-martingale with independent increments and A is a deterministic cad-lag process (or function) from $[0, \infty)$ into R^d_+ with $A(0) = 0$. On the other hand, if X is an integrable (cad-lag!) process with independent increments, i.e., $\mathbb{E}\{|X(t)|\} < \infty$ for every $t \ge 0$, and $(\mathcal{F}(t) : t \ge 0)$ is its natural filtration then

$$\mathbb{E}\{X(t) \,|\, \mathcal{F}(s)\} = \mathbb{E}\{X(t) - X(s)\} + X(s), \quad \text{a.s.},$$

for any $t \ge s \ge 0$. Hence, X is a (super-/sub-) martingale if and only if $\mathbb{E}\{X(t) - X(s)\} = 0$ $(\le 0 / \ge 0)$ for any $t \ge s \ge 0$.

The only reason that a semi-martingale may not be special is essentially the non-integrability of large jumps. If X is a semi-martingale satisfying $|X(t) - X(t-)| \le c$ for any $t > 0$ and for some positive (deterministic) constant $c > 0$, then X is special. Indeed, if we define $\tau_n := \inf\{t \ge 0 : |X(t) - X(0)| > n\}$ then

$\tau_n \to \infty$ as $n \to \infty$ and $\sup_{0\le s\le\tau_n} |X(s) - X(0)| \le n + c$. Thus X is a special semi-martingale and its canonical decomposition $X = X(0) + A + M$ satisfies $|A(t) - A(t-)| \le c$ and $|M(t) - M(t-)| \le 2c$, for any $t > 0$.

Similar to (3.31), another very useful estimate is the Lenglart's inequality: If X and A are two cad-lag adapted processes such that A is monotone increasing and $\mathbb{E}\{|X_\tau|\} \le \mathbb{E}\{A_\tau\}$, for every bounded stopping time τ, then for every stopping time τ and constants $\varepsilon, \eta > 0$ we have

$$\left\{ \begin{array}{l} P\big\{ \sup\limits_{t\le\tau} |X_t| \ge \varepsilon \big\} \le \dfrac{1}{\varepsilon}\Big[\eta + \mathbb{E}\big\{ \sup\limits_{t\le\tau}(A_t - A_{t-})\big\}\Big] + \\ \qquad\qquad\qquad\qquad\qquad\qquad + P\big\{A_\tau \ge \eta\big\}, \end{array} \right. \tag{3.32}$$

and if A is also predictable then the term with the jump $(A_t - A_{t-})$ is removed from the above estimate. A simple way to prove this inequality is first to reduce to the case where the stopping time τ is bounded. Then, defining $\theta := \inf\{s \ge 0 : |X_s| > \varepsilon\}$ and $\varrho := \inf\{s \ge 0 : A_s > \eta\}$, since A is not necessarily continuous, we have $A_{\varrho-} \le \eta$ and

$$\begin{aligned} &A_{\theta\wedge\tau\wedge\varrho} \le \eta + \sup_{t\le\tau}(A_t - A_{t-}), \\ &\big\{ \sup_{t\le\tau} |X_t| > \varepsilon \big\} \subset \big\{\theta \le \tau < \varrho\big\} \cup \big\{A_\tau \ge \eta\big\}. \end{aligned}$$

Hence, by means of the inequality

$$P\big\{\theta \le \tau < \varrho\big\} \le P\big\{|X_{\theta\wedge\tau\wedge\varrho}| \ge \varepsilon\big\} \le \frac{1}{\varepsilon}\,\mathbb{E}\big\{A_{\theta\wedge\tau\wedge\varrho}\big\},$$

we obtain (3.32). However, if A is predictable then ϱ is a predictable time, and there is a sequence of stopping times $(\varrho_k : k \ge 1)$ converging to ϱ such that $\varrho_k < \varrho$ if $\varrho > 0$. Thus $A_{\theta\wedge\tau\wedge\varrho} \le A_{\varrho-}$ almost surely, which completes the argument. Given a local martingale M, a good use of (3.32) is when the predictable compensator process $\langle M\rangle$ is continuous, and therefore $[M] = \langle M\rangle$, so that

$$P\big\{ \sup_{t\le\tau} |M_t| \ge \varepsilon \big\} \le \frac{\eta}{\varepsilon^2} + P\big\{\langle M\rangle_\tau \ge \eta\big\}, \quad \forall \varepsilon, \eta > 0, \tag{3.33}$$

for any stopping time τ. This is the case of a quasi-left continuous local martingale M.

For a comprehensive treatment with proofs and comments, the reader is referred to the books by Dellacherie and Meyer [58, Chapters V–VIII], Liptser and Shiryayev [158, Chapters 2–4, pp. 85–360]. Rogers and Williams [214, Section II.5, pp. 163–200], among others. A treatment of semi-martingale directly related with stochastic integral can be found in Protter [206].

Let us insist on the following concept, which characterize a large class of Markov processes suitable for modelling.

Definition 3.15 (quasi-left continuous)**.** As mentioned previously, a filtration $(\mathcal{F}(t) :\ge 0)$ of a probability space $(\Omega, \mathcal{F}, P)$ is called *quasi-left continuous* or

regular if for any increasing sequence of stopping time $\{\tau_n : n = 1, 2, \dots\}$ almost surely strictly convergent to τ, $P(\tau_n \le \tau_{n+1} < \tau < \infty, \tau > 0) = 1$, the σ-algebra $\mathcal{F}(\tau)$ is the minimal σ-algebra containing the sequence of σ-algebra $\{\mathcal{F}(\tau_n) : n = 1, 2, \dots\}$. This is equivalent to the condition $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ for any predictable stopping time, recall that a stopping time τ is predictable if there exists an announcing sequence of stopping times $\{\tau_n : n = 1, 2, \dots\}$, i.e., τ_n increases to τ and $P(\tau_n < \tau, \tau > 0) = 1$. A (cad-lag) integrable stochastic process X adapted to a filtration $\{\mathcal{F}(t) : t \ge 0\}$ is called *quasi-left continuous* or *regular* if $X(\tau_n)$ converges to $X(\tau)$ almost surely for any announcing sequence of stopping times τ_n convergent to τ. □

It can be proved (e.g., Rogers and Williams [214, Chapter VI, Theorems 18.1-2, pp. 346-347]) that a filtration $\{\mathcal{F}(t) : t \ge 0\}$ is quasi-left continuous if and only if every uniformly integrable martingale M relative to $\{\mathcal{F}(t) : t \ge 0\}$ satisfies $M(\tau) = M(\tau-)$ for any predictable stopping time and that any *Markov-Feller process* (also called Feller-Dynkin process) is regular with respect to its natural filtration, the discussion goes as follows.

Let E be a locally compact Polish (i.e., complete separable metric) space (usually, E is an open or closed subset of $\mathbb{R}^d$). A *Markov-Feller process* with states in E possesses a Feller semigroup $\{P(t) : t \ge 0\}$ in $C_0(E)$, with infinitesimal generator A with domain $D(A) \subset C_0(E)$, see Definition 2.40 in Chapter 2. Its transition function $P(t, x, \mathrm{d}y)$ can be defined on a compact base space $\dot{E}$, the one-point compactification of E, by $P(t, x, \{\infty\}) := 1 - P(t, x, E)$, so that $P(t, x, \dot{E}) = 1$. For any initial distribution on $\dot{E}$, we denote by P the (complete) probability measure induced by the transition function $P(t, x, \mathrm{d}y)$ on the canonical space $D([0, \infty), \dot{E})$ with its Borel σ-algebra $\mathcal{B}$, its canonical process $X(t) := \omega(t)$ and its filtration $\{\mathcal{F}(t) : t \ge 0\}$, see Definition 1.28 and Theorem 1.32 in Chapter 1. Note that the probability measure P and the completion necessary to generate the filtration $\{\mathcal{F}(t) : t \ge 0\}$ depend on the initial distribution $r \to P\{X(0) \le r\}$. All these elements constitute a standard *realization* of a Markov-Feller or Feller-Dynkin process with state in E (strictly speaking in $\dot{E}$).

Now, for a function f in $C_0(\dot{E})$ and $\lambda > 0$ the resolvent operator is given by

$$R_\lambda f(x) := \int_0^\infty \mathrm{e}^{-\lambda t} P(t) f(x) \, \mathrm{d}t = \int_0^\infty \mathrm{d}t \int_E \mathrm{e}^{-\lambda t} f(y) \, P(t, x, \mathrm{d}y),$$

and satisfies

$$R_\lambda f(x) = \mathbb{E}\{\xi \,|\, X(0) = x\}, \qquad \text{with } \xi := \int_0^\infty \mathrm{e}^{-\lambda t} f(X(t)) \mathrm{d}t.$$

Denoting by $\mathbb{E}^x\{\cdot\}$ the (conditional) expectation with respect to the probability measure P with the Dirac measure at x as the initial distribution and applying Markov property, we find that

$$\mathbb{E}^x\{\xi \,|\, \mathcal{F}(t)\} = \int_0^t \mathrm{e}^{-\lambda s} f(X(s)) \mathrm{d}s + \mathrm{e}^{-\lambda t} R_\lambda f(X(t)), \quad \text{a.s.} \quad \forall t > 0,$$

which proves that the right-hand side is a uniformly integrable martingale. Thus the Optional-Stopping Theorem 3.9, part (b), yields

$$\mathbb{E}^x\Big\{\int_0^\tau \mathrm{e}^{-\lambda t}\, f(X(t))\mathrm{d}t\Big\} + \mathbb{E}^x\{\mathrm{e}^{-\lambda\tau} R_\lambda f(X(\tau))\} = R_\lambda f(x).$$

Hence, if $f = (\lambda - A)\varphi$ for some φ in the domain $D(A)$ of the infinitesimal generator A we deduce that the process $Y(t) = Y(t, \varphi, \lambda)$ given by

$$Y(t) := \mathrm{e}^{-\lambda t}\,\varphi(X(t)) + \varphi(x) + \int_0^t \mathrm{e}^{-\lambda s}\,(\lambda - A)\varphi(X(s))\mathrm{d}s,$$

is a uniformly integrable martingale relative to $(P^x, \mathcal{F}(t) : t \geq 0)$. Therefore, the following identity, so-called *Dynkin's formula*

$$\mathbb{E}\{\varphi(X_\theta)\} = \mathbb{E}\{\varphi(X_\tau)\} + \mathbb{E}\{\int_\tau^\theta A\varphi(X(t))\,\mathrm{d}t\} \tag{3.34}$$

holds for any function φ in $D(A)$ and any stopping time satisfying $P\{\tau \leq \theta < \infty\} = 1$. Moreover, the filtration $\{\mathcal{F}(t) : t \geq 0\}$ is quasi-left continuous, i.e. $\mathcal{F}(\tau) = \mathcal{F}(\tau-)$ and $X(\tau) = X(\tau-)$ (almost surely) for any predictable stopping time τ.

For instance, we refer to Rogers and Williams [214, Chapter III, pp. 227–349], Dellacherie and Meyer [58, Chapters XI–XVI], Dynkin [70], among others. Also, the reader interested in a comprehensive study on the theory of martingales may consult the books He et al. [105] or Liptser and Shiryayev [158].

3.6 Strong Markov Processes

Starting from a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$, we may consider stochastic processes X with values in some Polish space E (complete separable metric space, usually locally compact) as (1) a family of E-valued random variables $\{X(t) : t \geq 0\}$, (2) a function on a product space $X : [0, \infty) \times \Omega \to E$, (3) a function space valued random variable, i.e., either a random variable with values in some sub-space of $E^{[0,\infty)}$ or a mapping from $[0, \infty)$ into the space of E-valued random variables. Except when explicitly mentioned, we are looking at a stochastic process as a random variable with values in some function space, a Polish space non-locally compact which most of the cases is either $D([0, \infty), E)$ or $C([0, \infty), E)$, with E being an Borel (usually open or closed) subset of $\mathbb{R}^d$.

A stochastic process X with values in a Polish space E (even more general, E could be a Lusin space, i.e., a topological space homeomorphic to a Borel subset of a complete separable metric space) is called a *Markov process* in the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ if the Markov property is satisfied, i.e.,

$$\mathbb{E}\{f(X(t) \,|\, \mathcal{F}(s)\} = \mathbb{E}\{f(X(t) \,|\, X(s)\}, \tag{3.35}$$

for every $t \geq s$ and any bounded Borel real function f on E. This is an almost surely equality due to the use of conditional probability. It means that the only

information relevant for evaluating the behavior of the process beyond time s is the value of the current state $X(s)$. This implies in particular that X is adapted. Points x in E are called *states* and E is the *state space* of X.

A Markov process can be identified by its *transition function*, which is defined by taking a particular class of function f in (3.35), namely characteristic or indicator functions $f = \mathbb{1}_B$ of Borel subsets B of E, i.e., with B in $\mathcal{B}(E)$. The transition function $p(s, x, t, B)$ is defined with following (minimal) regularity conditions:

(1) for fixed $0 \le s \le t$, x in E, the function $B \mapsto p(s, x, t, B)$ is a probability measure on $\big(E, \mathcal{B}(E)\big)$,

(2) for fixed $0 \le s \le t$, B in $\mathcal{B}(E)$ the function $x \mapsto p(s, x, t, B)$ is Borel measurable,

(3) for every $0 \le s \le t$, x in E, B in $\mathcal{B}(E)$ we have the identity $p(s, X(s), t, B) := \mathbb{E}\{\mathbb{1}_B(X(t)) \,|\, \mathcal{F}(s)\}$, almost surely.

Really, (1) and (2) are necessary conditions to make sense to the key condition (3). However, the Markov property alone is not sufficient to define the transition function. Condition (3) implies that for every $s \ge 0$, x in E, B in $\mathcal{B}(E)$ we have $p(s, x, s, B) = \mathbb{1}_B(x)$ and standard properties of the conditional probability yield the Chapman-Kolmogorov identity

$$p(s, x, t, B) = \int_E p(r, y, t, B) p(s, x, r, \mathrm{d}y), \tag{3.36}$$

valid for any $0 \le s < r < t$, x in E and B in $\mathcal{B}(E)$.

Markov processes are mathematical model for phenomena which evolve in time, in a random way and following some dynamic or evolution law. Most often, statistical experiments or physical considerations give only information about the so-called finite-dimensional distributions of a process. This means that for a given initial probability measure μ on $\big(E, \mathcal{B}(E)\big)$ and times $0 \le t_0 < t_1 < \cdots < t_n$ the probabilities $P_{t_0, t_1, \ldots, t_n}$ on E^{n+1} defined by

$$\begin{cases} P_{t_0, t_1, \ldots, t_n}(B_0 \times B_1 \times \cdots \times B_n) = \\ \quad = \displaystyle\int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(t_0, x_0, t_1, \mathrm{d}x_1) \int_{B_2} p(t_1, x_1, t_2, \mathrm{d}x_2) \cdots \\ \quad \cdots \displaystyle\int_{B_{n-1}} p(t_{n-2}, x_{n-2}, t_{n-1}, \mathrm{d}x_{n-1})\, p(t_{n-1}, x_{n-1}, t_n, B_n) \end{cases} \tag{3.37}$$

are the finite-dimensional distributions. Thus, starting from a function p satisfying the properties (1) and (2) of a transition function, and if the function p satisfies the Chapman-Kolmogorov identity (3.36), then the above relation (3.37) defines a consistent family of finite-dimensional distributions on the canonical product space $E^{[0,\infty)}$. Note that the Dirac measure $\delta(x_0)$, i.e., the unit mass concentrated at x_0, is the typical initial distribution at time t_0. For simplicity, let us discuss *homogeneous* Markov process, i.e., the case where the transition function is time invariant, i.e., $p(s, x, t, B) = p(0, x, t - s, B)$ for every $t \ge s$, x in E and B in $\mathcal{B}(E)$. Hence, the transition function can be taken as $p(x, t, B)$,

with $t \geq 0$. Remark that by adding an extra variable (the time), we can always reduce to homogeneous case. Thus, Kolmogorov's existence theorem can be used to construct a Markov process with the given transition function p, for each initial probability measure μ at time $t_0 = 0$, and then we have a family of Markov processes. Therefore, by a *realization* of Markov process with transition probability function p we mean a collection $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P_x, t \geq 0, x \in E)$ where P_x is the probability measure constructed as above with initial probability $\mu = \delta_x$, the Dirac measure at x. In Kolmogorov's construction, the process $X_t(\omega) = X(t, \omega) := \omega(t)$ is the coordinate (or identity) mapping and $\mathcal{F}(t)$ is the natural filtration associated with the process $X(t)$, which is not always right-continuous. Some difficulties appear since $\mathcal{F}(t)$ should be completed with respect to the probability measure P_x, given a completed filtration $\mathcal{F}_x(t)$ or $\mathcal{F}_\mu(t)$, which depend on initial parameter x or μ.

By means of the transition probability function, we may re-write the Markov property (3.35) as

$$P\{X(s+t) \in B \mid \mathcal{F}(s)\} = p(X(s), t, B), \tag{3.38}$$

for every $t \geq s \geq 0$ and any Borel subset B of E. Now, a *strong Markov process* is one for which the Markov property holds at stopping times of the filtration $\{\mathcal{F}(t) : t \geq 0\}$, i.e.,

$$P\{X(T+t) \in B \mid \mathcal{F}(T)\}\mathbb{1}_{T<\infty} = p(X(T), t, B)\mathbb{1}_{T<\infty}, \tag{3.39}$$

for every $t \geq 0$, any stopping time T and any Borel subset B of E. This says that the probabilistic evolution of the process after the stopping time T is just that of another process *restarted* at T, i.e., *the process restarts at stopping time.* The reader is referred to Doob [60, Theorems 8 and 9, pp. 556-560], see Theorem 1.29 in Chapter 1, for conditions ensuring the right-continuity of the filtration and the strong Markov property. In the statement (3.39), we remark the interest in using a filtration satisfying the usual condition, in particular the need of having a *completed* σ-algebra $\mathcal{F}(0)$. A useful definition in this context is the so-called *universally completed* filtration, which is constructed as follows. First, let $\{\mathcal{F}(t) : t \geq 0\}$ be the filtration (history) generated by the canonical process $X(t, \omega) = \omega(t)$, not necessarily satisfying the usual conditions. Denote by $\{\mathcal{F}^\mu(t) : t \geq 0\}$ the filtration which is obtained by completing $\mathcal{F}(0)$ with respect to the probability measure P_μ. Now the universally completed filtration is $\{\mathcal{F}^0(t) : t \geq 0\}$, where $\mathcal{F}^0(t) := \cap_\mu \mathcal{F}^\mu(t)$, for every $t \geq 0$. Note that the filtration $\{\mathcal{F}^0(t) : t \geq 0\}$, does not necessarily satisfies the usual conditions, but it is right-continuous if the initial filtration $\{\mathcal{F}(t) : t \geq 0\}$ is so.

As discussed earlier, the product space $E^{[0,\infty)}$ does not provide a suitable mathematical setting, we need to use the Polish sample space $D([0, \infty), E)$ or $C([0, \infty), E)$. This imposes more conditions on the transition function p, and eventually we are lead to the study of Markov-Feller processes and semigroups.

The reader may consult the classic references Blumental and Getoor [28], Dynkin [70] or more recent books, e.g., Davis [56], Rogers and Williams [214].

One of the most simple Markov processes in continuous time is the Poisson process. If $\{\tau_n : n, n = 1, 2, \ldots\}$ is a sequence of independent exponentially

distributed (with parameter λ) random variables, then the random variable $\theta_n := \tau_1 + \cdots + \tau_n$ has a Γ-distribution with parameters λ and $n-1$, for $n = 1, 2, \ldots$, i.e.,

$$P\{\theta_n \le t\} = \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1} \mathrm{e}^{-\lambda x} \mathrm{d}s, \quad \forall\, t \ge 0,$$

and the *counting process* defined by

$$p(t, \omega) := \sum_{n=1}^{\infty} \mathbb{1}_{\theta_n(\omega) \le t}, \quad \forall\, t \ge 0 \tag{3.40}$$

is a Poisson process, i.e., $p(0) = 0$, $p(t) - p(s)$ is a Poisson variable with mean $\lambda(t-s)$, namely

$$P\{p(t) - p(s) = n\} = \big[\lambda(t-s)\big]^n \exp\big[-\lambda(t-s)\big],$$

for every $n = 0, 1, \ldots$, and for any $0 \le t_0 < t_1 < \cdots < t_n$ the family $\{p(t_0), p(t_k) - p(t_{k-1}) : k = 1, 2, \ldots, n\}$ is a set of independent random variables. The parameter λ is usually called *jump rate*.

In a *compound* Poisson process the construction (3.40) is modified as follows

$$p_c(t, \omega) := \sum_{k=1}^{\infty} \eta_n(\omega) \mathbb{1}_{\theta_n(\omega) \le t}, \quad \forall\, t \ge 0, \tag{3.41}$$

where $\{\eta_n : n = 1, 2, \ldots\}$ is a sequence of independent identically distributed (with distribution law ν and independent of the $\{\tau_n\}$) $\mathbb{R}^d$-valued random variables. A integer-valued measure process can be associated, namely

$$\rho_c(t, B, \omega) := \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \le t} \mathbb{1}_{\eta_k(\omega) \in B}, \quad \forall\, t \ge 0,\ B \in \mathcal{B}(\mathbb{R}^d), \tag{3.42}$$

which captures all features of the compound process and extends to the so-called *Poisson measures*. Note that $\mathbb{E}\{\rho_c(t, B, \omega)\} = t\, \lambda\, \nu(B)$. The parameters λ and ν yield the integral operator

$$Ih(x) := \lambda \int_{\mathbb{R}} \big[h(x+y) - h(x)\big]\, \nu(\mathrm{d}y), \quad \forall\, x \in \mathbb{R}^d, \tag{3.43}$$

which is a characteristic element of the compound Poisson process. This integral operator is the infinitesimal generator of the Markov process, which in turn is determined by its kernel, the Lévy measure $M(\mathrm{d}y) := \lambda\, \nu(\mathrm{d}y)$. Note that to make the expression (3.42) interesting, we assume $\nu(\{0\}) = 0$ and then the mass of the origin $M(\{0\})$ does not play any role in the definition of I, thus the Lévy measure is on $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$.

All these examples are time and spatially homogeneous Markov processes. To relax the homogeneity, we must allow the Lévy measure to depend on t and x. For instance, we take $M(x, \mathrm{d}y)$ in the expression (3.43) of the integral

operator. The dependency on x of the kernel could be very general and in some cases hard to track. A typical assumption is the representation

$$M(x,B) = \lambda(x)\,\ell(\{\zeta \in [0,1] : x + j(x,\zeta) \in B\}), \tag{3.44}$$

for every x in $\mathbb{R}^d$ and B in $\mathcal{B}(\mathbb{R}^d_*)$, where $([0,1],\mathcal{L},\ell)$ is the canonical Lebesgue probability measure space, $\lambda : \mathbb{R}^d \to [0,\infty)$ and $j : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d_*$ are measurable functions, on which some regularity (such as continuity) in x may be required.

If $\{Z_n, U_n : n = 1,2,\dots\}$ are double sequence of independent uniformly distributed random variables in $([0,1],\mathcal{L},\ell)$, then the transformation

$$\Theta(x,u) := \inf\big\{t \geq 0 : \exp[-t\,\lambda(x)] \leq u\big\}, \tag{3.45}$$

with $\Theta(x,0) := +\infty$, yields the construction of the following Markov jump process by induction. Given θ_{k-1} and x_{k-1} we define

$$\theta_k := \theta_{k-1} + \Theta(x_{k-1}, U_k),$$
$$x_k := x_{k-1} + j(x_{k-1}, Z_k)$$

and for any t in the stochastic interval $[\![\theta_{k-1}, \theta_k[\![$ set $x(t) := x_k$. Naturally, we can start from any initial time θ_0 and state x_0, but we use $\theta_0 = 0$ and any given x_0. Assuming that $\theta_n \to \infty$ (e.g., this hold if $\lambda(\cdot)$ is bounded) the process $x(t)$ is defined for every time $t \geq 0$. Its associated integer-valued measure process is given by

$$\rho(t,B,\omega) := \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega)\leq t}\,\mathbb{1}_{x_k(\omega)\in B}, \quad \forall\, t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d). \tag{3.46}$$

The integral operator becomes

$$Ih(x) := \lambda(x) \int_{[0,1]} \big[h(x + j(x,\zeta)) - h(x)\big]\,\ell(\mathrm{d}\zeta), \quad \forall\, x \in \mathbb{R}^d, \tag{3.47}$$

which make sense for any bounded Borel measurable function h. The process $\{x(t) : t \geq 0\}$ a cad-lag realization (and piecewise constant) of a strong Markov process. Several other variations are possible.

3.7 Extended Generators

Let E be a Borel subset of Polish space, let $B(E)$ be the Banach space of bounded Borel measurable functions f from E into $\mathbb{R}$ with sup-norm $\|\cdot\|$, and let $(\Omega,\mathcal{F},\mathcal{F}_t,X_t,P_x,t \geq 0, x \in E)$ be a (strong) Markov process. For $t \geq 0$, define an operator $P(t) : B(E) \to B(E)$ by $P(t)f(x) := \mathbb{E}_x\{f(X(t)\}$, where $\mathbb{E}_x\{\cdot\}$ denotes the mathematical expectation relative to P_x. It is clear that $P(t)$ is a contraction, i.e. $\|P(t)f\| \leq \|f\|$, for every $t \geq 0$, and that the Chapman-Kolmogorov identity (3.36) are equivalent to the *semigroup property* $P(t)P(s) = P(s+t)$, for every $t,s \geq 0$.

Denote by B_0 the subset of $B(E)$ consisting of those functions f for which $\|P(t)f - f\|$ vanishes as t goes to zero. The contraction property shows that B_0 is a closed subspace of $B(E)$ and $\{P(t) : t \geq 0\}$ is called *strongly continuous* on B_0. Moreover, (1) B_0 is invariant under $P(t)$, for every $t \geq 0$, and (2) for every f in B_0 (which is itself a Banach space), the function $t \mapsto P(t)f$ is continuous from $[0, \infty)$ into B_0.

Now, let A be the *strong infinitesimal generator* of $\{P(t) : t \geq 0\}$ with domain $\mathcal{D}(A) \subset B_0 \subset B(E)$, i.e, f belong to $\mathcal{D}(A)$ and $Af = g$ if and only if

$$\lim_{t \to 0} \left\| \frac{P(t)f - f}{t} - g \right\| = 0.$$

Note that the domain $\mathcal{D}(A)$ is as important as the expression of A, there are examples of two different Markov process with the same expression for the infinitesimal generator A but with disjoint domains $\mathcal{D}(A)$, see Davis [56, Chapter 2].

Based on properties of derivatives and Riemann integrals of continuous functions with values in a Banach space, we can establish:

(1) if $f \in B_0$ and $t \geq 0$ then

$$\int_0^t P(s)f \mathrm{d}s \in \mathcal{D}(A) \quad \text{and} \quad A \int_0^t P(s)f \mathrm{d}s = P(t)f - f,$$

(2) if $f \in \mathcal{D}(A)$ and $t \geq 0$ then $P(t) \in \mathcal{D}(A)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t)f = AP(t)f = P(t)Af,$$
$$P(t)f - f = \int_0^t AP(s)f \mathrm{d}s = \int_0^t P(s)Af \mathrm{d}s.$$

In probabilistic terms, if $u(t) = u(x,t) := \mathbb{E}_x\{f(X(t))\} = P(t)f(x)$ with f in $\mathcal{D}(A)$ then u satisfies

$$\partial_t u(t) = Au(t), \quad u(0) = f, \tag{3.48}$$

which is an abstract version of the so-called *Kolmogorov backward* equation. The semigroup is determined by (3.48) and this determines the transition (probability) functions $p(x,t,B)$, which determines the finite-distributions and hence the probability measure P_x, i.e., the Markov process itself. Certainly, some technical conditions are required to turn this calculation into reality. For practical calculations it is more important the second expression in (2) which can be written as

$$\mathbb{E}_x\{f(X(t)\} = f(x) + \mathbb{E}_x\Big\{ \int_0^t Af(X(s) \mathrm{d}s \Big\}, \tag{3.49}$$

for every f in $\mathcal{D}(A)$, which is known as *Dynkin formula*.

Let f be in $\mathcal{D}(A)$ and define the real-valued process $\{M_f(t) : t \geq 0\}$ by

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t Af(X(s)) \mathrm{d}s. \tag{3.50}$$

By using the semigroup property and conditional expectation arguments, we can show that for every x in E the process $\{M_f(t) : t \geq 0\}$ is a martingale in $(\Omega, \mathcal{F}, P_x, \mathcal{F}(t), t \geq 0)$, i.e.,

$$\mathbb{E}_x\{M_f(t) \mid \mathcal{F}(s)\} = M_f(s), \quad \forall t \geq s \geq 0.$$

A natural extension of the domain $\mathcal{D}(A)$ of the (strong) infinitesimal generator is as follows.

Definition 3.16 (extended generator)**.** Let $B^*(E)$ be the space of all Borel measurable functions, not necessarily bounded, from E into $\mathbb{R}$. We say that a function f belongs to the domain of the *extended (infinitesimal) generator* if there exists another function g in $B^*(E)$ such that $t \mapsto g(X(t))$ is locally integrable P_x-almost surely and the process $\{M_f(t) : t \geq 0\}$ defined by

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t g(X(s))\mathrm{d}s$$

is a local martingale, i.e., there exists an increasing sequence of stopping times $\{\tau_n : n = 1, 2, \ldots\}$, with $\tau_n \to +\infty$ such that the stopped process $M_f^n(t) := M_f(t \wedge \tau_n)$ is a uniformly integrable martingale for each n. We use the notation $\mathcal{D}(\bar{A})$ for the extended domain and $\bar{A}f := g$ for the extended generator. □

Note that $\mathcal{D}(A) \subset \mathcal{D}(\bar{A})$ and that $\bar{A}f$ is uniquely defined (module subset of potential zero). Indeed, if $f = 0$ then the process $\{M_f(t) : t \geq 0\}$ is a continuous martingale with locally bounded variation, therefore $M_f(t) = M_f(0)$ is the constant process zero. Hence, $Af = 0$ except possibly on some measurable set B of E such that

$$\int_0^\infty \mathbb{1}_B(X(t))\mathrm{d}t = 0, \quad P_x - \text{a.s.},$$

for every x in E. Such a set B is said to have *potential zero*. The process $\{X(t) : t \geq 0\}$ *spend no time* in B, regardless of the starting point, so the process $\{M_f(t) : t \geq 0\}$ does not depend on the values of Af for x in B, and Af is unique up to sets of zero potential.

When $\{M_f(t) : t \geq 0\}$ is a martingale, Dynkin formula (3.49) holds. Usually, it is quite difficult to characterize $\mathcal{D}(\bar{A})$ but in most of the cases, there are easily checked sufficient conditions for membership in the extended domain $\mathcal{D}(\bar{A})$. For instance, the reader is refereed to the books by Davis [56, Chapter 1], Ethier and Kurtz [76, Chapter 4] for more details.

Let us go back to the examples in the previous section. For the particular case of the Poisson process $(p(t) : t \geq 0)$ given by (3.40), the extended infinitesimal generator is

$$Af(x) = \lambda[f(x+1) - f(x)], \quad \forall\, x \in \mathbb{R}$$

while for the compound Poisson process $(p_c(t) : t \geq 0)$ (3.40), A is the integral operator I given by (3.43). What is perhaps more relevant is the extended

domain $\mathcal{D}(\bar{A})$, which have not restriction at all (i.e., all real-valued function defined on $\mathbb{R}$) for the Poisson process, while a condition on local integrability, i.e., a measurable function $f : \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{D}(\bar{A})$, for the compound Poisson process with parameters λ and ν, if and only if

$$\mathbb{E}\{\sum_{i=1}^{\infty} |f(x+\eta_i) - f(x)|\, \mathbb{1}_{\theta_i<\sigma_n}\} < \infty, \quad \forall\, x, n,$$

where σ_n is a sequence of stopping times with $\sigma_n \to \infty$ almost surely.

For the class of Markov jump process constructed by induction, see (3.45) and (3.46), the full description of the extended domain $\mathcal{D}(\bar{A})$, with $A = I$ as in (3.47), is as follow. First, we say that a process $\{h(x,t,\omega) : t \geq 0,\, x \in \mathbb{R}^d\}$ belongs to $L^1(\rho)$, (where ρ is the integer-valued measure process) if

$$\mathbb{E}\Big\{\sum_{i=1}^{\infty} h(x_k, \theta_k, \omega)\Big\} < \infty.$$

Similarly, h belongs to $L^1_{\rm loc}(\rho)$, if there exists a sequence $\{\sigma_k : k \geq 0\}$ of stopping times with $\sigma_n \to \infty$ almost surely such that

$$\mathbb{E}\Big\{\sum_{i=1}^{\infty} h(x_k, \theta_k \wedge \sigma_n, \omega)\Big\} < \infty, \quad \forall\, n.$$

Now, a measurable function f belongs to $\mathcal{D}(\bar{A})$ if the process $h(x,t,\omega) := f(x) - f(x(t-,\omega))$ belongs to $L^1_{\rm loc}(\rho)$. This is particular case of Davis [56, Theorem 26.14, pp. 69–74].

3.8 Poisson Processes and Queues

In a practical way, a stochastic process is a mathematical model of a probabilistic experiment that generates a sequence of numerical values as it evolves in time. Each numerical value in the sequence is modelled by a random variable, so a stochastic process is simply a (finite or infinite) sequence of random variables. However, the properties of the evolution in time become essential when the focus is on the dependencies in the sequence of values generated by the process. Typically, arrival-type or outcome-type processes occur very frequently ("arrival" of such as message receptions at a receiver, job completions in a manufacturing cell, customer purchases at a store, trials of coin tosses, etc), where the focus is on modeling the "inter-arrival" (times between successive arrivals) are independent variables. These processes become Markov processes as the dimension is increased. In Markov processes, the experiments that evolve in time exhibit a very special type of dependence: the next value depends on the past values only through the current value, the present.

Clearly, the way how time is measured is critical. Essentially, there are only two ways, "discrete time" (where a unit time is identified and used, i.e., integer numbers are the model) and "continuous time" (where the time goes

continuously, i.e., real numbers are the model). For instance, if the arrivals occur in discrete time and the inter-arrival times are geometrically distributed, this is the *Bernoulli process* described as a sequence $\{x_1, x_2, \dots\}$ of independent random variables x_i with $P\{x_i = 1\} = p$ (arrival occurs in the i trial with probability $0 < p < 1$) and $P\{x_i = 0\} = p$ (arrival does not occur in the i trial). Here, arrival also means success in the outcome under consideration. Standard calculations show that if $S_n := x_1 + x_2 + \cdots + x_n$ denotes the number of arrivals in n independent trials then

$$P\{S_n = k\} = \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, \dots, n,$$

with a mean $\mathbb{E}\{S_n\} = np$ and a variance $\mathbb{E}\{S_n - np\}^2 = np(1-p)$, i.e., a binomial distribution with parameters p and n. Similarly, if $T := \inf\{i \geq 0 : x_i = 1\}$ denotes the number of trials up to (and including) the first arrival then

$$P\{T = k\} = (1-p)^{k-1} p, \quad k = 1, 2, \dots,$$

with mean $\mathbb{E}\{T\} = 1/p$ and a variance $\mathbb{E}\{S_n - 1/p\}^2 = (1-p)/p^2$, i.e., a geometric distribution with parameter p. This yields the *memoryless* fact that the sequence of random variables $\{x_{n+1}, x_{n+2}, \dots\}$ (the future after n) is also a Bernoulli process, which is independent of $\{x_1, \dots, x_n\}$. Also, the *fresh-start property* holds, i.e., for a given n define $T_n := \inf\{i \geq n : x_i = 1\}$ then $T_n - n$ has a geometric distribution with parameter p, and is independent of the random variables $\{x_1, \dots, x_n\}$.

The equivalent of this in continuous time is the Poisson process, where the inter-arrival times are exponentially distributed, i.e., given a sequence of independent identically exponentially distributed random variables $\{\tau_1, \tau_2, \dots\}$, the counting process

$$p(t) = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^{n} \tau_i \leq t < \sum_{i=1}^{n+1} \tau_i \end{cases}$$

with values in $\{0, 1, 2, \dots\}$, is called a Poisson process. A realization of this process requires some properties on the probability space.

Perhaps the simplest example of an uncountable probability space is the unit interval with the Lebesgue measure $([0,1], \mathcal{L}, \ell)$, where $\mathcal{L}$ is the Lebesgue σ-algebra. The real-valued random variable $U(\omega) := \omega$ satisfies $\ell(U \leq r) = (r \wedge 1) \vee 0$ for every r in $\mathbb{R}^d$, which is referred to as the *uniform distribution* on $[0,1]$. More general the Hilbert cube $\Omega := [0,1]^{\{1,2,\dots\}}$, i.e., the space of sequences $\omega = (\omega_1, \omega_2, \dots)$ with values in $[0,1]$, endowed with the product σ-algebra $\mathcal{F} := \mathcal{L}^{\{1,2,\dots\}}$ and the product measure $P := \ell^{\{1,2,\dots\}}$, provides a canonical space for a sequence of independent random variables $(U_1, U_2, \dots)$, each having uniform distribution on $[0,1]$, defined by $U_i(\omega) := \omega_i$, for every $\omega = (\omega_1, \omega_2, \dots)$. In theory, almost all statistical simulation is based on this probability space. Random number generator in computers produce sequences of numbers which are statistically indistinguishable (as much as possible) from

samples $(U_1, U_2, \ldots)$. Random variables with other distributions are then produced by well-known transformations. For instance, given a distribution F in the real line, i.e., a function $F : \mathbb{R} \to [0,1]$ monotonically increasing and right-continuous with $F(-\infty) = 0$ and $F(+\infty) = 1$, its inverse function defined by $F^{-1}(\rho) := \inf\{r : F(r) \geq \rho\}$ satisfies $F^{-1}(\rho) \leq r$ if and only if $\rho \leq F(r)$. Thus, if U is a random variable uniformly distributed in $[0,1]$ then $V := F^{-1}(U)$ satisfies $P(V \leq r) = F(r)$, i.e., F is the distribution of V.

Given a Borel subset E of $\mathbb{R}^d$, it is possible to construct a one-to-one Borel function $\phi : E \to [0,1]$ such that $\phi(E)$ is a Borel subset of $[0,1]$ and $\phi^{-1} : \phi(E) \to E$ is Borel measurable. From this we deduce that for any measure μ on a Borel subset E of R^d there exists a measurable function $\Upsilon : [0,1] \to E$ such that $\ell(\Upsilon^{-1}(B)) = \mu(B)$ for every B in $\mathcal{B}(E)$, the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ restricted to E. Indeed, setting $F(r) := \mu(\phi^{-1}([0,r]))$ and $F^{-1}(\rho) := \inf\{r : F(r) \geq \rho\}$ as above we may take $\Upsilon(\rho) := \phi^{-1}(F^{-1}(\rho))$ for any $F^{-1}(\rho)$ belongs to $\phi(E)$ and $\Upsilon(\rho) := 0$ otherwise.

One of the advantages of stochastic modeling is that calculations are greatly facilitated if the model is formulated as a Markov process, so that general methods for computing distributions and expectations (based on the Dynkin formula and the Kolmogorov backward equation) are available. If the randomness is in the form of point events then the prototype is the Poisson process. A non-negative real random variable T is exponentially distributed if its survivor function $F(t) = F_T(t) := P(T > t)$, for every $t \geq 0$, has the form $F(t) = \mathrm{e}^{-\lambda t}$, for some constant $\lambda > 0$. The mean and the standard deviation of T are both equal to $1/\lambda$. The *memoryless* property of the exponential distribution relative to the conditional distribution, i.e.,

$$P(T > t + s \mid T > s) = \frac{F(t+s)}{F(s)} = F(t) = \mathrm{e}^{-\lambda t},$$

make T a prototype of a (Markov) stopping time. Thus the conditional distribution of the remaining time (i.e., given $T > s$) is just the same as the unconditional distribution of T, regardless of the elapsed time s. Another way of expressing this is in terms of the *hazard rate*, which is by definition a function $h(t)$ satisfying

$$\lim_{\delta \to 0} \frac{P(T \in]s, s+\delta] \mid T > s) - h(s)}{\delta} = 0,$$

i.e., $h(s)\delta$ expresses, to first order, the probability that T occurs 'now' given that it has not occurred 'so far'. In the exponential case we have $P(T \in]s, s+\delta] \mid T > s) = 1 - \mathrm{e}^{-\lambda\delta}$, so that the hazard rate is constant, $h(t) = \lambda$. For a non-negative random variable with a general density function ψ the hazard rate is given by

$$h(s) = \frac{\psi(s)}{\Psi(s)}, \quad \forall s \in [0, c[, \quad \text{with}$$

$$\Psi(s) := \int_s^\infty \psi(r)\mathrm{d}r, \quad c := \inf\{r : \Psi(r) = 0\},$$

where $\Psi(s)$ is the corresponding survivor function. In fact, there is a one-to-one correspondence between h and Ψ based on the ordinary differential equation $\dot{\Psi} = -h\Psi$ and the initial condition $\Psi(0) = 1$. Thus the exponential is the only distribution with constant hazard rate.

Let us construct a sequence of independent identically distributed (exponential with parameter $\lambda > 0$) random variables $(\tau_1, \tau_2, \ldots)$ in the canonical Hilbert cube $(\Omega, \mathcal{F}, P)$. Let $(U_1, U_2, \ldots)$ be the canonical sequence of independent random variables each having uniform distribution on $[0, 1]$ as above. Then setting $\Psi(t) := \mathrm{e}^{-\lambda t}$ and $\Psi^{-1}(u) := \inf\{t \geq 0 : \Psi(t) \leq u\}$, for every $u > 0$, we define $\tau_i := \Psi^{-1}(U_i(\omega)) = -\ln[U_i(\omega)]/\lambda$, for every $i \geq 1$, which satisfies $P(\tau_i > t) = \Psi(t)$, i.e., exponentially (with parameter λ) distributed and independent.

Now define $\theta_0 = 0$, $\theta_n := \tau_1 + \tau_2 + \ldots + \tau_n$, which has $\Gamma(\lambda, n)$, i.e., $P(\theta_n \in \mathrm{d}t) = \big(\lambda^n s^{(n-1)}/(n-1)!\big)\mathrm{e}^{-\lambda t}\mathrm{d}t$, and

$$N(t) := \sum_{i=1}^{\infty} \mathbb{1}_{t \geq \theta_i}, \qquad \text{i.e.,} \quad N(t) := n \quad \text{if} \quad \theta_n \leq t < \theta_{n+1}.$$

The sample functions of $(N(t) : t \geq 0)$ are right-continuous step functions with jumps of height 1 at each τ_i, in particular it is cad-lag, belonging to the canonical sample space $D([0, \infty[)$. The random variable $N(t)$ has a Poisson distribution $P(N(t) = n) = \mathrm{e}^{-\lambda t}(\lambda t)^n/n!$, with mean $\mathbb{E}\{N(t)\} = \lambda t$. Denote by $(\mathcal{F}_t : t \geq 0)$ it natural filtration, i.e., $\mathcal{F}_t$ is the σ-algebra generated by the random variables $\{N(s) : 0 \leq s \leq t\}$. Fix $t > 0$ and denote by θ^t the last jump time before t, i.e, $\theta^t(\omega) = \theta_n(\omega)$, with $\theta_n(\omega) \leq t < \theta_{n+1}(\omega)$ and $n = n^t(\omega)$. In view of the memoryless property of the exponential, if $\tau_1^* := \theta_{n+1} - t$ and $\tau_i^* := \tau_{n+i}$, for $i \geq 2$, with $n = n^t(\omega)$, then the conditional distribution of τ_1^* given $\mathcal{F}_t$ (or equivalently, given that $\tau_{n+1} > t - \theta_n$) is exponential, $P(\tau_1^* > s \mid \mathcal{F}_t) = \mathrm{e}^{-\lambda t}$, and so the sequence $\{\tau_1^*, \tau_2^*, \ldots\}$ is independent identically distributed (exponential with parameter $\lambda > 0$). It follows that $N^*(s) := N(t + s) - N(t)$, $s \geq 0$ is a Poisson process independent of $\mathcal{F}_t$, i.e., the process 'restart' at time t. In particular, it has independent increments, i.e., $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3)$ are independent variables for any $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$. This implies that $(N(t) : t \geq 0)$ is a Markov process, indeed, for any bounded and Borel measurable function and $t > s \geq 0$ we have

$$\mathbb{E}\{f(N(t)) \mid \mathcal{F}_s\} = \mathrm{e}^{\lambda(t-s)} \sum_{k=0}^{\infty} f(k + N_s) \frac{[\lambda(t-s)]^k}{k!},$$

since $N(t) - N(s)$ is Poisson distributed with mean $\lambda(t - s)$. The Poisson process may be considered as a Markov process in either the integer numbers $E = \{0, \pm 1, \pm 2, \ldots\}$ or the non-negative integer numbers $E = \{0, 1, 2, \ldots\}$. The process $i + N(t)$ yields probability measure P_i and the transition function is

$$p(i, t, j) := \begin{cases} \mathrm{e}^{\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, & \text{if } j \geq i, \\ 0 & \text{otherwise,} \end{cases}$$

for any i, j in E. This defines a semigroup $P(t) := \mathbb{E}\{f(x+N(t))\}$ on the space $B(E)$ of real (Borel) bounded functions on E. The infinitesimal generator A is

$$Af(x) := \lim_{t\to 0} \frac{\mathbb{E}\{f(x+N(t))\}}{t} = \lambda[f(x+1) - f(x)], \quad \forall x \in E,$$

where the domain $\mathcal{D}(A)$ of the strong infinitesimal generator is the space functions f for which the above limit exists uniformly in x. Consider the process

$$M_x^f(t) := f(x+N(t)) - f(x) - \\ -\lambda \int_0^t [f(x+N(r)+1) - f(x+N(r))]\mathrm{d}r, \quad t \geq 0$$

for any f in $B(E)$ and x in E. In view of the independent increment property and the fact that $N(t) - N(s)$ is Poisson distributed, we have for $t > s$

$$\mathbb{E}\{f(x+N(t)) - f(x+N(s)) \mid \mathcal{F}_s\} = \\ = \mathrm{e}^{-\lambda(t-s)} \sum_{k=0}^{\infty} [f(k+x) - f(x)] \frac{[\lambda(t-s)]^k}{k!}$$

and

$$\mathbb{E}\{\lambda \int_s^t [f(x+N(r)+1) - f(x+N(r))]\mathrm{d}r\} = \\ = \lambda \sum_{k=0}^{\infty} [f(x+k+1) - f(x+k)] \int_s^t \mathrm{e}^{-\lambda r} \frac{(\lambda r)^k}{k!}\mathrm{d}r,$$

which yields $\mathbb{E}\{M_x^f(t) - M_x^f(s) \mid \mathcal{F}_s\} = 0$, i.e., $(M_x^f(t) : t \geq 0)$ is a martingale. Actually, this calculation remains valid for any function (because E is countable, all functions are Borel measurable) such that $\mathbb{E}\{|f(x+N(t))|\} < \infty$, for every x in E and $t \geq 0$. By the optional sampling theorem, the process $M_x^{f,n}(t) := M_x^f(t \wedge n \wedge \theta_n)$ is also a martingale, since $n \wedge \theta_n$ is a bounded stopping time. However, the process $(M_x^{f,n}(t) : t \geq 0)$ involves only the values of f on the finite set $\{x, x+1, \ldots, x+n\}$. Therefore, the process $(M_x^f(t) : t \geq 0)$ is a local martingale for any function. Thus the domain $\mathcal{D}(\bar{A})$ of the extended infinitesimal generator $\bar{A}$ consists of all functions $f : E \to \mathbb{R}$ with not restriction at all.

The *renewal process* is closely related to the Poisson process. It is a point process $(N(t) : t \geq 0)$ defined in a similar way to the Poisson process but with the inter-arrival time τ_i now being a sequence of independent identically distributed random variables with some density function ψ on $[0, \infty)$, not necessarily exponential. The process clearly 'restarts' at each 'renewal time' θ_i and the well-known renewal equation

$$m(t) := \mathbb{E}\{N(t)\}, \qquad m(0) = 0,$$
$$m(t) = \int_0^t [1 + m(t-r)]\psi(r)\mathrm{d}r, \quad \forall t \geq 0,$$

which can be solved by the Laplace transform methods.

The sequences of inter-arrival time $\{\tau_1, \tau_2, \ldots\}$ is now constructed as follows. Then setting $\Psi(t) := \int_t^\infty \psi(r)\mathrm{d}r$ and $\Psi^{-1}(u) := \inf\{t \geq 0 : \Psi(t) \leq u\}$, for every $u > 0$, we define $\tau_i := \Psi^{-1}(U_i(\omega))$, for every $i \geq 1$, which satisfies $P(\tau_i > t) = \Psi(t)$, i.e., independent identically distributed with density ψ.

When τ_i is not exponentially distributed, the memoryless property does not hold and the conditional distribution of the residual time $\tau_1^* := \theta_{n+1} - t$ (as defined above for the Poisson process) given $\mathcal{F}_t$ depends on the time $t - \theta_n$ since the last jump. Therefore, the renewal process itself is not a Markov process, if we add a new variable $S(t) := t - \theta_n$, the time since the last jump where $\theta_n \leq t < \theta_{n+1}$, then the new two-component process $X := \{(N(t), S(t)) : t \geq 0\}$ is a Markov process on $E := \{0, 1, 2, \ldots\} \times [0, \infty)$. Its evolution can be simulate as follows. For a fixed (n, s) and with $\psi(t)$, $\Psi(t)$ as above, we set first

$$\lambda(r) := \begin{cases} \dfrac{\psi(r)}{\Psi(r)} & \text{if } \ 0 \leq r < c_\Psi, \\ 0 & \text{otherwise,} \end{cases}$$

with $c_\Psi := \inf\{r : \Psi(r) = 0\}$, and then

$$\Psi(s, t) := \exp\Big(- \int_s^t \lambda(r)\mathrm{d}r \Big), \qquad \forall t \geq s \geq 0.$$

Note that $\Psi(0, t) = \Psi(t)$, for every $t \geq 0$. Thus, we re-define

$$\Psi^{-1}(s, u) := \inf\{t \geq 0 : \Psi(s, t) \leq u\}, \qquad \tau_1(\omega) = \theta_1(\omega) := \Psi^{-1}(s, U_1(\omega)),$$

with the convention that $\Psi^{-1}(s, u) := +\infty$ if $\Psi(s, t) > u$ for every $t \geq 0$. The sample path $X(t, \omega)$ is then $(n, s + t)$ if $0 \leq t < \theta_1(\omega)$ and $(n + 1, 0)$ if $t = \theta_1(\omega) < \infty$. Next, if $\theta_1(\omega) < \infty$ we restart with the initial state $(n + 1, 0)$ and the same recipe. This is $\tau_2(\omega) := \Psi^{-1}(0, U_2(\omega))$, $\theta_2(\omega) := \theta_1(\omega) + \tau_2(\omega)$ and

$$X(t, \omega) := \begin{cases} (n + 1, t - \theta_1(\omega)) & \text{if } \theta_1(\omega) \leq t < \theta_2(\omega), \\ (n + 2, 0) & \text{if } t = \theta_2(\omega) < \infty, \end{cases}$$

and so on. The key point is that this construction generalizes to a much more general situation.

The formal expression of the infinitesimal generator is

$$Af(n, s) := \lim_{t \to 0} \frac{\mathbb{E}\{f(n + N(t), s + S(t)\} - f(n, s)}{t} = \\ = \partial_s f(n, s) + \lambda(s)[f(n + 1, 0) - f(n, s)],$$

where the hazard rate $\lambda(t) := \psi(t)/\Psi(t)$ and ∂_s means the partial derivative in the second variable, i.e., in s. The domain $\mathcal{D}(A)$ of the strong infinitesimal generator should include conditions to ensure that the above limit exists uniformly in (n, s), in particular $f(n, s)$ should be differentiable in s. However, the

domain $\mathcal{D}(\bar{A})$ of the extended infinitesimal generator would only impose that $s \mapsto f(n,t)$ be absolutely continuous.

Another typical example is a *single-server queue*. Customers arrive at a queue at random times $\{\theta_1 \le \theta_2 \le \cdots\}$ which require a service time $\{\varsigma_1, \varsigma_2, \ldots\}$, measured in units of time for processing. The total service load presented up to time t is $L(t) := L_0 + \sum_i \varsigma_i \mathbb{1}_{\theta_i \le t}$, where $L_0 \ge 0$ is the service load existing at time 0. The virtual waiting time $V(t)$ is the unique solution of the equation

$$V(t) = L(t) - \int_0^t \mathbb{1}_{V(r)>0} \mathrm{d}r, \quad \forall t \ge 0,$$

and represents the time a customer arriving at time t waits for service to begin, or equivalently, the amount of unprocessed load at time t.

A similar way to describe a queueing system is by means of the relation $Q(t) = Q_0 + A(t) - D(t)$, where $(A(t) : t \ge 0)$ and $(D(t) : t \ge 0)$ are non-explosive point processes without common jumps, i.e., $A(t) = n$ for $\theta_n \le t < \theta_{n+1}$ and $D(t) = n$ for $\vartheta_n \le t < \vartheta_{n+1}$, $\theta_0 = \vartheta_0 = 0$, $\theta_n, \vartheta_n \to \infty$ as $n \to \infty$ and $P(\theta_i = \vartheta_j) = 0$, for every i, j. The random variable Q_0 is the initial state and the state process satisfies $Q(t) \ge 0$, for any $t \ge 0$, which is interpreted as the number of customer waiting in line or being attended by the server (i.e., in the system). The processes A and D are called arrival and departure processes. Thus, if A and B are two nonexplosive point processes without common jumps then to achieve the condition $Q(t) \ge 0$ we set $Y(t) := Q_0 + A(t) - B(t)$ and $m(t) := \min\{Y(r) \wedge 0 : r \in [0,t]\}$. Hence, a simple queueing system $Q(t) + Q_0 + A(t) - D(t)$ can be constructed with $Q(t) := Y(t) - m(t)$, $D(t) := \int_0^t \mathbb{1}_{Q(r-)>0} dB(r)$, where also $m(t) = \int_0^t \mathbb{1}_{Q(r-)=0} dB(r)$.

There is a conventional classification $A/B/n$ of queueing systems, where A refers to the arrival process (i.e., statistics of the increasing sequence of random variables $\{\theta_1, \theta_2, \ldots\}$), B to the service process (i.e., statistics of the sequence of random variables $\{\varsigma_1, \varsigma_2, \ldots\}$) and n is the number of servers. For instance, consider a $M/G/1$ queue, i.e., the letter M (for Markov) means that arrival are independent and exponential, i.e., from a Poisson process, and G (for general) means that the service time independent identically distributed with some arbitrary distribution on $(0, \infty)$.

A variable ν indicates whether the queue is busy $\nu = 1$ or empty $\nu = 0$. This means that ν vanishes, $\nu = 0$, if and only if the virtual waiting time vanishes, $v = 0$. Thus, starting from a time $t_0 \ge 0$ with $\nu(t_0) = 1$ and $V(t_0) = v$, the process $v(t) := V(t)$ decreases at unit rate until it hits zero, say at time t_1. Then $\nu(t_1)$ becomes zero, $\nu(t_1) = 0$, and $v(t) = 0$ until a new arrival $t_2 > t_1$ which takes an exponential time, and (ν, v) jumps to $(1, \vartheta)$, i.e., $\nu(t_2-) = 0$, $\nu(t_2+) = 1$, $v(t_2-) = 0$ and $v(t_2+) = \varsigma$. The state (ν, v) has a Markov evolution on the set $E = \{(0,0)\} \cup \{1\} \times (0, \infty)$, which is normalized to be a cad-lag process. Roughly speaking, if the initial state is $(1, v)$ then after a short time δ the state becomes $(1, v - \delta)$ with probability $(1 - \lambda\delta)$, while with probability $\lambda\delta$ the Markov process jumps to the new state $(1, v + \varsigma - \delta)$, where ς has the distribution on $(0, \infty)$ of the services time, namely F_ϑ, and λ is the parameter

of the exponential distribution of the arrival times. With this in mind, the infinitesimal generator has the expression

$$Af(1,v) = -\partial_v f(1,v) + \lambda \int_{(0,\infty)} [f(1,v+z) - f(1,v)] F_\vartheta(\mathrm{d}z),$$
$$Af(0,0) = \lambda \int_{(0,\infty)} [f(1,z) - f(0,0)] F_\vartheta(\mathrm{d}z),$$

for any v in $(0,\infty)$. It is clear that this formula of the infinitesimal generator A does not include the fact that the process jumps from $(1,0)$ to $(0,0)$, immediately after hitting $(1,0)$. This is the *boundary conditions*

$$f(1,0) = \int_{[0,\infty)} f(1,z) F_\vartheta(\mathrm{d}z)$$

added to the strong domain $\mathcal{D}(A)$ or extended $\mathcal{D}(\bar{A})$.

A construction of the Markov process starting at $x := (\nu, v)$ is described in the canonical Hilbert cube $(\Omega, \mathcal{F}, P)$, where $(U_1, U_2, \ldots)$ is sequence of independent random variables each having uniform distribution on $[0,1]$. First we set

$$Q(x,B) := \int_{[v,\infty)} \mathbb{1}_B(z) F_\varsigma(\mathrm{d}z),$$

$$\mathtt{X}(x,t) := \begin{cases} (\nu, v-t) & \text{if } \nu = 1, \\ (\nu, v) & \text{otherwise,} \end{cases}$$

and

$$\mathtt{T}(x) := \begin{cases} v & \text{if } \nu = 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Given an initial state $x = (\nu, v)$ in $E = \{(0,0)\} \cup \{1\} \times (0,\infty)$ we define

$$\Psi(x,t) := \mathbb{1}_{t<\mathtt{T}(x)} \mathrm{e}^{-\lambda t},$$

the survivor function of the first jump time θ_1 of the process and its (generalized) inverse

$$\Psi^{-1}(x,u) := \inf\{t \geq 0 : \Psi(x,t) \leq u\},$$
$$\tau_1(\omega) = \theta_1(\omega) := \Psi^{-1}(x, U_1(\omega)),$$

with the convention that $\Psi^{-1}(x,u) := +\infty$ if $\Psi(x,t) > u$ for every $t \geq 0$. This yields $P(\theta_1 > t) = \Psi(x,t)$. As mentioned above, we are working in the canonical Hilbert cube and there exist a measurable function Υ from $E \times [0,1]$ into $\mathring{E} := \{1\} \times (0,\infty)$ such that $\ell(\{u : \Upsilon(x,u) \in B\}) = Q(x,B)$, for every B

in $\mathcal{B}(E)$, where ℓ is the Lebesgue measure on $[0,1]$. The sample path $X(t,\omega)$ is defined up to the first jump as follows:

$$X(t,\omega) := \mathtt{X}(x,t), \quad \text{if} \quad 0 \le t < \theta_1(\omega),$$
$$X(\theta_1(\omega),\omega) := \Upsilon(\mathtt{X}(x,\theta_1(\omega)),\omega), U_2(\omega)), \quad \text{if} \quad \theta_1(\omega) < \infty.$$

Note that when $\nu = 1$, as long as $t < \theta_1(\omega)$ we have $v - t > 0$. If $t = \theta_1(\omega) < \infty$ then $v - t \ge 0$. On the other hand, when $\nu = 0$ we have $\theta_1(\omega) < \infty$ and $X(t,\omega) = (0,v)$ for every $t < \theta_1(\omega)$. In any case, the definition of Υ ensure that $X(\theta_1(\omega),\omega)$ belongs to $\mathring{E} = \{1\} \times (0,\infty)$. Now, if $\theta_1(\omega) < \infty$ the process restarts from $X(\theta_1(\omega),\omega)$ according to the same recipe. Thus, if $\theta_1(\omega) < \infty$ we define

$$\tau_2(\omega) := \Psi^{-1}(X(\theta_1(\omega),\omega), U_3(\omega)),$$
$$\theta_2(\omega) := \theta_1(\omega) + \tau_2(\omega)$$

and the sample path $X(t,\omega)$ up to the next jump is given by

$$X(t,\omega) := \mathtt{X}(x, t - \theta_1(\omega)), \quad \text{if} \quad \theta_1(\omega) \le t < \theta_2(\omega),$$
$$X(\theta_2(\omega),\omega) := \Upsilon(\mathtt{X}(x,\tau_2(\omega)),\omega), U_4(\omega)), \quad \text{if} \quad \theta_2(\omega) < \infty,$$

and so on.

This procedure define the sample path $X(t,\omega)$ if the sequence $\theta_k(\omega) \to \infty$. Hence, a common assumption is to impose that

$$\mathbb{E}\{N(t)\} < \infty, \quad \text{with} \quad N(t) := \sum_k \mathbb{1}_{t \ge \theta_k},$$

which yield some condition on the distribution $F_\varsigma(t)$ of the sequence $\{\varsigma_1, \varsigma_2, \ldots\}$ associated with the service time. Since

$$P(t < \tau_i < \infty) = \mathbb{1}_{t < \mathtt{T}(x)} \mathrm{e}^{-\lambda t}$$

a condition on the type $P(\varsigma > \varepsilon) = 1$ for some positive ε, on the service time distribution ensures the required assumption.

If the arrival process is a renewal process instead of a Poisson process then we need one more variable to have a Markov process, the time since the last jump s, so that the state is $x = (\nu, v, s)$ in $E = \{(0,0)\} \times [0,\infty) \cup \{1\} \times (0,\infty) \times [0,\infty)$ and an *intensity* or hazard rate $\lambda(t) := \psi(t)/\Psi(t)$, where ψ is the density of the arrival times. The previous simulation can be adapted, where $s + t$ is the evolution in last variable, which is reset to zero each time a new arrival occurs.

It is clear that the above technique can be used for more general situation, e.g., a $G/G/n$ queue system and many other stochastic models.

Another type of common jump process is the so-called *doubly stochastic Poisson process* or *conditional Poisson process* $(N(t) : t \ge 0)$ with stochastic intensity $\lambda = \lambda(t,\omega) \ge 0$ relative (i.e., adapted) to the filtration $(\mathcal{F}_t : t \ge 0)$ on a probability space $(\Omega, \mathcal{F}, P)$, which is defined by the condition

$$\mathbb{E}\big\{\mathrm{e}^{i\zeta[N(t)-N(s)]} \mid \mathcal{F}_s\big\} = \exp\Big[\big(\mathrm{e}^{iu} - 1\big) \int_s^t \lambda(r)\mathrm{d}r\Big],$$

for every $t \geq s \geq 0$, where $\lambda(r)$ is an $\mathcal{F}_0$-measurable and almost surely integrable function in $[0,t]$. Usually, the intensity takes the form $\lambda(t,\omega) = \lambda(t, Y(t,\omega))$, where the process $(Y(t) : t \geq 0)$ is $\mathcal{F}_0$-measurable and the function $(t,y) \mapsto \lambda(t,m)$ is a nonnegative Borel measurable with some appropriate integrability conditions. This means that on one hand we have a Poisson process with intensity $\lambda(t,y)$ where the parameter y is replaced by an independent process $y = Y(t)$ and the σ-algebra $\mathcal{F}_0$ is augmented with the σ-algebra generated by $(Y(t) : t \geq 0)$ to meet the $\mathcal{F}_0$-measurability condition. A conditional Poisson process $(N(t) : t \geq 0)$ is characterized by the property

$$\mathbb{E}\Big\{ \int_0^\infty X(t)\mathrm{d}N(t) \Big\} = \mathbb{E}\Big\{ \int_0^\infty X(t)\lambda(t)\mathrm{d}t \Big\},$$

valid for any $\mathcal{F}_t$-predictable process $(X(t) : t \geq 0)$, c.f., Bremaud [32, Chapter 2, Theorem T4]. Thus its predictable jumps compensator is the integral process $\langle N \rangle(t) = \int_0^t \lambda(r,\omega)\mathrm{d}r$. Conditional Poisson processes are in general not Markov processes, however, if the intensity function $\lambda(t,\omega) = \lambda(Y(t))$, where $(Y(t) : t \geq 0)$ is a $\mathcal{F}_0$-measurable Markov process, then the couple (N,Y) becomes a Markov process with a suitable filtration, actually a compound Poisson process.

To end this section let us mention the so-called *multivariate point processes* which are defined by means of two sequences $\{\theta_0, \theta_1, \theta_2, \ldots\}$ and $\{\zeta_1, \zeta_2, \ldots\}$ of random variables with values in $[0,\infty]$ and $\{1, 2, \ldots, d\}$, respectively, and satisfying $\theta_0 = 0$, if $\theta_n < \infty$ then $\theta_n < \theta_{n+1}$, and called *nonexplosive* when $\theta_\infty := \lim_n \theta_n = \infty$. The sample path is defined by the *d-counting* process $(N_i(t) : t \geq 0)$, $i = 1, 2, \ldots, d$

$$N_i(t) := \sum_{n=1}^{\infty} \mathbb{1}_{\theta_n \leq t}\, \mathbb{1}_{\zeta_n = i}, \forall t \geq 0.$$

Both the d-vector process $(N(t) : t \geq 0)$ with nonnegative integer-values components as above and the double sequence $\{\theta_n, \zeta_n : n = 1, 2, \ldots\}$ are called d-variate point process. Note each component $(N_i(t) : t \geq 0)$ is a (simple or univariate) point process and that only one component jumps at a given time, i.e., there is not common jumps among all the processes $(N_i(t) : t \geq 0)$, for $i = 1, 2, \ldots, d$. For instance, if the double sequence $\{\theta_n, \zeta_n : n = 1, 2, \ldots\}$ is independent identically distributed, θ_n $\Gamma(\lambda, n)$ distributed and ζ_n such that $P(\zeta_n = 1) = p$, $P(\zeta_n = 2) = 1 - p$, with some constant p in $(0,1)$, then interpreting $N_1(t)$ as the births and $N_2(t)$ as the deaths up to time t of a given population, the expression $N_1(t) - N_2(t)$ is a birth-and-death process.

For instance, the reader is refereed to the books by Bremaud [32, 33], Daley and Vere-Jones [55], Davis [56, Chapter 1], Revuz and Yor [212, Section XII.1, pp. 471–480], among others.

3.9 Piecewise Deterministic Processes

Non-diffusion stochastic models called piecewise-deterministic Markov processes (PDP) is proposed as a general framework for studying problems involving

non-diffusion continuous-time dynamical systems whose deterministic motion is punctuated by random jumps. A great number of applications in engineering systems, operations research, management science, economic and applied probability show the importance of these systems. Queuing systems, investment planning, stochastic scheduling and inventory systems are some examples, we refer to Davis [56] for a comprehensive study.

3.9.1 Vector Fields and ODE

Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a globally Lipschitz continuous function, i.e., there exists a constant $M > 0$ such that $|g(x) - g(x')| \leq M|x - x'|$ for every x, x' in $\mathbb{R}^d$. It is well-know that the ordinary differential equation (ODE) relative to g, i.e., the initial value problem (IVP) $\dot{x}(t) = g(x(t))$, $x(0) = x$, has a unique solution defined for every x in $\mathbb{R}^d$. We denote by $\mathtt{X}(x,t)$ its solution, i.e.,

$$\partial_t \mathtt{X}(x,t) = g(\mathtt{X}(x,t)), \quad \forall t \in \mathbb{R}, \qquad \mathtt{X}(x,0) = x, \quad \forall x \in \mathbb{R}^d,$$

which has the properties:

(1) the map $\mathtt{X}_t : x \mapsto \mathtt{X}(x,t)$ is Lipschitz continuous (uniformly in t), one-to-one and onto, indeed, $\mathtt{X}^{-1}(x,t) = \mathtt{X}(x,-t)$, for every x in $\mathbb{R}^d$ and t in $\mathbb{R}$,

(2) the family $\{\mathtt{X}_t : t \in \mathbb{R}\}$ is a group, i.e., $\mathtt{X}_t \circ \mathtt{X}_s = \mathtt{X}_{t+s}$, or more explicitly $\mathtt{X}(x, t+s) = \mathtt{X}(\mathtt{X}(x,s),t)$, for every x in $\mathbb{R}^d$ and t, s in $\mathbb{R}$.

This is referred to as an *homogeneous Lipschitz flow* in $\mathbb{R}^d$.

If f is a real valued continuously differentiable function, i.e., f in $C^1(\mathbb{R}^d)$, then we may consider $\mathtt{X}$ as a first-order differential operator, $\mathtt{G} : C^1(\mathbb{R}^d) \to C^0(\mathbb{R}^d)$, defined by

$$\mathtt{G}f(x) := \sum_{i=1}^{d} g_i(x) \partial^i f(x), \quad \forall x \in \mathbb{R}^d,$$

where $\partial^i f$ means the first partial derivative with respect to the variable x_i. By means of the chain rule we deduce that $t \mapsto x(t)$ is a solution of the ODE relative to g if and only if

$$\mathtt{G}f(x(t)) := \sum_{i=1}^{d} g_i(x(t)) \partial^i f(x(t)), \quad \forall t \in \mathbb{R},, \quad \forall f \in C^1(\mathbb{R}^d),$$

which is a coordinates-free form of the differential equation, the operator $\mathtt{G}$ (and the function g) is know as the *vector field* associated with the flow $\mathtt{X} = \{\mathtt{X}_t : t \in \mathbb{R}^d\}$.

If the function g is continuously differentiable then the function homeomorphism $x \mapsto \mathtt{X}(x,t)$ is indeed a diffeomorphism and it satisfies the linear system of ODEs

$$\partial_t \mathtt{X}_{i,j}(x,t) = \sum_{k=1}^{d} g_{i,k}[\mathtt{X}(x,t)]\, \mathtt{X}_{k,j}(x,t),$$

for every $i, j = 1, \dots, d$, x in $\mathbb{R}^d$ and $t \in \mathbb{R}$, where the subscript i, j as in $\mathtt{X}_{i,j}$ denotes the the first partial derivative in the x_j variable of the i component of $\mathtt{X}(x,t)$, i.e., $\partial_j \mathtt{X}_i(x,t)$.

This analysis can be extended to non-flat manifolds and the assumption on g can be weakened. Of particular interest for us is the case where $\mathbb{R}^d$ is replaced by a finite intersection of nonempty domain D having a representation of the form

$$\begin{cases} D := \{x \in \mathbb{R}^d : \phi(x) < 0\}, \\ \phi \in C^1(\mathbb{R}^d), \quad |\nabla\phi(x)| \geq 1, \quad \forall x \in \partial D, \end{cases} \tag{3.51}$$

which implies that D is an open set with an outward unit normal vector given by $\nabla\phi(x)/|\nabla\phi(x)|$ on the boundary ∂D. The function g defining the vector field $\mathtt{G}$ and the flow $\mathtt{X}$ is usually assumed locally Lipschitz continuous in $\bar{D}$ and with linear growth when D is unbounded. Local uniqueness and existence of solution to the IVP is ensure by the local Lipschitz continuity, and so the solution is extended to its maximum interval of existence with bounds (which may be infinite) $t^{\wedge}_{\bar{D},x} < t^{\vee}_{\bar{D},x}$. If $t^{\vee}_{\bar{D},x} < \infty$ (or $t^{\wedge}_{\bar{D},x} < \infty$) then we assume that there exist $t < t^{\vee}_{\bar{D}}$ ($t > t^{\wedge}_{\bar{D}}$) such that $\mathtt{X}(x,t)$ does not belongs to $\bar{D}$. This *non-explosive* condition effectively rule out the case where an explosion occurs in closure $\bar{D}$. The linear growth condition implies this non-explosion assumption, but it is not necessary. Thus under this non-explosion condition we can define the first exit time from any Borel subset E of $\bar{D}$ as follows

$$\tau_E(x) := \inf\{t \geq 0 : \mathtt{X}(x,t) \notin E\}, \quad \forall x \in \bar{D},$$

with the convention that $\tau_E(x) = \infty$ if $\mathtt{X}(x,t)$ belongs to E for every $t \geq 0$. It is clear that $\tau_E(x) = 0$ for any x in $\bar{D} \setminus E$ and that $x \mapsto \tau_E(x)$ is a measurable $[0,\infty]$-valued function. Also, if $g(x) \cdot \nu(x) > 0$, with $\nu(x) := \nabla\phi(x)/|\nabla\phi(x)|$ the exterior unit normal vector at x in ∂D, then $\tau_{\bar{D}}(x) = 0$. On the other hand, if $g(x) \cdot \nu(x) \leq 0$ for every x in ∂D then $\tau_{\bar{D}}(x) > 0$ for any x in ∂D, i.e., $\mathtt{X}(t,x)$ belongs to $\bar{D}$ for every x in $\bar{D}$ and $t \geq 0$.

It is convenient to divide the flow $\mathtt{X} = \{\mathtt{X}_t : t \in \mathbb{R}^d\}$ into two flows, a forward flow $\mathtt{X} = \{\mathtt{X}_t : t \geq 0\}$ and a backward flow $\mathtt{X} = \{\mathtt{X}_t : t \leq 0\}$.

Theorem 3.17. *Let $g : \mathbb{R}^d \to \mathbb{R}^d$ be a locally Lipschitz continuous function which yields a forward flow $\mathtt{X} = \{\mathtt{X}_t : t \geq 0\}$ without explosions in $\bar{D}$ as above, so that the first exist times τ_D from the open set D and $\tau_{\bar{D}}$ from the closed set $\bar{D}$. Then the functions $x \mapsto \tau_D(x)$ and $x \mapsto \tau_{\bar{D}}(x)$ are lower and upper semi-continuous, respectively. If $\partial^0 D := \{x \in \partial D : \tau_{\bar{D}}(x) = 0\}$ then $\tau_{\bar{D}}(\cdot)$ is continuous if and only if $\partial^0 D$ is closed.*

Proof. Take x and s such that $\tau_{\bar{D}}(x) < s$. Then there exists s' such that $\mathtt{X}(x,s')$ does not belongs to $\bar{D}$. If $x_n \to x$, by the continuity of $\mathtt{X}(\cdot,s')$ and because $\mathbb{R}^d \setminus \bar{D}$ is open, there exists N such that $\mathtt{X}(x_n,s')$ does not belongs to $\bar{D}$, for any $n \geq N$. Thus $\tau_{\bar{D}}(x_n) \leq s'$. This proves that $\limsup_n \tau_{\bar{D}}(x_n) \leq \tau_{\bar{D}}(x)$, i.e., $x \mapsto \tau_{\bar{D}}(x)$ is upper semi-continuous.

Similarly, take x and s such that $\tau_D(x) > s$. Then the closed set $\{\mathtt{X}(x,t) : 0 \le t \le s\}$ is contained in the open set D. If $x_n \to x$, again by continuity, there exists N such that $\{\mathtt{X}(x_n,t) : 0 \le t \le s\}$ contained in D, for any $n \ge N$. Thus $\tau_D(x_n) \ge s$. This proves that $\liminf_n \tau_D(x_n) \ge \tau_D(x)$, i.e., $x \mapsto \tau_D(x)$ is lower semi-continuous.

If $\tau_{\bar D}(\cdot)$ is continuous then it is clear that $\partial^0 D$ is closed. On the other hand, take x in $\bar D$ such that $\tau_{\bar D}(x) < \infty$. Since the composition property yields $\mathtt{X}(x, t + \tau_{\bar D}(x)) = \mathtt{X}(\mathtt{X}(x, \tau_{\bar D}(x)), t)$, we deduce that $\mathtt{X}(x, \tau_{\bar D}(x))$ must belongs to $\partial^0 D$. Hence the forward flow exists necessarily though $\partial^0 D$. If $\tau_{\bar D}(x) > s$ then the closed set $\{\mathtt{X}(x,t) : 0 \le t \le s\}$ has a positive distance to the closed set $\partial^0 D$. As in the case of $\tau_D(\cdot)$ we deduce that $\tau_{\bar D}(\cdot)$ is lower semi-continuous. □

We state for further reference the following concept.

Definition 3.18 (locally Lipschitz continuous forward flow)**.** Let E be a set in $\mathbb{R}^d$ having the following property, either

(1) E is the union of an open set $\mathring{E}$ in $\mathbb{R}^d$ and a relative open part $\partial E \setminus \partial^0 E$ of its boundary ∂E (which is the *non-active boundary*, so that the active boundary $\partial^0 E$ is closed); the interior set $\mathring{E}$ is a finite intersection of nonempty domain D having a representation of the form (3.51), or

(2) after a permutation of coordinates the set E has the form $E = E_1 \times E_2$, where E_1 is as in (1) but relative to $\mathbb{R}^{d_1}$ with $d_1 < d$ and E_2 is a single point in $\mathbb{R}^{d-d_1}$ and the corresponding vector field g has only d_1 non-zero components, i.e., g can be considered as a function from $\mathbb{R}^d$ into $\mathbb{R}^{d_1}$.

If g is a locally Lipschitz function from the closure $\bar E$ into $\mathbb{R}^d$ then the following ODE $\dot x(t) = g(x(t))$, for any $t > 0$, can be uniquely solved for any given initial condition $x(0) = x$ in $\mathbb{R}^d$ and locally define flow $(x,t) \mapsto \mathtt{X}(x,t)$ as the solution of the above IVP on the maximal interval of existence $[0, t_x^\vee[$, i.e., for any x in $\mathbb{R}^d$ the solution $x(t) = \mathtt{X}(x,t)$ is defined for every $0 \le t < t_x^\vee \le +\infty$. We say that $\mathtt{X} = \{\mathtt{X}(x,t) : t \ge 0,\ x \in E\}$ is a *locally Lipschitz forward flow* associated to the vector field g on $E \subset \mathbb{R}^d$ or to first-order differential operator $\mathtt{G}$ on $C^1(\mathbb{R}^d)$, with *active boundary* $\partial^0 E$ if for any x in E we have either *(1)* $t_x^\vee = +\infty$ and $\mathtt{X}(x,t)$ belongs to E for any $t \ge 0$ or *(2)* $\mathtt{X}(x,t)$ belongs to $\partial^0 E$ for some t in $]0, t_x^\vee[$. The forward flow $\mathtt{X} = \{\mathtt{X}(x,t) : t \ge 0\}$ is locally Lipschitz continuous in x in E, locally uniformly in t and the first exit time $\tau_E(x)$ from E is denoted by $\mathtt{T}(x)$. We may use the notation $\mathtt{X}^E$ and $\mathtt{T}^E$ to emphasize the dependency on the domain E. □

If the boundary ∂E of E (or E_1 if necessary) is smooth (e.g., $E := \{x \in \mathbb{R}^d : \phi(x) < 0\}$ as above) so that the outward normal vector $\nu(x)$ at x can be defined then every x in ∂E satisfying $g(x) \cdot \nu(x) > 0$ belongs to $\partial^0 E$ (active boundary) while and $g(x) \cdot \nu(x) < 0$ implies that x in the non-active boundary. The fact that we assume $\partial^0 E$ closed or $\partial E \setminus \partial^0 E$ relatively open ensure that the $\partial^0 E$ is the closure of x in ∂E satisfying $g(x) \cdot \nu(x) > 0$. Moreover, a key property is the continuity of the first exit time from E, i.e.,

$$\mathtt{T}(x) := \inf\{t > 0 : \mathtt{X}(x,t) \notin E\} = \inf\{t > 0 : \mathtt{X}(x,t) \in \partial^0 E\},$$

which is defined for any x in E, with the convention that $\mathtt{T}(x) = \infty$ if $\mathtt{X}(x,t)$ belongs to E for every $t > 0$. This means that the sets $\{x : \mathtt{T}(x) < \infty\}$ and $\{x : \mathtt{T}(x) = \infty\}$ are disjointed (one of them may be empty), and that the function $x \mapsto \mathtt{T}(x)$ is continuous on $\{x : \mathtt{T}(x) < \infty\}$. Note that the part of the boundary $\partial E \setminus \partial^0 E$ plays no role, and it is actually ignored. Two interesting cases are when either (1) $\partial^0 E = \emptyset$ so that E is closed, $E = \bar{E}$, or (2) $\partial^0 E = \partial E$ so that E is relative open $E = \mathring{E}_1 \times E_2$, where E_2 is a single point in $\mathbb{R}^{d-d_1}$. Moreover, the simplest situation is when $E = \mathbb{R}^{d_1} \times E_2$ and g is locally Lipschitz in $\mathbb{R}^d$ with a linear growth.

Given a locally Lipschitz continuous forward flow $\mathtt{X} = \{\mathtt{X}(x,t) : t \geq 0,\, x \in E\}$ with (closed!) active boundary $\partial^0 E$ then

$$P(t)f(x) := f(\mathtt{X}(x, t \wedge \mathtt{T}(x))), \quad \forall t \geq 0,\ x \in E,$$

defines a *semigroup* on the $C_*(E)$, real-valued bounded continuous functions on E vanishing on $\partial^0 E$. If f is a C^1 function then

$$P(t)f(x) - f(x) = \int_0^t P(s)\mathtt{G}f(x)\mathrm{d}s = \int_0^t \mathtt{G}P(s)f(x)\mathrm{d}s,$$

for every $t \geq 0$ and x in E. Thus the extended infinitesimal generator of the semigroup $\{P(t) : t \geq 0\}$, denoted by $\bar{\mathtt{G}}$ has the domain $\mathcal{D}(\bar{\mathtt{G}})$ which are all real-valued measurable functions f on E vanishing on $\partial^0 E$ such that the function $t \mapsto P(t)f(x)$ is absolutely continuous on $[0,T]$ for every positive real number $T \leq \mathtt{T}(x)$ and for every x in E. In this case, the expression $P(s)\mathtt{G}f(x)$ is only defined almost every where in s relative to the Lebesgue measure. The function f is only continuous along the flow, not necessarily continuous in other directions.

In same situations, it is important to single out the time variable so that weaker assumptions can be made. For instance, we may call $\mathtt{X} = \{\mathtt{X}(s,x,t) : t \geq s \geq 0,\, x \in E\}$ a nonhomogeneous locally Lipschitz forward flow associated with the time-variant vector fields $\mathtt{G} = \{\mathtt{G}(t) : t \geq 0\}$ the solution of the IVP $\dot{x}(t) = g(x(t),t)$, $x(s) = x$. Typical assumptions are (1) the function $x \mapsto g(x,t)$ is locally Lipschitz continuous with linear growth, uniformly in t, i.e., for any compact subset K of $\bar{E}$ and any $T > 0$ there exists a constant M such that $|g(x,t) - g(x',t)| \leq M|x - x'|$ for every x, x' in K, and t in $[0,T]$, and if E is unbounded then there exist a constant C such that $|g(x,t)| \leq C(1 + |x|)$, for any x in $\bar{E}$ and $t \geq 0$, and the function $t \mapsto g(x,t)$ is (Borel) measurable. Under this assumptions, the IVP has only absolutely continuous solutions and the flow have the *composition* property $\mathtt{X}_{r,t} \circ \mathtt{X}_{s,r} = \mathtt{X}_{s,t}$ or equivalent $\mathtt{X}(s,x,t) = \mathtt{X}(s, \mathtt{X}(r,x,t), r)$, for every $t \geq r \geq s$. Differentiating with respect to the initial data, we deduce the well-known relations between the flow and the vector field, namely $\partial_s \mathtt{X}(s,x,t) = -\mathtt{G}(s)\mathtt{X}(s,x,t)$, i.e.,

$$\partial_s \mathtt{X}_i(s,x,t) = -g_j(x,s)\, \partial_j \mathtt{X}_i(s,x,t), \quad \forall t \geq s \geq 0,\ x \in E,$$

and $\partial_t \mathtt{X}(s,x,t) = -\mathtt{G}^*(t)\mathtt{X}(s,x,t)$, i.e.,

$$\partial_t \mathtt{X}_i(s,x,t) = \partial_j \big[g_j(x,t)\, \mathtt{X}_i(s,x,t)\big], \quad \forall t \geq s \geq 0,\ x \in E,$$

which are the deterministic equivalent of Kolmogorov backward and forward equations. Also, we refer the reader to the book by Ladde and Lakshmikantham [146] for a complete treatment of ODE with random coefficients.

3.9.2 Definition of PDP

Again, this is a generalization of the Poisson process in the direction of strong Markov processes as in a queueing system. Essentially, this stochastic process has deterministic evolutions between two consecutive jumps, instead of being constant as a Poisson process. The deterministic dynamic are characterized by a family vector fields, namely $g(n,x)$ defined for every x in $E \subset \mathbb{R}^d$ and each n in a countable set N. The jump mechanism has a jump rate $\lambda(n,x)$ and a transition distribution $Q(n,x,\cdot)$. Note that the jump rate determines *when to jump*, so a vanishing jump rate (i.e., $\lambda = 0$) means not jump at that particular position, while an infinite jump rate (i.e., $\lambda = +\infty$) translates into a instantaneous jump. The transition distribution rate $Q(n,x,\cdot)$ determines *where to jump*, also called transition probability measures of jumps.

Let $\mathtt{X} = \{\mathtt{X}(n,x,t) : t \geq 0,\ x \in E,\ n = 0,1,\ldots\}$ be a family (indexed by n) of locally Lipschitz forward flows associated to vector fields $g(n,\cdot)$ on $E \subset \mathbb{R}^d$ or to first-order differential operator $\mathtt{G}(n,\cdot)$ on $C^1(\mathbb{R}^d)$, with active boundary $\partial_n^0 E$ (which may depend on n) and first exit from E time $\mathtt{T}(n,x)$ (or first hitting time to $\partial^0 E_n$), see Definition 3.18. Occasionally, we may use $\mathtt{X}_n(x,t) := \mathtt{X}(n,x,t)$ to emphasize the countable index n. Note that the dimension d is fixed and generally large, since $g(n,\cdot)$ may have several zero components which change with n. Clearly, the case where only finitely many n are used may be defined as a *module* operation in the variable n. Moreover, it may be useful to allow the set E to depend on n, i.e., $E_n \subset \mathbb{R}^{d_n}$, but we chose to disregard this case for now.

There are two type of jumps: (1) *interior jumps*, which are produced while inside region $E \setminus \partial_n^0 E$ and (2) *boundary jumps*, which are produced while on the active boundary $\partial_n^0 E$. Note that a point x belongs to the active boundary $\partial_n^0 E$ if and only if $\mathtt{T}(n,x) = 0$. The forward flow $\mathtt{X}(n,x,t)$ is defined for t in $[0, \mathtt{T}(n,x)]$ and for every (n,x) the backward flow $\mathtt{X}(n,x,-t)$ belongs to the inside $E \setminus \partial_n^0 E$ for t sufficiently small. Also, the functions $x \to \mathtt{T}(n,x)$ and $(x,t) \to \mathtt{X}(n,x,t)$ are continuous for every n. The interior jumps have the same exogenous origin as the one produced in the Poisson process, but the boundary jumps are forced or imposed by the continuous dynamic of the forward flow $\mathtt{X}$.

The state space of this piecewise deterministic process is $N \times E$, where N is a countable set (possible finite). Sufficient conditions on g to construct a locally Lipschitz forward flow have been discussed in the previous section, and the assumption of the jump are the following:

(1) the map $(n,x) \mapsto \lambda(n,x)$ is a Borel measurable function from $N \times E$ into $[0,\infty]$ and for each (n,x) there exists $\varepsilon = \varepsilon(n,x) > 0$ such that $t \mapsto \lambda(n, \mathtt{X}(n,x,t))$ is integrable in the interval $[0,\varepsilon)$,

(2) for each B in $\mathcal{B}(N \times E)$ the map $(n,x) \mapsto Q(n,x,B)$ is measurable, and

for each (n,x) in $N \times E$ the map $B \mapsto Q(n,x,B)$ is a probability measure on $N \times E$ satisfying $Q(n,x,\{(n,x)\}) = 0$.

Note that the integrability condition of (1) ensure that after any jumps, we do have a continuous evolution following the forward flow for a positive time, while the last condition of (2) ensure a state discontinuity at every jump time. Piecewise deterministic processes viewed as Markov processes have state in $N \times E$, a discrete (piecewise constant in t) component $\{n(t) : t \geq 0\}$ in N and a continuous (piecewise continuous in t) component $\{x(t) : t \geq 0\}$ in $E \subset \mathbb{R}^d$. These conditions are mainly necessary to make sense to the jump mechanism, however, we need another condition to forbid the accumulation of boundary jumps.

Perhaps, the most typical situation in *hybrid system modelling*, including most of the queueing systems, is the case of finitely many n, i.e., the state space is $\{0,1,2,\dots,K\} \times E$, and the locally Lipschitz forward flows $\{\mathtt{X}_n(x,t) : t \geq 0,\ x \in E\}$ are indexed by $n = 0,1,\dots,N$. The particular case when with only one n, namely $n = 0$, is essentially different from the others, the discrete component is useless, and we may work directly on E. This is, we do have a Markov process in E which generalize the Poisson process, without adding a discrete component. Even in this simple situation, we do not have a Feller process. The active boundary introduces instantaneous predictable jumps, producing a *deterministic* discontinuity. Thus, unless there is not active boundary, a piecewise deterministic process is not a Feller process, but we do have a strong Markov process. According to the definition and assumptions on the locally Lipschitz forward flow $\mathtt{X}_n$ in Definition 3.18, the active boundary $\partial_n^0 E$ is closed and contains all reachable points from the inside $E \setminus \partial_n^0 E$, i.e., defining

$$\begin{cases} \mathtt{T}(n,x) := \inf\{t > 0 : \mathtt{X}(n,x,t) \notin E\} = \\ \qquad\qquad = \inf\{t > 0 : \mathtt{X}(n,x,t) \in \partial_n^0 E\}, \end{cases} \tag{3.52}$$

which is defined for any x in E, with the convention that $\mathtt{T}(n,x) = \infty$ if $\mathtt{X}(n,x,t)$ belongs to E for every $t > 0$. The sets $\{x : \mathtt{T}(n,x) < \infty\}$ and $\{x : \mathtt{T}(n,x) = \infty\}$ are disjointed (one of them may be empty), and that the function $x \mapsto \tau(n,x)$ is continuous on $\{x : \mathtt{T}(n,x) < \infty\}$. To effectively rule out the accumulation of boundary jumps, see Davis [56, Proposition 24.6, pp 60–61], we may assume λ is bounded and that there exists $\varepsilon > 0$ such that

$$Q(n,x,\{(n,x) : \mathtt{T}(n,x) \geq \varepsilon\}) = 1, \ \ \forall\,(n,x),\ x \in \partial_n^0 E, \tag{3.53}$$

which include the particular case of an empty active boundary, i.e.,

$$\mathtt{T}(n,x) = \infty, \ \ \forall\,(n,x) \in N \times E.$$

Let us discuss a realization (or construction) of piecewise deterministic processes are described above, similarly to Section 3.8, but to emphasize the two components (discrete and continuous), we use the notation (n,x) instead of x. Thus the Markov process with sample path $(n(t),x(t) : t \geq 0)$, starting form a fixed initial point $(n(0),x(0))$, is realized in the canonical Hilbert cube $(\Omega,\mathcal{F},P)$,

where $(U_1, U_2, \ldots)$ is sequence of independent random variables each having uniform distribution on $[0,1]$. We suppose given the characteristics g, λ and Q on the state space $N \times E$ which determine the flow $\mathtt{X}$, $\mathtt{T}$ as in Definition 3.18 satisfying (3.52). The jump mechanism satisfies the conditions (1) and (2) mentioned above and assumption (3.53).

First we define

$$\Psi(n,x,t) := \mathbb{1}_{t<\mathtt{T}(n,x)} \exp\Big[-\int_0^t \lambda(n, \mathtt{X}(n,x,s))\mathrm{d}s\Big], \tag{3.54}$$

the survivor function of jumps times and its (generalized) inverse

$$\Psi^{-1}(n,x,u) := \inf\{t \geq 0 : \Psi(n,x,t) \leq u\}, \tag{3.55}$$

with the convention that $\Psi^{-1}(n,x,u) := +\infty$ if $\Psi(n,x,t) > u$ for every $t \geq 0$. Note that $P\{\Psi^{-1}(n,x,U_k) > t\} = \Psi(n,x,t)$, for any k. Secondly, as mentioned above, we are working in the canonical Hilbert cube and there exists a measurable function satisfying

$$\begin{cases} \Upsilon : N \times E \times [0,1] \longrightarrow \{(n,x) : \mathtt{T}(n,x) \geq \varepsilon\} \subset N \times E, \\ \ell(\{u : \Upsilon(n,x,u) \in B\}) = Q(n,x,B), \quad \forall\, B \in \mathcal{B}(N \times E), \end{cases} \tag{3.56}$$

where ε is as in (3.53) and ℓ is the Lebesgue measure on $[0,1]$.

Now the sample path $\big(n(t,\omega), x(t,\omega) : t \geq 0\big)$ is defined by induction as follows. Given θ_{k-1}, n_{k-1} and x_{k-1}, with $k = 1, 2, \ldots$, we set

$$\theta_k(\omega) := \theta_{k-1}(\omega) + \Psi^{-1}(n_{k-1}(\omega), x_{k-1}(\omega), U_{2k-1}(\omega)), \tag{3.57}$$

and if $\theta_{k-1}(\omega) \leq t < \theta_k(\omega)$ then

$$n(t,\omega) := n_{k-1}(\omega), \qquad x(t,\omega) := \mathtt{X}(n_{k-1}(\omega), x_{k-1}(\omega), t), \tag{3.58}$$

and if $\theta_k(\omega) < \infty$ then

$$(n_k, x_k) := \Upsilon\big(n_{k-1}(\omega), x(\theta_k(\omega)-,\omega), U_{2k}(\omega)\big), \tag{3.59}$$

where $x(s-,\omega)$ is the left-hand limit at a time s. If $\theta_{k-1}(\omega) = \theta_k(\omega)$ then we skip (3.58), define $x(\theta_k(\omega)-,\omega) = x_{k-1}(\omega)$ and go to (3.59). Therefore, if $\theta_k(\omega) = \infty$ then we have define the sample path for every time $t \geq \theta_{k-1}$, otherwise we have define the sample path in the stochastic interval $[\![\theta_{k-1}, \theta_k[\![$ as well as θ_k, n_k and x_k, and we can iterate (3.57),...,(3.59) with the initial $\theta_0 = 0$.

To actually see that the sample path is defined for every time, we will show that

$$\lim_{k\to\infty} \theta_k = \infty, \quad \text{a.s..} \tag{3.60}$$

Indeed, define the *counting jump process* $N_t := k$ if $\theta_{k-1} \leq t < \theta_k$ with $N_0 := 0$ and $k \geq 1$. By means of (3.54) and the fact that λ is bounded ($\lambda \leq c$), we can construct a Poisson process $(N_t^p : t \geq 0)$ with rate c, that dominate the counting process $(N_t : t \geq 0)$, i.e, $N_t \leq N_t^p$. Thus $\mathbb{E}\{N_t\} \leq \mathbb{E}\{N_t^p\} = ct$, which implies the condition (3.60).

We will only consider the class of piecewise deterministic processes satisfying the condition (3.60) is satisfied, or even more if $\mathbb{E}\{N_t\} < \infty$ for every , besides the assumptions made on the vector fields $\mathtt{X}_n$, $\mathtt{T}_n$ and (1) and (2) above on the jump mechanism.

Since $(n(t) : t \geq 0)$ is piecewise constant, $(x(t) : t \geq 0)$ is piecewise continuous and both are right-continuous, we take pass to the canonical sample space $\mathcal{D}([0,\infty), N \times E)$ and define the integer-valued measure process

$$\rho(t, B, \omega) := \sum_{k=1}^{\infty} \mathbb{1}_{\theta_k(\omega) \leq t} \mathbb{1}_{x_k(\omega) \in B}, \quad \forall\, t \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d) \tag{3.61}$$

associated with the piecewise deterministic process constructed above. When necessary, we may write $\rho_{nx}(t, \omega, B)$ to indicate dependency on the initial state $n_0 = n$, $x_0 = x$ at time $\theta_0 = 0$.

It is proved in Davis [56, Sections 25, 26, Theorems 24.3, 25.5, 26.14] that the filtration (history) generated by the piecewise deterministic processes is right-continuous, that $\big(n(t,\omega), x(t,\omega) : t \geq 0\big)$ is an homogeneous strong Markov process in the canonical sample space $\mathcal{D}([0,\infty), N \times E)$. Also, the extended infinitesimal generator has complete description as follows.

The expression of the extended infinitesimal generator

$$\begin{cases} \bar{A} f(n,x) := \mathtt{G}_n f(n,x) + \mathtt{I} f(n,x), \\ \mathtt{G}_n h(x) := \displaystyle\sum_{i=1}^{d} g_i(n,x) \partial_i h(x), \\ \mathtt{I} f(n,x) := \lambda(n,x) \displaystyle\int_{N \times E} [f(\eta,\xi) - f(n,x)]\, Q(n,x,\mathrm{d}\eta \times \mathrm{d}\xi) \end{cases} \tag{3.62}$$

where the first-order differential operator $\mathtt{G}_n$ is acting only on the continuous variable x, while the integral operator $\mathtt{I}$ may involves both variable n and x.

To full describe the extended domain $\mathcal{D}(\bar{A})$ we need the following concept. Now we say that a process $\{h(n,x,t,\omega) : n \in N,\ x \in \mathbb{R}^d,\ t \geq 0,\}$ belongs to $L^1(\rho)$, with $\rho = \rho(t,\omega,B)$ given by (3.61) if

$$\mathbb{E}\Big\{ \int_{N \times E \times \mathbb{R}^+} h \mathrm{d}\rho \Big\} := \mathbb{E}\Big\{ \sum_{i=1}^{\infty} h(n_k, x_k, \theta_k, \cdot) \Big\} < \infty. \tag{3.63}$$

Similarly, h belongs to $L^1_{\mathrm{loc}}(\rho)$, if there exists a sequence $\{\sigma_k : k \geq 0\}$ of stopping times with $\sigma_n \to \infty$ almost surely such that

$$\mathbb{E}\Big\{ \sum_{i=1}^{\infty} h(n_k, x_k, \theta_k \wedge \sigma_n, \cdot) \Big\} < \infty, \quad \forall\, n,$$

i.e., $h(n,x,t,\cdot)\, \mathbb{1}_{t < \sigma_n}$ belongs to $L^1(\rho)$.

A measurable function $f \colon N \times E \to \mathbb{R}$ belongs to $\mathcal{D}(\bar{A})$ if and only if several conditions are met:

(1) we have $f(n, \mathtt{X}(n,x,-t)) \to f(n,x)$ as $t \to 0$ and

$$f(n,x) = \int_{N\times E} f(\eta,\xi)\, Q(n,x,\mathrm{d}\eta \times \mathrm{d}\xi)$$

for every n in N, x in E such that $\mathtt{T}(n,x) = 0$,

(2) the function $t \to f(n, \mathtt{X}(n,x,t))$ is absolutely continuous on $[0, \mathtt{T}(n,x) \wedge T[$, for every $T > 0$, n in N and x in E,

(3) for every n in N and x in E, the process

$$h(n,x,t) := f(n,x) - f(n(t-,\omega), x(t-,\omega)), \quad \forall\, t > 0,$$

with $h(n,x,0) = 0$, belongs to $L^1_{\mathrm{loc}}(\rho_{nx})$.

Property (1) is called *boundary condition* since $\mathtt{T}(n,x) = 0$ if and only if x belongs to the active boundary $\partial^0_n E$, which is mainly related to the discrete variable n. Condition (2) involves only the continuous variable x and provided a weak sense to the differential operator $\mathtt{G}f(n(t),x(t))$ as the derivative (almost every t, the discrete variable n and ω are regarded as parameters) of the function $t \to f(n, \mathtt{X}(n,x,t))$. Property (3) can be re-written as

$$\mathbb{E}\Big\{ \sum_{k=1}^{n} |f(n_k,x_k) - f(n_{k-1}, \mathtt{X}(n_{k-1},x_{k-1},\theta_k)| \mathbb{1}_{\theta_k} \le n \Big\} < \infty,$$

for every $n \ge 1$, by taken $\sigma_n := \theta_n \wedge n$, which is certainly verified if f is bounded.

To complete this discussion let as mention that when $\mathtt{T}(n,x) = \infty$ for every (n,x) and the jump rate $\lambda(n,x)$ is bounded then a piecewise deterministic process is a Feller process. In general, a piecewise deterministic process is a Borel right process. This means that (a) the state space $N \times E$ is topological homeomorphic to a Borel subset of a compact metric space, (b) the semigroup $P(t)h := \mathbb{E}\{h(n(t),x(t)\}$ maps the bounded Borel functions into itself, (c) the sample paths $t \to (n(t),x(t))$ are right continuous almost surely, (d) if f is an α-excessive function for $\{P(t) : t \ge 0\}$ then the function $t \to f(n(t),x(t))$ is right continuous almost surely. Recall that a non-negative function f is called α-excessive (with $\alpha \ge 0$) if $\exp(-t\alpha)P(t)f \le f$ for every $t \ge 0$ and $\exp(-t\alpha)P(t)f \to f$ as $t \to 0$.

As mentioned early, a comprehensive study on piecewise deterministic process can be found in Davis [56].

3.10 Lévy Processes

Random walks capture most of the relevant features found in sequences of random variables while Lévy processes can be thought are their equivalent in continuous times, i.e., they are stochastic processes with independent and stationary increments. The best well known examples are the Poisson process and the Brownian motion. They form the class of space-time homogeneous Markov processes and they are the prototypes of semi-martingales.

Definition 3.19. A $\mathbb{R}^d$-valued or d-dimensional *Lévy process* is a random variable X in a complete probability space $(\Omega, \mathcal{F}, P)$ with values in the canonical $D([0,\infty), \mathbb{R}^d)$ such that

(1) for any $n \geq 1$ and $0 \leq t_0 < t_1 < \cdots < t_n$ the $\mathbb{R}^d$-valued random variables $X(t_0)$, $X(t_1) - X(t_2)$,...,$X(t_n) - X(t_{n-1})$ are independent (i.e., independent increments),

(2) for any $s > 0$ the $\mathbb{R}^d$-valued random variables $X(t)-X(0)$ and $X(t+s)-X(s)$ have the same distributions (i.e., stationary increments),

(3) for any $s \geq 0$ and $\varepsilon > 0$ we have $P(|X(t) - X(s)| \geq \varepsilon) \to 0$ as $t \to s$ (i.e., stochastically continuous) and

(4) $P(X(0) = 0) = 1$.

An *additive process* is defined by means of the same properties except that condition (2) on stationary increments is removed. □

Usually the fact that the paths of a Lévy process are almost surely cad-lag is deduced from conditions (1),...,(4) after a modification of the given process. However, we prefer to impose a priori the cad-lag regularity. It is clear that under conditions (2) (stationary increments) and (4) we may replace condition (3) (on stochastically continuous paths) by condition $P(|X(t)| \geq \varepsilon) \to 0$ as $t \to 0$, for every $\varepsilon > 0$.

A classic tool to analyze distributions in $\mathbb{R}^d$ is *characteristic functions* (or Fourier transform). Thus, for a given distribution μ of a random variable ξ in $\mathbb{R}^d$, the characteristic function $\hat{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ is defined by

$$\hat{\mu}(y) := \int_{\mathbb{R}^d} \mathrm{e}^{\mathtt{i}\, x\cdot y}\, \mu(\mathrm{d}x) = \mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y\cdot \xi}\}.$$

Several properties relating characteristic functions are known, e.g., if $\hat{\mu}$ is integrable in $\mathbb{R}^d$ then μ is absolutely continuous with respect to the Lebesgue measure, with a bounded continuous density g given by the inversion formula

$$g(x) := (2\pi)^{-d} \int_{\mathbb{R}^d} \mathrm{e}^{-\mathtt{i}\, x\cdot y}\, \hat{\mu}(y)\mathrm{d}y.$$

In particular $y \mapsto \hat{\mu}(y)$ is uniformly continuous from $\mathbb{R}^d$ into the complex plane $\mathbb{C}$, $\hat{\mu}(0) = 1$ and $\hat{\mu}$ is *positive definite*, i.e., for any $k = 1, 2, \ldots$, $\zeta_1, \ldots, \zeta_k$ in $\mathbb{C}$ and $x_1, \ldots, x_k$ in $\mathbb{R}^d$ we have $\sum_{i,j=1}^k \hat{\mu}(x_i - x_j)\zeta_i \bar{\zeta}_j \geq 0$. Moreover, Bochner Theorem tell us that the converse, i.e., any complex-valued continuous function φ in R^d with $\varphi(0) = 1$ and positive definite is the characteristic function of a distribution, e.g., Da Prato and Zabczyk [51, Theorem I.2.3, pp. 48–52] for a proof valid in separable Hilbert spaces.

If μ_1 and μ_2 are the distributions of two $\mathbb{R}^d$-valued independent random variables ξ_1 and ξ_2 then the *convolution* $\mu_1 \star \mu_2$ defined by

$$(\mu_1 \star \mu_2)(B) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_B(x + y)\, \mu_1(\mathrm{d}x)\, \mu_2(\mathrm{d}y), \quad \forall B \in \mathcal{B}(\mathbb{R}^d)$$

is the distribution of the sum $\xi_1 + \xi_2$. We have $\widehat{\mu_1 \star \mu_2} = \hat\mu_1\,\hat\mu_2$, and therefore, the characteristic functions of independence of random variables is product of characteristic function of each variable.

If X is a Lévy process then we may consider the characteristic function of the R^d-valued random variable $X(1)$, i.e.,

$$\hat\mu(y) := \mathbb{E}\{\mathrm{e}^{\mathtt{i}\,y\cdot X(1)}\}.$$

Since $X(1) = X(1/n) + [X(2/n) - X(1/n)] + \cdots + [X(1) - X(1-1/n)]$, the random variable $X(1)$ can be expressed as the sum of n independent identically distributed random variables. Therefore, μ is the n-fold convolution of some distribution μ_n, i.e., $\mu = \mu_n^{n\star}$, μ_n is the distribution of $X(1/n)$. A distribution μ with the above property is called *infinitely divisible*. For instance, Gaussian, Cauchy and Dirac-δ distributions on $\mathbb{R}^d$, as well as Poisson, exponential and Γ distributions on $\mathbb{R}$, are infinitely divisible, for instance see Stroock [238, Section 3.2, pp. 139–153].

Any infinitely divisible distribution μ has a never vanishing characteristic function $\hat\mu$ which can be expressed as an exponential function, i.e.,

$$\hat\mu(y) = \exp[-\phi(y)], \quad \forall y \in \mathbb{R}^d,$$

where ϕ is uniquely determined as a complex-valued continuous function in $\mathbb{R}^d$ with $\phi(0) = 1$, which is called *characteristic exponent* or the Lévy symbol. Thus, we have $\mathbb{E}\{\mathrm{e}^{\mathtt{i}\,y\cdot X(t)}\} = \exp[-t\phi(y)]$ for t rational and by continuity for any $t \geq 0$. Since the Fourier transform is one-to-one, the expression

$$\widehat{\mu^{\star t}}(y) := \exp[-t\phi(y)], \quad \forall y \in \mathbb{R}^d,\ t > 0,$$

define the $^\star t$-convolution. Moreover, $\mu^{\star t}$ is also an infinitely divisible distribution.

A key result is Lévy-Khintchine formula states that a complex-valued function ϕ is the characteristic exponent of an infinitely divisible distributions μ if and only if

$$\phi(y) = \mathtt{i}\,g\cdot y + \frac{1}{2}\,Qy\cdot y + \int_{\mathbb{R}^d_*} \big[1 - \mathrm{e}^{\mathtt{i}\,y\cdot x} + \mathtt{i}\,y\cdot x\,\mathbb{1}_{|x|<1}\big]\,m(\mathrm{d}x),$$

for every y in $\mathbb{R}^d$, where g belongs to $\mathbb{R}^d$, Q is a non-negative semi-definite $d\times d$-matrix and m is a Radon measure on $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$ which integrates the function $x \mapsto |x|^2 \wedge 1$. The representation of ϕ by (g, Q, m) is unique. However, the cut-off function $\mathbb{1}_{|x|<1}$ may be replaced by a bounded smooth function which is equal to 1 at the origin, e.g. $(1+|x|^2)^{-1}$. In this case, the parameter g changes and we have for every y in $\mathbb{R}^d$,

$$\phi(y) = \mathtt{i}\,f\cdot y + \frac{1}{2}\,Qy\cdot y + \int_{\mathbb{R}^d_*} \big[1 - \mathrm{e}^{\mathtt{i}\,y\cdot x} + \mathtt{i}\,\frac{y\cdot x}{1+|x|^2}\big]\,m(\mathrm{d}x),$$
$$f = g + \int_{\mathbb{R}^d} x\big[\frac{1}{1+|x|^2} - \mathbb{1}_{|x|<1}\big]\,m(\mathrm{d}x).$$

We may also use $\sin x$ as in Krylov [142, Section 5.2, pp. 137–144], for the one-dimensional case.

An important class of Lévy processes are the so-called (compound) Poisson processes. A Lévy process X is called a *Poisson process* with parameter $c > 0$, if $X(t)$ has a Poisson distribution with mean $c\,t$, for every $t \geq 0$. Similarly, a Lévy process X is called a *compound Poisson* process with parameters (c, γ), where $c > 0$ and γ is a distribution in $\mathbb{R}^d$ with $\gamma(\{0\}) = 0$ (i.e., γ is a distribution in $\mathbb{R}^d_*$), if $\mathbb{E}\{e^{i\,y\cdot X(t)}\} = \exp[-t\,c(\hat{\gamma}(y) - 1)]$, for any $t \geq 0$ and y in $\mathbb{R}^d$. The parameters (c, γ) are uniquely determined by X and a simple construction is given as follows. If $\{\zeta_n : n = 1, 2, \dots\}$ is a sequence of independent identically distributed (with distribution law γ) random variables, and $\{\tau_n : n = 1, 2, \dots\}$ is another sequence of independent exponentially distributed (with parameter c) random variables, with $\{\zeta_n : n = 1, 2, \dots\}$ independent of $\{\tau_n : n = 1, 2, \dots\}$, then for $\theta_n := \tau_1 + \tau_2 + \cdots + \tau_n$ (which has a Gamma distribution with parameters γ and n), the expressions

$$
\begin{aligned}
&X(t) := \sum_{n=1}^{\infty} \zeta_n \mathbb{1}_{t \geq \theta_n}, \quad \text{with} \quad \delta X(t) := X(t) - X(t-) \\
&\delta X(\theta_n) = \zeta_n, \quad \text{and} \quad \delta X(t) = 0 \ \text{ if } t \neq \theta_n, \ \forall n, \ \text{or equivalently} \\
&X(t) := \zeta_1 + \zeta_2 + \cdots + \zeta_n \ \text{ if } \ \sum_{i=1}^{n} \tau_i = \theta_n \leq t < \theta_{n+1} = \sum_{i=1}^{n+1} \tau_i,
\end{aligned}
$$

are realizations of a compound Poisson process and its associate point (or jump) process. Indeed, for any integer k, any $0 \leq t_0 < t_1 < \cdots < t_k$ and any Borel subsets $B_0, B_1, \dots, B_k$ of $\mathbb{R}^d$ we can calculate the finite-dimensional distributions of X by the formula

$$
\begin{aligned}
P(X(t_0) \in B_0,\, X(t_1) - X(t_0) \in B_1, \dots, X(t_k) - X(t_{k-1}) \in B_k) = \\
= P\big(X(t_0) \in B_0\big)\, P\big(X(t_1) - X(t_0) \in B_1\big) \dots \\
\dots P\big(X(t_k) - X(t_{k-1}) \in B_k\big).
\end{aligned}
$$

This yields the expression

$$
\mathbb{E}\{e^{i\,y\cdot X(t)}\} = \exp[-t\,c\,(1 - \hat{\gamma}(y))], \quad \forall y \in \mathbb{R}^d, t \geq 0,
$$

which is continuous in t. Then, all conditions in Definition 3.19, including the stochastic continuity of path (3), are satisfied. Note that for a pairwise disjoint family of Borel sets of the form $]s_i, t_i] \times B_i$, with $0 \leq s_i < t_i$, B_i in $\mathcal{B}(\mathbb{R}^d)$, $i = 1, 2, \dots, k$ the integer-valued random variables

$$
\nu(]s_i, t_i] \times B_i) := \sum_{n=1}^{\infty} \mathbb{1}_{s_i < \theta_n \leq t_i}\, \mathbb{1}_{\zeta_n \in B_i}, \quad \forall i = 1, 2, \dots, k
$$

are independent identically Poisson distributed, with parameter (or mean) $c\,(t_i - s_i)\gamma(B_i)$.

An interesting point is the fact that a compound Poisson process in $\mathbb{R}$, with parameters (c, σ) such that $c > 0$ and σ is a distribution in $(0, \infty)$, is increasing in t and its Laplace transform is given by

$$\mathbb{E}\{\mathrm{e}^{-\xi X(t)}\} = \exp\Big[- t\, c \int_{(0,\infty)} (\mathrm{e}^{-\xi x} - 1)\, \sigma(\mathrm{d}x)\Big], \quad \forall \xi \in \mathbb{R}, t \geq 0.$$

These processes are called *subordinator* and are used to model random *time changes*, possible discontinuous. Moreover, the Lévy measure m of any Lévy process with increasing path satisfies

$$\int_{\mathbb{R}^1_*} |x|\, m(\mathrm{d}x) = \int_0^\infty x\, m(\mathrm{d}x) < \infty,$$

e.g., see books Bertoin [21, Chapter III, pp. 71-102], Itô [112, Section 1.11] and Sato [220, Chapter 6, pp. 197-236].

Another interesting case is the so-called *symmetric* Lévy processes where the characteristic exponent (also called Lévy exponent) $\phi(y)$ (defined early) satisfies

$$\int_{\mathbb{R}^d} \Big|\frac{1}{1+\phi(y)}\Big|\, \mathrm{d}y < \infty,$$

which implies that $\phi(y)$ is a positive real-valued even function. Moreover, the only possible case occurs when the dimension $d = 1$, and actually, $\phi(y)$ takes the form

$$\phi(y) = \frac{1}{2} Q y^2 + 2 \int_{\mathbb{R}^d_*} \big[1 - \cos(xy)\big] m(\mathrm{d}x), \quad \forall y \in \mathbb{R}$$

for some nonnegative constant Q and some measure m on $\mathbb{R}$ which integrates the function $(1 \wedge x^2)$. In this one-dimensional case, the resolvent (measure)

$$\int_{\mathbb{R}} f(y)\, R(\lambda, \mathrm{d}y) = \int_0^\infty \mathrm{e}^{-\lambda t} \mathbb{E}\big\{ f\big(X(t) + x\big)\big\} \mathrm{d}t,$$

has the density

$$r(\lambda, y) = \frac{1}{\pi} \int_0^\infty \frac{\cos(xy)}{\lambda + \phi(y)}\, \mathrm{d}t, \quad \forall y \in \mathbb{R},$$

while the transition densities are given by

$$p(t, x) = \frac{1}{\pi} \int_{\mathbb{R}} \mathrm{e}^{-xy}\, \mathrm{e}^{-t\phi(y)} \mathrm{d}y, \quad \forall t > 0,\, x \in \mathbb{R},$$

see the book by Marcus and Rosen [166, Section 4.1, pp. 135-144] for details. The interested reader, may consult the book by Applebaum [6], which discuss Lévy process at a very accessible level.

The next typical class Lévy processes is the Wiener processes or Brownian motions. A Lévy process X is called a *Brownian motion* or *Wiener* process in $\mathbb{R}^d$, with (vector) drift b in $\mathbb{R}^d$ and (matrix) co-variance σ^2, a nonnegative-definite $d \times d$ matrix, if $\mathbb{E}\{\mathrm{e}^{y \cdot X(t)}\} = \exp\left[-t(|\sigma y|^2/2 - i\, b)\right]$, for any $t \geq 0$ and y

in $\mathbb{R}^d$, i.e., if $X(t)$ has a Gaussian distribution with (vector) mean $\mathbb{E}\{X(t)\} = bt$ and (matrix) co-variance $\mathbb{E}\{(X(t) - bt)^*(X(t) - bt)\} = t\sigma^2$. A *standard* Wiener process is when $b = 0$ and $\sigma^2 = 1$, the identity matrix. The construction of a Wiener process is a somehow technical and usually details are given for the standard Wiener process with t in a bounded interval. The general case is an appropriate transformation of this special case. First, let $\{\xi_n : n = 1, 2, \dots\}$ be a sequence of independent identically normally distributed (i.e., Gaussian with zero-mean and co-variance 1) random variables in $\mathbb{R}^d$ and let $\{e_n : n = 1, 2, \dots\}$ be a complete orthonormal sequence in $L^2(]0, \pi[)$, e.g., $e_n(t) = \sqrt{2/\pi}\cos(nt)$. Define

$$X(t) := \sum_{n=1}^{\infty} \xi_n \int_0^t e_n(s)\mathrm{d}s, \quad t \in [0, \pi].$$

It is not hard to show that X satisfies all conditions of a Wiener process, except for the stochastic continuity and the cad-lag sample property of paths. Next, essentially based on the (analytic) estimate: for any constants $\alpha, \beta > 0$ there exists a positive constant $C = C(\alpha, \beta)$ such that

$$|X(t) - X(s)|^\alpha \le C\,|t - s|^\beta \int_0^\pi \mathrm{d}t \int_0^\pi |X(t) - X(s)|^\alpha\,|t - s|^{-\beta-2}\mathrm{d}s,$$

for every t, s in $[0, \pi]$, we may establish that that series defining the process X converges uniformly in $[0, \pi]$ almost surely. Indeed, if X_k denotes the k partial sum defining the process X then an explicit calculations show that

$$\mathbb{E}\{|X_k(t) - X_\ell(s)|^4\} = \mathbb{E}\Big\{\Big|\sum_{n=\ell+1}^{k} \xi_n \int_s^t e_n(r)\mathrm{d}r\Big|^4\Big\} \le 3|t - s|^2,$$

for every $t \ge s \ge 0$ and $k > \ell \ge 1$. After using the previous estimate with $\alpha = 4$ and $1 < \beta < 2$ we get

$$\mathbb{E}\{\sup_{|t-s|\le\delta} |X_k(t) - X_\ell(s)|^4\} \le C\,\delta^\beta, \quad \forall \delta > 0,\ k > \ell \ge 1,$$

for a some constant $C > 0$. This proves that X is a Wiener process with continuous paths. Next, the transformation $t\,X(1/t)$ (or patching k independent copies, i.e., $X_k(t)$ if $(k-1)\pi \le t < k\pi$, for $k \ge 1$.) produces a standard Wiener process in $[0, \infty)$ and the process $b\,t + \sigma\,X(t)$ yields a Wiener process with parameters b and σ.

The above estimate is valid even when t is multidimensional and a proof can be found in Da Prato and Zabczyk [52, Theorem B.1.5, pp. 311–316]. For more details on the construct arguments, see, e.g., Friedman [90] or Krylov [141].

We are ready to state the general existence result

Theorem 3.20 (construction)**.** *Let m be a Radon measure on $\mathbb{R}^d_*$ such that*

$$\int_{\mathbb{R}^d_*} |x|^2 \wedge 1\, m(\mathrm{d}x) < \infty,$$

Q be a nonnegative-definite $d \times d$ matrix and g be a vector in $\mathbb{R}^d$. Then there exists a unique probability measure P on the canonical probability space $\Omega = D([0,\infty),\mathbb{R}^d)$ such that the canonical process $(X(t) := \omega(t) : t \geq 0)$ is a Lévy process with characteristic *(g, Q, m), i.e.,*

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y\cdot X(t)}\} = \exp[-t\,\phi(y)], \quad \forall y \in \mathbb{R}^d,\ t \geq 0, \qquad \textit{with}$$
$$\phi(y) := \mathtt{i}\, g\cdot y + \frac{1}{2}\, Qy\cdot y + \int_{\mathbb{R}^d_*} \left[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y\cdot x \mathbb{1}_{|x|<1}\right] m(\mathrm{d}x).$$

Conversely, given a Lévy process X the characteristic (g, Q, m) are uniquely determined through the above formula.

Proof. Only some details are given. First, consider the case where $Q = 0$, which corresponds to *Poisson measures* and *point processes*, a step further from the compound Poisson processes. Essentially, a point process is the jumps process constructed from a cad-lag process. Poisson measures are particular case of integer-valued measures, which are the distribution of the jumps of cad-lag processes. More extensive comments are given in Section 4.1.3, here we recall a couple of arguments used to construct a Poisson measure. Let m be a Radon measure in $\mathbb{R}^d_*$ (which integrates the function $|x|^2 \wedge 1$ is used later) and write $m = \sum_k m_k$, where $m_k(B) := m(B \cap R_k)$, $\mathbb{R}^d_* = \cup_k R_k$, $m(R_k) < \infty$ and $R_k \cap R_\ell = \emptyset$ if $k \neq \ell$. To each m_k we may associate a compound Poisson and point processes

$$Y_k(t) := \sum_{n=1}^{\infty} \zeta_{n,k} \mathbb{1}_{t\geq\theta_{n,k}} \qquad \text{or} \qquad Y_k(t) = Z_{n,k} \ \text{ if } \ \theta_{n-1,k} < t \leq \theta_{n,k},$$
$$\delta Y_k(t) := Y_k(t) - Y_k(t-) = \zeta_{n,k}\mathbb{1}_{t=\theta_{n,k}}, \quad \forall t \geq 0,$$

where $\theta_{n,k} := \tau_{1,n,k} + \tau_{2,n,k} + \cdots + \tau_{n,n,k}$, $\{\tau_{i,n,k} : i = 1,\dots,n,\ n = 1,2,\dots\}$ is a sequence of independent exponentially distributed (with parameter $m(R_k) = c_k$) random variables, and $Z_{n,k} := \zeta_{1,k} + \zeta_{2,k} + \cdots + \zeta_{n,k}$, $\{\zeta_{n,k} : n = 1,2,\dots\}$ is another sequence of independent identically distributed (with distribution law m_k/c_k) random variables, the family $\{\tau_{i,n,k},\ \zeta_{n,k} : i = 1,\dots,n,\ n,k \geq 1\}$ is independent. Since the processes $\{Y_k : k \geq 1\}$ are independent, the characteristic function of the point process $Y := \sum_k Y_k$ is the product of those of Y_k, which should reconstruct the measure m. The independence property and the diffuse character (non atoms) of the exponential distribution ensure that there are no simultaneous jumps among the $\{Y_k : k \geq 1\}$. Hence, the jump process $\delta Y = \sum_k \delta Y_k$ is indeed a Poisson point process with characteristic measure m, i.e.,

$$\nu(]s,t] \times B) := \sum_{n,k=1}^{\infty} \mathbb{1}_{s<\theta_{n,k}\leq t}\, \mathbb{1}_{\zeta_{n,k}\in B}, \quad \forall t > s \geq 0,\ B \in \mathcal{B}(\mathbb{R}^d_*)$$

is a Poisson random measure with intensity measure $\mathbb{E}\{\nu(]s,t] \times B)\} = (t - s)m(B)$. In general, we cannot re-order the jumps in a increasing manner as those of a compound Poisson process. Next, some *martingale* estimates are

necessary to establish good behavior of the process Y. First the exponential formula, for any complex-valued Borel function f on $\mathbb{R}^d_*$

$$\text{if} \quad \int_{R^d_*} |1 - \mathrm{e}^{f(x)}|\, m(\mathrm{d}x) < \infty \quad \text{then} \quad \forall t \geq 0 \quad \text{we have}$$

$$\mathbb{E}\Big\{ \exp\Big[\sum_{0\leq s\leq t} f(\delta Y(s))\Big]\Big\} = \exp\Big\{ - t \int_{R^d_*} [1 - \mathrm{e}^{f(x)}]\, m(\mathrm{d}x)\Big\}.$$

Secondly, the Doob's maximal inequality for the compensated jumps

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T} \Big| \sum_{0\leq s\leq t} f(\delta Y(s)) - t \int_{R^d_*} f(x)\, m(\mathrm{d}x)\Big|^2\Big\} \leq$$
$$\leq 4T \int_{R^d_*} |f(x)|^2\, m(\mathrm{d}x),$$

valid for any real valued Borel function f on $\mathbb{R}^d_*$ and any $T > 0$, see Sections 3.3 and 3.4 for more detail and references.

After *compensate* the small jumps, this Poisson measure ν and its associated jump process δY yield a Lévy process with characteristic $(0, 0, m)$. Indeed, define

$$X_1(t) := \int_{]0,t]\times\{|x|\geq 1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x) = \sum_{s\leq t} \delta Y(s)\, \mathbb{1}_{|\delta Y(s)|\geq 1},$$

and

$$X_2^\varepsilon(t) := \int_{]0,t]\times\{\varepsilon\leq|x|<1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x) -$$
$$-\mathbb{E}\Big\{ \int_{]0,t]\times\{\varepsilon\leq|x|<1\}} x\, \nu(\mathrm{d}s \times \mathrm{d}x)\Big\} =$$
$$= \sum_{s\leq t} \delta Y(s)\, \mathbb{1}_{\varepsilon<|\delta Y(s)|\leq 1} - t \int_{\varepsilon\leq|x|<1} x\, m(\mathrm{d}x),$$

for $t, \varepsilon > 0$, which are two compound Poisson processes with characteristic exponents

$$\phi_1(y) = \int_{|x|\geq 1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x}\big]\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

$$\phi_2^\varepsilon(y) = \int_{\varepsilon\leq|x|<1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y\cdot x\big]\, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

respectively, after using the above exponential formula. In view of the martingale inequality

$$\mathbb{E}\big\{ \sup_{t\leq T} |X_2^\varepsilon(t) - X_2^\delta(t)|^2\big\} \leq 4T \int_{\delta\leq|x|<\varepsilon} |x|^2\, m(\mathrm{d}x),$$

for every $T > 0$, $\varepsilon > \delta > 0$, and because the intensity measure m integrates the function $|x|^2 \wedge 1$, the family of processes $\{X_2^\varepsilon : \varepsilon > 0\}$ converges to a process X_2 and

$$\mathbb{E}\big\{ \sup_{t\le T} |X_2^\varepsilon(t) - X_2(t)|^2 \big\} \le 4T \int_{|x|<\varepsilon} |x|^2 \, m(\mathrm{d}x),$$

for every $T > 0$ and $\varepsilon > \delta > 0$. This cad-lag process X_2 has stationary independent increments, so a Lévy process with characteristic exponent

$$\phi_2(y) = \int_{|x|<1} \big[1 - \mathrm{e}^{\mathtt{i}\, y\cdot x} + \mathtt{i}\, y \cdot x\big] \, m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d.$$

Therefore, $X_1 + X_2$ is a Lévy process with characteristic $(0, 0, m)$.

Now to conclude, consider given a drift g in $\mathbb{R}^d$ and a co-variance Q (non-negative-definite $d \times d$ matrix). If B is a standard Wiener process independent of (X_1, X_2) (i.e., of the previous construction) then define $X_3 := (\sqrt{Q}\, B(t) - g\, t : t \ge 0)$, which is a Lévy process with characteristic exponent

$$\phi_3(y) = \mathtt{i}\, g \cdot x + \frac{1}{2}\, Qy \cdot y, \quad \forall y \in \mathbb{R}^d.$$

Finally, $X := X_1 + X_2 + X_3$ is a Lévy process with the desired characteristic (g, Q, m). The converse follows from Lévy-Khintchine formula for infinitely divisible distributions.

An important point to remark is that the above construction shows that any Lévy process is a Wiener process plus the limit of a sequence of compound Poisson processes. Also note that any infinitely divisible probability measure on $\mathbb{R}^d$ can be viewed as the distribution of a Lévy process evaluated at time 1. □

It is perhaps relevant to remark that even if any Lévy process can be expressed as a limit of compound Poisson processes, the structure of a typical graph of Levy process eludes us. For instance, almost surely, the jumping times $J = \{t : X(t,\omega) \ne X(t-,\omega)\}$ are countable, and (a) if the Levy measure satisfies $m(\mathbb{R}^d) = \infty$ then J is dense in $[0, \infty)$ while (b) if $m(\mathbb{R}^d) < \infty$ then J can be written as an increasing sequence $\{\tau_k : k \ge 1\}$, $\tau_k \le \tau_{k+1}$, of independent random variables having exponential distributions with mean $1/m(\mathbb{R}^d)$, see Sato [220, Theorem 21.3, pp. 136–137].

Note that for a given Lévy process X with the characteristic (g, Q, m) we can define $\delta X(t) := X(t) - X(t-)$ and the integer-valued (random) measure ν_X associated with the jumps δX of X, and its martingale measure $\mu_X := \nu_X - m$, since $m = \nu_X^p$ is the predictable jump compensator (which is actually deterministic). To make sense to discontinuous (purely jumps) part $X_d(t)$ of X, which is the *compensated* sum of all jumps $\sum_{s\le t} \delta X(s)$, and therefore define its continuous part as $X_c = X - X_d$ we proceed essentially as above. For $t, \varepsilon > 0$, consider

$$X_1(t) := \int_{]0,t]\times\{|x|\ge 1\}} x \, \nu_X(\mathrm{d}s \times \mathrm{d}x) = \sum_{s\le t} \delta X(s)\, \mathbb{1}_{|\delta X(s)|\ge 1},$$

and

$$X_2^\varepsilon(t) := \int_{]0,t]\times\{\varepsilon\le|x|<1\}} x\,\mu_X(\mathrm{d}s\times\mathrm{d}x) = \\ = \sum_{s\le t}\delta X(s)\,\mathbb{1}_{\varepsilon<|\delta X(s)|\le 1} - t\int_{\varepsilon\le|x|<1} x\,m(\mathrm{d}x)$$

are compound Poisson process, and using the canonical semi-martingale decomposition, the limit process $X_d := X_1 + \lim_{\varepsilon\to 0} X_2^\varepsilon$ is a Lévy process with characteristic exponent

$$\phi_\delta(y) = \int_{\mathbb{R}^d_*}\left[1 - \mathrm{e}^{\mathtt{i}\,y\cdot x} + i\,y\cdot x\right]\mathbb{1}_{|x|<1}\,m(\mathrm{d}x), \quad \forall y\in\mathbb{R}^d,$$

which is called the discontinuous (purely jumps) part of X. If the jumps are not of bounded variation, the series $\sum_{s\le t}\delta X(s)$ is meaningless, unless it is compensated with the *small jumps* (we used jumps greater than 1, but it suffices greater than some positive constant). Sometimes we *ignore* large jumps (by assuming that the Lévy measure m integrates z at infinite), and so the Lévy process corresponding to the characteristic exponent

$$\phi_d(y) = \int_{\mathbb{R}^d_*}\left[1 - \mathrm{e}^{\mathtt{i}\,y\cdot x} + i\,y\cdot x\right]m(\mathrm{d}x), \quad \forall y\in\mathbb{R}^d.$$

is uniquely determined could be used as X_d.

Since X is quasi-left continuous, we have $\delta X = \delta X_d$. Due to the independence of increments, $X(t) - X(t-)$ results independent of $X(t-) = X_c(t)$, i.e., the processes X_c and X_d are independent and the characteristic exponent of X_c must be

$$\phi_c(y) = \mathtt{i}\,g\cdot x + \frac{1}{2}\,Qy\cdot y, \quad \forall y\in\mathbb{R}^d.$$

Thus X_c is a Wiener process and the characteristic (g, Q, m) can also be found as

$$g := -\mathbb{E}\{X_c(1)\}, \qquad Q := \mathbb{E}\{[X_c(1)+g]^*[X_c(1)+g]\}, \\ m(B) := \mathbb{E}\Big\{\sum_{0<t\le 1}\mathbb{1}_{\delta X(t)\in B}\Big\}, \quad \forall B\in\mathcal{B}(\mathbb{R}^d),$$

where $\delta X(t) := X(t) - X(t-)$, $X_c := X - X_d$ and $(\cdot)^*$ is the transpose operator.

On the other hand, if A is a $n\times d$ matrix then $AX := (AX(t) : t\ge 0)$ is a n dimensional Lévy process with characteristic

$$g = Ag + \int_{\mathbb{R}^d_*} Ax[\mathbb{1}_{|Ax|<1} - \mathbb{1}_{|x|<1}]\,m(\mathrm{d}x), \\ Q_A = AQA^*, \qquad m_A = mA^{-1},$$

where $mA^{-1}(B) = m(\{x : Ax\in B\})$.

If $Q = 0$ and the Lévy measure m integrates the function $|x| \wedge 1$ then the characteristic exponent may be re-written in a simpler way, as

$$\phi(y) = -\mathtt{i}\, g \cdot x + \int_{\mathbb{R}^d_*} \left[1 - \mathrm{e}^{\mathtt{i}\, y \cdot x}\right] m(\mathrm{d}x), \quad \forall y \in \mathbb{R}^d,$$

where g is now referred to as the *drift coefficient*. In this case, the Lévy process has locally bounded variation, not necessarily integrable, unless m integrates the function $|x|$. Certainly, if X has a finite Lévy measure m on $\mathbb{R}^d_*$ then X is a compound Poisson process plus a drift.

Now we take a look at the resolvent operators associated with Lévy processes. Let $\{P(t) : t \geq 0\}$ be the semigroup associated with a Lévy process, i.e.,

$$P(t) : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d), \qquad P(t)f(x) := \mathbb{E}\{f(X(t) + x)\}, \tag{3.64}$$

where $C_0(\mathbb{R}^d)$ is the Banach space of continuous functions vanishing at infinity. Then, the family $\{R(\lambda) : \lambda > 0\}$ of linear and bounded operators from $C_0(\mathbb{R}^d)$ into itself and the family of $\{R(\lambda, dy) : \lambda > 0\}$ of finite measures on $\mathbb{R}^d$, defined by

$$\begin{cases} R(\lambda)f(x) := \displaystyle\int_0^\infty \mathrm{e}^{-\lambda\, t}\, P(t)f(x)\mathrm{d}t, \\ \displaystyle\int_{R^d} f(y)R(\lambda, dy) := \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda\, t}\, f(X(t))\mathrm{d}t\Big\}, \end{cases} \tag{3.65}$$

which satisfies

$$R(\lambda)f(x) = \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda\, t}\, f(X(t) + x)\mathrm{d}t\Big\}$$
$$\int_{R^d} f(y)R(\lambda, dy) = R(\lambda)f(0),$$

are called the *resolvent operators* and the *resolvent kernel* associated with the Lévy process X. It is also clear that $R(\lambda)$ is a convolution operator, i.e.,

$$\begin{aligned} &\text{if} && \int_{R^d} f(y)\check{R}(\lambda, dy) := \mathbb{E}\Big\{ \int_0^\infty \mathrm{e}^{-\lambda\, t}\, f(-X(t))\mathrm{d}t\Big\} \\ &\text{then} && (\check{R}(\lambda, \cdot) \star f)(x) := \int_{R^d} f(x - y)\check{R}(\lambda, dy) = R(\lambda)f(x). \end{aligned}$$

The resolvent operators describe the distribution of the Lévy process evaluated at independent exponential times, i.e., if $\tau = \tau(\lambda)$ is an independent (of X) random variable having an exponential law with parameter $\lambda > 0$, then $\mathbb{E}\{f(X(\tau) + x)\} = \lambda\, R(\lambda)f(x)$.

The semigroup property yields the identity

$$\begin{cases} R(\lambda) - R(\mu) = (\mu - \lambda)R(\lambda)R(\mu), \quad \forall \lambda, \mu > 0 \quad \text{or} \\ \check{R}(\lambda, \cdot) - \check{R}(\mu, \cdot) = (\mu - \lambda)\check{R}(\lambda, \cdot) \star \check{R}(\mu, \cdot) \end{cases} \tag{3.66}$$

so-called *resolvent equation*. Thus the image of $C_0(\mathbb{R}^d)$ under $R(\lambda)$, denoted by $\mathcal{D}$, does not depend on $\lambda > 0$. Since

$$\lambda R(\lambda)f(x) - f(x) = \int_0^\infty \mathrm{e}^{-s}[P(s/\lambda)f(x) - f(x)]\mathrm{d}s$$

we deduce that $\lambda R(\lambda)f \to f$ in $C_0(\mathbb{R}^d)$ as $\lambda \to \infty$. Therefore, $\mathcal{D}$ is a dense subspace of $C_0(\mathbb{R}^d)$. Moreover, if $R(\lambda)f = R(\lambda)g$ for some $\lambda > 0$, the resolvent equation shows that $R(\lambda)f = R(\lambda)g$ for any $\lambda > 0$ and then, as $\lambda \to \infty$ we deduce $f = g$, i.e., $R(\lambda)$ is a one-to-one mapping from $C_0(\mathbb{R}^d)$ onto $\mathcal{D}$. The infinitesimal generator A from $\mathcal{D}$ into $C_0(\mathbb{R}^d)$ is defined by the relation

$$R(\lambda)(\lambda I - A) = I \quad \text{or equivalently} \quad A := \lambda I - [R(\lambda)]^{-1}, \tag{3.67}$$

where I is the identity mapping, and $\mathcal{D} = \mathcal{D}_A$ is called the *domain* of the infinitesimal generator A.

The Fourier transform for f in $L^1(\mathbb{R}^d)$, namely

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\,x\cdot\xi} f(x)\mathrm{d}x, \quad \forall \xi \in \mathbb{R}^d,$$

yields simple expressions of these operators in term of the characteristic exponent

$$\phi(\xi) := -\ln(\mathbb{E}\{\mathrm{e}^{\mathrm{i}\,X(1)\cdot\xi}\})$$

of the Lévy process X. We have for any f in $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$

$$\begin{cases} \widehat{P(t)f}(\xi) = \mathrm{e}^{-t\phi(-\xi)}\widehat{f}(\xi), \quad \forall t \geq 0,\ \xi \in \mathbb{R}^d, \\ \widehat{R(\lambda)f}(\xi) = [\lambda + \phi(-\xi)]^{-1}\widehat{f}(\xi), \quad \forall \lambda > 0,\ \xi \in \mathbb{R}^d, \end{cases} \tag{3.68}$$

and, for any f in $\mathcal{D}_A$ such that Af belongs to $L^1(\mathbb{R}^d)$

$$\widehat{Af}(\xi) = -\phi(-\xi)\widehat{f}(\xi), \quad \forall \xi \in \mathbb{R}^d. \tag{3.69}$$

Hence, Lévy-Khintchine formula and the inversion of Fourier transform yield the following expression for the infinitesimal generator of Lévy processes

$$\begin{cases} Af(x) := -g \cdot \nabla f + \dfrac{1}{2}\nabla \cdot Q\nabla f + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_*} \left[f(\cdot + y) - f - \mathbb{1}_{|y|<1}\, y \cdot \nabla f\right] m(dy), \end{cases} \tag{3.70}$$

for any smooth function f, e.g., twice-continuously differentiable and bounded function f.

Let us mention a result from Blumental and Getoor [28]. If the resolvent kernel $R(\lambda, x + dy)$ is absolutely continuous (with respect to the Lebesgue measure) for some $\lambda > 0$ and some x in R^d then it is absolutely continuous for every $\lambda > 0$ and every x in R^d. Moreover this is equivalent to the so-called *strong Feller property* of the resolvent operators, namely, for any $\lambda > 0$ and f in $L^\infty(R^d)$ the function $x \mapsto R(\lambda)f(x)$ is continuous.

To conclude this section we briefly discuss the so-called *local time* associated with a Levy process, full details can be found in the book Bertoin [21, Chapter V, pp. 125–154]. Let $X(t)$ a one dimensional Lévy process, then its characteristic function is given by

$$\mathbb{E}\{\exp(\mathtt{i}\xi X(t))\} = e^{-t\psi(\xi)}$$

where ψ is characterized by

$$\psi(\xi) = -\mathtt{i}\gamma\xi + \tfrac{1}{2}\sigma^2\xi^2 - \int_{\mathbb{R}} \Big(e^{\mathtt{i}\xi y} - 1 - \mathtt{i}\xi y \mathbb{1}_{\{|y|<1\}}\Big)\mathrm{d}m(y)$$

where $\mathrm{d}m(y)$ is the Lévy measure. For instance, the choice $\gamma = 0$, $\sigma = 1$ and $\mathrm{d}m = 0$ yields the Brownian motion, while $\gamma = 0$ and $\sigma = 0$ produces a pure jump processes. In particular, $m = \delta_1$ corresponds to the Poisson process and for $\mathrm{d}m(y) = \frac{1}{|y|^{1+\alpha}}\mathrm{d}y$, with $0 < \alpha < 2$, we get the so-called *α-stable Lévy processes* with $\psi(\xi) = \frac{2}{c}|\xi|^\alpha$, where

$$|\xi|^\alpha = -\frac{c}{2}\int_{\mathbb{R}} \Big(e^{\mathtt{i}\xi y} - 1 - \mathtt{i}\xi y \mathbb{1}_{\{|y|<1\}}\Big)m(\mathrm{d}y), \qquad \frac{1}{c} = \int_0^\infty \frac{1-\cos s}{s^{1+\alpha}}\mathrm{d}s,$$

i.e., $c = \frac{2}{\pi}\Gamma(1+\alpha)\sin(\frac{\pi\alpha}{2})$.

For any $t > 0$, the *occupation measure* $\mu(t, \mathrm{d}x)$ on the time interval $[0, t]$ of the Lévy process X is defined as

$$\mu(t, B, \omega) := \int_0^t \mathbb{1}_{\{X(s,\omega)\in B\}}\mathrm{d}s,$$

for every Borel subset of $\mathbb{R}$.

Theorem 3.21 (occupation measure)**.** *For any $t \geq 0$, the occupation measure $\mu(t, \mathrm{d}x)$ is absolutely continuous with respect to the Lebesgue measure with a density in $L^2(\mathrm{d}y \times P)$ if and only if*

$$\int_{\mathbb{R}} \Re\Big\{\frac{1}{1+\psi(y)}\Big\}\mathrm{d}y < \infty. \tag{3.71}$$

Moreover, if the above condition fails, then $\mu(t, \mathrm{d}x)$ is singular with respect to the Lebesgue measure for any $t > 0$ and with probability 1. □

Brownian motions and α-stable Lévy processes with $1 < \alpha < 2$ satisfy condition (3.71) and therefore theirs occupation measures have densities with respect to the Lebesgue measure. While Poisson processes and α-stable Lévy processes with $0 < \alpha \leq 1$ do not satisfy (3.71).

Now, if condition (3.71) is satisfied then for every $t > 0$ and any x in $\mathbb{R}$ we can define

$$\ell(t, x) = \limsup_{\epsilon\to 0} \tfrac{1}{2\epsilon}\int_0^t \mathbb{1}_{\{|X(s)-x|<\epsilon\}}\mathrm{d}s,$$

which is referred to as the *local time* at the level x and time t for the Lévy process X. It is clear that $\{\ell(t, x) : x \in \mathbb{R}\}$ serves as a $\mathcal{F}(t)$-measurable version

of the density of $\mu(t, \mathrm{d}x)$. Note that for every x, the process $\ell(\cdot, x)$ is (cad-lag) nondecreasing, which may increase only when $X = x$. Thus, ℓ is jointly measurable.

To end this section, let us take a look at the path-regularity of the Lévy processes. If we drop the cad-lag condition in the Definition 3.19 then we use the previous expressions (for either Lévy or additive processes *in law*) to show that there exits a cad-lag version, see Sato [220, Theorem 11.5, p. 65], which is actually indistinguishable if the initial Lévy or additive process was a separable process.

Proposition 3.22. *Let y be an additive process in law on a (non-necessarily completed) probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{F}_t^0(y)$ denote the σ-algebra generated by the random variables $\{y(s) : 0 \le s \le t\}$. Define $\mathcal{F}_t(y) = \mathcal{F}_t^0(y) \vee \mathcal{N}$, the minimal σ-algebra containing both $\mathcal{F}_t^0(y)$ and $\mathcal{N}$, where $\mathcal{N} = \{N \in \mathcal{F} : P(N) = 0\}$. Then $\mathcal{F}_t(y) = \cap_{s>t}\mathcal{F}_s(y)$, for any $t \ge 0$.*

Proof. Set $\mathcal{F}_t^+(y) = \cap_{s>t}\mathcal{F}_s(y)$ and $\mathcal{F}_\infty^0(y) = \vee_{t\ge 0}\mathcal{F}_t^0(y)$. Since both σ-algebras contain all null sets in $\mathcal{F}$, we should prove that $\mathbb{E}(Z \mid \mathcal{F}_t^+(y)) = \mathbb{E}(Z \mid \mathcal{F}_t(y))$ for any $\mathcal{F}_\infty^0(y)$-measurable bounded random variable Z, to get the right-continuity of the filtration. Actually, it suffices to establish that

$$\mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t^+(y)\} = \mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t(y)\}$$

for any choice of $0 \le s_1 \le s_2 \le \ldots \le s_n$, $(r_1, r_2, \ldots, r_n)$, and n. Moreover, only the case $s_1 > t$ need to be considered. To this purpose, we use the characteristic function $f_t(r) = \mathbb{E}\{e^{\mathrm{i}ry(t)}\}$ which satisfies $f_{t+s}(r) = f_t(r)f_s(r)$, and the martingale property of $M_t(r) = e^{\mathrm{i}ry(t)}/f_t(r)$ with respect to $\mathcal{F}_t(y)$.

Now, let $s_1 > t' \ge t$ and consider

$$\begin{aligned}
\mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_{t'}(y)\} &= f_{s_n}(r_n)\mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^{n-1} r_j y(s_j)} M_{s_n}(r_n) \mid \mathcal{F}_{t'}(y)\} = \\
&= f_{s_n}(r_n)\mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^{n-1} r_j y(s_j)} M_{s_{n-1}}(r_n) \mid \mathcal{F}_{t'}(y)\} = \\
&= f_{s_n - s_{n-1}}(r_n) f_{s_{n-1}}(r_{n-1} + r_n) \times \\
&\qquad \times \mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^{n-2} r_j y(s_j)} M_{s_{n-1}}(r_{n-1} + r_n) \mid \mathcal{F}_{t'}(y)\}) = \\
&= \ldots = f_{s_n - s_{n-1}}(r_n) f_{s_{n-1} - s_{n-2}}(r_{n-1} + r_n) \times \\
&\qquad \times f_{s_{n-2} - s_{n-3}}(r_{n-2} + r_{n-1} + r_n) \times \\
&\qquad \times \ldots \times f_{s_2 - s_1}(r_2 + \cdots + r_{n-2} + r_{n-1} + r_n) e^{\mathrm{i}r_1 y(s_1)},
\end{aligned}$$

i.e., we have

$$\mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_{t+\varepsilon}(y)\} = \mathbb{E}\{e^{\mathrm{i}\sum_{j=1}^n r_j y(s_j)} \mid \mathcal{F}_t(y)\}, \quad \forall \varepsilon > 0.$$

and the proof is finished by passing to the limit as $\varepsilon \to 0$. □

• *Remark* 3.23. Sometimes, an adapted process y (not necessarily cad-lag) is called *additive* with respect to a filtration $\mathbb{F}$ (non necessarily right-continuous or complete) if the random variable $y(s) - y(t)$ is independent of $\mathcal{F}(t)$, for any $s > t \ge 0$. Because y is adapted and $\mathcal{F}(t)$ increasing, this is equivalent to a stronger condition, namely, the σ-algebra $\mathcal{G}(t)$ generated by $\{y(s_2) - y(s_1) :$

$s_2 > s_1 \geq t\}$ is independent of $\mathcal{F}(t)$ for any $t \geq 0$. Now, let $\mathcal{N}$ be the σ-algebra of all null sets in $\mathcal{F}$ and set $\mathcal{F}(t+) = \cap_{\varepsilon>0}\mathcal{F}(t+\varepsilon)$. If y is right-continuous in probability then we want show that $\mathbb{E}\{\cdot \,|\, \mathcal{F}(t+)\} = \mathbb{E}\{\cdot \,|\, \mathcal{F}(t)\}$. Indeed, for any t there is a sequence $\{t_n\}$, $t_n > t$ convergent to t and a set of measure null such that $y(t_n,\omega) \to y(t,\omega)$, for every ω in $\Omega \setminus N$. Since $y(s) - y(t_n)$, $s > t$, is independent of $\mathcal{F}(t_n) \supset \mathcal{F}(t+)$, we have

$$\mathbb{E}\big\{f\big(y(s) - y(t_n)\big)\mathbb{1}_F\big\} = \mathbb{E}\big\{f\big(y(s) - y(t_n)\big)\big\}\,\mathbb{E}\{1_F\}, \quad \forall F \in \mathcal{F}(t+),$$

for every continuous function f. Hence, $y(s)-y(t)$, $s > t$ is independent of $\mathcal{F}(t+)$, i.e., $\mathcal{G}(t) \vee \mathcal{N}$ is independent of $\mathcal{F}(t+)$, for every $t \geq 0$. Now, if A is in $\mathcal{F}(t)$ and B in $\mathcal{G}(t) \vee \mathcal{N}$ then the $\mathcal{F}(t)$-measurable random variable $\mathbb{1}_A\, P(B)$ is a version of the conditional expectation $\mathbb{E}\{\mathbb{1}_A \mathbb{1}_B \,|\, \mathcal{F}(t+)\}$, and a class monotone argument shows that for any bounded and $\mathcal{F}(t) \vee \mathcal{G}(t) \vee \mathcal{N}$-measurable random variable h we have a $\mathcal{F}(t)$-measurable version of the $\mathbb{E}\{h \,|\, \mathcal{F}(t+)\}$. This proves that $\mathcal{F}(t+) = \mathcal{F}(t) \vee \mathcal{N}$, i.e., another way of proving the previous Proposition 3.22. This proof is inspired by Letta [152], based on a personal communication. □

The reader is referred to the books by Bremaud [32], Elliott [73], Protter [206]), and the comprehensive works by Bertoin [21, Chapters O and I, pp. 1–42] and Sato [220, Chapters 1 and 2, pp. 1–68].

3.11 Transition Functions

Now we focus on the transition functions of spatially homogeneous Markov processes or additive processes. There are several aspects of a Markov Process, depending on the specific emphasis given to the discussion, one of the following elements is first studied and then other elements are derived. A Markov process with valued in $\mathbb{R}^d$ may be presented as

(a) a family of $\mathbb{R}^d$-valued stochastic processes $X = X_{sx}$ indexed by the initial distribution $X(s) = x$, $s \geq 0$,

(b) a probability transition function $P(s,x,t,A)$ with $t > s \geq 0$, $x \in \mathbb{R}^d$ and A a Borel subset of $\mathbb{R}^d$,

(c) a family of linear and bounded evolution operators $\Phi(t,s)$ from $B(\mathbb{R}^d)$, the Banach space of bounded Borel real-valued function on $\mathbb{R}^d$ into itself, indexed by $t \geq s \geq 0$,

(d) a family of linear and bounded operators $R(\lambda)$ from $B(\mathbb{R}^d)$ into itself, indexed by $\lambda > 0$,

(e) a family of linear possible unbounded (infinitesimal generator) operators $A(t)$ defined in a subspace $\mathcal{D}(A(t))$ of $B(\mathbb{R}^d)$ into $B(\mathbb{R}^d)$, indexed by $t \geq 0$.

Certainly, each of these (a),. . . ,(e) elements should satisfy some specific conditions to yield a Markov process.

The elements $R(\lambda)$ in (d) are called *resolvent* operators and are mainly used with *time-homogeneous* Markov processes, i.e., when (a) $X_{sx} = X_{0x}$ for any $s > 0$ or (b) $P(s,x,t,A) = P(0,x,t-s,A)$ for any $t > s \geq 0$ or (c) the

evolution operators $\Phi(t,s) = \Phi(t-s)$ for any $t > s \geq 0$ or (e) $A(t) = A$ for any $t \geq 0$. It is clear that by adding a new dimension to $\mathbb{R}^d$ we may always assume we are in the time-homogeneous, however, in most of the cases, we prefer to live the special time variable t with its preferential role and to work with non-time-homogeneous Markov processes. It is possible to use a Polish (separable complete metric space) $\mathcal{O}$ instead of the Euclidean space $\mathbb{R}^d$, usually $\mathcal{O}$ is locally compact since the infinite-dimensional case needs some special care.

The principle stating that *the future is independent of the past given the present* is called *Markov property* and formally is written as

$$P\{X(t) \in B \mid X(r),\, r \leq s\} = P\{X(t) \in B \mid X(s)\}, \tag{3.72}$$

for every $t > s \geq 0$ and $B \in \mathcal{B}(\mathbb{R}^d)$, which should be satisfied by the family of processes. This same property viewed by the transition function is called the *Chapman-Kolmogorov identity*,

$$P(s,x,t,B) = \int_{\mathbb{R}^d} P(s,x,r,\mathrm{d}y)\, P(r,y,t,B), \tag{3.73}$$

for every $t > r > s$, x in $\mathbb{R}^d$ and B in $\mathcal{B}(\mathbb{R}^d)$. For the evolution operators this is called the *semigroup property* are written as

$$\Phi(t,s) = \Phi(t,r)\Phi(r,s) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall t > r > s > 0, \tag{3.74}$$

and in the case of time-homogeneous Markov processes, the resolvent operators satisfy the so-called *resolvent equation*, namely

$$R(\lambda) - R(\nu) = (\nu - \lambda)R(\lambda)R(\nu) \ \text{ in } \ B(\mathbb{R}^d), \quad \forall \lambda, \nu > 0. \tag{3.75}$$

The resolvent $\{R(\lambda) : \lambda > 0\}$ is mainly used in potential theory, the semi-group $\{\Phi(t) : t \geq 0\}$ and the infinitesimal generator A are well know in analysis, while the family of stochastic processes X and the transition function $P(s,x,t,B)$ are more probabilistic tools. At this general level, we ramark that the Markov property (3.72) is almost surely satisfied, i.e., only *version* of the stochastic processes are involved and therefore a property on the *sample path* should be added. The evolution and resolvent operators are defined on $B(\mathbb{R}^d)$, which is a non-separable Banach space, so that in general the theory is very delicate.

Out interest is in Markov-Feller or Feller-Dynkin processes, instead of the large space $B(\mathbb{R}^d)$ we use the separable Banach space $C_0(\mathbb{R}^d)$, of all continuous functions vanishing at infinity (i.e., for any $\varepsilon > 0$ there exists a compact subset K of $\mathbb{R}^d$ such that $|\varphi(x)| \leq \varepsilon$ for every x in $\mathbb{R}^d \setminus K$). Thus, after a one-point compactification method, we are reduced to $C(\bar{\mathbb{R}}^d)$, with $\bar{\mathbb{R}}^d = \mathbb{R}^d \cup \{\infty\}$ being a compact Polish space. For the family of stochastic processes X_x, this yields a *cad-lag condition* on the sample path. Regarding the Chapman-Kolmogorov identity (3.73) we have

Definition 3.24 (transition function)**.** A *(Markov) transition function* on the Borel space $(\mathbb{R}^d, \mathcal{B})$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$, is a function $P(s,x,t,B)$ defined for $t > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ such that

(a) for each $t > s \geq 0$ and x in $\mathbb{R}^d$ the function $B \mapsto P(s, x, t, B)$ is a positive measure on $(\mathbb{R}^d, \mathcal{B})$, with $P(s, x, t, \mathbb{R}^d) \leq 1$,

(b) for each $t > 0$ and B in $\mathcal{B}$ the function $(s, x) \mapsto P(s, x, t, B)$ is a measurable,

(c) for any $s \geq 0$, for any compact subset K of $\mathbb{R}^d$ and any $\varepsilon > 0$ we have

$$\lim_{t \to s} \sup_{x \in K} \left[1 - P(s, x, t, \{y \in \mathbb{R}^d : |y - x| \leq \varepsilon\})\right] = 0,$$

so-called uniformly stochastic continuous,

(d) for each $t > r > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ we have

$$P(s, x, t, B) = \int_{\mathbb{R}^d} P(s, x, r, \mathrm{d}y)\, P(r, y, t, B),$$

i.e., Chapman-Kolmogorov identity.

These properties can be rephrased in term of linear non-negative operators from $B(\mathbb{R}^d)$, the space of real-valued bounded and Borel functions on $\mathbb{R}^d$, into itself, defined by

$$P(t, s)\varphi(x) := \int_{\mathbb{R}^d} \varphi(y) P(s, x, t, \mathrm{d}y) = P(s, x, t, \varphi), \tag{3.76}$$

for every $t > s \geq 0$ and x in $\mathbb{R}^d$, which satisfies

(a') for each $t > s \geq 0$ and φ in $B(\mathbb{R}^d)$ with $0 \leq \varphi \leq 1$ we have $0 \leq P(t, s)\varphi \leq 1$,

(b') for each $t > s \geq 0$ and x in $\mathbb{R}^d$ the mapping $B \mapsto P(t, s)\mathbb{1}_B(x)$ is σ-additive on $\mathcal{B}(\mathbb{R}^d)$,

(c') for any $s \geq 0$ and φ in $C_0(\mathbb{R}^d)$, continuous functions on $\mathbb{R}^d$ vanishing at infinity, we have

$$\lim_{t \to s} P(t, s)\varphi(x) = \varphi(x), \quad \forall x \in \mathbb{R}^d,$$

i.e., the *stochastic continuity property* , a weaker version of *(c)*,

(d') for each $t > r > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$ we have

$$P(t, s) = P(t, r)\, P(r, s), \quad \text{in} \quad B(\mathbb{R}^d),$$

usually referred to as the *semigroup property*, and the transition function is called a *Feller transition* if the following condition (e) , so-called *Feller property*, is satisfied

(e) for each $t > s \geq 0$ and φ in $C_0(\mathbb{R}^d)$ we have $P(t, s)\varphi$ in $C_0(\mathbb{R}^d)$, i.e., $P(t, s)$ can be considered as acting on $C_0(\mathbb{R}^d)$.

It is called *time-homogeneous* if $P(s, x, t, B) = P(0, x, t - s, B)$ and *spatially-homogeneous* if $P(s, x, t, B) = P(s, 0, t, B - x)$, for any $t > s \geq 0$, x in $\mathbb{R}^d$ and B in $\mathcal{B}$. It is called a transition *probability* function if $P(s, x, t, \mathbb{R}^d) = 1$, for any $t > s \geq 0$ and x in R^d. □

Certainly, to define a transition function we only need a measurable space $(E, \mathcal{E})$ and t belonging to some set T with a complete order, instead of the Euclidean space $\mathbb{R}^d$ and the real semi-line $[0, \infty)$. However, for time-homogeneous transition function, essentially we need the semi-line $[0, \infty)$ and for the spatially-homogeneous transition function E has to be a vector space, e.g., $\mathbb{R}^d$.

Condition (b') is satisfied when E is locally compact, i.e., $\mathbb{R}^d$, but it is mentioned above as a difficulty when considering the infinite-dimensional case. Instead of the transition function in the form $P(s, x, t, B)$ we may look at the family of linear non-negative operators $P(t, s)$ from $C_0(\mathbb{R}^d)$ into itself as a two-parameter C_0-semigroup, which satisfies $0 \le P(t, s)\varphi \le 1$ for any $0 \le \varphi \le 1$.

For instance, the reader is referred to Stroock and Varadhan [241, Chapter 9, pp. 208–247] for some useful estimates on the transition probability functions for diffusion processes in $\mathbb{R}^d$.

In either of these two equivalent forms of transition function we complete the definition by using the one-point compactification of E, say $\bar{E} = E \cup \{\infty\}$ with $P(s, x, t, \{\infty\}) = 1 - P(s, x, t, \mathbb{R}^d)$, so that $P(s, x, t, B)$ is a transition function in compact Polish space $\bar{E}$. Thus, time-homogeneous means $P(t, s) = P(t - s)$ while spatially-homogeneous means that $P(t, s)$ commutes with the translations operators $T_h\varphi(x) := \varphi(x - h)$, i.e., for any $t > s \ge 0$ and h in $\mathbb{R}^d$ we have $P(t, s)\, T_h = T_h\, P(t, s)$ in $C_0(\mathbb{R}^d)$.

Condition (c) or (c') means that the Markov process X is stochastically continuous, i.e., for any $\varepsilon > 0$ and $s \ge 0$ there is a $\delta > 0$ such that $P\{|X(t) - X(s)| \ge \varepsilon\} < \varepsilon$ for any t in $](s - \delta) \wedge 0, s + \delta[$. On a bounded interval, this is equivalent to a *uniformly stochastically continuous* property, namely for any $\varepsilon > 0$ there is a $\delta > 0$ such that $P\{|X(t) - X(s)| \ge \varepsilon\} < \varepsilon$ for any t, s in $[0, 1/\varepsilon]$ satisfying $|t - s| \le \delta$. Actually, because the Polish space E is locally compact, both conditions (c) and (c') are equivalent under the Feller assumption (d).

The relation between a transition function and the evolution operators (or semigroup) is clearly (3.76) with $\Phi(t, s) = P(t, s)$. In the time-homogeneous case, this relates with the resolvent operators by

$$\left\{ \begin{aligned} R(\lambda)\varphi(x) &= \int_0^\infty \mathrm{e}^{-t\,\lambda}\, \Phi(t)\varphi(x)\mathrm{d}t = \\ &= \int_0^\infty \mathrm{e}^{-t\,\lambda}\mathrm{d}t \int_{\mathbb{R}^d} \varphi(y) P(t, x, \mathrm{d}y), \ \ \forall x \in \mathbb{R}^d, \end{aligned} \right. \tag{3.77}$$

which may be generalized to the non-homogeneous case.

A crucial relation between the transition function $P(s, x, t, B)$ and the family of stochastic processes $X = X_{sx}$ is the equality

$$P\{X(t) \in B \mid X(r),\, r \le s\} = P(s, X(s), t, B), \tag{3.78}$$

for every $t > s \ge 0$ and B in $\mathcal{B}(\mathbb{R}^d)$, which is the Markov property itself. This is the primary building block, in the sense that when the family of stochastic processes X is given first, some property on their paths is necessary to construct the transition function, condition (3.72) is not sufficient. The general theory of Markov processes is rather delicate, so that we prefer to limit ourself to the case

of standard Markov processes, i.e., cad-lag path and stochastically continuous in a filtered spaces (satisfying the usual conditions).

Generally, a Markov process is used for modelling the dynamic of a motion (e.g., of a particle). Intuitively, the Markov property expresses a *prediction* of subsequent motion (of a particle), knowing its position at time t, does not depend on what has been observed during the time interval $[0, t]$. In most of the cases, the above (simple) Markov property is not sufficient, this *starting afresh* property need to be used with stopping times. This is called the *strong Markov property* and written as

$$P\{X(t+\tau) \in B \mid X(r+\tau),\, r \le 0\} = P(\tau, t, X(\tau), B), \tag{3.79}$$

for every $t \ge 0$, B in $\mathcal{B}(\mathbb{R}^d)$, and every stopping time. It is clear that any Markov process with cad-lag paths and a Feller transition satisfies the strong Markov property (3.79).

Only in very particular cases the transition function is explicitly known, such as a Wiener or a Poisson process, see (1.38) or (1.39) in Chapter 1. In most of the cases, the transition function is constructed from a family of linear possible unbounded (infinitesimal generator) operators $A(t)$ defined in a domain $\mathcal{D}(A(t))$ and indexed in $t \ge 0$. Moreover, what is actually known is the expression to the operators $A(t)$ for *smooth* or *test* functions, e.g., $A(t)$ is a second order elliptic differential operator with given coefficients, or more general an integro-differential operator of a particular form. The semigroup theory or the theory of evolution operators address this question, i.e., (1) if a semigroup $\{\Phi(t) : t \ge 0\}$ is given then characteristic properties on its so-called infinitesimal generator A are listed and (2) if a given operator A satisfies the characteristic properties of an infinitesimal generator then a semigroup $\{\Phi(t) : t \ge 0\}$ can be constructed. For a linear and bounded operator A the arguments go back to the exponential function, i.e.,

$$A\varphi := \lim_{t \to 0} \frac{\Phi(t)\varphi - \varphi}{t} \quad \text{and} \quad \Phi(t) := \sum_{n=0}^{\infty} \frac{(t\,A)^n}{n!} = \mathrm{e}^{tA}.$$

In general, a much more sophisticated argument is necessary, Conditions (a') and (e') of the Definition 3.24 are characteristic properties of the so-called Markov-Feller (or Feller-Dynkin) semigroups, which is the main tool we use to model stochastic dynamics. Clearly, assumption (e') imposes a certain type of regularity, while (a') translates into the so-called *maximum principle* satisfied by its infinitesimal generator, see Chapter 2 for an overview of the semigroup $\Phi(t)$ and its infinitesimal generator A

For a given transition probability function $P(s, x, t, B)$ as in Definition 3.24, since $P(s, x, t, B)$ and an initial distribution determine the finite-dimensional of the Markov process, we may use Kolmogorov's construction to define a family of $\mathbb{R}^d$-valued random variables $\{X_{sx}(t) : t \ge 0\}$ for each initial time $s \ge 0$ and initial distribution x in $\mathbb{R}^d$ such that the Markov property (3.78) is satisfied, i.e., for any real numbers $s < t_1 < \cdots < t_n$ and Borel subsets $B_1, \ldots, B_n$ of R^d

the family of probability measures

$$P_{sx,t_1,\dots,t_n}(B_1\times\ldots\times B_n):=\int_{B_1}P(s,x,t_1,\mathrm{d}x_1)\times$$
$$\times\int_{B_2}P(t_1,x_1,t_2,\mathrm{d}x_2)\dots\int_{B_n}P(t_{n-1},x_{n-1},t_n,\mathrm{d}x_n),$$

for any $s<t_1<\cdots<t_n$, has the consistency property. Therefore there exists a unique probability measure P_{sx} on the space Ω of all functions from $[s,\infty)$ into $\mathbb{R}^d$ such that $P_{sx}\{X(t)\in B\}=P(s,x,t,B)$ for any $t>0$ and B in $\mathcal{B}(\mathbb{R}^d)$, where X is the canonical (coordinate or projection) process, namely $X(t,\omega):=\omega(t)$ for any ω in Ω. Besides this, for any bounded and measurable function $f(x_1,\dots,x_n)$ we have

$$\mathbb{E}_{sx}\{f(X(t_1),\dots,X(t_n))\}=\int P(s,x,t_1,\mathrm{d}x_1)\times$$
$$\times\int P(t_1,x_1,t_2,\mathrm{d}x_2)\dots\int f(x_1,\dots,x_n)\,P(t_{n-1},x_{n-1},t_n,\mathrm{d}x_n).$$

Thus, the Markov property (3.78) holds true for this construction. Since no condition on the paths is assumed, this is referred to as a Markov process *in law*, where the crucial Markov property may be re-written as

$$\mathbb{E}_{sx}\{f(X(s_1),\dots,X(s_m))\,g(X(r+t_1),\dots,X(r+t_n))\}=$$
$$=\mathbb{E}_{sx}\{f(X(s_1),\dots,X(s_m))\,h(X(r))\},$$

where $h(\xi):=\mathbb{E}_{r\xi}\{g(X(r+t_1),\dots,X(r+t_n))\}$ and $s<s_1<\ldots<s_m\le r\le t_1<\ldots<t_n$. Note that only conditions (a), (b) and (d) in Definition 3.24 of transition function are used to construct a Markov process in law. As mentioned previously, if the transition function $P(s,x,t,B)$ is not a full probability, i.e., $P(s,x,t,\mathbb{R}^d)\le 1$ then we need to use the one-point compactification $\bar{\mathbb{R}}^d$ of R^d and define $P(s,x,t,\{\infty\})=1-P(s,x,t,\mathbb{R}^d)$ and $P(s,\infty,t,\{\infty\})=1$. In this case, the above random variables $\{X_{sx}(t):t\ge 0\}$ take values in $\bar{\mathbb{R}}^d$.

Given a transition function $P(s,x,t,B)$ we define the modulus of stochastic continuity by

$$\begin{cases}\alpha(\varepsilon,T,\delta,K):=\sup\big\{1-P\big(s,x,t,\{y:|y-x|\le\varepsilon\}\big)\ :\\ \qquad\qquad :\forall x\in K,\ s,t\in[0,T],\ 0<t-s\le\delta\big\},\end{cases}\tag{3.80}$$

where $K\subset\mathbb{R}^d$. Because of assumption (c) or (c') on a transition function we know that for any $\varepsilon,T>0$ and any x in $\mathbb{R}^d$ we have $\alpha(\varepsilon,T,\delta,\{x\})\to 0$ as $\delta\to 0$. However, we need to assume that

$$\lim_{\delta\to 0}\alpha(\varepsilon,T,\delta,\mathbb{R}^d)=0,\quad\forall\varepsilon,T>0,\tag{3.81}$$

This condition (3.81) is satisfied for a Feller transition.

The following result addresses the construction of standard Markov processes

Theorem 3.25. *Let $P(s,x,t,B)$ be a transition probability function satisfying (3.81). Then for any initial condition (s,x) there exists a probability measure P_{sx} on the canonical space $D([0,\infty),\mathbb{R}^d)$ such that the canonical process $X(t,\omega):=\omega(t)$ is a Markov process with transition function $P(s,x,t,B)$, which satisfies $P_{sx}\{X(t)=x,\ t\le s\}=1$. Moreover, if the transition function satisfies*

$$\lim_{\delta\to 0}\frac{\alpha(\varepsilon,T,\delta,\mathbb{R}^d)}{\delta}=0,\quad \forall\varepsilon,T>0, \tag{3.82}$$

then the support of the measure P_{sx} is the canonical space $C([0,\infty),\mathbb{R}^d)$. Furthermore, if $P(s,x,t,B)$ is a Feller transition function then the strong Markov property relative to the canonical filtration $(\mathcal{F}(t):t\ge 0)$ (universally completed with respect to the family $\{P_{sx}:(s,x)\}$ and right-continuous), i.e.,

$$P_{sx}\{X(\theta)\in B\mid\mathcal{F}(\tau)\}=P(\tau,X(\tau),\theta,B),\ \ \forall B\in\mathcal{B}(\mathbb{R}^d), \tag{3.83}$$

for any finite stopping times $\theta\ge\tau\ge s$, and the filtration $(\mathcal{F}(t):t\ge 0)$ is quasi-left continuous.

Proof. Since this is a classic result for the construction of Markov processes, only the key points will be discussed here, for instance, reader may consult the book by Dellacherie and Meyer [58, Section XIV.24, pp. 169–172] or Sato [220, Theorem 11.1, pp. 59–63] for details.

First, we need some notation. Let R be a subset of times in $[0,\infty)$ and $\varepsilon>0$. We say that a family $X=\{X(t):t\ge 0\}$ of $\mathbb{R}^d$-valued random variables (1) has ε-oscillations n-times in R for a fixed ω if there exist $t_0<t_1<\cdots<t_n$ in R such that $|X(t_i)-X(t_{i-1})|>\varepsilon$ for any $i=1,\dots,n$, or (2) has ε-oscillations infinitely often in R for a fixed ω if for any n the family X has ε-oscillations n-times in R. Denote by $B^X(n,\varepsilon,R)$ and $B^X(\infty,\varepsilon,R)$ the set of ω where X has ε-oscillations n-times and infinitely often in R, respectively.

Most of the arguments is to find a modification of the Markov process in law constructed above. To that effect, denote by Ω_2 the set of ω such that the one-sided limits

$$\lim_{s\to t,\ s<t\ s\in\mathbb{Q}}X(s,\omega)\qquad\text{and}\qquad\lim_{s\to t,\ s>t\ s\in\mathbb{Q}}X(s,\omega)$$

exist in $\mathbb{R}^d$ for any $t\ge 0$. Note that for any strictly decreasing sequence $\{t_n\}$ to t, of rational numbers in $[0,\ell]$, there exists $N=N(\varepsilon,\ell)$ such that $|X(t_n,\omega)-X(t_N,\omega)|\le\varepsilon$ for any $n\ge N$ and ω in $\Omega\setminus B^X(\infty,\varepsilon,[0,\ell]\cap\mathbb{Q})$. This shows that Ω_2 contains the set

$$\Omega_2^*:=\Omega\setminus\bigcup_{\ell=1}^{\infty}\bigcup_{k=1}^{\infty}B^X(\infty,4/k,[0,\ell]\cap\mathbb{Q}),$$

which is measurable since $\mathbb{Q}$ is countable.

The following modification, $X^*(t,\omega):=0$ for every $\omega\in\Omega\setminus\Omega_2^*$, and

$$X^*(t,\omega):=\lim_{s\to t,\ s<t\ s\in\mathbb{Q}}X(s,\omega),\quad\forall\omega\in\Omega_2^*,$$

has cad-lag paths and because the stochastically continuity we obtain

$$P\{X(t,\omega) = X^*(t,\omega),\ \omega \in \Omega_2^*\} = 1.$$

To complete this cad-lag modification we need to show that $P(\Omega_2^*) = 1$.

The following estimate, proved by induction on the integer n, yields the result as discussed below. If $0 \le s_1 < \cdots < s_m \le r \le t_1 < \cdots < t_k < r + \delta \le T$ and $R := \{t_1, \ldots, t_k\}$ then we have

$$\mathbb{E}\{Z\, \mathbb{1}_{B^X(n,4\varepsilon,R)}\} \le \mathbb{E}\{Z\}\, [2\alpha(\varepsilon, T, \delta, \mathbb{R}^d)]^n, \tag{3.84}$$

for every $Z = f(X(s_1), \ldots, X(s_\ell))$ with a nonnegative measurable function f, and where $\alpha(\varepsilon, T, \delta, \mathbb{R}^d)$ is defined by (3.80). A key point is the fact that the right-hand side does not depend on k.

Thus, to show that $P(\Omega_2^*) = 1$ we will prove that $P\{B^X(\infty, 4/k, [0,\ell] \cap \mathbb{Q})\} = 0$ for any integer k and ℓ. Indeed, by making a subdivision of $[0,\ell]$ into j equal intervals, we obtain

$$P\{B^X(\infty, 4/k, [0,\ell] \cap \mathbb{Q})\} \le \\ \le \sum_{i=1}^{j} \lim_{n\to\infty} P\{B^X(n, 4/k, [(i-1)\ell/j, i\ell/j] \cap \mathbb{Q}),$$

and from the above estimate (3.84) with $\{t_1, t_2, \ldots\} := [(i-1)\ell/j, i\ell/j] \cap \mathbb{Q}$ deduce

$$P\{B^X(n, 4/k, [(i-1)\ell/j, i\ell/j] \cap \mathbb{Q})\} \le [2\alpha(1/k, \ell, \ell/j, \mathbb{R}^d)]^n,$$

for every $n = geq1$. In view of condition (3.81), for a given ℓ we can select the integer j sufficiently large so that $2\alpha(1/k, \ell, \ell/j, \mathbb{R}^d) < 1$. Hence, as $n \to \infty$ we get $P\{B^X(n, 4/k, [(i-1)\ell/j, i\ell/j] \cap \mathbb{Q})\} = 0$, which implies $P(\Omega_2^*) = 1$.

When condition (3.82) is satisfied, we have to find a measurable set Ω_1^* with $P(\Omega_1^*) = 1$ and such that $X^*(t,\omega) = X^*(t-,\omega)$ for any $t > 0$ and ω in Ω_1^*. Indeed, for a given $\ell > 0$, consider the set $R(n,\varepsilon,\omega)$, with $n = 1, 2, \ldots$ and $\varepsilon > 0$, defined as the number of $i = 1, \ldots, n$ such that $|X^*(i\ell/n, \omega) - X^*((i-1)\ell/n, \omega)| > \varepsilon$. Then, $\omega \mapsto R(n,\varepsilon,\omega)$ is measurable and

$$\mathbb{E}\{R(n,\varepsilon,\cdot)\} \le n\, \alpha(\varepsilon, \ell, \ell/n).$$

Hence, condition (3.82) and Fatou's lemma yield $\mathbb{E}\{\liminf_{n\to\infty} R(n,\varepsilon,\cdot)\} = 0$ and therefore the set

$$\Omega_1^*(\ell) := \bigcap_{k=1}^{\infty} \big\{\omega : \liminf_{n\to\infty} R(n, 1/k, \omega) = 0\big\}$$

is measurable with full measure, i.e., $P\{\Omega_1^*(\ell)\} = 1$. Moreover, if ω is in $\Omega_1^*(\ell)$ then for any t in $(0,\ell]$ we have $|X^*(t,\omega) - X^*(t-,\omega)| \le \varepsilon$, for every $\varepsilon > 0$. Thus $\Omega_1^* := \bigcap_\ell \Omega_1^*(\ell)$ has the desired property.

It is clear that once a cad-lag version, namely X^*, has been found, we can take the image probability measure in the canonical space to produce P_{sx} as

required. On the other hand, the stochastic continuity and the cad-lag regularity of the paths imply that $P\{X^*(t) = X^*(t-)\} = 1$ for any $t > s$.

The right-continuity of paths ensures that the process X^* is adapted to $\mathcal{F}(t) := \mathcal{F}^{sx}(t+) = \bigcap_{\varepsilon>0} \mathcal{F}^{sx}(t)$, where $\mathcal{F}^{sx}(t)$ is the σ-algebra generated by the canonical process and P-null sets. Thus (3.83) is satisfied after using the continuity of the transition probability function and approximating any finite stopping time.

Regarding the quasi-left continuity we proceed as follows. Let $\{\tau_n : n \geq 1\}$ be a sequence of stopping times convergence almost surely to τ, with $P(\tau_n < \tau < \infty, \tau > s) = 1$. For any two functions f and g in $C_0(\mathbb{R}^d)$ we have

$$\lim_{t\to 0} \lim_{n\to\infty} \mathbb{E}\{f(X^*(\tau_n))\, g(X^*(\tau_n + t))\} =$$

$$= \lim_{t\to 0} \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau + t-))\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau))\},$$

because the right-continuity of the paths. On the other hand, the strong Markov property (3.79) and the Feller property imply

$$\lim_{n\to\infty} \mathbb{E}\{f(X^*(\tau_n))\, g(X^*(\tau_n + t))\} = \mathbb{E}\{f(X^*(\tau-))\, P(\tau, \tau + t, X^*(\tau-), g)\}$$

and

$$\lim_{t\to 0} \mathbb{E}\{f(X^*(\tau-))\, P(\tau, \tau + t, X^*(\tau-), g)\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau-))\}.$$

Hence,

$$\mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau))\} = \mathbb{E}\{f(X^*(\tau-))\, g(X^*(\tau-))\},$$

i.e., $P\{X^*(\tau) = X^*(\tau-)\} = 1$ and X^* is almost surely continuous at τ. □

Usually, condition (3.81) is replaced by

$$\begin{cases} \text{(a)} & \lim\limits_{|x|\to\infty} \sup\limits_{0\leq s<t\leq T} P(s, x, t, K) = 0, \\ \text{(b)} & \lim\limits_{\delta\to 0} \alpha(\varepsilon, T, \delta, K) = 0, \quad \forall \varepsilon, T > 0, \end{cases} \tag{3.85}$$

for any compact subset K of $\mathbb{R}^d$, and assumption (3.82) can be substituted by

$$\lim_{\delta\to 0} \frac{\alpha(\varepsilon, T, \delta, K)}{\delta} = 0, \quad \forall \varepsilon, T > 0, \quad \text{any compact } K \subset \mathbb{R}^d, \tag{3.86}$$

and in general this construction ie valid for a transition function, without the probability condition $P(s, x, t, \mathbb{R}^d) = 1$, see Taira [243, Chapter 9 and 10, pp. 273–424].

To properly handle the strong Markov property, we need to use the *universally complete* σ-algebra, i.e., first we remark that the above construction can be used with any initial law μ at any time 0 and the corresponding filtration is $\{\mathcal{F}^{0\mu}(t) : t \geq 0\}$. Thus $\mathcal{F}^0(t) := \bigcap_\mu \mathcal{F}^{0\mu}(t)$, which is not necessarily complete with respect to $P^{0\mu}$, but it satisfies $\mathcal{F}^0(t+) = \mathcal{F}^0(t)$, i.e., it is right-continuous,

and the so called Blumenthal's zero-one law, i.e., $P(A) = 0$ or $P(A) = 1$ for any A in $\bigcap_{t>0} \mathcal{F}^0(t)$.

Let us look at the particular case of additive processes, see Definition 3.19, which include the Lévy processes. The transition function of an additive process is spatially homogeneous, i.e., if $P(s, x, t, B)$ is the transition function of an additive process X then $P(s, x, t, B) = P(s, 0, t, B - x)$ and we only have to consider transition functions of the form $P(s, t, B)$. Thus, any additive process X yields a transition function $P(s, t, B) := P\{X(t) - X(s) \in B\}$, for any $t > s \geq 0$ and B in $\mathcal{B}(R^d)$ so that X is a (stochastically continuous) Markov process in R^d stating at 0. Its associated semigroup is called a *convolution* semigroup, i.e.,

$$P(t, s)\varphi(x) := \int_{\mathbb{R}^d} \varphi(x + y) P(s, t, \mathrm{d}y)$$

and Chapman-Kolmogorov identity is re-written as

$$P(s, t, B) = \int_{\mathbb{R}^d} P(s, r, \mathrm{d}x)\, P(r, t, B - x),$$

for every $t > r > s \geq 0$ and B in $\mathbb{R}^d$. It is also clear that the previous Theorem 3.25 applies to this case, to obtain a cad-lag of additive processes in law. Because the transition function $P(s, t, B)$ is spatially homogeneous, it satisfies the Feller conditions and the process is quasi-left continuous, see Definition 3.15.

Lévy processes X are also time-homogeneous and its semigroup is a true convolution and the infinitely divisible distribution $\mu := X(1)$ completely determines the process, see Section 3.10. Thus to each infinitely divisible distribution μ there corresponds a Lévy process. For instance, Poisson and compound Poisson processes correspond to Poisson and compound Poisson distributions. The Lévy process on $\mathbb{R}^d$ corresponding to a Cauchy distribution with parameters γ in $\mathbb{R}^d$ and $c > 0$, namely, for any B in $\mathcal{B}(\mathbb{R}^d)$,

$$\begin{cases} \mu(B) := \pi^{-(d+1)/2} \Gamma(\frac{d+1}{2})\, c \displaystyle\int_B \left(|x - \gamma|^2 + c^2\right)^{-(d+1)/2} \mathrm{d}x, \\ \text{and} \quad \hat{\mu}(y) = \mathrm{e}^{-c|y| + \mathtt{i}\,\gamma \cdot y}, \quad \forall y \in \mathbb{R}^d, \end{cases} \tag{3.87}$$

is called a *Cauchy process*. However, the Lévy process on $\mathbb{R}$ corresponding to an exponential distribution is called a Γ-process, since it has a Γ distribution at any $t > 0$.

If X is an additive process on $\mathbb{R}^d$ with a Gaussian distribution at each t, then X has continuous paths almost surely, see Sato [220, Theorem 11.7, pp. 63-64]. For instance, for dimension $d = 1$, the characteristic function is

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\, y \cdot X(t)}\} = \mathrm{e}^{-t\, y^2/2}, \quad \forall t \geq 0,\ y \in \mathbb{R}^d,$$

and a simple calculation shows that condition (3.82) of Theorem 3.25 is satisfied. Actually, the only additive process with continuous paths are Wiener processes.

For a given additive process X we consider the σ-algebra $\mathcal{F}(t)$ generated by all null sets and the family of random variables $X(s)$ with $s \leq t$. Because of the independence of increments, an application of Kolmogorov's zero-one law

to a *tail* σ-algebra shows that $\mathcal{F}(t)$ is already right-continuous, so that it is the filtration associated with X.

The reader is referred to the books by Blumental and Getoor [28], Dellacherie and Meyer [58, Chapters XI–XVI], Ethier and Kurtz [76], Sato [220, Chapter 1 and 2, pp. 1–68], among others.

3.12 Hunt and Standard Processes

In the modern theory of Markov processes, the emphasis is put on Markov transition functions $p(t, x, A)$, where $t \geq 0$, $x \in E$, a locally compact Hausdorff space, and A is any element of the Borel σ-algebra in E, as described in the previous Section 3.11 for the case $E = \mathbb{R}^d$. Thus, starting from a Markov transition function (or its Laplace transform, the resolvent), the actual construction of a Markov process having the prescribed transition functions is known a *realization* of the Markov process. Certainly, Kolmogorov construction and path regularity is the natural approach. Moreover, the strong Markov property is a highly desired. Therefore, continuous time Markov processes are usually constructed in the canonical cad-lag sample space $\Omega = D([0, \infty), E)$, i.e., we construct a probability measure P on Ω such that the canonical process $X_t(\omega) = \omega(t)$ is the desired Markov process. Furthermore, to simplify the notation, only the time-homogenous Markov process is considered, since by adding one dimension to the space (i.e., using $E \times [0, \infty)$ instead of E) all assumptions can be transported to the time-dependent case. Clearly, it is not necessary to have an explicit form for the transition function (or resolvent). Only a number of properties are involved, which can be obtained from a given semigroup. The semigroup is described in term of its infinitesimal generator or its Dirichlet form.

Essentially based on super-median functions and super-martingales arguments, the general theory of processes (e.g. Dellacherie and Meyer [58, Section XIV.24, pp. 169–172]) shows that a cad-lag realization can be constructed for any Markov transition function, i.e., satisfying (a), (b) and (d) of Definition 3.24, see also Section 1.9 in Chapter 1 and Section 2.8 in Chapter 2, i.e.,

(a) for each $t > 0$ and x in E the function $B \mapsto P(t, x, B)$ is a probability measure on $(E, \mathcal{B}(E))$,

(b) for each B in $\mathcal{B}(E)$ the function $(t, x) \mapsto P(t, x, B)$ is a measurable,

(c) for any x in E and B in $\mathcal{B}(E)$ we have

$$\lim_{t \to 0} P(t, x, B) = \delta_x(A),$$

i.e., the limit is equal to 1 if x belongs to A, and 0 otherwise,

(d) for each $s, t > 0$, x in E and B in $\mathcal{B}(E)$ we have

$$P(s + t, x, B) = \int_E P(s, x, \mathrm{d}y)\, P(t, y, B),$$

i.e., the Chapman-Kolmogorov identity holds.

Condition (c) is not actually necessary, but certainly is a natural complement to (d), which becomes necessary for the sub-Markov transitions, see below. Note that the Chapman-Kolmogorov identity combined with the inequality $0 \leq P(t,x,B) \leq 1$ show that the limit in (c) is monotone decreasing.

Besides the Feller property, there is another key property (necessary to build a nice theory), the so called *quasi-left continuity*. Moreover, the one-point compactification $\bar{E} = E \cup \{\infty\}$ (where the symbol ∞ does not belong to E) of a locally compact Hausdorff space E (the state space) and the concept of a *coffin state* and a *lifetime* are necessary (recall that for a locally compact space E, a point ∞ is adjoined to E as the point at infinity if K is not compact, and as an isolated point if K is compact). Typically, given a stopping time ς and an adapted E-valued cad-lag process Y on a filtered probability space $(\Omega, \mathbb{F}, P)$, $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$, we define a new process X as follows

$$X(t) = \begin{cases} Y(t) & \text{if } t < \varsigma, \\ \infty & \text{if } t \geq \varsigma, \end{cases}$$

which is $\bar{E}$-valued with lifetime ς. Clearly, the process Y needs only to be defined on the semi-open stochastic interval $[\![0, \varsigma[\![$. Now, a cad-lag (up to its lifetime) process X with values in $\bar{E}$ has the lifetime $\varsigma := \inf\{t \geq 0 : X(t) = \infty\}$ if $X(t) = \infty$ for every $t \geq \varsigma$. It is important to observe that X with values in $\bar{E}$ may not be "fully" cad-lag, the limit as $t \to \varsigma$ does not necessarily exist. Thus, the canonical cad-lag space $D([0,\infty), E)$ cannot be used with $\bar{E}$ instead of E, this requires some adjustment, via the so-called Ray resolvent. This means that the coffin state and one-point compactification state space are just convenient notations, the key elements are the lifetime ς and the E-valued process X, which is cad-lag only up to ς, i.e., the pathwise left-hand limit $X(\varsigma-)$ may not exist, even on $\varsigma < \infty$.

Definition 3.26 (quasi-left continuity)**.** Let ς be a stopping time on a filtered probability space $(\Omega, \mathbb{F}, P)$, $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$. A cad-lag adapted process X with valued in E is called *quasi-left continuous* on $[\![0, \varsigma[\![$ if $X(\tau_n) \to X(\tau)$ almost surely on $\{\tau < \varsigma\}$, for any increasing sequence $\{\tau_n\}$ of stopping times converging to τ, with $\tau_n < \tau$ almost surely. □

Note that for a (cad-lag) quasi left-continuous we have $P\{X(t) = X(t-)\} = 0$ for every t in $[\![0, \varsigma[\![$ and the set of probability zero, where X_{τ_n} fails to converge, depends on the stopping time τ and the sequence $\{\tau_n\}$. Hence, in general the process is not continuous on the left, because it may not be possible to find a common set of probability zero for all times, i.e., a continuous modification of X does not necessarily exist.

Recall also that a stopping time τ is predictable if and only if that there exists an increasing sequence of stopping times τ_n with $\tau_n < \tau$ almost surely and converging to τ. Also, a stopping time τ is called *totally inaccessible* if for any predictable stopping time θ we have that $P(\tau = \theta < \infty) = 0$.

It is proved in Jacod and Shiryaev [117, Propositions 1.32, 2.26], that the random set of discontinuities of any cad-lag process X, i.e., the set of random

jumps $J_X = \{(t,\omega), X(t_-,\omega) \neq X(t,\omega)\}$ has the form $J_X = \bigcup_n [\![\tau_n]\!]$ where τ_n is a sequence of stopping times, which is called a sequence that exhausts the jumps of X. Moreover, the following statements are equivalent:

(1) the process X is quasi left-continuous on $[\![0,\infty[\![$,

(2) we have $X(\tau^-) = X(\tau)$ almost surely on $\{\tau < \infty\}$ for every predictable time τ,

(3) there exists a sequence of totally inaccessible stopping times that exhausts the jumps of X.

3.12.1 Digression on Markov processes

As mentioned early and mainly for the notation simplicity, details are given only for time-homogeneous Markov processes.

Definition 3.27 (Markov processes)**.** A cad-lag *Markov process* with values in a Hausdorff space E is composed by the following elements:

(1) the canonical cad-lag space $D = D([0,\infty), E)$ endowed with σ-algebra $\mathcal{F}$, the σ-algebra generated by the canonical process $X_t(\omega) = \omega(t)$,

(2) a right-continuous increasing family of σ-algebras $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$ on $(D, \mathcal{F})$, i.e., a not necessarily completed filtration,

(3) a family of cad-lag processes (P^μ, X) or (P, X^μ) (depending on what is to be emphasized) indexed by μ, i.e., P^μ is a probability measure on $(D, \mathcal{F})$ and X^μ a measurable function from D into itself, such that the E-valued random variable $X(t)$ is $\mathcal{F}_t$-measurable, for every $t \geq 0$, and μ is the initial distribution on the state space E, i.e., $P\{X(0) \in B\} = \mu(B)$, for every B in $\mathcal{B}(E)$,

(4) a transition function p on E, i.e., (a),..., (d) above are satisfied,

All these elements come together in the Markov property

$$P(X(t) \in B \mid \mathcal{F}_s) = p(t-s, X(s), B), \quad \forall t \geq s \geq 0,\ \forall B \in \mathcal{B}(E), \tag{3.88}$$

which is an almost surely equality. □

Observe that the left hand side in the Markov property is only defined almost surely, while the right hand side is well defined everywhere. This means that for every μ there exists a subset $N = N^\mu$ of the canonical space D with $P^\mu(N) = 0$ such that equality (3.88) holds outside of N.

The above definition may take place in an arbitrary probability space $(\Omega, \mathcal{F})$, but since the processes are cad-lag, the canonical sample can be used to concretize arguments. Moreover, with the notation (P^μ, X) the probability measure is emphasized with $X(t,\omega) = \omega(t)$ the canonical process, while (P, X^μ) suggests a fixed probability with a selectable process X^μ.

The indexation in term of distributions μ on the state space is not actually necessary, a smaller class suffices. Indeed, if the Markov process can be constructed for any $\mu = \delta_x$ with x in E and some measurability conditions are imposed on the mapping $x \mapsto (P_x, X)$ or $x \mapsto (P, X_x)$, where P_x and X_x correspond

to the initial distribution δ_x, e.g., $P_x\{X(0) = x\} = 1$ or $P\{X_x(0) = x\} = 1$. Clearly, we have $P^\mu = P_x\,\mu(\mathrm{d}x)$.

The filtration $\mathbb{F}$ needs to be *universally completed* with respect to the probability measures P^μ or P_x. This is to enlarge each $\mathcal{F}_0$ (and consequently $\mathcal{F}_t$ for every $t > 0$) with sets which have measure zero relative to each P^μ, i.e., first define $\mathcal{F}_t^\mu$ by completing $\mathcal{F}_t$ with respect to P^μ, next set $\mathcal{F}_t^0 = \cap_\mu \mathcal{F}_t^\mu$ and finally, take $\bar{\mathcal{F}}_t = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}^0$ to make it right-continuous if necessary, i.e., $\bar{\mathbb{F}} := (\bar{\mathcal{F}}_t : t \geq 0)$. In most of the cases, the initial filtration $\mathbb{F}$ is just the history of the process X, which may depends on μ if we insist in the (P, X^μ) setting. So, by using the embedding in the canonical space (P^μ, X) with $X(t,\omega) = \omega(t)$ the canonical processes, the initial filtration is fixed and $\bar{\mathbb{F}}$ is usefully.

For this universally completed filtration $\bar{\mathbb{F}}$, we consider an equality similar to (3.88), namely, for every almost surely finite stopping time τ relative to $\bar{\mathbb{F}}$ assume

$$P(X(\tau + t) \in B \mid \bar{\mathcal{F}}_\tau) = p(t, X(\tau), B), \quad \forall t \geq 0,\ \forall B \in \mathcal{B}(E), \tag{3.89}$$

which is again an almost surely equality, and where $\bar{\mathcal{F}}_\tau$ is the σ-algebra generated by the stopping time τ. This is referred as the *strong Markov property* and a Markov process satisfying this condition is called a *strong Markov process*. Certainly, condition (3.89) can be relative to the initial filtration $\mathbb{F}$, but some technical reasons lead to the universally completed filtration $\bar{\mathbb{F}}$.

For the canonical cad-lag space, the so-called *shift operator* $\vartheta_t(\omega) := \omega(\cdot + t)$ is viewed as mapping D into itself, and satisfying

$$X_s \circ \vartheta_t = X_t \circ \vartheta_s = X_{s+t}, \quad \forall t, s \geq 0,$$

where the composition $X_s \circ \vartheta_t(\omega) = X_s(\vartheta_t(\omega))$ is used. In the abstract setting the shift map ϑ is postulate with the above properties. The strong Markov property takes the form

$$P(X_t \circ \vartheta_\tau \in B \mid \bar{\mathcal{F}}_\tau) = p(t, X_\tau, B), \quad \forall t \geq 0,\ \forall B \in \mathcal{B}(E),$$

for every almost surely finite stopping time τ relative to $\bar{\mathbb{F}}$.

In all the above, it is clear that the transition function is the main element in the definition of Markov processes. The construction of a Markov transition function is quite delicate and several ways are known. Starting from a infinitesimal generator or a Dirichlet form, functional analysis and in particular the semigroup theory are used to obtain suitable transition functions. Its probabilistic counterpart starts with simple (or known) Markov processes and via some transformations a transition function is obtain, the key arguments are stochastic differential equations and its generalizations. In some cases, the transition function obtained is not quite a Markov transition function, it is what is called a *sub-Markov transition function*, i.e., all conditions are satisfied, except for (a) where $p(t, x, \cdot)$ is only a sub-probability, i.e., $B \mapsto p(t, x, B)$ is a measure with $p(t, x, E) \leq 1$. In this case, the condition (c) is more important.

Some restrictions on the state space E are necessary to deal with sub-Markov transition function, namely, E is now a locally compact Hausdorff. This is

necessary to consider its one-point compactification $\bar{E} = E \cup \{\infty\}$ and the companion argument about coffin state and lifetime of a process. This sub-Markov case is reduced to the preceding Markov theory by extending the given sub-Markov transition function in E to a Markov transition function on $\bar{E}$. Indeed, set $\bar{p}(t,x,B) = p(t,x,B)$ for any Borel set B in E, $\bar{p}(t,x,\{\infty\}) = 1 - p(t,x,E)$, $\bar{p}(t,\infty,\{\infty\}) = 1$ and $\bar{p}(t,\infty,E) = 0$. As mentioned early, this extra point is often called the coffin state. The extra coffin state ∞ does not belong to E, it is the point at "infinity" when E is non compact and it is an isolated point when E is compact.

Hence, the sub-Markov case is reduced to a Markov case on a compact space $\bar{E}$, which has a particular isolated point. For the Markov process X corresponding to the Markov transition function $\bar{p}$ on the one-point compactification $\bar{E}$, the *lifetime functional*

$$\varsigma = \inf\big\{t \geq 0 : X(t) = \infty\big\} \tag{3.90}$$

acting on the canonical space plays a fundamental role. Again, observe that the canonical cad-lag space still being $D([0,\infty),E)$ and not $D([0,\infty),\bar{E})$, because $\bar{X}$ with values in $\bar{E}$ is not "fully" cad-lag, the limit as $t \to \varsigma$ does not necessarily exist. Hence, a Markov process with values in the compact space $\bar{E}$ with lifetime $\varsigma = \infty$ (almost surely for every P_μ) is actually a realization of a Markov transition on the initial locally compact space E. A sub-Markov transition function produces either a E-valued cad-lag *sub-Markov process* with a lifetime $\varsigma < \infty$ with positive probability or equivalently a $\bar{E}$-valued cad-lag (up to its lifetime) Markov process.

Most of the interest is on state spaces E, which are actually Polish spaces (complete separable metric spaces) so that the canonical cad-lag sample space $D([0,\infty),E)$ is also a Polish space. The locally compact character is used when dealing with sub-Markov transition functions.

The canonical process $X_t(\omega) = X(t,\omega) := \omega(t)$ and the shift map $\vartheta_t(\omega) = \vartheta(t,\omega) := \omega(\cdot + t)$, which are defined on the canonical space $D([0,\infty),E)$, may be considered as E-valued and D-valued cad-lag processes, respectively (note that $X = \{X(t) : t \geq 0\}$ can be regarded as a D-valued random variable). If we work on the one-point compactification state space $\bar{E}$ and the canonical filtration $\mathbb{F}$ is used (i.e., generated by the canonical process) then the lifetime functional (3.90) can be interpreted as a stopping time. There are other type of functionals or processes that we may consider

Definition 3.28 (functional)**.** Let A and L be measurable functions from the canonical cad-lag space $D([0,\infty),E)$ into $D([0,\infty),\mathbb{R})$. Then A is called an *additive* functional if $A_0(\omega) = 0$ and $A_t(\omega) - A_s(\omega) = A_{t-s}(\omega(\cdot + s))$, for every $t \geq s \geq 0$ and ω in $D([0,\infty),E)$. Also L is called a *multiplicative* functional if $L_0(\omega) = 1$ and $L_t(\omega)L_s(\omega) = L_{t-s}(\omega(\cdot + s))$, for every $t \geq s \geq 0$ and ω in $D([0,\infty),E)$. Similarly, an additive functional A is called *increasing* (or nondecreasing) if $A_t - A_s \geq 0$, for every $t > s \geq 0$, and a multiplicative functional L is called *positive* (or nonnegative) if $L_t \geq 0$, for every $t > 0$. If A or L is considered on the one-point compactification state space $\bar{E}$ then we also require

A or L to be constant on the stochastic interval $[\![\varsigma, \infty[\![$, i.e., it has been extended to $[0, \infty)$ with values in $\bar{E}$. □

It is possible to generalize and suppose that A_t has bounded variation trajectories, still having the possibility to integrate with respect to A_t in the Stieltjes sense. Typical examples are

$$A_t = \int_0^t c(X_s)\mathrm{d}s \quad \text{and} \quad L_t = \exp\Big\{ \int_0^t c(X_s)\mathrm{d}s \Big\},$$

where c is a positive (measurable) function defined on E, and in the case of $\bar{E}$ we suppose $c(\infty) = 0$.

If one is working on an abstract measurable space $(\Omega, \mathcal{F})$ then the shift operator ϑ is used to re-write the conditions as $A_t - A_s = A_{t-s} \circ \vartheta_s$ and $L_t L_s = L_{t-s} \circ \vartheta_s$, for every $t > s \geq 0$. In this case, functionals are regarded as either $\mathbb{R}$-valued cad-lag processes or $D([0, \infty), \mathbb{R})$-valued random variables.

As usually, the difficulty appears as soon as a probability (or a family of probabilities, as in the case of Markov processes) is assigned on the canonical space $D([0, \infty), E)$. The functionals are almost surely defined and we are interested in having "good" versions (or modifications) of them.

3.12.2 Classification of Markov processes

Concrete examples of Markov processes like diffusion processes, jump processes and Lévy processes have many properties in common besides the Markov property. They have all the Feller property, i.e., the semigroup associated to the transition functions is a (strongly) continuous on $C_0(E)$, in particular, it maps the function space $C_0(E)$ in itself or equivalent the Feller property is satisfied. Here the state space E is a locally compact Hausdorff space (usually a Polish space), and $C_0(E)$ is the space of continuous functions null at infinity. Note that the Feller processes are defined just starting from the transition functions, see Section 2.8 in Chapter 2.

It can be proved that given a Feller transition function $p(t, x, A)$ in E, there exist strong Markov processes (with cad-lag quasi left-continuous trajectories) having p as its transition function, e.g. see Rogers and Williams [214, Chapters III and VI]. The class of processes defined by these properties is the class of Hunt processes, which contains the Feller processes. A Hunt process can be constructed from any regular symmetric Dirichlet form (see Fukushima et al. [92]). However, to extend this result to regular non-symmetric Dirichlet forms (see Ma and Röckner [161]), it is necessary to consider a light generalization of Hunt process, that of *standard process*, introduced by Blumenthal and Getoor [28].

Roughly speaking, we can schematize

$$\text{Feller processes} \subset \begin{matrix} \text{Hunt processes} \subset \text{standard processes} \subset \text{right processes} \\ \text{Ray processes} \end{matrix}$$

where formal definitions are given below.

Definition 3.29 (Hunt and standard)**.** Let us be a time-homogeneous cad-lag Markov process (as in Definition 3.27) with values in a Hausdorff space E (i.e., obtained form a Markov transition function on E where its lifetime functional $\varsigma = \infty$ by definition) or with values in the one-point compactification $\bar{E}$ of a locally compact Hausdorff space E (i.e., obtained form a sub-Markov transition function on E where ς denotes its lifetime functional).

(a) It is called a *Hunt process* if the strong Markov property (3.89) is satisfied and the paths are quasi left-continuous on $[0, \infty)$, with respect to any probability P^μ.

(b) It is called a *standard process* if the strong Markov property (3.89) is satisfied and the paths are quasi left-continuous on $[0, \varsigma)$, with respect to any probability P^μ.

(c) Finally, it is *special standard* if also the random variable X_τ is $\bigvee_n \mathcal{F}_{\tau_n}$-measurable, for any increasing sequence $\{\tau_n\}$ of stopping times converging to τ, with $\tau_n < \tau$ almost surely. □

Here $\bigvee_n \mathcal{F}_{\tau_n}$ is the smaller σ-algebra which contains $\mathcal{F}_{\tau_n}$, for every $n \geq 1$. Note the small difference with Hunt and Standard processes on the quasi left-continuous property that is valid almost surely on $\{\tau < \zeta\}$, instead of $\{\tau < \infty\}$. Thus a standard process is a Hunt process if its lifetime is infinite or is it is quasi left-continuous at ς, i.e., a Hunt process is realized in the canonical space $D([0, \infty), E)$ or $D([0, \infty], \bar{E})$. Clearly, it can be proved that any Hunt process is a special standard process.

The following result holds:

Theorem 3.30. *A Hunt process admits a Lévy system, i.e., there exists a continuous additive functional H and a family of kernel $N(x, \mathrm{d}y)$ on E, such that $t \mapsto H_t$ is a continuous, $N(x, \{x\}) = 0$ for any $x \in E$, and*

$$\begin{cases} \mathbb{E}^\mu\Big(\displaystyle\sum_{0<s\leq t} f(X_{s-}, X_s) \mathbb{1}_{J_X}(s, \omega) \Big) = \\ \qquad = \mathbb{E}^\mu\Big(\displaystyle\int_0^t \mathrm{d}H_s \int_E f(X_s, y) N(X_s, \mathrm{d}y) \Big) \end{cases} \tag{3.91}$$

for any Borel positive f, defined on $E \times E$, and any initial distribution μ on E, where $J_X = \{(t, \omega), X(t_-, \omega) \neq X(t, \omega)\}$. □

Recall that a kernel $N(x, \mathrm{d}y)$ on E means (a) for each x in E the map $B \mapsto N(x, B)$ is a (σ-finite) measure on $\mathcal{B}(E)$ and (b) for each B in $\mathcal{B}(E)$ the map $x \mapsto N(x, B)$ is measurable. For instance, a proof can be found in Benveniste and Jacod [20] or Sharpe [225, Section 73, pp. 342–350].

Note that if we consider any Borel positive f, defined on $E \times E$ with $f(x, x) = 0$ then (3.91) reads

$$\mathbb{E}^\mu\Big(\sum_{0<s\leq t} f(X_{s-}, X_s) \Big) = \mathbb{E}^\mu\Big(\int_0^t \mathrm{d}H_s \int_E f(X_s, y) N(X_s, \mathrm{d}y) \Big),$$

and for $f = 1$ out of diagonal we get that $\mathbb{E}^\mu\{\text{number of jumps in } (0,t]\} = \mathbb{E}^\mu\{H_t\}$.

Following the discussion in Rogers and Williams [214, Chapter III], the class of Ray processes (defined in a axiomatic way) is the most convenient, among the various classes of Markov processes, to cope all the (time-continuous) Markov chains. They are in a sense "equivalent" to the class of right processes, introduced by Meyer (*la classe droite*), see also Dellacherie and Meyer [58, Chapters XI–XVI] or Sharpe [225, Chapter I].

As already pointed out, to give a Markov transition function $p(t,x,A)$ on E (which is jointly measurable in t and x) is equivalent to give the Laplace transform, called in this context a *resolvent kernel*

$$R_\lambda(x,A) = \int_0^\infty \mathrm{e}^{-\lambda t}\, p(t,x,A)\, \mathrm{d}t$$

For any bounded Borel function f on E we consider also the resolvent operator

$$R_\lambda f(x) = \int_E f(y)\, R_\lambda(x, \mathrm{d}y)$$

that verifies the resolvent identity $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$ and $\lambda R_\lambda 1 = 1$. If it happens that the resolvent operator maps $C_0(E)$ into itself and $\lambda R_\lambda \to I$ in the uniform convergence of $C_0(E)$, then the Markov process associated is Feller process.

Ray weakened the condition of strong continuity in zero, introducing the so called Ray resolvent. To this end, let us introduce the concept of *α-super-median function*, i.e. a positive, continuous and bounded function on E such that

$$\lambda R_{\lambda+\alpha} f \le f, \qquad \forall \lambda > 0,$$

and let us denote the *cone* of α-super-median functions by CSM^α.

Now, $\{R_\lambda\}$ is called a *Ray resolvent* on E if each R_λ maps $C_b(E) \to C_b(E)$ and

$$\bigcup_{\alpha \ge 0} \mathsf{CSM}^\alpha \text{ separates points of } E.$$

In other words, given two points x, x' in E, there exist a α and a α-super-median function f such that $f(x) \neq f(x')$. The canonical Markov process associated to a Ray resolvent admits a cad-lag modification that has the strong Markov property.

• *Remark* 3.31. In potential theory the resolvent operator R_λ is called the *λ-potential* operator. There is a concept similar to α-super-median functions in term of the semigroup P_t associated to the Markov transition function $p(t,x,A)$, namely, a positive Borel function f is called *α-super-mean-valued* (also called *α-super-averaging*, see Chung [44, p. 45]) if $\mathrm{e}^{-\alpha t} P_t f \le f$ for all $t \ge 0$. A α-super-mean-valued function is also a α-super-median function, but, in general, the

converse is not true, i.e., a α-super-median function is not necessarily a α-super-mean-valued function. In potential theory it is important also the concept of α-*excessive* function f, i.e., a α-super-mean-valued function f such that $\mathrm{e}^{-\alpha t}P_t f \uparrow f$ as $t \downarrow 0$. This class of functions are useful generalization of *super-harmonic* functions in classical potential theory. □

Definition 3.32 (right and Ray processes)**.** The class of *right processes* with state space E is the class of time-homogeneous cad-lag strong Markov processes, such that the process $f(X_t)$ is almost surely right continuous for any $\alpha > 0$ and any α-super-mean-valued function f. On the other hand, a *Ray process* is one with Ray resolvent, i.e., the super-median functions $\cup_{\alpha\geq 0}\mathsf{CSM}^\alpha$ separates points of E. □

A useful result is the fact that any standard process is a right process, which is not so simple to prove. As mentioned early, Ray processes are (time-homogeneous) cad-lag strong Markov processes.

3.12.3 Some Examples and Remarks

First we give a couple of examples.

Example 3.1. Consider jump Markov processes, for instance, see the books Bremaud [32], Feller [81, II, Section X.3, pp. 316–320], Ethier-Kurtz [76, Section 8.1, pp. 376–382], Sharpe [225, Section 72, pp. 339–342]. Let $\{X_t\}$ be a cad-lag piecewise constant process. In order that $\{X_t\}$ be a Markov process, we introduce a kernel $Q(x, B)$ on E, which describes the probability distribution of the points where the process jumps away from x, and a function $\lambda(x) \geq 0$, that describes the rate of jumps, i.e. $P^x(\tau > t) = \mathrm{e}^{-\lambda(x)t}$, where $\tau = \inf\{t : X_t \neq X_0\}$. We set $\tau_0 = 0$ and define recursively $\tau_n = \tau_{n-1} + \tau(\vartheta_{\tau_{n-1}})$ the subsequent jump times. Alternatively we describe the process by means of infinitesimal generator

$$Af(x) = \lambda(x)\int_E \big[f(y) - f(x)\big]\, Q(x, \mathrm{d}y).$$

In this case, if λ is bounded then then Lévy system can be taken as

$$H_t = \int_0^t \lambda(X_s)\mathrm{d}s, \qquad N(x, B) = Q(x, B).$$

If λ is not bounded we can take

$$H_t = t, \qquad N(x, B) = \lambda(x)\mathbb{1}_{\lambda(x)<\infty}Q(x, B).$$

In general the Lévy system for a Markov process is not uniquely determined. □

Example 3.2. In the case of a Lévy process in $\mathbb{R}^n$, with Lévy measure ν, then we choice for a Lévy system can be

$$H_t = t, \qquad N(x, B) = \nu(B - x).$$

For example, consider the symmetric Cauchy process, i.e. a purely jump Lévy process with Lévy measure

$$\nu(\mathrm{d}y) = \frac{1}{\pi}\frac{1}{y^2}\mathrm{d}y$$

on $\mathbb{R} \setminus \{0\}$. □

Now, let us show some application of additive functionals. Given a Hunt process $\{X_t\}$ with Markov transition function $p(t, x, A)$, consider a continuous positive additive functional A_t and the multiplicative functional

$$L_t = \mathrm{e}^{-A_t}$$

as in Definition 3.28. We will consider two other Hunt processes, starting from X_t.

The first is X_t^A, the *canonical subprocess* of X_t with respect to the multiplicative functional L_t. It is constructed starting from the Markov transition function $p^A(t, x, B)$, given by the semigroup

$$P_t^A f(x) = \mathbb{E}^x\{L_t f(X_t))\}, \quad x \in E,\ f \text{ Borel and positive.}$$

It is possible also to construct X_t^A by introducing a new "lifetime"

$$\varsigma^A = \inf\{t < \zeta : A_t \geq \zeta\},$$

where ζ is a random variable exponentially distributed with mean value 1 and independent of the Hunt process X_t with respect to P^x, for every $x \in E$. Then define

$$X_t^A = X_t, \quad \text{for } t < \varsigma^A$$

and $X_t^A = \infty$ otherwise. For details see Fukushima et al. [92, pag. 326].

The second process is given by

$$\check{X}_t = X_{\tau_t}, \qquad \tau_t(\omega) = \inf\{s > 0 : A_s(\omega) > t\}.$$

with its natural filtration $\check{\mathcal{F}}_t = \mathcal{F}_{\tau_t}$, for $t \geq 0$. It can proved that the process $\{\check{X}_t\}$ is a strong Markov process, with respect to the filtration $\{\check{\mathcal{F}}_t : t \geq 0\}$. Moreover, if A_t is a strictly positive continuous additive functional, then $\{\check{X}_t\}$ is a Hunt process.

There are other functionals (besides additive or multiplicative) that can be used to transform Markov processes, e.g., the first exit time from a region (usually a smooth open or closed subset of $\mathbb{R}^d$). For instance, if $\mathcal{O}$ is an open connected subset of $\mathbb{R}^d$ (which is the interior of its closure) then the procedure of stopping the process at the first exit time from the closure of $\mathcal{O}$, namely,

$$X_t^{\mathcal{O}} = X_{t\wedge\tau}, \qquad \tau(\omega) = \inf\{s \geq 0 : X_s(\omega) \in \mathbb{R}^d \setminus \mathcal{O}\},$$

produces a Hunt Markov process if the initial we were so. However, $X^{\mathcal{O}}$ may not be a Feller process when X is Feller. Note that the Hunt character of the process

has to do mainly with time-regularity of the paths, which is clearly preserved by the above functional (even if $\mathcal{O}$ is only a Borel set), while the Feller character involves the study of $\tau = \tau_x$ as a functional depending on the initial condition $X_0 = x$. The reader may consult, among other sources, the papers Stroock and Varadhan [239, 240] for a complete study on the (degenerate) diffusion processes with boundary conditions.

Perhaps, the prototype of Hunt processes are the diffusions process (with jumps) in infinite dimension or with boundary conditions. For instance, a Levy or Wiener process in infinite dimension does not produce a strongly continuous semigroup in C_0, even if the Feller property (i.e., mapping C_0 into itself) is satisfied. If we stop a degenerate diffusion in $\mathbb{R}^d$ at the first exit time from a smooth domain then we produce a simple example of a Markov process without the Feller property, which is stochastically continuous and produces a Hunt process. Moreover, piecewise deterministic processes (see Davis [56] and references therein) may not be Feller processes (this is mainly due to the nature of the boundary conditions or jump-mechanism, they may have predictable jumps, so they are not necessarily quasi left-continuous, see Section 3.9), but they are cad-lag strong Markov processes, actually, Ray and right processes. If a Markov transition function yields a Hunt process then we may expect that a sub-Markov transition functions (with the same degree of regularities) should yield a special standard process. On the other hand, most examples of sub-Markov transition functions are obtained form Markov processes and multiplicative functional, so that really they produce Hunt processes. More representative are stochastic differential equations with unbounded coefficients (e.g., see Stroock and Varadhan [241, Chapter 10, pp. 248–260], which may yield explosions or solutions with a finite lifetime, i.e., special standard processes which are not Hunt processes.

3.13 Random Orthogonal Measures

Before going further, we take a look at the $\mathcal{L}^p$ and L^p spaces, for $1 \leq p < \infty$. Let μ be a complete σ-finite measure on the measurable space $(S, \mathcal{B})$ and Π be a total π-system of finite measure sets, i.e., (a) if F and G belong to Π then $F \cap G$ also belongs to Π, (b) if F is in Π then $m(F) < \infty$, and (c) there exists a sequence $\{S_k : k \geq 1\} \subset \Pi$ such that $S_k \subset S_{k+1}$ and $S = \bigcup_k S_k$. For any measurable function f with values in the extended real line $[-\infty, +\infty]$ we may define the quantity

$$\|f\|_p := \Big(\int |f|^p \, \mathrm{d}\mu \Big)^{1/p},$$

which may be infinite. The set of step or elementary functions $\mathcal{E}(\Pi, \mu)$ is defined as all functions of the form

$$e := \sum_{i=1}^{n} c_i \mathbb{1}_{A_i},$$

where c_i are real numbers and A_i belongs to Π for every $i = 1, \ldots, n$, i.e., the function e assumes a finite number of non-zero real values on sets in Π. Denote

by $\mathcal{L}^p(\Pi,\mu)$ the sets of $\mathcal{B}$-measurable functions f with values in $[-\infty,+\infty]$ for which there exists a sequence $(e_1,e_2,\ldots)$ of step functions $\mathcal{E}(\Pi,\mu)$ such that $\|f-e_n\|_p\to 0$ as $n\to\infty$. Since

$$|f|^p\le 2^{p-1}|e_n|^p+2^{p-1}|f-e_n|^p,$$

all functions in $\mathcal{L}^p(\Pi,\mu)$ satisfies $\|f\|_p<\infty$, and in view of the triangle inequality

$$\|f+g\|_p\le\|f\|_p+\|g\|_p,\quad \forall f,g\in\mathcal{L}^p(\Pi,\mu),$$

the map $f\mapsto\|\cdot\|_p$ is a semi-norm. For $p=2$, we may use the bilinear form

$$(f,g):=\int f\,g\,\mathrm{d}\mu,\quad \forall f,g\in\mathcal{L}^2(\Pi,\mu)$$

as a semi-inner product, which yields the semi-norm $\|\cdot\|_2$.

If f,f_n belong to $\mathcal{L}^p(\Pi,\mu)$ and $\|f-f_n\|_p\to 0$ as $n\to\infty$ we say that f_n converges to f in $\mathcal{L}^p(\Pi,\mu)$. Also if f_m,f_n belong to $\mathcal{L}^p(\Pi,\mu)$ and $\|f_m-f_n\|_p\to 0$ as $m,n\to\infty$ we say that $\{f_n\}$ is a Cauchy sequence in $\mathcal{L}^p(\Pi,\mu)$. It is clear that any Cauchy sequence in $\mathcal{L}^p(\Pi,\mu)$ has a almost everywhere convergent sub-sequence. Next, essentially based on the triangular inequality and Fatou's Lemma, we deduce that $\mathcal{L}^p(\Pi,\mu)$ is a complete vector space, i.e., (1) for any f,g in $\mathcal{L}^p(\Pi,\mu)$ and any a,b in $\mathbb{R}$ the function $af+bg$ is in $\mathcal{L}^p(\Pi,\mu)$ and (2) any Cauchy sequence in $\mathcal{L}^p(\Pi,\mu)$ converges to a function in $\mathcal{L}^p(\Pi,\mu)$. Thus, if $\sigma_\mu(\Pi)$ is the smaller sub-σ-algebra of $\mathcal{B}$ containing Π and all μ-null sets then $\mathcal{L}^p(\Pi,\mu)=\mathcal{L}^p(\sigma_\mu(\Pi),\mu)$, after using a monotone class argument.

If we should identify functions which are equals almost everywhere, i.e., use classes of equivalence $f\sim g$ if and only if $f=g$ almost everywhere, then the quotient space $L^p(\Pi,\mu):=\mathcal{L}^p(\Pi,\mu)/_\sim$ is a Banach (Hilbert for $p=2$) space.

Definition 3.33 (random orthogonal measure)**.** A family of real-valued random variables $\{\zeta(A):A\in\Pi\}$ on a complete probability space $(\Omega,\mathcal{F},P)$ is called a *random orthogonal measure with structural measure* μ if

(a) $\mathbb{E}\{|\zeta(A)|^2\}<\infty$ for any A in Π,

(b) $\mathbb{E}\{\zeta(A)\,\zeta(B)\}=\mu(A\cap B)$ for any A,B in Π.

Note that the random variables $\zeta(A)$ are almost surely defined, i.e., they are elements in $L^2(\Omega,\mathcal{F},P)$, and the measure μ and the π-system Π are as above. □

Clearly, the above condition (b) translates the *orthogonal* condition, whist the word *measure* can be justified as follows: if A is a disjoint union of sets in Π, i.e., $A=\cup_i A_i$, $A_i\cap A_j=\emptyset$ if $i\neq j$, then

$$\Big(\mathbb{1}_A-\sum_i\mathbb{1}_{A_i}\Big)^2=\mathbb{1}_{A\smallsetminus\cup_i A_i}=\mathbb{1}_A-\sum_i\mathbb{1}_{A_i},$$

which yields

$$\mathbb{E}\Big\{\Big(\mathbb{1}_A-\sum_i\mathbb{1}_{A_i}\Big)^2\Big\}=\mu\Big(\mathbb{1}_{A\smallsetminus\cup_i A_i}\Big),$$

i.e., for each sequence $\{A_i\}$ as above, there exists a set Ω_0 in $\mathcal{F}$ with $P(\Omega_0) = 1$ such that $\zeta(A, \omega) = \sum_i \zeta(A_i, \omega)$ for every ω in Ω_0. This is not to say that a *regular* selection exists, i.e., to show that (except for set of probability zero) the mapping $A \mapsto \zeta(A)$ can be extended to a measure in $\sigma(\Pi)$, which involves a countable generated π-system Π and some topology on ω, as in the case of regular conditional probability measures.

Let as define the operator $e \mapsto I(e)$ from the set of elementary (or step) functions $\mathcal{E}(\Pi, \mu)$ into the Hilbert space $L^2(\Omega, \mathcal{F}, P) = L^2(\mathcal{F}, P)$ by the formula

$$\text{if} \quad e = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i} \quad \text{then} \quad I(e) = \int e \mathrm{d}\zeta := \sum_{i=1}^{n} c_i \, \zeta(A_i), \tag{3.92}$$

which is clearly independent of the particular representation of the given elementary function. Thus, we have

$$(I(e), \zeta(A))_{\mathcal{F}} = (e, \mathbb{1}_A)_{\Pi}, \quad \forall A \in \Pi,$$
$$\|I(e)\|_{2,\mathcal{F}} = \|e\|_{2,\Pi},$$

where $(\cdot, \cdot)_{\mathcal{F}}$ and $(\cdot, \cdot)_{\Pi}$ denote the inner or scalar products in the Hilbert spaces $L^2(\Omega, \mathcal{F}, P)$ and $L^2(\Pi, \mu) = L^2(\sigma_\mu(\Pi), \mu)$, respectively. Next, by linearity the above definition is extended to the vector space generated by $\mathcal{E}(\Pi, \mu)$, and by continuity to the whole Hilbert space $L^2(\Pi, \mu)$. Hence, this procedure constructs a linear isometry map between the Hilbert spaces $L^2(\Pi, \mu)$ and $L^2(\mathcal{F}, P)$ satisfying

$$\begin{cases} I : f \mapsto \displaystyle\int f \mathrm{d}\zeta, \quad \forall f \in L^2(\Pi, \mu), \\ (I(f), I(g))_{\mathcal{F}} = (f, g)_{\Pi}, \quad \forall f, g \in L^2(\Pi, \mu). \end{cases} \tag{3.93}$$

Certainly, there is only some obvious changes if we allow integrand functions with complex values, and if the spaces $L^p(\Pi, \mu)$ are defined with complex valued functions, and so, the inner product in L^2 need to use the complex-conjugation operation.

The above construction does not give a preferential role to the *time* variable as in the case of stochastic processes, and as mentioned in the book by Krylov [141, Section III.1, pp. 77-84], this procedure is used in several opportunities, not only for the stochastic integral. The interested reader may consult Gikhman and Skorokhod [98, Section V.2] for a detailed analysis on (vector valued) orthogonal measures.

3.13.1 Orthogonal or Uncorrelated Increments

Random orthogonal measures is a generalization of stochastic processes with orthogonal (or uncorrelated) increments, the reader is referred to the classic book Doob [59, Chapter IX, pp. 425–451] for more details. A $\mathbb{R}^d$-valued (for complex valued use conjugate) x is said to have *uncorrelated increments* if the increments are square-integrable and uncorrelated, i.e., if (a) $\mathbb{E}\{|x(t) - x(s)|^2\} <$

∞, for every $t > s \geq 0$ and (b) $\mathbb{E}\{(x(t_1) - x(s_1))\,(x(t_2) - x(s_2))\} = \mathbb{E}\{x(t_1) - x(s_1)\}\,\mathbb{E}\{x(t_2) - x(s_2)\}$ for any $0 \leq s_1 < t_1 \leq s_2 < t_2$. Similarly, x has *orthogonal increments* if $\mathbb{E}\{(x(t_1) - x(s_1))\,(x(t_2) - x(s_2))\} = 0$. It is clear that a process with independent increments is also a process with uncorrelated increments and that we may convert a process x with uncorrelated increments into a process with orthogonal (and uncorrelated) increments y by subtraction its means, i.e., $y(t) = x(t) - \mathbb{E}\{x(t)\}$. Thus, we will discuss only orthogonal increments.

If y is a process with orthogonal increments then we can define the (deterministic) monotone increasing function $F_y(t) = \mathbb{E}\{|y(t) - y(0)|^2\}$, for any $t \geq 0$, with the property that $\mathbb{E}\{|y(t) - y(s)|^2\} = F_y(t) - F_y(s)$, for every $t \geq s \geq 0$. Because the function F_y has a countable number of discontinuities, the mean-square left $y(t-)$ and right $y(t+)$ limit of $y(t)$ exist at any $t \geq 0$ and, except for a countable number of times $y(t-) = y(t) = y(t+)$. Therefore, we can define real-valued random variables $\{\zeta(A) : A \in \Pi_+\}$, where Π_+ is the π-system of semi-open intervals $(a, b]$, $b \geq a \geq 0$ and

$$\zeta(A) = y(b+) - y(a+), \quad A = (a, b],$$

which is a random orthogonal measure with structural measure μ, the Lebesgue-Stieltjes measure generated by F_y, i.e., $\mu(A) = F_y(b+) - F_y(a+)$, for any $A = (a, b]$. Certainly, we may use the π-system Π_- of semi-open intervals $[a, b)$, $b \geq a \geq 0$ and $\zeta(A) = y(b-) - y(a-)$, with $A = [a, b)$, or the combination of the above π-system, and we get the same structural measure (and same extension of the orthogonal measure ζ). Moreover, we may even use only the π-system of interval of the form $[0, b)$ (or $(0, b]$) to initially define the random orthogonal measure.

Now, applying the previous we can define the stochastic integral for any (deterministic) function in $L^2(\sigma_\mu(\Pi), \mu)$

$$\int_{\mathbb{R}} f(t)\mathrm{d}y(t) = \int f\mathrm{d}\zeta$$

as an equivalence class of square-integrable random variables, even if we actually think of a particular member. Moreover, the way how this is defined (via limit of elementary or step functions) allows us to that the stochastic integral process

$$\Phi(s) = \int_{\mathbb{R}} \varphi(s, t)\mathrm{d}y(t)$$

can be chosen measurable if φ is a measurable function with respect to the Lebesgue-Stieltjes measure $\mathrm{d}s\,\mathrm{d}F(t)$ satisfying

$$\int_{\mathbb{R}} \varphi(s, t)\mathrm{d}F(t),$$

all s except in a set of zero Lebesgue measure. Clearly, the stochastic integral over a Borel (even μ-measurable) set of time A can be define as

$$\int_A f(t)\mathrm{d}y(t) = \int f\mathbb{1}_A\mathrm{d}\zeta.$$

A Fubini type theorem holds, for the double integral, and in particular, if h is an absolutely continuous function and $\mathbb{1}_{\{s\leq t\}}$ denotes the function equal to 1 when $s \leq t$ and equal to 0 otherwise, then exchanging the order of integration we deduce

$$\int_a^b \mathrm{d}s \int_{(a,b]} h'(s)\mathbb{1}_{\{s\leq t\}}\mathrm{d}y(s) = \int_{(a,b]} [h(t)-h(a)]\mathrm{d}y(t) = \\ = [h(b)-h(a)][y(b+)-y(a+)] - \int_a^b [y(t)-y(a+)]\mathrm{d}t,$$

for any $b > a \geq 0$.

3.13.2 Typical Examples

There two typical constructions of random orthogonal measures, based on the Poisson and the Gaussian distributions, or equivalent on the Poisson process and the Wiener process, both are processes with independent increments.

Perhaps a simple (constructed) example of a random orthogonal measure begins with a given (structural) finite measure m on $S = \mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$, where the π-system Π plays almost not role. Let $\{\tau_n, z_n : n \geq 1\}$ be a sequence of independent random variables in a probability space $(\Omega, \mathcal{F}, P)$, such that each τ_n is exponentially distributed with parameter $m(\mathbb{R}^d_*)$ and z_n has the distribution law $A \mapsto m(A)/m(\mathbb{R}^d_*)$. Define the compound Poisson process $p_t = \sum_n z_n\, \mathbb{1}_{t\geq\theta_n}$, with $\theta_n = \tau_1 + \cdots + \tau_n$. This can be written as $p_t = \sum_{n=1}^{N_t} z_n$, where $N_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is the Poisson process counting the jumps, which has a Poisson distribution with intensity $\lambda = m(\mathbb{R}^d_*)$, i.e., $P\{N_t = n\} = \mathrm{e}^{-\lambda t}(\lambda t)^n/n!$, $n = 0, 1, 2, \ldots$, and thus $E\{N_t\} = \lambda t$ and $E\{(N_t - \lambda t)^2\} = \lambda t$.

If the emphasis is only on the jumps then the series defining the Poisson process p_t is regarded as the sum-of-jumps of the sequence of jumps $\{z_n, \theta_n : n \geq 1\}$, which is referred to as a Poisson point process, where z_n is the size of the jump at the time θ_n. Note that if initially the measure m is given on $\mathbb{R}^d$ and $m(\{0\}) \neq 0$ then the above expression of N_t does not count the actual jumps of the compound Poisson process p_t, i.e., the random process $q_t = \sum_n \mathbb{1}_{z_n=0}\mathbb{1}_{t\geq\theta_n}$ intervenes.

The independence of the random variables $\{z_n\}$ and $\{\theta_n\}$ and the fact all random variables z_n have the same distribution, imply that

$$\mathbb{E}\{p_t\} = \mathbb{E}\{z\}\sum_n \mathbb{E}\{\mathbb{1}_{t\geq\theta_n}\} = m(z)t,$$

where $m(z)$ means the integral of the function $z \mapsto z$ with respect to the measure m, i.e., $m(z) = \mathbb{E}\{z_1\}m(\mathbb{R}^d_*)$. Similarly, if $m(|z|^2) = \mathbb{E}\{|z_1|^2\}m(\mathbb{R}^d_*)$ then more calculations show that the variance $\mathbb{E}\{|p_t - m(z)t|^2\} = m(|z|^2)t$, and also

$$\mathbb{E}\{\mathrm{e}^{irp_t}\} = \exp\big[m(\mathbb{R}^d_*)t\big(\mathbb{E}\{\mathrm{e}^{irz_1}\} - 1\big)\big] = \exp\big[tm(\mathrm{e}^{irz} - 1)\big]$$

is its characteristic function. Moreover, these distributions also imply that

$$\mathbb{E}\{\mathbb{1}_{z_n\in A}\} = \frac{m(A)}{m(\mathbb{R}^d_*)} \qquad \text{and} \qquad \sum_k \mathbb{E}\{\mathbb{1}_{\theta_{n+k}\leq t}\} = m(\mathbb{R}^d_*)t,$$

for every $t \geq 0$ and A in Π. Therefore, this yields the Poisson orthogonal measure

$$\zeta_t(A) = \sum_n \big[\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n} - \mathbb{E}\{\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n}\}\big], \quad \forall A \in \Pi.$$

Indeed, by construction $\mathbb{E}\{\zeta_t(A)\} = 0$, $\sum_n \mathbb{E}\{\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n}\} = m(A)t$, and

$$\mathbb{E}\{\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n}\, \mathbb{1}_{z_k \in B}\, \mathbb{1}_{t \geq \theta_k}\} = \frac{m(A)\, m(B)}{m(\mathbb{R}^d_*)}\, \mathbb{E}\{\mathbb{1}_{t \geq \theta_{n \vee k}}\},$$
$$\mathbb{E}\{\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n}\, \mathbb{1}_{z_n \in B}\, \mathbb{1}_{t \geq \theta_n}\} = \frac{m(A \cap B)}{m(\mathbb{R}^d_*)}\, \mathbb{E}\{\mathbb{1}_{t \geq \theta_n}\}, \quad \forall n, k,$$

and because $\sum_{n,k} = \sum_n + 2\sum_n \sum_{k=n+1}$ we have

$$\sum_{n,k} \mathbb{E}\{\mathbb{1}_{z_n \in A}\, \mathbb{1}_{t \geq \theta_n}\, \mathbb{1}_{z_k \in B}\, \mathbb{1}_{t \geq \theta_k}\} = m(A \cap B)t + 2m(A)m(B)t$$

which yields

$$\mathbb{E}\{\zeta_t(A)\zeta_t(B)\} = \sum_n \mathbb{E}\{\mathbb{1}_{z_n \in A \cap B}\, \mathbb{1}_{t \geq \theta_n}\} = t\, m(A \cap B),$$

as desired. Recall that the mapping $A \mapsto \zeta(A, \omega)$ is regarded as defined on for any A in Π and taking values in $L^2(\Omega, \mathcal{F}, P)$, i.e., properly saying the symbol $\zeta(A)$ is a class of equivalence of square-integrable random variables.

In general, if m is measure in $\mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$ that integrates the function $x \mapsto 1 \wedge |x|^2$ and $\{R_k : k \geq 1\}$ is a countable partition of finite m-measure, i.e., $\mathbb{R}^d_* = \bigcup_k R_k$ with $m(R_k) < \infty$ and $R_k \cap R_n = \emptyset$, for $k \neq n$, then we repeat the previous procedure with the finite measure $A \mapsto m(A \cap R_k)$ to construct an independent sequence of compound Poisson processes $\{p_t(R_k) : k \geq 1\}$, which yields the independent sequence Poisson orthogonal measures $\{\zeta_t(R_k) : k \geq 1\}$. Since $\mathbb{E}\{\zeta_t(R_k)\} = 0$, the sequence of Poisson orthogonal measures is an orthogonal system in $L^2(\Omega, \mathcal{F}, P)$, and so the series $\zeta_t(A) = \sum_k \zeta_k(A)$, for every A in Π, defines a Poisson orthogonal measure with structural measure $A \mapsto t\, m(A)$. Summing-up, if for a fixed $k = 1, 2, \ldots$, $(N^k_t, t \geq 0)$ is the Poisson process and $\{z^k_n : n \geq 1\}$ is the iid sequence with distribution $m(\cdot \cap R_k)$ then the compound Poisson processes $p_t(R_1), p_t(R_2), \ldots$ are independent and the series of jumps $\sum_k \sum_{n=1}^{N^k_t} z^k_n$ defines a Poisson point process with Lévy measure m, which yields the same Poisson orthogonal measure, namely,

$$\zeta_t(A) = \sum_k \Big[\sum_{n=1}^{N^k_t} \mathbb{1}_{z^k_n \in A} - \mathbb{E}\big\{ \sum_{n=1}^{N^k_t} \mathbb{1}_{z^k_n \in A} \big\} \Big], \quad \forall A \in \Pi,\ t \geq 0, \tag{3.94}$$

where the series (in the variable k) converges in the L^2-norm, i.e., for each k the series in n reduces to a finite sum for each ω, but the series in k defines $\zeta_t(A)$ as an element in $L^2(\Omega, \mathcal{F}, P)$. Note that in this construction, the variable t is considered fixed, and that $A \mapsto \mu(A) = tm(A)$ is the structural measure

associated with the Poisson orthogonal measure $A \mapsto \zeta_t(A)$. Therefore, any square-integrable (deterministic) function f, i.e., any element in $L^2(\Pi, \mu) = L^2(\sigma_\mu(\Pi), \mu)$.

As seen in the previous section, any process with orthogonal increments yields a random orthogonal measure, in particular, a one-dimensional standard Wiener process $(w(t), t \geq 0)$ (i.e., $w(t)$ is a standard normal variables, $t \mapsto w(t)$ is almost surely continuous, and $\mathbb{E}\{w(t) \wedge w(s)\} = t \wedge s$) has independent increments and thus the expression $\zeta(]a, b]) = w(b) - w(a)$ defines a random orthogonal measure on the π-system of semi-open intervals $\Pi_+ = \{]a, b] : a, b \in \mathbb{R}\}$ with the Lebesgue measure as its structural measure, i.e., $\mathbb{E}\{\zeta(]a, b])\} = b - a$.

Similarly, the Poisson orthogonal measure $\zeta_t(A)$ defined previously can be regarded as a random orthogonal measure on π-system Π (which is composed by all subsets of $S = \mathbb{R}^d_* \times (0, \infty)$ having the form $K \times (0, t]$ for a compact set and a real number $t \geq 0$) with structural measure $\mu = m \times \mathrm{d}t$, where $\mathrm{d}t$ is the Lebesgue measure.

With this argument, we are able to define the stochastic integral of an (deterministic) integrand function $L^2(\sigma_\mu(\Pi), \mu)$ with respect to a random orthogonal measure constructed form either a Poisson point process with Lévy measure m or a (standard) Wiener process, which are denoted by either

$$\zeta(K \times (0, t]) = \tilde{p}(A \times (0, t]), \quad \text{and} \quad \int_{\mathbb{R}^d_* \times]0,T} f(t) \tilde{p}(\mathrm{d}z, \mathrm{d}t),$$

or

$$\zeta(]a, b]) = w(]a, b]), \quad \text{and} \quad \int_a^b f(t) \mathrm{d}w(t).$$

Note that this is not a pathwise integral, e.g., the paths of the Wiener process are almost surely of unbounded variation on any bounded time-interval and something similar holds true for the Poisson point process depending on the Lévy measure.

Perhaps a simple construction of a Wiener process begins with a sequence of independent standard normally distributed random variables $\{e_{i,n} : i = 1, 2, \ldots, 4^n, n \geq 1\}$. Since each $e_{i,n}$ has zero mean and are independent of each other, the sequence is orthogonal in $L^2 = L^2(\Omega, \mathcal{F}, P)$, actually, it is an orthonormal system since all variances are equal to 1. Recalling the dyadic expressions that if $t = k2^{-m} = (k2^{n-m})2^{-n}$, $1 \leq k \leq 4^m$ then $k2^{n-m} \leq 4^n$, $\mathbb{1}_{i2^{-n} \leq t} = 1$ if and only if $i = 1, \ldots, k2^{n-m}$, which yields $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = k2^{n-m} = t2^n$ if $k2^{n-m} = t2^n \geq 1$, we deduce $t = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t}$, so that the random variable

$$w(t) = \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq t}, \tag{3.95}$$

is defined as a convergent series in $L^2(\Omega, \mathcal{F}, P)$, for every $t > 0$. Indeed, regard the expression as an orthogonal series expansion, and set $w(0) = 0$, for any

$t \geq s \geq 0$, to have

$$\mathbb{E}\{|w(t) - w(s)|^2\} = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{E}\{|e_{i,n}|^2\} \mathbb{1}_{s<i2^{-n}\leq t} = \\ = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{s<i2^{-n}\leq t} = (t-s).$$

Thus, $t \mapsto w(t)$ provides a L^2-norm continuous random process satisfying (a) $w(t)$ is a Gaussian random variable with mean $\mathbb{E}\{w(t)\} = 0$ and variance $\mathbb{E}\{|w(t)|^2\} = t$, and (b) $w(s)$ is independent of $w(t) - w(s)$ for every $t > s$. The fact that there is a continuous version of the limiting process $(w(t) : t \geq 0)$, which is called a Wiener process plays not an important role in this analysis. Indeed, the expressions (3.95) of a Wiener process and (3.94) of a centered Poisson point process are cad-lag and therefore, the corresponding random orthogonal measures are measures, for every fixed ω almost surely.

Certainly, for dimension d higher than 1 we should use the covariance matrix, i.e., $\mathbb{E}\{w(t)w^*(t)\} = tI_d$, with I_d the identity matrix. In this case, this could take us to discuss vector-valued random orthogonal measure, or simply consider a sum of independent Wiener processes and their corresponding orthogonal measures.

However, with little effort, an index $j = 1, \ldots, d$ could be added to the iid sequence $\{e^j_{i,n}\}$, so that d-intervals $(0,t] = (0,t_1] \times \cdots \times (0,t_d]$ on $S = (0,+\infty)^d$ could be used to define

$$w(t_1, \ldots, t_d) = \sum_n 2^{-n} \sum_{i=1}^{4^n} \sum_{j=1}^{d} e^j_{i,n} \mathbb{1}_{i2^{-n}\leq t_1, \ldots, i2^{-n}\leq t_d}, \tag{3.96}$$

as a convergent series in $L^2(\Omega, \mathcal{F}, P)$. Besides being a Gaussian random variable with mean zero, note that

$$\mathbb{1}_{i2^{-n}\leq t_1} \cdots \mathbb{1}_{i2^{-n}\leq t_d} = \mathbb{1}_{i2^{-n}\leq t_1, \ldots, i2^{-n}\leq t_d}$$

implies

$$\mathbb{E}\{|w(t_1, \ldots, t_d)|^2\} = \sum_n 4^{-n} \sum_{i=1}^{4^n} \sum_{i=1}^{d} \mathbb{E}\{|e^j_{i,n}|^2\} \mathbb{1}_{i2^{-n}\leq t_1, \ldots, i2^{-n}\leq t_d} = \\ = \prod_{j=1}^{d} \Big[\sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n}\leq t_j} \Big] = \prod_{j=1}^{d} t_j,$$

which yields the (random) Gaussian orthogonal measure $\zeta(]0,t]) := w(t_1, \ldots, t_d)$ in $\mathbb{R}^d$, with the Lebesgue measure on $(0,\infty)^d$.

Clearly, this last example is related with the so-called *white noise measure*, and *Brownian sheet* or *space-time* Brownian motion, e.g., see Kallianpur and Xiong [123, Section 3.2, pp. 93–109].

3.13.3 Filtration and Martingales

At this point, only deterministic integrand can be taken when the integrator is a standard Wiener process or a Poisson point process with Lévy measure m. To allow stochastic integrand a deeper analysis is needed to modify the π-system. Indeed, the two typical examples of either the Poisson or the Gaussian orthogonal measure suggests a π-system of the form either $\Pi = \{K\times]0,\tau] \subset \Omega\times\mathbb{R}^d_*\times(0,\infty)\}$, with the structural measure $\mu(K\times]0,\tau]) = \mathbb{E}\{\tau\}m(K)$ for the underlying product measure $P\times m\times \mathrm{d}t$, or $\Pi = \{]0,\tau] \subset \Omega\times(0,\infty)\}$, with the structural measure $\mu(]0,\tau]) = \mathbb{E}\{\tau\}$ for the underlying product measure $P\times \mathrm{d}t$, for a compact set K of $\mathbb{R}^d_*$ and a *bounded stopping time* τ. This means that there is defined a filtration $\mathbb{F}$ in the probability space $(\Omega,\mathcal{F},P)$, i.e., a family of sub σ-algebras $\mathcal{F}_t \subset \mathcal{F}$ such that (a) $\mathcal{F}_t \subset \mathcal{F}_s$ if $s > t \geq 0$, (b) $\mathcal{F}_t = \bigcap_{s>t}\mathcal{F}_s$ if $t \geq 0$, and (c) N belongs to $\mathcal{F}_0$ if N is in $\mathcal{F}$ and $P(N) = 0$. Therefore, of relevant interest is to provide some more details on the square-integrable functions that can be approximated by a sequence of Π-step functions, i.e., the Hilbert space $L^2(\Pi,\mu)$ or better $\mathcal{L}^2(\Pi,\mu)$.

This filtration $\mathbb{F}$ should be such that either $t \mapsto \zeta(K\times]0,t])$ or $t \mapsto \zeta(]0,t])$ is a $\mathbb{F}$-martingale. Because, both expressions (3.95) of a Wiener process and (3.94) of a Poisson point process have zero-mean with independent increments, the martingale condition reduces to either $\zeta(K\times]0,t])$ or $\zeta(]0,t])$ being $\mathcal{F}_t$-measurable, i.e., adapted to the filtration $\mathbb{F}$.

Under this $\mathbb{F}$-martingale condition, either the Poisson or the Gaussian orthogonal measure can be considered as defined on the above Π with structural (product) measure either $P\times m\times \mathrm{d}t$ or $P\times \mathrm{d}t$, i.e., just replacing a deterministic time t with a bounded stopping time τ. All this requires some work. In particular, a key role is played by the so-called predictable σ-algebra $\mathcal{P}$ in either $\Omega\times\mathbb{R}^d_*\times(0,\infty)$ or $\Omega\times(0,\infty)$, which is the σ-algebra generated by the π-system Π, and eventually completed with respect to the structural measure μ. For instance, in this setting, a real-valued process $(f(t), t \geq 0)$ is an integrand (i.e., it is an element in $L^2(\Pi,\mu)$) if and only if (a) it is square-integrable, (i.e., it belongs to $L^2(\mathcal{F}\times\mathcal{B},\mu)$, $\mathcal{B}$ is the Borel σ-algebra either in $\mathbb{R}^d_*\times]0,\infty[$ or in $]0,\infty[$), and (b) its μ-equivalence contains a predictable representative. In other words, square-integrable predictable process are the good integrand, and therefore its corresponding class of μ-equivalence. Sometimes, stochastic intervals are denoted by $\rrbracket a,b\rrbracket$ (or $[\![a,b[\![$) to stress the randomness involved. Certainly, this argument also applies to the multi-dimensional Gaussian orthogonal measures (or Brownian sheet). On the other hand, the martingale technique is used to define the stochastic integral with respect to a martingale (non-necessarily with orthogonal), and various definitions are proposed. In any way, the stochastic integral becomes very useful due to the stochastic calculus that follows.

Among other sources, regarding random orthogonal measures and processes, the reader may consult the books by Krylov [141, Section III.1, pp. 77-84], Doob [59, Section IX.5, pp. 436–451], Gikhman and Skorokhod [98, Section V.2] and references therein for a deeper analysis.

Part II

Stochastic Differential Equations with Jumps

Chapter 4

Stochastic Calculus

Let $(X_t = X(t,\omega) : t \geq 0)$ be a family of $\mathbb{R}^d$-valued random variables in a given (complete) probability space $(\Omega, \mathcal{F}, P)$. Certainly, measurability is a first difficulty encountered, i.e., when the whole family is bounded we expect that an operation like $\sup_{t\geq 0} X_t$ or $\inf_{t\geq 0} X_t$ should produce a random variable, which is not the case in general. A way to fix this problem is to use a *separable* family of random variables, i.e., there exist a countable dense subset I of $[0,\infty)$ and a null measurable set N, $P(N) = 0$, such that for every open subset O of $[0,\infty)$ and any closed closed subset C of $\mathbb{R}^d$ the set $\{\omega \in \Omega : X_t(\omega) \in C,\ \forall t \in O \setminus I\}$ is a subset of N. In this approach, instead of working in the continuum set of indexes $[0,\infty)$ we are actually working in a countable set I, which is called a separant. Many situations are resolved with this analysis, but any limit operation in the variable t forces us to complete the space and go back to $[0,\infty)$. Moreover, this condition depends on the particular version used, i.e., if X is separable and Y is another process satisfying $P\{X_t = Y_t\} = 1$ for every $t \geq 0$ then Y may not not be separable. However, we can prove that any process X has a separable version Y, but Y have values in the one-point compactification $\mathbb{R}^d \cup \{\infty\}$, with $P\{Y(t) = \infty\} = 0$ for every t.

To make aware the reader of some difficulties that may arrive in the theory of general processes, we discuss some initial issues. A *(stochastic) process* is a family of $\mathbb{R}^d$-valued random variables where some regularity in the t-variable index have been imposed. For instance:

(1) a stochastically (left or right) continuous process $(X_t : t \geq 0)$ satisfies $X_s \to X_t$ in probability as $s \to t$ ($s < t$ or $s > t$ for left or right continuous) for any $t \geq 0$, or

(2) a (left or right) continuous process has almost surely (left or right) continuous paths, i.e., $t \mapsto X_t(\omega)$ is (left or right) continuous for any ω outside of a null set.

Note that condition (2) implies condition (1). However, these two conditions are essentially very different one from the other. Condition (1) is intrinsic to the joint finite-dimensional distributions of the family of random variables $\{X(t) :$

$t \geq 0\}$ and therefore remains valid for any version of the process $(X_t : t \geq 0)$, while condition (2) is attached to the particular version of the process, say a *pathwise* condition. In the first case (1), we are looking at the process as a function from $[0, \infty)$ into the set of $\mathbb{R}^d$-valued random variables, while in the second case (2), we have random variables with values in the space of (left or right) continuous $\mathbb{R}^d$-valued functions, almost surely. Both concept are intended to address the difficulty presented by the fact that the conditions

$$\text{(a)} \quad P\{X_t = Y_t\} = 0, \ \forall t \geq 0,$$

$$\text{(b)} \quad P\{X_t = Y_t, \ \forall t \geq 0\} = 0,$$

are not equivalent, since t ranges on a uncountable set. If both processes $(X_t : t \geq 0)$ and $(Y_t : t \geq 0)$ are left or right continuous (or separable) then (a) and (b) are actually equivalent. Indeed, take a countable dense set I and consider the event $N = \bigcup_{t \in I} \{\omega : X_t(\omega) \neq Y_t(\omega)\}$ for two processes satisfying (a). Since the union is countable, $P(N) = 0$ and the continuity of their paths imply that $X_t(\omega) = Y_t(\omega)$ for any ω in $\Omega \setminus N$ and any t. If both processes are only separable (see Definition 1.1) then we take $I = I_X \cup I_Y$ (where I_X or I_Y are a separant set associated with X or Y) and proceed along the line of the previous argument.

On the other hand, if the processes are only stochastically right (or left) continuous then (a) and (b) may not be equivalent. However, a simple argument shows that given a separable stochastically right (or left) continuous process X then any countable dense set is separant. Indeed, for any countable dense set $I = \{t_1, t_2, \ldots\}$ we can find a sequence of positive numbers $\{\delta_1, \delta_2, \ldots\}$ such that $P\{|X(t) - X(t_n)| \geq 2^{-n}\} < 2^{-n}$ for any t in $[t_n, t_n + \delta_n]$. By the Borel-Cantelli lemma the set

$$N_t := \bigcap_m \bigcup_{n \geq m} \{\omega : \ |X(t, \omega) - X(t_n, \omega)| \geq 2^{-n}\}$$

has probability zero. Since $\mathbb{R} = \cup_n [t_n, t_n + \delta_n]$, for any t in $\mathbb{R}$ and any ω in $\Omega \setminus N_t$ there is a sequence of indexes in I such that $X(t_k, \omega)$ converges to $X(t, \omega)$. Because X is separable, there is countable dense set J and null set N, $P(N) = 0$ such that for any t in $\mathbb{R}$ and ω in $\Omega \setminus N$ the previous convergence holds with indexes in J. Therefore, for ω outside of the null set $\bar{N} = N \cup_{t \in J} N_t$, we can find a is a sequence of indexes in I such that $X(t_k, \omega)$ converges to $X(t, \omega)$. Moreover, for the given process X, this argument shows that there exists a separable process Y satisfying (a), but not necessarily (b). Indeed, it suffices to define $Y_t(\omega) = X_t(\omega)$ for any t and ω such that ω belongs to $\Omega \setminus N_t$ and $Y_t(\omega) = 0$ otherwise.

In a typical example we consider the Lebesgue measure on $[0, 1]$, two processes $X_t(\omega) = t$ for any t, ω in $[0, 1]$ and $Y_t(\omega) = t$ for $\omega \neq t$ and $Y_t(\omega) = 0$ otherwise. It is clear that condition (a) is satisfied, but (b) does not hold. The process X is continuous (as in (2), sometimes referred to as pathwise continuity), but Y is only stochastically continuous (as in (1), sometimes referred to as continuous in probability), since is clearly almost sure continuous. Also, note

that a stochastic process $(X_t : t \geq 0)$ is (right or left) continuous if its restriction to a separant set is so.

Another key issue is the filtration, i.e., a family of sub σ-algebras $(\mathcal{F}_t : t \geq 0)$ of $\mathcal{F}$, such that $\mathcal{F}_s \subset \mathcal{F}_t$ for every $t > s \geq 0$. As long as the probability P is unchanged, we may complete the $\mathcal{F}$ and $\mathcal{F}_0$ with all the subsets of measure zero. However, in the case of Markov processes, the probability $P = P_\mu$ depends on the initial distribution μ and the *universally completed* filtration is used to properly express the *strong Markov property*. On the other hand, the right-continuity of the filtration, i.e., the property $F_t = F_{t+}$, for every $t \geq 0$, where $\mathcal{F}_{t+} := \cap_{s>t} F_s$, is a desirable condition at the point that by *filtration* we understand a right-continuous increasing family of sub σ-algebras $(\mathcal{F}_t : t \geq 0)$ of $\mathcal{F}$ as above. Usually, the filtration $(\mathcal{F}_t : t \geq 0)$ is attached to a stochastic process $(X_t : t \geq 0)$ in the sense that the random variables $(X_s : s \leq t)$ are $\mathcal{F}_t$-measurable. The filtration generated by a process (or the *history* of the process, i.e, $\mathcal{F}_t = \mathcal{H}_t$ is the smaller sub σ-algebra of $\mathcal{F}$ such that all random variables $(X_s : s \leq t)$ are measurable) represents the *information* obtained by observing the process. The new information is related to the *innovation*, which is defined as the decreasing family of sub σ-algebras $(\mathcal{I}_t : t \geq 0)$, where $\mathcal{I}_t = \mathcal{F}_t^\perp$ is the smaller sub σ-algebra of $\mathcal{F}$ containing all set independent of $\mathcal{F}_t$, i.e., a bounded function f is $\mathcal{F}_t^\perp$-measurable if and only if $\mathbb{E}\{f\, g\} = \mathbb{E}\{f\}\,\mathbb{E}\{g\}$ for any integrable g in $\mathcal{F}_t$-measurable. Hence, another stochastic process $(Y_t : t \geq 0)$ is called *adapted* if Y_t is $\mathcal{F}_t$-measurable for every $t \geq 0$ and *non-anticipating* (or *non-anticipative*) if Y_t is independent of the innovation $\mathcal{I}$, which is equivalent to say that Y_t is $\mathcal{I}_t^\perp$-measurable or $\mathcal{F}_t^{\perp\perp}$-measurable, i.e., $\mathbb{E}\{\varphi(Y_t)\, g\} = \mathbb{E}\{\varphi(Y_t)\}\,\mathbb{E}\{g\}$ for any bounded real Borel measurable function φ and any integrable g satisfying $\mathbb{E}\{f\, g\} = \mathbb{E}\{f\}\,\mathbb{E}\{g\}$ for every integrable f which is $\mathcal{F}_t$-measurable. Notice that the filtration $(\mathcal{F}_t : t \geq 0)$, the process or the concept adapted can be defined in a measurable space $(\Omega, \mathcal{F})$, but the innovation $(\mathcal{I}_t : t \geq 0)$ or the concept of non-anticipative requires a probability space $(\Omega, \mathcal{F}, P)$, which involves the regularity in the t-variable index discussed above. Thus, for a *filtered* space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, we understand a probability space endowed with a filtration, which is always right-continuous. As long as P is fixed, we may assume that $\mathcal{F}_0$ is complete, even more that $\mathcal{F}_t = \mathcal{F}_t^{\perp\perp}$ for every $t \geq 0$ and $\mathcal{F} = \vee_{t\geq 0}\mathcal{F}_t$. Sometimes we may change the probability P, but the filtration may change only when the whole measurable space is changed, except that it may be completed with all null sets as needed.

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$ be a filtered space. A minimum condition we require of a stochastic process is to be *measurable*, i.e., the function $(t, \omega) \mapsto X(t, \omega)$ is measurable with respect to the product σ-algebra $\mathcal{B} \times \mathcal{F}$, where $\mathcal{B} := \mathcal{B}([0, \infty[)$ is the Borel σ-algebra in $[0, \infty[= [0, \infty)$. When general processes are involved, subsets N of $\mathcal{B} \times \mathcal{F}$ with the property that $P\big(\cup_{t\geq 0}\, \{\omega : (t, \omega) \in N\}\big) = 0$ are called *evanescent* and two processes which differ in an evanescent set are considered equals, every concepts and results are valid *except an evanescent set*, without special mention. As mentioned above, if the processes have some extra *path regularity*, such as separable or stochastically left (or right) continuous, then this is the same as *modification* or *version* of the process. However, the

standard technique is to make a regular (e.g., cad-lag) modification of a general process and refer always to this version. Related to the adapted processes are the *progressively measurable* processes, which are stochastic processes such that the function $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to the product σ-algebra $\mathcal{B}([0,T]) \times \mathcal{F}$, when considered as a mapping from $\Omega \times [0,T]$, for every T in $[0,\infty[$. There are a couple of useful sub σ-algebras of $\mathcal{B} \times \mathcal{F}$:

(1) the *predictable* σ-algebra $\mathcal{P}$, generated by sets of the form $\{0\} \times F_0$ and $(s,t] \times F_s$ for any F_s in $\mathcal{F}_s$, any $t > s \geq 0$

(2) the *optional* (or *well measurable*) σ-algebra $\mathcal{O}$, generated by sets of the form $\{0\} \times F_0$ and $[s,t) \times F_s$ for any F_s in $\mathcal{F}_s$, any $t > s \geq 0$.

For the sake of convenience and as long as no confusion may arrive, we may exchange the order of the variables t and ω so that $\Omega \times [0,\infty)$ or $[0,\infty) \times \Omega$ are regarded as the same. Clearly $\mathcal{P} \subset \mathcal{O} \subset \mathcal{B}([0,\infty[) \times \mathcal{F}$, where in general the inclusions are strict. It can be proved that $\mathcal{P}$ is the σ-algebra generated by continuous (or left continuous) adapted processes, and that $\mathcal{O}$ is generated by right continuous (or cad-lag) adapted processes. A stochastic process X is called predictable (or optional) if the function $(t,\omega) \mapsto X(t,\omega)$ is measurable with respect to $\mathcal{P}$ (or $\mathcal{O}$). However, a $\mathcal{F}$-measurable function from Ω into $[0,\infty]$ is called an optional (or stopping) time if $\{\tau \leq t\}$ (or $\{\tau < t\}$ because $\mathcal{F}_t = \mathcal{F}_{t+}$) is in $\mathcal{F}_t$ for every $t \geq 0$ and $\mathcal{F}_\tau$ is the σ-algebra of all sets A in $\mathcal{F}_\infty := \vee_{t\geq 0}\mathcal{F}_t$ such that $A \cap \{\tau \leq t\}$ belongs to $\mathcal{F}_t$ for every $t \geq 0$. If τ and θ are optional times then stochastic intervals of the form $[0,\tau]$ and $(\theta,\tau]$ are predictable. A stopping time is called predictable if there exists a (announcing) sequence of stopping time $\{\tau_1 \leq \tau_2 \leq \tau_k < \tau\}$ convergent to τ. It can be proved that τ is optional (or predictable) if and only if the function $(t,\omega) \mapsto \mathbb{1}_{t\geq\tau}$ is an optional (or predictable) process. Notice that if two processes X and Y are equals except in an evanescent set then X is predictable (or optional or progressively measurable or adapted) if and only if Y is so. These measurability properties are not preserved when using versions of the same process. For instance, if X is a stochastically left continuous adapted process then for every $t, \varepsilon > 0$ there exists $\delta = \delta(t,\varepsilon)$ such that $P\{|X(t) - X(s)| \geq \varepsilon\} \leq \varepsilon$, for any s in $[t-\delta, t]$. Thus, for every sequence of partitions $\pi_n := \{0 = t_{0,n} < t_{1,n} < \cdots < t_{k,n} < \cdots\}$, with $\sup_k(t_{k,n} - t_{k-1,n})$ vanishing as $n \to \infty$, we can define

$$X_n(t,\omega) := \begin{cases} X(0,\omega) & \text{if } t = 0, \\ X(t_{k-1,n},\omega) & \text{if } t_{k-1,n} < t \leq t_{k,n}, \quad k \geq 1. \end{cases}$$

It is clear that X_n is predictable and so is the subset A of $\Omega \times [0,\infty)$, where the sequence $X_n(t,\omega)$ is convergent is also predictable. Therefore the limit

$$Y(t,\omega) := \begin{cases} \lim_n X_n(t,\omega) & \text{for } (t,\omega) \in A, \\ 0 & \text{otherwise,} \end{cases}$$

is also a predictable process. By Borel-Cantelli lemma the set

$$N_t := \bigcap_m \bigcup_{n\geq m} \{\omega : \exists\, k \text{ such that } t_{k-1,n} < t \leq t_{k,n}, \\ , |X(t,\omega) - X(t_{k,n},\omega)| \geq 2^{-n}\}$$

has probability zero for every $t > 0$. Hence, for any ω in $\Omega \setminus N_t$ the sequence $X_n(t,\omega)$ is convergent to $X(t,\omega)$, i.e., $P\{X(t) = Y(t)\} = 1$, for every $t \geq 0$. Thus any stochastically left continuous adapted process has a predictable version. It is clear that X and Y does not necessarily differ on an evanescent set, i.e., the complement of A is not an evanescent set.

As discussed later, the integrand of a stochastic integral is thought as an equivalence class with respect to a product measure in $(0,\infty)\times\Omega$ of the form $\mu = \mathrm{d}\alpha(t,\omega)P(\mathrm{d}\omega)$, where $\alpha(t,\omega)$ is an integrable nondecreasing process. In this case, two processes may belong to the same μ-equivalence class without being a version of each other. For instance, if $\alpha(t,\omega)$ is constant on a subinterval (a,b) then two processes x and y may belong to the same μ-equivalence class even if $x(t,\omega) \neq y(t,\omega)$ for every (t,ω) in $(a,b)\times\Omega$. Conversely, two processes, which are versions of each other, may not belong to the same μ-equivalence class. For instance, if $\mathrm{d}\alpha(t,\omega) = \alpha_1(t,\omega)\mathrm{d}\beta(t)$ for some deterministic measure $\mathrm{d}\beta(t)$ then Fubini's Theorem ensures that two processes x and y belong to the same μ-equivalence class whenever we have a version of the other. However, independently of the form of $\mathrm{d}\alpha(t,\omega)$, two undistinguishable processes (i.e., differing on an evanescent set) must belong to the same μ-equivalence class. Moreover, a measure μ in the product space $(0,\infty)\times\Omega$ vanishes on every evanescent set if and only if it has the product form $\mu = \mathrm{d}\alpha(t,\omega)P(\mathrm{d}\omega)$ for some integrable nondecreasing process α, e.g., see Dellacherie and Meyer [58, Section VII.1]).

Martingales plays a key role in stochastic analysis. Recall that a martingale is a cad-lag process X with the following property relative to the conditional expectation

$$\mathbb{E}\{X(t) \mid X(r),\, 0 \leq r \leq s\} = X(s), \quad \forall t \geq s > 0.$$

When the $=$ sign replaced by the $\geq$ sign in the above property, the process X is called a sub-martingale, and similarly a super-martingale with the $\leq$ sign. The conditional expectation requires an integrable process, i.e., $\mathbb{E}\{|X(t)|\} < \infty$ for every $t \geq 0$ (for sub-martingale $\mathbb{E}\{[X(t)]^+\} < \infty$ and for super-martingale $\mathbb{E}\{[X(t)]^-\} < \infty$ are sufficient). Moreover, only a version of the process X is characterized by this property, so that a condition on the paths is also required. A minimal condition is to have a separable process X, but this theory becomes very useful when working with cad-lag process X. We adopted this point of view, so in this context, a *martingale is always a cad-lag integrable process.* Most of the time we replace the conditional expectation property with a more general statement, namely

$$\mathbb{E}\{X(t) \mid \mathcal{F}(s)\} = X(s), \quad \forall t \geq s > 0,$$

where now X is a cad-lag integrable process adapted to the filtration $(\mathcal{F}(t) : t \geq 0)$, which is always assumed right-continuous and even completed when

necessary. However, the concept of martingale is independent of the filtration as soon as some regularity on the paths is given. Actually, the conditional expectation property is equivalent to the condition

$$\mathbb{E}\Big\{X(t)\prod_{i=1}^{n}h_i\big((X(s_i)\big)\Big\}=\mathbb{E}\Big\{X(s)\prod_{i=1}^{n}h_i\big(X(s_i)\big)\Big\},$$

for every $0\le s_1<s_2\cdots\le s_n\le s<t$, any (real-valued) Borel and bounded functions h_i, $i=1,\dots,n$, any integer n. Nevertheless, to weaker the condition on integrability, a technical *localization* procedure is used, and a *local martingale* is a cad-lag process X such that $X_k:t\mapsto X(t\wedge\tau_k)-X(0)$ is a martingale for some increasing sequence of stopping times τ_k satisfying $\tau_k\to\infty$. This forces the use of a filtration.

Note the contrast of the previous property and the Markov property valid for a Markov process X: for any $n=1,2\dots$, any bounded measurable (actually continuous suffices) functions $f_1,\dots,f_n$, $g_1,\dots,g_n$, h, and times $s_1\le\cdots\le s_n\le t\le t_1\le\cdots\le t_n$ we have

$$\begin{aligned}\mathbb{E}\big\{h(X_t)\big(\prod_{i=1}^{n}f(X_{s_i})\big)\big(\prod_{i=1}^{n}g(X_{t_i})\big)\big\}&=\\&=\mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i})\,|\,X_t)\}\,\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t\}\big\},\end{aligned}$$

where $\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i}\,|\,X_t)\}$ and $\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t\}$ are X_t-measurable functions satisfying

$$\begin{aligned}\mathbb{E}\big\{h(X_t)\prod_{i=1}^{n}f(X_{s_i})\big\}&=\mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}f(X_{s_i})\,|\,X_t)\}\big\},\\\mathbb{E}\big\{h(X_t)\prod_{i=1}^{n}g(X_{t_i})\big\}&=\mathbb{E}\big\{h(X_t)\,\mathbb{E}\{\prod_{i=1}^{n}g(X_{t_i})\,|\,X_t)\}\big\},\end{aligned}$$

i.e., they are the conditional expectations with respect to the σ-algebra generated by the random variable X_t. This is briefly expressed by saying that the *past* and the *future* are independent given the *present*. Clearly, this condition involves only the finite-dimensional distributions of the process, and no condition on integrability for X is necessary for the above Markov property.

Actually, in this chapter we may use the concept of second-order processes, namely, to consider a process $X=\{X_t:t\ge0\}$ as a family of random variables with values in $L^2(\Omega,\mathcal{F},P)$ or better, as an element belonging to the Hilbert space $L^2\big((0,T);L^2(\Omega,\mathcal{F},P)\big)$ or the Banach space $C^0\big([0,T],L^2(\Omega,\mathcal{F},P)\big)$. A posteriori, we may take a continuous or cad-lag version, i.e., to consider elements in the less structured space of random variables with values in either $C([0,T])$ or $D([0,T])$.

For instance, the reader is referred to the books Chung and Williams [45], Bichteler [25], Dudley [62, Chapter 12, pp. 439–486], Durrett [67], Elliott [73], Kuo [145], Medvegyev [168], Protter [206], among others, for various presentations on stochastic analysis.

4.1 Probability Measures and Processes

We are interested in the law of two particular type of Lévy processes, the Wiener and the Poisson processes in Hilbert spaces. There is a rich literature on Gaussian processes, but less is known in Poisson processes, actually, we mean *compensated* Poisson processes. For stochastic integration we also use the Poisson random measures and in general integer random measures.

Definition 4.1 (Lévy Space)**.** For any nonnegative symmetric square matrix a and any σ-finite measure π in $\mathbb{R}^d_* = \mathbb{R}^d \smallsetminus \{0\}$ satisfying

$$\int_{\mathbb{R}^d_*} \big(|y|^2 \wedge 1\big)\pi(\mathrm{d}y) < \infty,$$

there exists a unique probability measure $P_{a,\pi}$, called Lévy *noise* space, on the space $\mathcal{S}'(\mathbb{R},\mathbb{R}^d)$ of Schwartz tempered distributions on $\mathbb{R}$ with values in $\mathbb{R}^d$ such that

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}\langle\cdot,\phi\rangle}\big\} = \exp\Big(-\frac{1}{2}\int_{\mathbb{R}} a\phi(t)\cdot\phi(t)\mathrm{d}t\Big)\times \\ \times \exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\phi(t)\cdot y} - 1 - \mathtt{i}\,\mathbb{1}_{\{|y|<1\}}\,\phi(t)\cdot y\big]\pi(\mathrm{d}y)\Big),$$

for any test function ϕ in $\mathcal{S}(\mathbb{R},\mathbb{R}^d)$. Therefore, a cad-lag version ℓ of the stochastic process $t \mapsto \langle\cdot, \mathbb{1}_{(0,t)}\rangle$ is well define and its law P on the canonical sample space $\mathbb{D} = D([0,\infty),\mathbb{R}^d)$ with the Skorokhod topology and its Borel σ-algebra $\mathcal{B}(\mathbb{D})$ is called the canonical *Lévy space* with parameters a and π, the diffusion covariance matrix a and the Lévy measure π. □

Clearly, ℓ is a Lévy process (see Section 3.10 in Chapter 3)

$$\langle\omega,\phi\rangle = \int_{\mathbb{R}} \omega(t)\cdot\phi(t)\,\mathrm{d}t, \quad \forall\omega\in\mathcal{S}'(\mathbb{R},\mathbb{R}^d),\ \phi\in\mathcal{S}(\mathbb{R},\mathbb{R}^d)$$

and $\cdot$ denotes the scalar product in the Euclidian space $\mathbb{R}^d$. To simplify notation and not to the use $\mathbb{1}_{\{|y|<1\}}$, we prefer to assume a stronger assumption on the Lévy measure, namely

$$\int_{\mathbb{R}^d_*} \big(|y|^2 \wedge |y|\big)\pi(\mathrm{d}y) < \infty,$$

and even to have a finite second moment, we assume

$$\int_{\mathbb{R}^d_*} |y|^2\pi(\mathrm{d}y) < \infty.$$

The existence of the probability $P_{a,\pi}$ was discussed in Section 1.14 of Chapter 1, and obtained via a Bochner's type theorem in the space of Schwartz tempered distributions (we may also use the Lebesgue space $L^2(]0,T[,\mathbb{R}^d)$, for $T>0$).

The expression of the characteristic function contains most of the properties of a Lévy space. For instance, we can be construct $P_{a,\pi}$ as the product $P_a \times P_\pi$

of two probabilities, one corresponding to the first exponential (called Wiener white noise, if a is the identity matrix)

$$\exp\Big(-\frac{1}{2}\int_{\mathbb{R}} ax(t)\cdot x(t)\mathrm{d}t\Big),$$

which has support in $C([0,\infty),\mathbb{R}^d)$, and another one corresponding to the second exponential (called compensated Poisson noise)

$$\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}x(t)\cdot y}-1-\mathtt{i}\mathbb{1}_{\{|y|<1\}}\,x(t)\cdot y\big]\pi(\mathrm{d}y)\Big).$$

The canonical process corresponding to P_a and P_π, denoted by $w(t)$ and $\bar p(t)$, are independent. Moreover, they may be assumed to take valued in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The topological space $\Omega = C([0,\infty),\mathbb{R}^n)\times D([0,\infty),\mathbb{R}^m)$ with the probability $P = P_w\times P_{\bar p}$ on the Borel σ-algebra $\mathcal{F}$ and the two canonical process w and $\bar p$ is called the canonical Wiener-Poisson space.

On the other hand, also the process

$$t\mapsto\frac{\exp\big[\mathtt{i}x\cdot\ell(t)\big]}{\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\ell(t)}\big\}}$$

is a complex-valued martingale, where

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}x\cdot\ell(t)}\big\}=\exp\Big(-\frac{t}{2}ax\cdot x+t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}x\cdot y}-1-\mathtt{i}\mathbb{1}_{\{|y|<1\}}\,x\cdot y\big]\pi(\mathrm{d}y)\Big),$$

for any x in $\mathbb{R}^d$. The process ℓ is a $\mathbb{R}^d$-valued martingale itself, with $\ell(0)=0$, and $\ell = w+\bar p$, where w is a Wiener process (continuous martingale) and $\tilde p$ is a *compensated Poisson process* (purely discontinuous martingale).

A generalization of this to infinite-dimensional spaces involves Sazonov's Theorem 1.41 and Minlos' Theorem 1.42, and the concept of nuclear operators, see Section 2.1.3 in Chapter 2. For instance, a Wiener random measure and a (compensated) Poisson random measure are constructed as follows, replacing $\mathbb{R}^d$ by and L^2 space. Given a (nonnegative) Radon measure m on $\mathbb{R}^d$, we get a probability measure P_m on $L^2_m = L^2(\mathbb{R}\times\mathbb{R}^d,\mathrm{d}t\times\mathrm{d}m)$ such that

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}(\cdot,h)}\big\}=\exp\Big(-\frac{1}{2}\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d}|h(t,x)|^2m(\mathrm{d}x)\Big),\quad\forall h\in L^2_m,$$

where $(\cdot,\cdot)$ denotes the scalar product in L^2_m. Then we choose a continuous version $w(t,B)$ of the stochastic process $(t,B)\mapsto(\cdot,\mathbb{1}_{(0,t)}\mathbb{1}_B)$, $t\ge 0$, B in $\mathcal{B}(\mathbb{R}^d)$ and bounded. Thus, $t\mapsto w(t,B)/m(B)$ is a standard Wiener process, and $B\mapsto w(t,B)$ is a (random) measure. Moreover, if $B_1,\dots,B_n$ are disjoint sets then $w(t,B_1),\dots,w(t,B_n)$ are independent processes. Similarly, given a σ-finite measure π in $\mathbb{R}^d_*$ as in Definition 4.1, we get a probability measure P_π on $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$ with the product measure $\mathrm{d}t\times\pi(\mathrm{d}y)$ such that

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}(\cdot,\phi)}\big\}=\exp\Big(\int_{\mathbb{R}}\mathrm{d}t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\phi(t,y)}-1-\mathtt{i}\phi(t,y)\big]\pi(\mathrm{d}y)\Big),$$

for any function ϕ in $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$, where now $(\cdot,\cdot)$ denotes the scalar product in $L^2_\pi(\mathbb{R}\times\mathbb{R}^d_*)$. Therefore, we can justify the use of $\phi(t,y)=\mathbb{1}_{(a,b)}(t)\mathbb{1}_B(y)$, and then we choose a cad-lag version $\tilde p(t,B)$ of the stochastic process $(t,B)\mapsto(\cdot,\mathbb{1}_{(0,t)}\,\mathbb{1}_B)$, $t\ge0$, B in $\mathcal{B}(\mathbb{R}^d_*)$, with $\bar B\cap\{0\}=\emptyset$, $\bar B$ is the closure. Moreover, $B\mapsto\tilde p(t,B)$ is a (random) measure, and if $B_1,\dots,B_n$ are disjoint sets then $\tilde p(t,B_1),\dots,\tilde p(t,B_n)$ are independent processes. Actually, $p(t,B)=\tilde p(t,B)+t\pi(B)$ is a (Poisson) integer-valued measure because

$$\mathbb{E}\{e^{\mathrm{i}rp(t,B)}\}=\exp\Big(t\pi(B)\big[e^{\mathrm{i}r}-1\big]\Big),\quad\forall r\in\mathbb{R},\tag{4.1}$$

for any B in $\mathcal{B}(\mathbb{R}^d_*)$, with $\bar B\cap\{0\}=\emptyset$, and any $t\ge0$.

• *Remark* 4.2. First recall the separability of the σ-algebra $\mathcal{B}(\mathbb{R}^d)$ or $\mathcal{B}(\mathbb{R}^d_*)$, i.e., there is an increasing sequence of finite σ-algebras $\{\mathcal{B}_k\}$ such that $\mathcal{B}=\vee_k\mathcal{B}_k$, e.g. see Malliavin [162, Section 6.1, pp. 219–220]. It is clear now that we are able to show that for any t in a countable set and for each ω outside of a null set, the function $B\mapsto w(t,B,\omega)$ (or $B\mapsto\tilde p(t,B,\omega)$) is a (positive) measure on any $\mathcal{B}_k$, $k\ge1$. Hence, we can take a *version* so that for any B in $\mathcal{B}$ the process $t\mapsto w(t,B)$ (or $t\mapsto\tilde p(t,B)$) is continuous or cad-lag, and for any $t\ge0$ the set function $B\mapsto w(t,B)$ (or $B\mapsto\tilde p(t,B)$) is a measure on $\mathcal{B}(\mathbb{R}^d)$ (or $\mathcal{B}(\mathbb{R}^d_*)$, respectively). Actually, w and $\tilde p$ are random measures in both variables, i.e., in $\mathbb{R}\times\mathbb{R}^d$. Note that sometimes it is convenient to use the notation $p(B,t)$, $\tilde p(B,t)$ and $\bar p(B,t)$, i.e., we may exchange the order of the variable t and B as long no confusion is made. □

As discussed later to study the jumps, we may use the construction of the $\mathbb{R}^d$-valued compensated Poisson process $\bar p(t)$ or the compensated Poisson "point" process if the emphasis is on the jumps $\delta p(s)=p(s)-p(s-)$. We define the $\mathbb{R}^d$-valued Poisson measure

$$\bar p(t,B)=\sum_{0<s\le t}[\bar p(s)-\bar p(s-)]\mathbb{1}_{\{\bar p(s)-\bar p(s-)\in B\}},\qquad\forall B\in\mathcal{B}_*,$$

where the sum has a finite number of terms and $\mathcal{B}_*$ denotes the ring of Borel sets B in $\mathcal{B}(\mathbb{R}^d_*)$ satisfying $\bar B\cap\{0\}=\emptyset$, $\bar B$ is the closure. We have

$$\mathbb{E}\{e^{\mathrm{i}x\cdot\bar p(t,B)}\}=\exp\Big(t\int_B\big[e^{\mathrm{i}x\cdot y}-1\big]\pi(\mathrm{d}y)\Big),\quad\forall x\in\mathbb{R}^d,\ B\in\mathcal{B}_*,$$

which implies

$$\mathbb{E}\{x\cdot\bar p(t,B\cap\{|y|<1\})\}=t\int_B x\cdot y\,\mathbb{1}_{\{|y|<1\}}\pi(\mathrm{d}y),\quad\forall x\in\mathbb{R}^d,\ B\in\mathcal{B}_*,$$

for any $t\ge0$.

Sometimes, instead of using the (Poisson) point processes $\bar p(t)$ or (Poisson) vector-valued measure $\bar p(t,B)$, we prefer to use the (Poisson) counting (integer) measure

$$p(t,B)=p(]0,t]\times B)=\sum_{0<s\le t}\mathbb{1}_{\{\bar p(s)-\bar p(s-)\in B\}},\quad\forall B\in\mathcal{B}_*,$$

which is a Poisson process with parameter $\pi(B)$, i.e., (4.1) holds for any B in $\mathcal{B}_*$, or equivalently

$$P\{p(t,B)=n\}=\frac{\big(t\pi(B)\big)^n}{n!}\mathrm{e}^{-t\pi(B)},\quad \forall B\in\mathcal{B}_*,\ n=0,1,\dots,$$

for any $t>0$. Moreover, because there are a finite number of jumps within B, the integral

$$\bar{p}(t,B)=\int_{]0,t]\times B} zp(\mathrm{d}t,\mathrm{d}z),\quad \forall B\in\mathcal{B}_*,\ t>0$$

is finite and reproduces the $\mathbb{R}^d$-valued Poisson measure initially defined. To reproduce $\bar{p}(t)$ on this context, we should make sense to the limit

$$\bar{p}(t)=\bar{p}(t,\{|y|\geq 1\})+\lim_{\varepsilon\to 0}\big[\bar{p}(t,\{\varepsilon\leq|y|<1\})-t\pi(\{\varepsilon\leq|y|<1\})\big],$$

by means of the stochastic integral. All theses facts are particular cases of the theory of random measures, martingale theory and stochastic integral.

4.1.1 Gaussian Processes

A $\mathbb{R}^d$-valued random variable ξ is *Gaussian distributed* (also called *normally distributed*) with parameters (c,C) if its (complex-valued) characteristic function has the following form

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda\cdot\xi)\}=\exp(\mathtt{i}\,\lambda\cdot c-\lambda\cdot C\lambda/2),\quad \forall\lambda\in\mathbb{R}^d,$$

or equivalently if its distribution has the form

$$P(\xi\in B)=\int_B\big[(2\pi)^n\sqrt{\det(C)}\big]^{-1/2}\exp\big(-\frac{C^{-1}(x-c)\cdot(x-c)}{2}\big)\mathrm{d}x,$$

for every Borel subset of $\mathbb{R}^d$, where c is the (vector) mean $\mathbb{E}\{\xi\}$ and C is the (matrix) covariance $\mathbb{E}\{(\xi-c)^2\}$. When $c=0$ the random variable ξ is called *centered* or symmetric. Notice that the expression with the characteristic function make sense even if C is only a symmetric nonnegative definite matrix, which is preferred as the definition of Gaussian variable. It is clear that a $\mathbb{R}^d$-valued Gaussian variable has moments of all orders and that a family of centered $\mathbb{R}^d$-valued Gaussian variables is independent if and only if the family is orthogonal in $L^2(\Omega,\mathcal{F},P)$. Next, an infinite sequence $(\xi_1,\xi_2,\dots)$ of real-valued (or $\mathbb{R}^d$-valued) random variables is called *Gaussian* if any (finite) linear combination $c_1\xi_1+\cdots+c_n\xi_n$ is a Gaussian variable. Finally, a probability measure μ on the Borel σ-algebra $\mathcal{B}$ of a separable Banach space B is called a *(centered) Gaussian measure* if any continuous linear functional h is (centered) Gaussian real-valued random variable when considered on the probability space $(B,\mathcal{B},\mu)$. If $B{=}H$ a separable Hilbert space then the mean c value and covariance C operator are well defined, namely,

$$(c,h)=\int_H(h,x)\mu(\mathrm{d}x),\quad \forall h\in H,$$
$$(Ch,k)=\int_H(h,x)(k,x)\mu(\mathrm{d}x)-(c,h)(c,k),\quad \forall h,k\in H,$$

where $(\cdot,\cdot)$ is the inner product in H. Moreover, the covariance C operator is a trace-class operator, i.e., for any (or some) orthonormal basis $\{e_n : n \geq n\}$ in H the series $\mathrm{Tr}(C) := \sum_n (Ce_n, e_n)$ converges.

A fundamental result is the following Fernique's bound

$$\int_B \mathrm{e}^{\lambda \|x\|^2}\, \mu(\mathrm{d}x) \leq \mathrm{e}^{16\,\lambda\, r^2} + \frac{\mathrm{e}^2}{\mathrm{e}^2 - 1}, \tag{4.2}$$

valid for any centered Gaussian measure μ on the separable Banach space $(B, \mathcal{B})$ and any λ, $r > 0$ such that

$$\ln\big(1 - \mu(\{x : \|x\| \leq r\})\big) + 32\,\lambda\, r \leq \ln\big(\mu(\{x : \|x\| \leq r\})\big) - 1,$$

where $\|\cdot\|$ is the norm in B.

In particular, any continuous linear functional φ on B has a finite second moment. Thus, the dual space B' of B can be identified with a subspace of $L^2(B, \mathcal{B}, \mu)$ and call $\bar{B}'$ the Hilbert space obtained as the closure of this subspace. Recalling that any φ in B' is a centered Gaussian variable with covariance $|\varphi|^2_{L^2}$, we define the mapping J by setting

$$\begin{aligned} &J : L^2(B, \mathcal{B}, \mu) \to B, \\ &J(\varphi) := \int_B x\, \varphi(x)\, \mu(\mathrm{d}x), \quad \forall \varphi \in L^2(B, \mathcal{B}, \mu), \end{aligned}$$

but we consider J only acting from $\bar{B}' \subset L^2(B, \mathcal{B}, \mu)$ into B. Since the linearity and continuity of φ and Hölder inequality yield

$$\begin{aligned} &\varphi(J(\varphi)) = \int_B |\varphi(x)|^2\, \mu(\mathrm{d}x), \\ &\|J(\varphi)\|^2 \leq |\varphi|^2_{L^2} \int_B \|x\|^2\, \mu(\mathrm{d}x), \end{aligned}$$

the mapping J is one-to-one, continuous and linear. The image $H = J(\bar{B}')$ is continuously embedded in B as a Hilbert space with the inner product

$$(f, g)_H = \int_B J^{-1}(f)(x)\, J^{-1}(g)(x)\, \mu(\mathrm{d}x), \quad \forall f, g \in H.$$

Moreover, any φ in the dual space B' is a centered Gaussian random variable on $(B, \mathcal{B}, \mu)$ with covariance $|\varphi|^2_H$, where the dual norm is given by $|\varphi|^2_H := \sup\{|\varphi(x)| : |x|_H \leq 1\}$. The space $H = H_\mu$ is called a *reproducing kernel space* for the centered Gaussian measure $(B, \mathcal{B}, \mu)$. Now, denote by H^0_μ the image of B' by J, i.e., $H^0_\mu := J(B')$, which is dense in H_μ.

Let $\{e_1, e_2, \ldots\}$ be a orthonormal basis in H_μ with elements in H^0_μ, and let $\{\xi_1, \xi_2, \ldots\}$ be a sequence of independent real-valued random variables with standard normal distribution (i.e., Gaussian with parameters $0, 1$) relative to a (complete) probability space $(\Omega, \mathcal{F}, P)$. Then, it can be proved that the sequence of partial sums $\{\sum_{k=1}^n \xi_k\, e_k : n = 1, 2, \ldots\}$ converges almost surely in B to a random variable ξ with law μ. Notice that the above series does not converges

almost surely in H_μ, but the map $h \mapsto X(h) := \sum_k \xi_k\,(h, e_k)$ is well defined for any h in H_μ, and called *white noise*, see Da Prato and Zabczyk [51, Theorems 1.2.6–12, pp. 37–48].

This procedure can be done backward, i.e., starting from the Hilbert space H. With respect to the previous construction, now H is the dual space of H_μ. Pick an orthonormal basis $\{e_1, e_2, \ldots\}$ in H and a sequence $\{\xi_1, \xi_2, \ldots\}$ of independent real-valued random variables with standard normal distribution in a (complete) probability space $(\Omega, \mathcal{F}, P)$. Since

$$\mathbb{E}\{[\sum_{k=\ell}^{n} \xi_k\,(h, e_k)]^2\} = \sum_{k=\ell}^{n} [(h, e_k)]^2 \mathbb{E}\{|\xi_k|^2\} = \sum_{k=\ell}^{n} [(h, e_k)]^2,$$

for every $n > \ell \geq 1$, we may define $X(h) := \sum_k \xi_k\,(h, e_k)$, for any h in H, as a convergent series in $L^2(\Omega, \mathcal{F}, P)$. The map $h \mapsto X(h)$ is linear, $X(h)$ is a centered Gaussian random variable with covariance $\mathbb{E}\{[X(h)]^2\} = |h|^2$, for any h in H. Actually, the series also converges almost surely and $X(h)$ is indeed an equivalence class. The space $\{X(h) : h \in H\}$ is a Gaussian subspace of $L^2(\Omega, \mathcal{F}, P)$, which is isomorphic to H. In particular $\mathbb{E}\{X(f)\,X(g)\} = (f, g)$, for any f and g in H. This show that $X(f)$ is independent of $X(g)$ if and only if f and g are orthogonal (because independence and orthogonality are equivalent in a Gaussian space). The family $\{X(h) : h \in H\}$ is called an *isonormal Gaussian* stochastic process. If $H = L^2(A, \mathcal{A}, \mu)$, where $(A, \mathcal{A}, \mu)$ is a σ-finite measure space, the mapping X is called a *Gaussian measure* or *white noise* with intensity μ on $(A, \mathcal{A})$. When F belongs to $\mathcal{A}$ we write $X(F) := X(\mathbb{1}_F)$. Thus, if F and G are sets with $\mu(F) < \infty$ and $\mu(G) < \infty$ then $\mathbb{E}\{X(F)\,X(G)\} = \mu(F \cap G)$, and so that $X(F)$ and $X(G)$ are independent when F and G are disjoint. Notice that if $\{F_k : k = 1, 2, \ldots\}$ is a pairwise disjoint sequence of subset in $\mathcal{A}$, $F = \cup_k F_k$ with $\mu(F) < \infty$ then $X(F) = \sum_k X(F_k)$ *almost surely* so that some regularity (as in the case of regular conditional probability) is need to ensure the existence of a good selection, in order that $F \mapsto X(F, \omega)$ is a measure for ω outside of a set of probability zero.

Sometimes, the initial point is a family of centered Gaussian random variables $X = \{X(h) : h \in H\}$ in a complete probability space $(\Omega, \mathcal{F}, P)$, where the index H is a separable Hilbert space, the σ-algebra $\mathcal{F}$ is the smallest complete σ-algebra such that $X(h)$ is measurable for any h in H and $\mathbb{E}\{X(f)\,X(g)\} = (f, g)_H$, for any f and g in H. This is called a Gaussian process on H. Notice that mapping $h \mapsto X(h)$ has to be linear and provides an isometry from H onto a closed subspace of $L^2(\Omega, \mathcal{F}, P)$, where all elements are zero-mean Gaussian random variables.

Consider the Hermite polynomials, which are defined by

$$h_0(x) := 1, \quad h_n(x) := \frac{(-1)^n}{n!} \mathrm{e}^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \mathrm{e}^{-x^2/2},\ n = 1, 2, \ldots,$$

which satisfies several properties, e.g.,

$$\exp\big[\frac{x^2}{2} - \frac{(x-t)^2}{2}\big] = \sum_{n=0}^{\infty} t^n\, h_n(x), \quad \forall t, x,$$

$h'_n = h_{n-1}$, $(n+1)h_{n+1}(x) = x\,h_n(x) - h_{n-1}(x)$, $h_n(-x) = (-1)^n\,h_n(x)$, $h_n(0) = 0$ if n is odd and $h_{2n}(0) = (-1)^n/(2^n\,n!)$. It is not hard to show that for any two random variables ξ and ζ with joint standard normal distribution we have $\mathbb{E}\{h_n(\xi)\,h_m(\zeta)\} = (\mathbb{E}\{\xi\,\zeta\})/n!$ if $n = m$ and $\mathbb{E}\{h_n(\xi)\,h_m(\zeta)\} = 0$ otherwise. Essentially based on the one-to-one relation between signed measures and their Laplace transforms, we deduce that only the null element ξ in $L^2(\Omega, \mathcal{F}, P)$ (recall that $\mathcal{F}$ is generated by $\{X(h) : h \in H\}$) satisfies $\mathbb{E}\{\xi\,\exp(X(h))\} = 0$, for any h in H. Hence, the space H can be decomposed into an infinite orthogonal sum of subspaces, i.e.,

$$L^2(\Omega, \mathcal{F}, P) = \oplus_{n=0}^{\infty} H_n,$$

where H_n is defined as the subspace of $L^2(\Omega, \mathcal{F}, P)$ generated by the family random variables $\{h_n(X(h)) : h \in H,\ |h|_H = 1\}$. Thus, H_0 is the subspace of constants and H_1 the subspace generated by $\{X(h) : h \in H\}$. This analysis continues with several applications, the interest reader is referred to Hida et al. [107], Kallianpur and Karandikar [122], Kuo [144], among others.

Going back to our main interest, we take $H = L^2(\mathbb{R}^+)$ with the Lebesgue measure, initially the Borel σ-algebra, and we construct the family of equivalence classes of centered Gaussian random variables $\{X(h) : h \in H\}$ as above. Thus we can pick a random variable $b(t)$ within the equivalence class $X([0,t]) = X(\mathbb{1}_{[0,t]})$. This stochastic process $b = (b(t) : t \geq 0)$ has the following properties:

(1) The process b has *independent increments*, i.e. for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the random variables $\{b(t_0),\ b(t_1) - b(t_0), \ldots,\ b(t_n) - b(t_{n-1})\}$ are independent. Indeed, they are independent since $b(t_k) - b(t_{k-1})$ is in the equivalence class $X(]t_{k-1}, t_k])$ which are independent because the interval $]t_{k-1}, t_k]$ are pairwise disjoint.

(2) The process b is a Gaussian process, i.e., for any sequence $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ the $\mathbb{R}^{n+1}$-valued random variable $(b(t_0), b(t_1), \ldots, b(t_n))$ is a Gaussian random variables. Indeed, this follows from the fact that $\{b(t_0),\ b(t_1) - b(t_0), \ldots,\ b(t_n) - b(t_{n-1})\}$ is a family of independent real-valued Gaussian random variable.

(3) For each $t > 0$ we have $\mathbb{E}\{b^2(t)\} = t$ and $b(0) = 0$ almost surely. Moreover, using the independence of increments we find that the covariance $\mathbb{E}\{b(t)\,b(s)\} = t \wedge s$.

(4) Given a function f in $L^2(\mathbb{R}^+)$ (i.e., in H) we may pick an element in the equivalence class $X(f\,\mathbb{1}_{[0,t]})$ and define the integral with respect to b, i.e., $X(f\,\mathbb{1}_{[0,t]})$.

(5) The hard part in to show that we may choose the random variables $b(t)$ in the equivalence class $X([0,t])$ in a way that the path $t \mapsto b(t,\omega)$ is continuous (or at least cad-lag) almost surely. A similar question arises when we try to show that $F \mapsto X(\mathbb{1}_F)$ is a measure almost surely. Because $b(t) - b(s)$ is Gaussian, a direct calculation show that $\mathbb{E}\{|b(t) - b(s)|^4\} = 3|t-s|^2$. Thus, Kolmogorov's continuity criterium (i.e., $\mathbb{E}\{|b(t) - b(s)|^\alpha\} \leq C|t-s|^{1+\beta}$ for some positive

constants α, β and C) is satisfied. This show the existence of a continuous stochastic process B as above, which is called *standard Brownian motion* or *standard Wiener process*. The same principle can be used with the integral $\langle f, b\rangle(t) = X(f\,\mathbb{1}_{[0,t]})$, as long as f belongs to $L^\infty(\mathbb{R}^+)$. This continuity holds true also for any f in $L^2(\mathbb{R}^+)$, by means of theory of stochastic integral as seen later.

It is clear that we may have several independent copies of a real-valued standard Brownian motion and then define a $\mathbb{R}^d$-valued standard Brownian motion. Moreover, if for instance, the space $L^2(\mathbb{R}, \mathcal{X})$, for some Hilbert $\mathcal{X}$ (or even co-nuclear) space, is used instead of $L^2(\mathbb{R})$ then we obtain the so called cylindrical Brownian motions or space-time Wiener processes, which is not considered here. We may look at B as a random variable with values in the canonical sample space $C = C([0,\infty), \mathbb{R}^d)$, of continuous functions with the locally uniform convergence (a separable metric space), and its Borel σ-algebra $\mathcal{B} = \mathcal{B}(C)$. The law of B in the canonical sample space C define a *unique* probability measure W such that the coordinate process $X(t) := \omega(t)$ is a standard Brownian motion, which is called the *Wiener measure*. Thus $(C, \mathcal{B}, W)$ is referred to as a *Wiener space*.

Generally, a *standard Wiener process* is defined as a real-valued continuous stochastic process $w = (w(t) : t \geq 0)$ such that (1) it has independent increments and (2) its increments $w(t) - w(s)$, $t > s \geq 0$, $k+1, 2, \ldots, d$ are normally distributed with zero-mean and variance $t - s$. This definition is extended to a d-dimensional process by coordinates, i.e., $\mathbb{R}^d$-valued where each coordinate w_k is a real-valued standard Wiener process and $\{w_1, w_2, \ldots, w_n\}$ is a family of independent processes. For any f in $L^\infty(\mathbb{R}^+)$, the integral with respect to the standard Wiener process $w = (w_1, \ldots, w_d)$ is defined as a $\mathbb{R}^d$-valued continuous centered Gaussian process with independent increments and independent coordinates such that for any $k = 1, 2, \ldots, d$

$$\int_0^t f(s)\,\mathrm{d}w_k(s) := X_k(f\,\mathbb{1}_{[0,t]}),$$
$$\mathbb{E}\Big\{\Big(\int_0^t f(s)\,\mathrm{d}w_k(s)\Big)^2\Big\} = \int_0^t f^2(s)\mathrm{d}s,$$

for any $t \geq 0$. Notice that the second equality specifies the covariance of the process.

Similarly, we can define the Gaussian-measure process $w(t, \cdot)$, by using the Hilbert space $L^2(\mathbb{R}^+ \times \mathbb{R}^d)$ with the product measure $\mathrm{d}t \times m(\mathrm{d}x)$, where $m(\mathrm{d}x)$ is a Radon measure on $\mathbb{R}^d$ (i.e., finite on compact subsets). In this case $w(t, K)$ is a Wiener process with diffusion $m(K)$ (and mean zero) and $w(t, K_1), \ldots, w(t, K_n)$ are independent if $K_1, \ldots, K_n$ are disjoint. Clearly, this is related with the so-called *white noise measure* (e.g., see Bichteler [25, Section 3.10, pp. 171–186]) and *Brownian sheet* or *space-time* Brownian motion. The reader is referred to Kallianpur and Xiong [123, Chapters 3 and 4, pp. 85–148] for the infinite dimensional case driven by a space-time Wiener process and a Poisson random

measure. This requires the study of martingales with values in Hilbert, Banach and co-nuclear spaces, see also Métivier [178].

The following Lévy's characterization of a Wiener process is a fundamental results, for instance see Revuz and Yor [212, Theorem IV.3.6, pp. 150].

Theorem 4.3 (Lévy). *Let X be an adapted $\mathbb{R}^d$-valued continuous stochastic process in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$. Then X is a Wiener if and only if X is a (continuous) local martingale and one of the two following conditions is satisfied:*

(1) *$X_i\, X_j$ and $X_i^2 - t$ are local martingales for any $i, j = 1, \dots, d$, $i \neq j$,*

(2) *for any $f_1, f_2, \dots, f_d$ functions in $L^\infty(\mathbb{R}^+)$ the (exponential) process*

$$Y_f(t) := \exp\Big\{\mathtt{i} \sum_k \int_0^t f_k(s)\, dX_k(s) + \frac{1}{2} \sum_k \int_0^t f_k^2(s) \mathrm{d}s\Big\},$$

defined for every $t \geq 0$, is a (bounded) complex-valued martingale. □

Clearly, condition (1) means that the (matrix-valued) predictable quadratic variation process $\langle X \rangle$ associated with X is such that $\langle X_i, X_i \rangle(t) = t$ and $\langle X_i, X_j \rangle(t) = 0$ for any $i, j = 1, \dots, d$, $i \neq j$. In condition (2) we may also take f_k in $L^2(\mathbb{R}^+)$ and even adapted processes. The assumption on *continuity* is essential to the above Lévy's theorem.

It can be proved that a Gaussian semi-martingale X is continuous if and only if it is stochastically continuous, i.e., $P(|X(t) - X(s)| > \varepsilon)$ goes to zero as $t \to s$, for any $\varepsilon > 0$. Moreover, a centered Gaussian local martingale X with $X(0) = 0$ and independent increments, is actually a locally square integrable and its predictable quadratic variation (non-necessarily continuous) satisfies $\langle X \rangle(t) \wedge \langle X \rangle(s) = \mathbb{E}\{X(t)\, X(s)\}$, for any $t \geq s \geq 0$. It is also clear that for a centered Gaussian martingale X with $X(0) = 0$, the covariance matrix $c(t) := (\mathbb{E}\{X_i(t)\, X_j(t)\} : t \geq 0,\, i, j = 1, 2, \dots, d)$ satisfies

$$\mathbb{E}\{\exp[\mathtt{i}\, \lambda \cdot (X(t) - X(s))]\} = \exp[-\lambda \cdot (c(t) - c(s))\lambda/2],$$

for every λ in $\mathbb{R}^d$ and $t \geq s \geq 0$. This property completely characterizes the finite distributions of X, see Liptser and Shiryayev [158, Section 4.9, pp. 270–306].

The Ornstein-Uhlenbeck process is another typical example of Gaussian process that is given by

$$X(t) := \exp(-\alpha t) X_0 + \int_0^t \exp[-\alpha(t - s)] \sigma \mathrm{d}w(s), \quad \forall t \geq 0,$$

where α and σ are matrices, α has positive eigenvalues, X_0 is an initial random variable normally distributed and w is an standard Wiener process. Even more general, if $\Phi(t, s)$ denotes the fundamental (matrix) solution of a linear ordinary differential equation with matrix $\alpha(t)$, i.e.,

$$\dot{\Phi}(t, s) = -\alpha(t)\Phi(t, s), \quad \forall t \neq s,$$
$$\Phi(s, s) = 1, \quad \forall s,$$

then

$$X(t) := \Phi(t,0)X_0 + \int_0^t \Phi(t,s)\sigma(s)\mathrm{d}w(s), \quad \forall t \geq 0,$$

is a Gaussian process with mean $m_i(t) := \mathbb{E}\{X_i(t)\}$ and covariance matrix $v_{ij}(s,t) := \mathbb{E}\{[X_i(s)-m_i(s)][X_j(t)-m_j(t)]\}$, which can be explicitly calculated. For instance, in the one-dimensional case with constant α and σ we have

$$\begin{aligned} m(t) &:= \mathbb{E}\{X(t)\} = \mathrm{e}^{-\alpha t}m(0), \\ v(s,t) &:= \mathbb{E}\{[X(s)-m(s)][X(t)-m(t)]\} = \\ &= \Big\{\frac{\sigma^2}{2\alpha}\big[\mathrm{e}^{2\alpha(s\wedge t)}-1\big] + v(0)\Big\}\mathrm{e}^{-\alpha(s+t)}. \end{aligned}$$

Therefore, if the initial random variable has mean zero and the variance is equal to $v_0 = \sigma^2/(2\alpha)$, then X is a stationary, zero-mean Gaussian process with covariance function $\rho(s,t) = v_0\exp(-\alpha|t-s|)$.

4.1.2 Compensated Poisson Processes

A $\mathbb{R}^d$-valued random variable ξ has a *compensated Poisson distributed* (also called *centered Poisson distributed*) with parameter π if its (complex-valued) characteristic function has the following form

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda\cdot\xi)\} = \exp\Big(\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x}-1-\mathtt{i}\,\lambda\cdot x\big]\pi(\mathrm{d}x)\Big), \quad \forall\lambda\in\mathbb{R}^d,$$

where π is a Radon measure on $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ satisfying

$$\int_{\mathbb{R}^d_*}|x|^2\pi(\mathrm{d}x) < \infty.$$

Usually, the arguments begin with a compound Poisson variable p in $\mathbb{R}^d$ (mainly, d=1) with a finite measure π as parameter, i.e.,

$$\mathbb{E}\{\exp(\mathtt{i}\,\lambda\cdot p)\} = \exp\Big(\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x}-1\big]\pi(\mathrm{d}x)\Big). \quad \forall\lambda\in\mathbb{R}^d,$$

Then define $\xi = p - \mathbb{E}\{p(t)\}$ as a centered Poisson distribution random variable. Next, the construction and properties of the compensated Poisson (or centered Poisson) random variable ξ are extended for characteristic measures π as above.

It is called *symmetric* if π satisfies

$$\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\,\lambda\cdot x}-1-\mathtt{i}\,\lambda\cdot x\big]\pi(\mathrm{d}x) = \int_{\mathbb{R}^d_*}\big[\mathrm{e}^{-\mathtt{i}\,\lambda\cdot x}-1+\mathtt{i}\,\lambda\cdot x\big]\pi(\mathrm{d}x),$$

for every λ in $\mathbb{R}^d$. It is clear that a $\mathbb{R}^d$-valued compensated Poisson variable ξ has finite first and second moments, i.e.,

$$\mathbb{E}\{|\xi|^2\} = \int_{\mathbb{R}^d_*}|x|^2\pi(\mathrm{d}x),$$

and if we add the condition

$$\int_{\mathbb{R}^d_*} |x|^p \pi(\mathrm{d}x) < \infty, \quad \forall p \geq 2.$$

then all moments are finite.

An infinite sequence $(\xi_1, \xi_2, \ldots)$ of real-valued (or $\mathbb{R}^d$-valued) random variables is called *compensated Poisson* process if any (finite) sum $\xi_1 + \cdots + \xi_n$ is a compensated Poisson variable (the sequence is necessarily independent). Next, given a (nuclear) countably Hilbertian space $\Phi = \bigcap_{n\geq 0} \Phi_n$, its dual space $\Phi' = \bigcup_{n\geq 0} \Phi_{-n}$ (see Section 2.1.3 in Chapter 2), a probability measure μ on the Borel σ-algebra $\mathcal{B}(\Phi')$ is called a *compensated Poisson measure* if $\langle \cdot, \varphi \rangle$ is a compensated Poisson real-valued random variable, for any φ in Φ, when considered on the probability space $(\Phi', \mathcal{B}(\Phi'), \mu)$, i.e., there exists a σ-finite measure π on $\Phi'_* = \Phi' \smallsetminus \{0\}$ such that

$$\mathbb{E}_\mu\{\mathrm{e}^{\mathtt{i}\langle \cdot, \varphi \rangle}\} = \exp\Big(\int_{\Phi'_*} \big[\mathrm{e}^{\mathtt{i}\langle \cdot, \varphi \rangle} - 1 - \mathtt{i}\langle \cdot, \varphi \rangle\big] \mathrm{d}\pi \Big), \quad \forall \varphi \in \Phi.$$

Similarly to the finite-dimensional case, besides the condition

$$\int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi < \infty, \quad \forall \varphi \in \Phi,$$

we assume that

$$\int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi \leq C_0 \|\varphi\|_n^2, \quad \forall \varphi \in \Phi, \tag{4.3}$$

for some $n \geq 0$ and some constant $C_0 > 0$.

• *Remark* 4.4. Minlos' Theorem 1.42 ensures the existence of a probability measure μ for any given σ-finite measure such that

$$\varphi \mapsto \int_{\Phi'_*} \big[\mathrm{e}^{\mathtt{i}\langle \cdot, \varphi \rangle} - 1 - \mathtt{i}\langle \cdot, \varphi \rangle\big] \mathrm{d}\pi$$

is continuous, in particular if (4.3) holds. Note that (4.3) is equivalent to the condition that

$$\varphi \mapsto \int_{\Phi'_*} |\langle \cdot, \varphi \rangle|^2 \mathrm{d}\pi$$

is continuous. However, if we wish to replace the space Φ by a Banach space B some difficulties appears and we cannot guarantee the existence of a probability measure μ, e.g., see Rudiger [219]. □

Under the assumption (4.3), there is a separable Hilbert space $\Phi \subset H \subset \Phi_0$, with continuous and dense inclusion, and a nonnegative symmetric trace-class operator R in $L_1(H)$ (i.e., $R^{1/2}$ is a Hilbert-Schmidt operator), such that the support of π is included in $R(H) \subset H \subset \Phi_0$, i.e.,

$$\pi\big(\{\chi \in \Phi' : \langle \chi, \varphi \rangle \leq r\}\big) = \pi\big(\{h \in R(H) : \langle h, \varphi \rangle \leq r\}\big), \quad \forall \varphi \in \Phi,\ r \in \mathbb{R},$$

and defining

$$\pi_0(B) = \pi(\{h \in H_* : R^{-1}h \in B\}), \quad \forall B \in \mathcal{B}(H_*)$$

or equivalently $\pi_0 = R\pi$, with $H_* = H \setminus \{0\}$, we have

$$\int_{\Phi'} \mathrm{e}^{\mathtt{i}\langle\cdot,\varphi\rangle}\mathrm{d}\mu = \exp\Big(\int_{H_*}\big[\mathrm{e}^{\mathtt{i}\langle R\cdot,\varphi\rangle} - 1 - \mathtt{i}\langle R\cdot,\varphi\rangle\big]\mathrm{d}\pi_0\Big), \quad \forall\varphi\in\Phi.$$

The integrability condition becomes

$$\int_{H_*} |(R\cdot,\varphi)|^2\mathrm{d}\pi_0 \le C_0\|\varphi\|_H^2, \quad \forall\varphi\in H,$$

for some constant $C_0 > 0$, which yields

$$\int_{H_*} (Rh,h)\pi_0(\mathrm{d}h) \le C_0\,\mathrm{Tr}(R) < \infty.$$

Hence, Sazonov's Theorem 1.41 shows that μ is actually supported in H, i.e., μ is a *compensated Poisson measure* with parameter $\pi = R^{-1}\pi_0$ satisfying

$$\int_H \mathrm{e}^{\mathtt{i}(h,k)}\mu(\mathrm{d}k) = \exp\Big(\int_{H_*}\big[\mathrm{e}^{\mathtt{i}(Rh,k)} - 1 - \mathtt{i}(Rh,k)\big]\pi_0(\mathrm{d}k)\Big), \quad \forall h\in H.$$

Thus, by working on a nuclear countably Hilbertian space we are reduced to the case of a Hilbert space. Now, we can justify

$$\mathbb{E}_\mu\big\{\langle\cdot,\varphi\rangle\big\} = 0 \quad \text{and} \quad \mathbb{E}_\mu\big\{|\langle\cdot,\varphi\rangle|^2\big\} = \int_{H_*}|\langle Rh,\varphi\rangle|^2\pi_0(\mathrm{d}h), \quad \forall\varphi\in\Phi,$$

actually, we may take φ in H, replace the duality $\langle\cdot,\cdot\rangle$ by $(\cdot,\cdot)$, and assume $H = \Phi_0$.

Hence, the map $\varphi \mapsto \langle\cdot,\varphi\rangle$ allows us to identify the space H with a subspace of $L^2(\Phi',\mathcal{B}(\Phi'),\mu) = L^2(H,\mathcal{B}(H),\mu)$ and then to call $\bar{H}$ the Hilbert space obtained as the closure of this subspace. Recalling that any φ in H the random variable $\langle\cdot,\varphi\rangle$ is a compensated Poisson variable with with parameter π, we define the mapping J by setting

$$J : L^2(H,\mathcal{B}(H),\mu) \to H,$$
$$J(\varphi) := \int_H h\,\varphi(h)\,\mu(\mathrm{d}h), \quad \forall\varphi\in L^2(H,\mathcal{B}(H),\mu),$$

but we consider J only as being from $\bar{H} \subset L^2(H,\mathcal{B}(H),\mu)$ into H. Since the linearity and continuity of φ and Hölder inequality yield

$$\varphi(J(\varphi)) = \int_H |\varphi(x)|^2\,\mu(\mathrm{d}x),$$
$$\|J(\varphi)\|_H^2 \le \Big(\int_H |\varphi(x)|^2\,\mu(\mathrm{d}x)\Big)\Big(\int_H \|h\|_H^2\,\mu(\mathrm{d}h)\Big),$$

the mapping J is one-to-one, continuous and linear. The image $H_\mu = J(\bar{H})$ is continuously embedded in H as a Hilbert space with the inner product

$$(f,g)_{_\mu} = \int_H J^{-1}(f)(h)\, J^{-1}(g)(h)\, \mu(\mathrm{d}h), \quad \forall f,g \in H_\mu.$$

Now, set $H_\mu^0 = J(H)$, which is dense in $H_\mu \subset H \subset \bar{H}$. Clearly, if f and g belong to H_μ^0 then $(f,g)_{_\mu} = (J^{-1}f, J^{-1}g) = \langle J^{-1}f, J^{-1}g\rangle$.

Let $\{e_1, e_2, \ldots\}$ be an orthonormal basis in H_μ with elements in H_μ^0, and for every h in H, consider the expression $X = \sum_j \langle h, J^{-1}e_j\rangle$, which is a sum of independent real-valued random variables $\xi_j(\omega) = \langle \omega, J^{-1}e_j\rangle$, with joint compensated Poisson distribution

$$\mathbb{E}_\mu\big\{e^{\mathtt{i}\sum_{j=1}^n c_j\xi_j}\big\} = \int_{\mathbb{R}_*^n} \big[\mathrm{e}^{\mathtt{i}\sum_{j=1}^n c_j s_j} - 1 - \mathtt{i}\sum_{j=1}^n c_j s_j\big]\bar{\pi}_n(\mathrm{d}s), \quad \forall c \in \mathbb{R}^n,$$

where $\bar{\pi}_n$ is the projection on $\mathbb{R}_*^n$ of π, i.e., with $h_j = (h, J^{-1}e_j)$,

$$\bar{\pi}_n(B) = \pi\big(\{h \in H_* : (h_1, \cdots, h_n) \in B,\, h_j = 0,\, \forall j > n\}\big), \quad \forall B \in \mathcal{B}(\mathbb{R}_*^n).$$

Thus

$$\mathbb{E}_\mu\{\sum_{j=1}^n |\xi_j|^2\} = \int_{\mathbb{R}_*^n} \Big(\sum_{j=1}^n s_j^2\Big)\bar{\pi}_n(\mathrm{d}s) = \int_{H_*} \langle Rh^n, h^n\rangle \pi_0(\mathrm{d}h),$$

where $h^n = \sum_{j=1}^n \langle h, J^{-1}e_j\rangle e_j$. Hence, the series $X = \sum_{j=1}^\infty \xi_j(\omega)e_j$ converges in $\bar{H} \subset L^2(H, \mathcal{B}(H), \mu)$, i.e., it can be considered as a $\bar{H}$-valued random variable on the probability space $(\Omega, \mathcal{F}, P) = (H, \mathcal{B}(H), \mu)$. Because $\{e_1, e_2, \ldots\}$ is an orthonormal basis in H_μ, the mapping

$$X(h) = \langle X, h\rangle = \sum_{j=1}^n \xi_j \,\langle J^{-1}e_j, h\rangle = \sum_{j=1}^n \xi_j \,(e_j, Jh)_\mu$$

is a H_μ-valued random variable (almost surely) well defined for any $h = J^{-1}Jh$ in H_μ, and called *Poisson white noise*.

Let $\{\xi_1, \xi_2, \ldots\}$ be a sequence of independent real-valued compensated Poisson random variables with parameters $\{\pi_1, \pi_2, \ldots\}$ in (complete) probability space $(\Omega, \mathcal{F}, P)$, i.e.,

$$\mathbb{E}\big\{e^{\mathtt{i}r\xi_j}\big\} = \exp\Big(\int_{\mathbb{R}_*} \big[\mathrm{e}^{\mathtt{i}rs} - 1 - \mathtt{i}rs\big]\pi_j(\mathrm{d}s)\Big), \quad \forall r \in \mathbb{R},\, j \geq 1,$$

with π_j satisfying

$$\int_{\mathbb{R}_*} s^2 \pi_j(\mathrm{d}s) \leq C_0, \quad \forall j \geq 1, \tag{4.4}$$

for some constant $C_0 > 0$. Now, for any given sequence of nonnegative real numbers $r = \{r_1, r_2, \ldots\}$, define the measures $\bar{\pi}_{r,n}$ and π_{j,r_j} on $\mathbb{R}^n$ as

$$\int_{\mathbb{R}^n} f(s)\bar{\pi}_{r,n}(\mathrm{d}s) = \sum_{j=1}^n \int_{\mathbb{R}} f_j(\sqrt{r_j}s_j)\pi_j(\mathrm{d}s_j) = \sum_{j=1}^n \int_{\mathbb{R}^n} f(s)\pi_{j,r_j}(\mathrm{d}s),$$

for any $n \geq 1$ and for every positive Borel function f in $\mathbb{R}^n$ satisfying $f(0) = 0$, where $s = (s_1, \dots, s_n)$ and $f_1(s_1) = f(s_1, 0, \dots, 0)$, $f_2(s_2) = f(0, s_2, \dots, 0)$, $\dots$, $f_n(s_n) = f(0, 0, \dots, s_n)$, i.e.,

$$\pi_{j,r_j}(\mathrm{d}s) = \delta_0(\mathrm{d}s_1) \dots \delta_0(\mathrm{d}s_{j-1}) \pi_j(r_j^{-1/2} \mathrm{d}s_j) \delta_0(\mathrm{d}s_{j+1}) \dots \delta_0(\mathrm{d}s_n),$$

where δ_0 is the Dirac measure at 0 and $\pi_j(r_j^{-1/2}\mathrm{d}s_j) = 0$ if $r_j = 0$. We can check that $\bar{\xi}_{r,n} = (\sqrt{r_1}\xi_1, \dots, \sqrt{r_n}\xi_n)$ has a compensated Poisson distribution with parameter $\bar{\pi}_{r,n}$, i.e.,

$$\mathbb{E}\{e^{\mathtt{i}c\cdot\bar{\xi}_{r,n}}\} = \int_{\mathbb{R}^n_*} \big[\mathrm{e}^{\mathtt{i}c\cdot s} - 1 - \mathtt{i}c \cdot s\big] \bar{\pi}_{r,n}(\mathrm{d}s), \quad \forall c \in \mathbb{R}^n,$$

where the dot "$\cdot$" denotes the scalar product in $\mathbb{R}^n$. Clearly, (4.4) implies

$$\sum_{j=1}^n \int_{\mathbb{R}^n_*} |s_j|^2 \bar{\pi}_{r,n}(\mathrm{d}s) \leq C_0 \sum_{j=1}^n r_j, \quad \forall n \geq 1,$$

with the same constant $C_0 > 0$.

Moreover, we may regard the measures $\bar{\pi}_{r,n}$ and π_{j,r_j} as being defined either on $\mathbb{R}^n$ or directly on the infinite product $\mathbb{R}^\infty$ (the space of all sequences), namely,

$$\int_{\mathbb{R}^\infty} f(s) \bar{\pi}_{r,n}(\mathrm{d}s) = \int_{\mathbb{R}^n} f(s_1, \dots, s_n, 0, 0, \dots) \bar{\pi}_{r,n}(\mathrm{d}s)$$

or equivalently,

$$\pi_{j,r_j}(\mathrm{d}s) = \delta_0(\mathrm{d}s_1) \dots \delta_0(\mathrm{d}s_{j-1}) \pi_j(r_j^{-1/2} \mathrm{d}s_j) \delta_0(\mathrm{d}s_{j+1}) \delta_0(\mathrm{d}s_{j+2}) \dots,$$

and $\bar{\pi}_{r,n} = \sum_{j=1}^n \pi_{j,r_j}$. Note the projection type property

$$\bar{\pi}_{r,n}(B) = \bar{\pi}_{r,n+k}\big(\{s \in \mathbb{R}^\infty : (s_1, \dots, s_n) \in B,\ s_j = 0,\ j > n\}\big),$$

for any B in $\mathcal{B}(\mathbb{R}^n)$. Therefore, the series $\bar{\pi}_r = \sum_{j=1}^\infty \pi_{j,r_j}$ defines a measure on $\mathbb{R}^\infty$. Hence, if the series $\sum_{j=1}^\infty r_j$ is convergent then

$$\int_{\mathbb{R}^\infty} |s|^2 \bar{\pi}_r(\mathrm{d}s) = \sum_{j=1}^\infty \int_{\mathbb{R}^n_*} |s_j|^2 \bar{\pi}_{r,n}(\mathrm{d}s) \leq C_0 \sum_{j=1}^\infty r_j < \infty, \tag{4.5}$$

i.e., $\bar{\pi}_r$ becomes a σ-finite measure on $\ell^2_* = \ell^2 \setminus \{0\}$, where ℓ^2 is the Hilbert space of square-convergent sequences. Also, we have

$$\int_{\ell^2_*} f(s) \bar{\pi}_r(\mathrm{d}s) = \lim_n \int_{\ell^2_*} f(s) \bar{\pi}_{r,n}(\mathrm{d}s) = \sum_{j=1}^\infty \int_{\ell^2_*} f(s) \pi_{j,r_j}(\mathrm{d}s),$$

for any continuous function f such that $|f(s)| \leq |s|^2$, for any s in ℓ^2_*. Moreover, since $r_j = 0$ implies $\pi_{j,r_j} = 0$ on ℓ^2_*, we also have $\pi_{j,r_j}(R^{-1}\{0\}) = 0$ for any j, where R is the nonnegative symmetric trace-class operator $s \mapsto (r_1 s_1, r_2 s_2, \dots)$.

Hence $\bar\pi_r(R^{-1}\{0\}) = 0$. This means that support of $\bar\pi_r$ is contained in $R(\ell^2_*)$ and we could define a new pre-image measure by setting $\bar\pi_0(B) = \bar\pi_r(RB)$, for any B in $\mathcal{B}(\ell^2_*)$ with the property

$$\int_{\ell^2_*} f(s)\bar\pi(\mathrm{d}s) = \int_{\ell^2_*} f(Rs)\bar\pi_0(\mathrm{d}s), \quad \forall f \ge 0 \text{ and measurable.}$$

It is clear that estimate (4.5) identifies the measures only on ℓ^2_* and so, we may (re)define all measures at $\{0\}$ by setting

$$\bar\pi_r(\{0\}) = \bar\pi_{r,n}(\{0\}) = \pi_{j,r_j}(\{0\}) = \bar\pi_0(\{0\}) = 0.$$

Then we can consider the measures as σ-finite defined either on ℓ^2 or on ℓ^2_*.

Now, let H be a separable Hilbert space, R be a nonnegative symmetric (trace-class) operator in $L_1(H)$, and $\{e_1, e_2, \dots\}$ be an orthonormal basis of eigenvectors of R, i.e., $Re_j = r_j e_j$, $(e_j, e_k) = 0$ if $j \ne k$, $|e_j| = 1$, for every j, and $\mathrm{Tr}(R) = \sum_{j=1}^\infty r_j < \infty$, $r_j \ge 0$. Note that the kernel of R may be of infinite dimension, i.e., there infinite many $r_j = 0$. Consider the product measure π on $H_* = H \setminus \{0\}$, with support in $R(H)$, defined as

$$\pi\big(\{h \in H_* : (h, e_j) \in Be_j,\ \forall j\}\big) = \bar\pi_r(B), \quad \forall B \in \mathcal{B}(H_*)$$

or equivalently

$$\int_{H_*} f(h)\pi(\mathrm{d}h) = \int_{\ell_*} f(s_1e_1 + \cdots + s_ne_n + \cdots)\bar\pi_r(\mathrm{d}s),$$

nonnegative Borel function f in H_*. In particular,

$$\int_H |h|^2\pi(\mathrm{d}h) = \int_{H_*} \Big|\sum_{j=1}^\infty s_je_j\Big|^2 \bar\pi_r(\mathrm{d}s) = \sum_{j=1}^\infty \int_{\ell_*} s_j^2 r_j \pi_j(\mathrm{d}s) \le C_0\,\mathrm{Tr}(R)$$

and if $\pi_0 = R\pi$, i.e., $\pi_0(B) = \pi(RB)$, for every B in $\mathcal{B}(H_*)$, then

$$\int_{H_*} f(h)\pi(\mathrm{d}h) = \int_{H_*} f(Rh)\pi_0(\mathrm{d}h),$$

for any nonnegative Borel measurable function f on H_*.

• *Remark* 4.5. Recall the following result, e.g., see Federer [79, Section 2.2.13, pp. 69]. Let X be a complete separable metric space, Y be a Hausdorff space, $f\colon X \to Y$ be a continuous function, and μ be a measure Y such that every closed subset of Y is μ measurable. Then the f image of every Borel subset of X is μ measurable. This result is classic in the general study of Borel σ-algebras, analytic sets and universally measurable sets, i.e., the fact that a projection of a Borel measurable set is analytic and so, it is universally measurable, e.g., see Dellacherie and Meyer [58, Section III.75–85, pp. 243–254], Dudley [62, Section 13.2, pp 493–499] or Parthasarathy [195, Section I.3]. It is now clear that the above measure π_0 can be defined in term of π, provided that π has support contained in $R(H_*)$. Note that for any orthonormal basis $\{e_j\}$ in H and any

measure m on H with $m(\{0\}) = 0$ we have $m(B) = \sum_j m(e_j B)$, for any B in $\mathcal{B}(H)$, where $e_j B$ is the (orthogonal) projection of B in the e_j direction, i.e., $e_j B = \{(b, e_j)e_j : b \in B\}$. Thus, for any integrable function f with $f(0) = 0$ we have

$$\int_H f(h)m(\mathrm{d}h) = \sum_j \int_H f(h)m(e_j \mathrm{d}h) = \sum_j \int_H f(e_j h)m(\mathrm{d}h),$$

where $f(e_j h) = f\big((h, e_j)e_j\big)$ and $m(e_j \mathrm{d}h)$ is the measure $B \mapsto m(e_j B)$. □

Therefore, the H-valued random variable

$$X = \sum_{j=1}^{\infty} \sqrt{r_j}\, \xi_j\, e_j$$

satisfies

$$\mathbb{E}\{|X|^2\} = \sum_{j=1}^{\infty} r_j\, \mathbb{E}\{|\xi_j|^2\} = \sum_{j=1}^{\infty} r_j \int_{\mathbb{R}_*} s^2 \pi_j(\mathrm{d}s),$$

and

$$\begin{aligned}\mathbb{E}\big\{e^{\mathtt{i}(h,X)}\big\} &= \prod_{j=1}^{\infty} \mathbb{E}\big\{e^{\mathtt{i}\sqrt{r_j}(h,e_j)\xi_j}\big\} = \\ &= \exp\Big(\sum_{j=1}^{\infty} \int_{\mathbb{R}_*} \big[\mathrm{e}^{\mathtt{i}\sqrt{r_j}(h,e_j)s_j} - 1 - \mathtt{i}\sqrt{r_j}(h,e_j)s_j\big]\pi_j(\mathrm{d}s_j)\Big) = \\ &= \exp\Big(\sum_{j=1}^{\infty} \int_{\ell_*^2} \big[\mathrm{e}^{\mathtt{i}(h,e_j)s} - 1 - \mathtt{i}(h,e_j)s\big]\bar{\pi}_r(\mathrm{d}s)\Big),\end{aligned}$$

i.e.,

$$\begin{aligned}\mathbb{E}\big\{e^{\mathtt{i}(h,X)}\big\} &= \exp\Big(\int_{H_*} \big[\mathrm{e}^{\mathtt{i}(h,k)} - 1 - \mathtt{i}(h,k)\big]\pi(\mathrm{d}k)\Big) = \\ &= \exp\Big(\int_{H_*} \big[\mathrm{e}^{\mathtt{i}(Rh,k)} - 1 - \mathtt{i}(Rh,k)\big]\pi_0(\mathrm{d}k)\Big),\end{aligned}$$

for every h in H. Thus, X is a compensated Poisson random variable with values in H and Lévy measure π in H_*. Next, the mapping

$$h \mapsto X(h) = \sum_{j=1}^{\infty} \sqrt{r_j}\, \xi_j\, (h, e_j)$$

from H into $L^2(\Omega, \mathcal{F}, P)$ is linear, $X(h)$ is a (real-valued) compensated Poisson random variable with covariance $\mathbb{E}\{[X(h)]^2\} = |h|^2$, for any h in H. Thus the space $\{X(h) : h \in H\}$ is a Poisson subspace of $L^2(\Omega, \mathcal{F}, P)$, which is isomorphic to H. In particular $\mathbb{E}\{X(f)\, X(g)\} = (f, g)$, for any f and g in H, and $X(f)$ is a compensated Poisson variable independent of $X(g)$ if $(f, g) = 0$, i.e., if f and g are orthogonal. The family $\{X(h) : h \in H\}$ is called an *compensated Poisson*

stochastic process. If $H = L^2(A, \mathcal{A}, \mu)$, where $(A, \mathcal{A}, \mu)$ is a σ-finite measure space, the mapping X is called a *Poisson measure* or *Poisson white noise* with intensity μ on $(A, \mathcal{A})$. When F belongs to $\mathcal{A}$ we write $X(F) := X(\mathbb{1}_F)$. Thus, if F and G are sets with $\mu(F) < \infty$ and $\mu(G) < \infty$ then $\mathbb{E}\{X(F)\,X(G)\} = \mu(F \cap G)$, and so that $X(F)$ and $X(G)$ are independent when F and G are disjoint. Notice that if $\{F_k : k = 1, 2, \ldots\}$ is a pairwise disjoint sequence of subset in $\mathcal{A}$, $F = \cup_k F_k$ with $\mu(F) < \infty$ then $X(F) = \sum_k X(F_k)$ *almost surely* so that some regularity (as in the case of regular conditional probability) is need to ensure the existence of a good selection, in order that $F \mapsto X(F, \omega)$ is a measure for ω outside of a set of probability zero.

Sometimes, the initial point is a family of compensated Poisson random variables $X = \{X(h) : h \in H\}$ in a complete probability space $(\Omega, \mathcal{F}, P)$, where the index H is a separable Hilbert space, the σ-algebra $\mathcal{F}$ is the smallest complete σ-algebra such that $X(h)$ is measurable for any h in H and $\mathbb{E}\{X(f)\,X(g)\} = (f, g)_H$, for any f and g in H. This is called a compensated Poisson process on H. For the particular case of a standard Poisson process (and some similar one, like symmetric jumps) we have the so-called Charlier polynomials $c_{n,\lambda}(x)$, an orthogonal basis in $L^2(\mathbb{R}^+)$ with the weight $\alpha(x) = \sum_{n=1}^{\infty} \mathbb{1}_{\{x \geq n\}} \mathrm{e}^{-\lambda} \lambda^n / n!$, $\lambda \neq 0$, which are the equivalent of Hermit polynomials in the case of a Wiener process. Charlier polynomials are defined by the generating function

$$t \mapsto \mathrm{e}^{-\lambda t}(1 + t)^x = \sum_{n=0}^{\infty} c_{n,\lambda}(x) \frac{t^n}{n!},$$

or explicitly by the expression

$$c_{n,\lambda}(x) = \sum_{k=0}^{n} \binom{n}{k} \binom{x}{k} k!\, (-\lambda)^{n-k}$$

and they satisfy the orthogonal relations

$$\int_0^{\infty} c_{m,\lambda}(x)\, c_{n,\lambda}(x)\, \mathrm{d}\alpha(x) = \sum_{k=1}^{\infty} c_{m,\lambda}(k)\, c_{n,\lambda}(k)\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = 0, \quad \text{if } m \neq n$$

and

$$\int_0^{\infty} c_{n,\lambda}(x)\, c_{n,\lambda}(x)\, \mathrm{d}\alpha(x) = \sum_{k=1}^{\infty} c_{n,\lambda}(k)\, c_{n,\lambda}(k)\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = \lambda^n n!.$$

Also the three-terms recurrence formula

$$c_{\lambda,n+1}(x) = (x - n - \lambda) c_{\lambda,n}(x) - n\lambda c_{\lambda,n-1}(x),$$

and several other properties, e.g., see Chihara [41] or Szegö[242].

The previous analysis applied to the *particular case* when a Lévy measure π is given on a separable Hilbert space H. The measure π is constructed from a nonnegative symmetric (trace-class) operator R in $L_1(H)$ with eigenvalues and

eigenvectors $\{r_j, e_j,\, j = 1, 2, \ldots\}$, where $\{e_j\}$ is a orthonormal basis in H and $\sum_j r_j < \infty$, and a sequence $\{\pi_j\}$ of Lévy measures on $\mathbb{R}_*$ satisfying (4.4). Thus, we define the Lévy measures $\bar\pi_{r,n}$ on $\mathbb{R}^n$ satisfying (4.5), which induces the Lévy measures $\bar\pi_r$ and $\bar\pi$ on ℓ^2_* and π and π_0 on H_*, satisfying

$$\int_{H_*} |h|^2_H \pi(\mathrm{d}h) = \int_{H_*} |R^{1/2}h|^2_H \pi_0(\mathrm{d}h) < \infty. \tag{4.6}$$

By means of Sazonov's Theorem 1.41, there is a probability measure P on $(\Omega, \mathcal{F})$, with $\Omega = L^2(\mathbb{R}, H_*)$ and $\mathcal{F} = \mathcal{B}(\Omega)$, such that for any ϕ in $L^2(\mathbb{R}, H_*)$ we have

$$\begin{aligned}\mathbb{E}\{e^{\mathtt{i}\langle\cdot,\phi\rangle}\} &= \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(\phi(t),h)} - 1 - \mathtt{i}(\phi(t), h)\big]\pi(\mathrm{d}h)\Big) = \\ &= \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(R\phi(t),h)} - 1 - \mathtt{i}(R\phi(t), h)\big]\pi_0(\mathrm{d}h)\Big),\end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2(\mathbb{R}, H_*)$ and $(\cdot,\cdot)$ is the inner product in H. Hence, we can pick a H_*-valued random variable $\bar p(t)$ in $\omega \mapsto (\omega, \cdot\, \mathbb{1}_{(0,t)})$ such that $t \mapsto \bar p(t)$ is a cad-lag stochastic process, called a (H_*-valued) compensated Poisson point process with Lévy measure π.

On the other hand, consider the space $\Omega = L^2_\pi(\mathbb{R}\times H_*)$ with the σ-finite product measure $\mathrm{d}t \times \pi(\mathrm{d}h)$ on $\mathbb{R} \times H_*$. Again, by means of Sazonov's Theorem 1.41 (remark that the condition (4.6) is not being used), there is a probability measure P on $(\Omega, \mathcal{F})$, with $\mathcal{F} = \mathcal{B}(\Omega)$, such that

$$\mathbb{E}\{e^{\mathtt{i}\langle\cdot,\varphi\rangle}\} = \exp\Big(\int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \big[\mathrm{e}^{\mathtt{i}\varphi(t,h)} - 1 - \mathtt{i}\varphi(t, h)\big]\pi(\mathrm{d}h)\Big),$$

for any φ in $L^2_\pi(\mathbb{R} \times H_*)$, where now $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2_\pi(\mathbb{R} \times H_*)$. Note that if $\{(t, y) : \varphi_1(t, y) \neq 0\}$ and $\{(t, y) : \varphi_2(t) \neq 0\}$ are disjoint in $\mathbb{R} \times H_*$ (except for a set of $\mathrm{d}t \times \pi(\mathrm{d}y)$ measure zero), then the random variables (ω, φ_1) and (ω, φ_2) are independent. Now, in particular, if $\varphi = \mathbb{1}_{(0,t)}\mathbb{1}_B$, $t > 0$ and B in $\mathcal{B}(H_*)$, with $\pi(B) < \infty$, we can pick a real-valued random variable $p(t, B)$ in

$$\omega \mapsto \int_{\mathbb{R}} \Big(\int_B \omega(t, h)\pi(\mathrm{d}h) + \pi(B)\Big)\varphi(t)\mathrm{d}t,$$

such that $t \mapsto p(t, B)$ is a cad-lag stochastic process and $B \mapsto p(t, B)$ is a (random) measure, called a Poisson (integer) measure. Actually, p is a measure in both variables. These stochastic process has the following properties:

(1) For any B in $\mathcal{B}(H_*)$, with $\pi(B) < \infty$, the real-valued process $p(\cdot, B)$ or the H-valued process $\bar p$ has independent increments.

(2) For any sequence of disjoint sets $B_1, \ldots, B_n$ in $\mathcal{B}(H_*)$ the stochastic processes $p(t, B_1), \ldots, p(t, B_n)$ are independent.

(3) The process $p(t, B)$ is a Poisson process with parameter $\pi(B)$ and $\bar p(t)$ is a compensated Poisson point process, i.e., for any sequences of disjoint sets

$B_1, \dots, B_n$ in $\mathcal{B}(H_*)$ with $\pi(B_i) < \infty$, and $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n$ we have

$$\mathbb{E}\Big\{ e^{\mathtt{i} \sum_{j=1}^n r_j (p(t_j, B_j) - p(t_{j-1}, B_j))} \Big\} = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1}) \pi(B_j) \big[\mathrm{e}^{\mathtt{i} r_j} - 1 \big] \Big),$$

for any sequence $r_1, \dots, r_n$ in $\mathbb{R}$, whilst for the H-valued process $\bar{p}(t)$ we obtain

$$\begin{aligned} \mathbb{E}\Big\{ e^{\mathtt{i} \sum_{j=1}^n (\bar{p}(t_j) - \bar{p}(t_{j-1}), h_j)} \Big\} = \\ = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1}) \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(h_j, h)} - 1 - \mathtt{i}(h_j, h) \big] \pi(\mathrm{d}h) \Big) = \\ = \exp\Big(\sum_{j=1}^n (t_j - t_{j-1}) \int_{H_*} \big[\mathrm{e}^{\mathtt{i}(R h_j, h)} - 1 - \mathtt{i}(R h_j, h) \big] \pi_0(\mathrm{d}h) \Big), \end{aligned}$$

for any sequence $h_1, \dots, h_n$ in H.

(4) For each $s > t \geq 0$, we have $\mathbb{E}\{\bar{p}(t)\} = 0$,

$$\mathbb{E}\big\{ |\bar{p}(s) - \bar{p}(t)|^2 \big\} = (s-t) \int_{H_*} |h|^2 \pi(\mathrm{d}h) = (s-t) \int_{H_*} |R^{1/2} h|^2 \pi_0(\mathrm{d}h),$$

and $p(0, B) = 0$ almost surely. Moreover, using the independence of increments we find that

$$\mathbb{E}\big\{ |\bar{p}(r) - \bar{p}(s)|^2 |\bar{p}(s) - \bar{p}(t)|^2 \big\} = (s-t)(r-s) \Big(\int_{H_*} |h|^2 \pi(\mathrm{d}h) \Big)^2,$$

for any $r > s > t \geq 0$.

(5) For any deterministic function φ in $L^2_\pi(\mathbb{R} \times H)$ and ϕ in $L^2(\mathbb{R}, H)$, we can define the (stochastic) integrals

$$\begin{aligned} \int_{\mathbb{R} \times H_*} \varphi(t, h) \tilde{p}(\mathrm{d}t, \mathrm{d}h) &= \langle \cdot, \varphi \rangle_{L^2_\pi(\mathbb{R} \times H)} = \int_{\mathbb{R}} \mathrm{d}t \int_{H_*} \omega(t, h) \varphi(t, h) \pi(\mathrm{d}h), \\ \int_{\mathbb{R}} \big(\phi(t), \bar{p}(\mathrm{d}t) \big)_H &= \langle \cdot, \phi \rangle_{L^2(\mathbb{R}, H)} = \int_{\mathbb{R}} \big(\omega(t), \phi(t) \big)_H \mathrm{d}t, \end{aligned}$$

where $\tilde{p}(t, B) = p(t, B) - t\pi(B)$. In particular, if we assume (4.6) then π integrates $h \mapsto |h|^2$, and we can define the stochastic integral

$$\omega \mapsto \int_{H_*} h p(t, \mathrm{d}h) = \int_{(0,t]} \mathrm{d}t \int_{H_*} \omega(t, h) h \pi(\mathrm{d}h),$$

which has the same distribution as the compensated Poisson point process $\bar{p}(t)$ obtained before.

The law of the process $\bar{p}$ on the canonical space either $D([0, \infty), H)$ or $D([0, \infty), H_*)$ is called a (H-valued) compensated Poisson measure with Lévy measure π.

4.1.3 Integer Random Measures

In the same way that a measure (or distribution) extends the idea of a function, random measures generalize the notion of stochastic processes. In terms of random noise, the model represents a noise distribution in time and some other *auxiliary space variable*, generalizing the model of noise distribution only in the time variable. Loosely speaking, we allow the index to be a measure. The particular class where the values of the measure are only positive integers is of particular interest to study the jumps of a random process.

Returning to the sample space, we know that an element ω in $D([0,\infty),\mathbb{R}^d)$ has at most a countable number of jumps, with only a finite number of jumps of size greater than a positive quantity. For any Borel set B in $\mathcal{B}(\mathbb{R}^d_*)$ with $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$ (so-called punctured d-space) the number of jumps before a time t and with values in B are finite if B is compact. Thus, for any (cad-lag) stochastic process with values in $\mathbb{R}^d$ or equivalently for any random variable X with values in $D([0,\infty),\mathbb{R}^d)$ we can define a measure ν_X with integer values, as the number of jumps in B within a bounded time interval, i.e.,

$$\nu_X(B\times]a,b],\omega) := \#\{t : a < t \le b,\ X(t,\omega) - X(t-,\omega) \in B\}, \tag{4.7}$$

for any $b > a \ge 0$, B in $\mathcal{B}(\mathbb{R}^d_*)$, and where $\#$ denotes the number of elements (which may be infinite) of a set. Sometime we use the notation $\nu_X(B,]a,b],\omega)$ and we may look at this operation as a *functional* on $D([0,\infty),\mathbb{R}^d)$, i.e., for every $b > a \ge 0$ and B in $\mathcal{B}(\mathbb{R}^d_*)$,

$$\nu(B,]a,b],\omega) := \sum_{a<t\le b} \mathbb{1}_B\big(\omega(t) - \omega(t-)\big),$$

so that $\nu_X(B\times]a,b],\omega) = \nu(B,]a,b],X(\cdot,\omega))$. For each ω, this is Radon measure on $\mathbb{R}^d_*\times(0,\infty)$ with integer values. By setting $\nu(\mathbb{R}^d_*\times\{0\}) := 0$ we may consider ν as a measure on $\mathbb{R}^d_*\times[0,\infty)$.

This measure ν is used as a characterization of the jumps $\delta X := (\delta X(t) = X(t) - X(t-) : t > 0)$, in the sense that ν vanishes if and only if the process X is continuous. Note that for any continuous function $f(t,x)$ which vanishes near $x = 0$ we have

$$\int_{\mathbb{R}^d_*\times(a,b]} f(x,t)\nu(\mathrm{d}\zeta,\mathrm{d}t) = \sum_{\delta X(t)\neq 0} \mathbb{1}_{\{a<t\le b\}} f\big(\delta X(t),t\big),$$

where the sum is finite. In this sense, the random measure ν contains all information about the jumps of the process X. Moreover, remark that ν is a sum of Dirac measures at $(\delta X(t),t)$, for $\delta X(t) \neq 0$. This sum is finite on any set separated from the origin, i.e., on any sets of the form

$$\big\{(x,t) \in \mathbb{R}^d_*\times[0,\infty) : t \in]a,b],\ |x| \ge \varepsilon\big\},$$

for every $b > a \ge 0$ and $\varepsilon > 0$.

Recall that the Skorokhod's topology, given by the family of functions defined for ω in $D([0,\infty),\mathbb{R}^d)$ by the expression

$$w(\omega,\delta,]a,b]) := \inf_{\{t_i\}} \sup_i \sup\{|\omega(t)-\omega(s)| \,:\, t_{i-1}\le s<t<t_i\}$$

where $\{t_i\}$ ranges over all partitions of the form $a=t_0<t_1<\cdots<t_{n-1}<b\le t_n$, with $t_i-t_{i-1}\ge\delta$ and $n\ge 1$, makes $D([0,\infty),\mathbb{R}^d)$ a complete separable metric space. Again, note that

$$\nu(\{z\in\mathbb{R}^d : |z|\ge w(\omega,\delta,]a,b])\},]a,b],\omega)\le\frac{b-a}{\delta},$$

for every $\delta>0$, and $b>a\ge 0$.

Another point is the following fact that for any set B in $\mathbb{R}^d_*$ with a positive distance to the origin, we can define the sequence of jump-times and jump-size as

$$\begin{aligned}
&\tau_0(B,]a,b],\omega) := a, \quad \zeta_0(B,]a,b],\omega) := 0,\\
&\tau_k(B,]a,b],\omega) := \inf\{t\in]\tau_{k-1},b] : \omega(t)-\omega(t-)\in B\}, \quad k\ge 1,\\
&\zeta_k(B,]a,b],\omega) := \omega(\tau_k)-\omega(\tau_k-), \quad k\ge 1,
\end{aligned}$$

for any $b\ge a\ge 0$ and $1\le k\le\nu(B,]a,b],\omega)$. Thus, if ω_n is a sequence converging to ω in $D([0,\infty),\mathbb{R}^d)$, and B is also an open set with boundary ∂B satisfying $\nu(\partial B,]a,b],\omega)=0$, and such that the first jump $\omega(a)-\omega(a-)$ and the last jump $\omega(b)-\omega(b-)$ have a positive distance to B, then

$$\begin{aligned}
&\nu(B,]a,b],\omega_n)\to\nu(B,]a,b],\omega),\\
&\tau_k(B,]a,b],\omega_n)\to\tau_k(B,]a,b],\omega),\\
&\zeta_k(B,]a,b],\omega_n)\to\zeta_k(B,]a,b],\omega),
\end{aligned}$$

for any $k=0,1,\ldots,\nu(B,]a,b],\omega)$.

Definition 4.6 (integer measure)**.** Let $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge 0)$ be a filtered space. A *random measure* on a Polish space E is a random variable ν with values in the space of σ-finite measures on the Borel σ-algebra $\mathcal{B}(E)$. In most of the cases, the Polish space E is locally compact and the random variable ν take values in the space of Radon (nonnegative) measures (finite on every compact sets) on $\mathcal{B}(E)$. If the time-variable is singled-out, e.g., $E=\mathbb{R}^m_*\times[0,\infty)$ then it is required that $\nu(\mathbb{R}^m_*\times\{0\})=0$. In this case a random measure on $\mathbb{R}^m_*\times[0,\infty)$ is called a *optional or predictable (respectively, locally integrable)* if for any stopping time $\tau<\infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t\mapsto\nu(K\times[0,t\wedge\tau])$ is optional or $t\mapsto\nu(K\times[0,t\wedge\tau[)$ is predictable (respectively, $\mathbb{E}\{\nu(K\times[0,t\wedge\tau])\}<\infty$ for every $t>0$). Moreover, an optional locally integrable random measure ν is called *integer measure* or *integer-valued random measure* if it takes values in $\{0,1,\ldots,\infty\}$, $\nu(\mathbb{R}^m_*\times\{0\})=0$ and $\nu(\mathbb{R}^m_*\times\{t\})=0$ or $=1$ for any $t>0$. □

When referring to an integer-valued random measure, the above definition implies that we mean an optional locally integrable integer-valued random measure. Moreover, the local integrability ensures that the *product measure* $\nu(\mathrm{d}x \times \mathrm{d}t, \omega)\, P(\mathrm{d}\omega)$ is σ-finite. It is clear that we may replace $\mathbb{R}^m_*$ by a locally compact Polish E. An essential point is the use of the following two properties: (1) the σ-algebra $\mathcal{E}$ is generated by a countable algebra and (2) any $(E, \mathcal{E})$-valued random variable x on a probability space $(\Omega, \mathcal{F}, P)$ admits a regular conditional distribution relative to a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. This *disintegration property* (2) can be restated as: for any positive and finite measure m on the product space $(E \times B, \mathcal{E} \times \mathcal{B})$ there exist a measurable kernel $k(\mathrm{d}x, b)$ such that $m(\mathrm{d}x, \mathrm{d}b) = k(\mathrm{d}x, b)\, m_B(\mathrm{d}b)$, where $m_B(\mathrm{d}b) := m(E, \mathrm{d}b)$ is the B-marginal distribution of m. Clearly, this is related to the conditional property, and this is used to define the *compensator*, a key instrument for the stochastic integral. These properties are satisfied by the so-called *Blackwell spaces*, see Dellacherie and Meyer [58]. Only the case of locally compact Polish spaces will be used here.

A typical example of optional (respectively, predictable) integer measure on $\mathbb{R}^m_*$ is the one constructed by (4.7) for an adapted (i.e., optional) (respectively, predictable) locally integrable stochastic process with values in $\mathbb{R}^m$. Notice that integrability at infinity is not an issue in the above definition of integer-valued measure, the key part is the integrability away of the origin, i.e., we may use $\mathbb{E}\{\nu(B \times [0,t])\} < \infty$, for any Borel subset B of $\mathbb{R}^m_*$ with a positive distance to the origin. Certainly, this can be viewed as a *localization* (via a sequence of stopping times) of the integral condition

$$\mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times [0,t]} (|\zeta|^2 \wedge 1) \nu(\mathrm{d}\zeta, \mathrm{d}t) \Big\} < \infty,$$

for every $t \geq 0$, which is used later for Lévy measures.

Given an integer-valued random measure ν on $\mathbb{R}^m_*$, the set $\{t : \nu(\mathbb{R}^m_* \times \{t\}) = 1\}$ is countable for any ω and can be written as a sequence $(\tau_n(\omega) : n = 1, 2, \ldots)$. Moreover, because ν assumes only integers values, there is a sequence $(a_n(\omega) : n = 1, 2, \ldots)$ such that $\nu(\{(a_n, \tau_n)\}) = 1$ and $\nu(\mathbb{R}^m_* \times [0, \infty) \smallsetminus \{(a_n, \tau_n)\}) = 0$. Because ν is finite on compact subsets of $\mathbb{R}^d_*$, for each ε, $t > 0$ there exists only a finite number of (a_n, τ_n) such that $\varepsilon \leq |a_n| \leq 1/\varepsilon$ and $\tau_n \leq t$. Hence we may always rewrite ν as

$$\begin{cases} \nu(B, \omega) = \sum_n \mathbb{1}_{(a_n(\omega), \tau_n(\omega)) \in B}, \quad \forall B \in \mathcal{B}(\mathbb{R}^m_* \times [0, \infty)), \\ A^\varepsilon_\nu(t, \omega) := \sum_n a_n(\omega)\, \mathbb{1}_{\varepsilon \leq |a_n| \leq 1/\varepsilon}\, \mathbb{1}_{\tau_n(\omega) \leq t}, \quad \forall t \geq 0, \end{cases} \tag{4.8}$$

this determines an optional locally integrable jump process A^ε_ν on $\mathbb{R}^m_*$, and so the following expression for every F in $\mathcal{F}(s)$ and $t \geq s \geq 0$,

$$\mu^\varepsilon(]s, t] \times F) := \mathbb{E}\big\{ [A^\varepsilon_\nu(t) - A^\varepsilon_\nu(s)]\, \mathbb{1}_F \big\}, \tag{4.9}$$

defines a bounded ($\mathbb{R}^d$-valued) measure on $[0, \infty) \times \Omega$.

If the jump processes $\{A^\varepsilon_\nu : \varepsilon > 0\}$ have a uniformly locally integrable bounded variation, i.e., $\mathbb{E}\{\sum_n |a_n|\} < \infty$,, then $A_\nu = (\sum_n a_n \mathbb{1}_{\tau_n \le t} : t \ge 0)$ has a locally integrable bounded variation (when $d = 1$ we have a signed measure μ^ε) and a measure μ (limit as $\varepsilon \to 0$, which is called Doléans measure), can be defined. To come back from this ($\mathbb{R}^d$-valued) measure μ^ε to the process A^ε (or to the integer-valued random measure ν), we need μ^ε to vanish for any evanescent set, i.e., $\mu^\varepsilon(N) = 0$ for any subset N of $[0,\infty) \times \Omega$ such that $P(\cup_t\{\omega : (t,\omega) \in N)\}) = 0$. The point is that the integer measure ν captures all the features of the family of processes A^ε, even when A can not be defined. In other words, if A^ε is a semi-martingale we will see that μ^ε may define a measure as ε vanishes.

Returning to the compensator, as in Definitions 3.5 (in Chapter 3) and 4.6, we have a unique dual predictable projection ν^p of any optional locally integrable random measure ν, characterized (almost surely) as being a predictable random measure such that $\mathbb{E}\{\nu(K \times [0, t \wedge \tau]) - \nu^p(K \times [0, t \wedge \tau])\} = 0$ for any for any stopping time $\tau < \infty$, any compact subset K of $\mathbb{R}^m_*$ and any $t > 0$, or equivalently the process $t \mapsto \nu(K\times]0,t]) - \nu^p(K\times]0,t])$ is a martingale. Hence, by a monotone class argument, we have

$$\mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times [0,\infty)} X(z,t)\, \nu(\mathrm{d}z, \mathrm{d}t)\Big\} = \mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times [0,\infty)} X(z,t)\, \nu^p(\mathrm{d}z, \mathrm{d}t)\Big\},$$

for any nonnegative function $(z,t,\omega) \mapsto X(z,t,\omega)$ measurable with respect to the product σ-algebra $\mathcal{B}(\mathbb{R}^m_*) \times \mathcal{O}$ (with $\mathcal{O}$ being the optional σ-algebra) where the product measure $\nu(\mathrm{d}z, \mathrm{d}t, \omega)\, P(\mathrm{d}\omega)$ is defined. Recall that we assume $\nu(\mathbb{R}^m_* \times \{0\}) = 0$, so that $\nu(K \times \{0\}) = \nu^p(K \times \{0\}) = 0$. Moreover, based on the disintegration property, the predictable compensator can be written as $\nu^p(\mathrm{d}z, \mathrm{d}t, \omega) = k(\mathrm{d}z, t, \omega)\mathrm{d}A(t, \omega)$, where A is a integrable predictable increasing process and $k(\mathrm{d}z, t, \omega)$ is a measurable kernel. We refer to Bichteler [25, Sections 3.10, 4.3, pp. 171–186, 221–232], He et al. [105], Jacod and Shiryaev [117, Section II.1, pp. 64–74], and Kallenberg [120] for a full discussion on random measures, only some results are reported here.

Theorem 4.7. *Let ν^p be compensator of an integer-valued random measure ν. Then the predictable random measure ν^p (which is not necessarily an integer-valued random measure) has the following properties. First* (a) *its predictable support, namely the set $\{(t,\omega) : 0 < \nu^p(\mathbb{R}^m_* \times \{t\}, \omega) \le 1\}$, can be written as a sequence of predictable stopping times, i.e., $\{(\tau^p_n(\omega), \omega) : n = 1, 2, \ldots\}$ with τ^p_n a predictable stopping time for any n, and $P(\{\omega : 0 < \nu^p(\mathbb{R}^m_* \times \{t\}, \omega) \le 1\}) = 1$, for any $t \ge 0$. Next* (b) *we have*

$$\nu^p(K \times \{\tau\}) = \mathbb{E}\Big\{ \sum_n \mathbb{1}_{a_n \in K} \,|\, \mathcal{F}(\tau-)\Big\},$$

on the predictable support, for any predictable stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$. Moreover, if ν is defined as the number of jumps (4.7)

of a (special) semi-martingale X then the predictable processes in $t > 0$,

$$\sqrt{\sum_{0<s\le t} \nu^p(\mathbb{R}^m_* \times \{s\})} \quad \textit{and} \quad \sqrt{\int_{\mathbb{R}^m_* \times]0,t]} (|z|^2 \wedge |z|)\, \nu^p(\mathrm{d}z, \mathrm{d}t)},$$

are locally integrable. They also are integrable or (locally) square integrable if the semi-martingale X has the same property. Furthermore, X is quasi-left continuous if and only if its predictable support is an empty set, i.e., $\nu^p(\mathbb{R}^m_* \times \{t\}) = 0$, for any $t \ge 0$. □

Note if $\nu(\mathrm{d}z, \mathrm{d}t, \omega)$ is a quasi-left continuous integer random measure then its predictable compensator can be written as $\nu^p(\mathrm{d}z, \mathrm{d}t, \omega) = k(\mathrm{d}z, t, \omega)\mathrm{d}A(t, \omega)$, where k is a measurable (predictable) kernel and A is a continuous increasing process.

To check the point regarding the quasi-left continuity for a square integrable martingale X, let $\tau < \theta < \infty$ be given two stopping times. Since, for any compact subset K of $\mathbb{R}^d_*$ the quantity

$$\mathbb{E}\Big\{ \sum_{\tau<t\le\theta} \mathbb{1}_{\delta X(t)\in K} |\delta X(t)|^2 \Big\} = \mathbb{E}\Big\{ \int_{K\times]\tau,\theta]} |z|^2\, \nu(\mathrm{d}z, \mathrm{d}t) \Big\}$$

is a finite, the number of jumps is finite for each ω and ν can be replaced by ν^p in the above equality, we deduce

$$\varepsilon^2\, \mathbb{E}\{\nu(K\times]\tau, \theta]) \,|\, \mathcal{F}(\tau)\} \le \mathbb{E}\Big\{ \int_{K\times]\tau,\theta]} |z|^2\, \nu(\mathrm{d}z, \mathrm{d}t) \,|\, \mathcal{F}(\tau) \Big\} \le$$
$$\le \mathbb{E}\{|X(\theta)|^2 - |X(\tau)|^2 \,|\, \mathcal{F}(\tau)\},$$

where $\{|z| < \varepsilon\} \cap K = \emptyset$, $\varepsilon > 0$. Hence, $\nu(K \times [0, t])$ and $\nu^p(K \times [0, t])$ are quasi-left continuous if and only if X is quasi-left continuous.

Note that the previous theorem selects a particular representation (or realization) of the compensator of an integer-valued random measure suitable for the stochastic integration theory. Thus, we always refer to the compensator satisfying the properties in Theorem 4.7. Moreover, given an integer-valued random measure ν the process $\nu_{qc}(]0, t \wedge \tau] \times K)$ given by the expression

$$\nu_{qc}(K\times]0, t\wedge\tau]) = \nu(K\times]0, t\wedge\tau]) - \sum_{0<s\le t\wedge\tau} \nu^p(K \times \{s\}),$$

is quasi-left continuous, and its compensator is the continuous part of the compensator ν^p, denoted by ν^p_c. Hence, for any stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t \mapsto \tilde{\nu}_{qc}(K\times]0, t \wedge \tau])$, with $\tilde{\nu}_{qc} := \nu_{qc} - \nu^p_c$ is a local (purely discontinuous) martingale, whose predictable quadratic variation process obtained via Doob-Meyer decomposition is actually the process $\nu^p_c(K\times]0, t \wedge \tau])$, i.e.,

$$\langle \tilde{\nu}_{qc}(K \times [0, \cdot \wedge \tau])\rangle(t) = \nu^p_c(K\times]0, t \wedge \tau]), \quad \forall\, t \ge 0.$$

Thus, the optional locally integrable random measure $\tilde{\nu} := \nu - \nu^p = \tilde{\nu}_{qc}$ is called the (local) *martingale random measure* associated with ν or with the cad-lag process X.

• *Remark* 4.8. It is clear that two (or more) random measures ν_1 and ν_2 are called independent if for any stopping time $\tau < \infty$ and any compact subset K of $\mathbb{R}^m_*$ the stochastic process $t \mapsto \nu_1(K \times [0, t \wedge \tau])$ and $t \mapsto \nu_2(K \times [0, t \wedge \tau])$ are independent. Therefore, if ν_1 and ν_2 are defined as the number of jumps (4.7) of two (or more) (special) semi-martingale X_1 and X_2 then the (purely) jumps processes δX_1 and δX_2 are independent if and only the random measures ν_1 and ν_2 (and therefore ν_1^p and ν_1^p) are so. However, the random measure associated via (4.7) with the jumps $(\delta X_1, \delta X_2)$ considered in $\mathbb{R}^{2m}_*$, and the almost *product* measure ν in $\mathbb{R}^{2m}_*$ defined by $\nu(K_1 \times K_2 \times]0, t]) = \nu_1(K_1 \times]0, t]) \, \nu_1(K_2 \times]0, t])$, for every K_1, K_2 compact subset of $\mathbb{R}^m_*$ and $t > 0$ may not agree. Certainly, they are the same if the process X_1 and X_2 do not jump simultaneously. In particular, if X_1 and X_2 are Poisson processes with respect to the same filtration then they are independent if and only if they never jump simultaneously. □

A fundamental example of jump process is the simple point process $(N(t) : t \geq 0)$ which is defined as a increasing adapted cad-lag process with nonnegative integer values and jumps equal to 1, i.e., $\delta N(t) = 0$ or $\delta N(t) = 1$ for every $t \geq 0$, and $N(t)$ represents the number of *events* occurring in the interval $(0, t]$ (and so more then one event cannot occur exactly a the same time). Given $(N(t) : t \geq 0)$ we can define a sequence $\{T_n : n \geq 0\}$ of stopping times $T_n := \{t \geq 0 : N(t) = n\}$. Notice that $T_0 = 0$, $T_n < T_{n+1}$ on the set $\{T_{n+1} < \infty\}$, and $T_n \to \infty$. Since

$$N(t) = \sum_{n=0}^{\infty} \mathbb{1}_{T_n \leq t}, \quad \forall t \geq 0,$$

the sequence of stopping times completely characterizes the process, and because $N(T_n) \leq n$, any point process is locally bounded. An *extended Poisson process* N is an adapted point process on the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ satisfying:

(1) $\mathbb{E}\{N(t)\} < \infty$, for every $t \geq 0$,

(2) $N(t) - N(s)$ is independent of $\mathcal{F}(s)$, for every $t \geq 0$,

The function $a(t) := \mathbb{E}\{N(t)\}$ is called intensity (of N). It can be proved that if the function a is continuous then N is a *Poisson process* and if $a(t) = t$ for every $t \geq 0$ then N is a *standard Poisson process*. In this example, the compensator can be calculated, it can be proved (e.g., Jacod and Shiryaev [117, Proposition I.3.27, pp. 34–35]) that the compensator of an extended Poisson process is equal to its intensity, i.e., $N^p(t) = \mathbb{E}\{N(t)\}$ and that N is quasi-left continuous if and only if it is a Poisson process (i.e., its intensity is continuous). In general, even though the jumps are always countable they can not be ordered as in a point process. This yields the notion of integer-valued random measures.

Our main interest is on integer-valued random measure ν_X associated with a quasi-left continuous semi-martingale X, so that $t \mapsto \nu_X^p(K \times]0, t \wedge \tau])$ is continuous and for $\tilde{\nu}_X := \nu_X - \nu_X^p$ we have the following expressions for the

optional and predictable quadratic variation processes

$$[\tilde\nu_X(K\times]0,\cdot\wedge\tau])](t) = \langle\tilde\nu_X(K\times]0,\cdot\wedge\tau])\rangle(t) = \nu^p_X(K\times]0,t\wedge\tau]), \quad (4.10)$$

for any $t>0$, any stopping time $\tau<\infty$ and any compact subset K of R^m_*. Ignoring the local character of the semi-martingale X, this yields the compensated jumps equality

$$\mathbb{E}\Big\{\Big|\int_{K\times]0,t\wedge\tau]}\varphi(z,s)\,\tilde\nu_X(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} = \\ = \mathbb{E}\Big\{\int_{K\times]0,t\wedge\tau]}|\varphi(z,s)|^2\,\nu^p_X(\mathrm{d}z,\mathrm{d}s)\Big\}$$

and estimate

$$\mathbb{E}\Big\{\sup_{0\le t\le T}\Big|\int_{K\times]0,t\wedge\tau]}\varphi(z,s)\,\tilde\nu_X(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} \le \\ \le 4\,\mathbb{E}\Big\{\int_{K\times]0,T\wedge\tau]}|\varphi(z,s)|^2\,\nu^p_X(\mathrm{d}z,\mathrm{d}s)\Big\},$$

for any Borel measurable function $\varphi(z,s)$ such that the right-hand side is finite. Thus, we can define the integral of φ with respect to $\tilde\nu_X$

$$\tilde\nu_X(\varphi\mathbb{1}_{]0,t\wedge\tau]}) := \lim_{\varepsilon\to0}\int_{\{|x|\ge\varepsilon\}\times]0,t\wedge\tau]}\varphi(z,s)[\nu_X(\mathrm{d}z,\mathrm{d}s)-\nu^p_X(\mathrm{d}z,\mathrm{d}s)], \quad (4.11)$$

where φ vanishes for $|z|$ large and for $|z|$ small. All this is developed with the stochastic integral, valid for any predictable process instead of $\varphi\mathbb{1}_{]0,t\wedge\tau]}$. The point here is that the integral

$$\int_{\{|x|<1\}\times]0,t\wedge\tau]}z\,\tilde\nu_X(\mathrm{d}z,\mathrm{d}s)$$

is meaningful as a limit in L^2 for every φ square integrable with respect to ν^p_X, and the compensated jumps estimate holds.

In this way, the stochastic process X and the filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\ge0)$ determine the predictable compensator ν^p_X. Starting from a given integer-valued random measure ν and by means of the previous Theorem 4.7, we can define its *compensated martingale random measure* $\tilde\nu := \nu-\nu^p$, where ν^p is the compensator. The Doléans measure on $\mathbb{R}^m_*\times[0,\infty)\times\Omega$ relative to the integer measure ν is defined as the *product measure* $\mu := \nu(\mathrm{d}z,\mathrm{d}s,\omega)\,P(\mathrm{d}\omega)$, i.e., associated with the jumps process Z_K induced by ν, namely, for every compact subset K of $\mathbb{R}^m_*$

$$Z_K(t,\omega) := \int_{K\times]0,t]}z\,\nu(\mathrm{d}z,\mathrm{d}s), \quad \forall t\ge0.$$

Therefore whenever ν integrate the function $z\mapsto|z|$ we can consider the process $Z_{\mathbb{R}^m_*}$ as in (4.9). Conversely, if a given (m-valued) Doléans measure μ vanishes on any evanescent set, i.e., $\mu(K\times N)=0$ for every compact K of $\mathbb{R}^m_*$ and for

any subset N of $[0,\infty)\times\Omega$ such that $P(\cup_t\{\omega : (t,\omega)\in N)\}) = 0$, then there is an optional ($\mathbb{R}^m$-valued) jump process A with integrable bounded variation associated with μ. This argument can be localized as long as we assume $\nu^p(]0,t\wedge\tau)\times K) < \infty$, for any compact K in $\mathbb{R}^m$ (not only in $\mathbb{R}^m_*$) to get a jump process A with locally integrable bounded variation path associated to ν. Now, for this jump process A we can defined an integer-valued measure ν with the same initial predictable compensator ν^p.

The following *canonical representation* of (special) semi-martingale holds. Let ν_X be the (random) integer measure associated with the semi-martingale X, namely, $\nu_X(B\times]a,b])$ is the number of jumps on the time interval $(a,b]$ of the process X with a value δX belonging to the set B, i.e. for every $b > a \geq 0$ and B in $\mathcal{B}(\mathbb{R}^d_*)$,

$$\nu_X(B\times]a,b]) := \#\{t : a < t \leq b,\ X(t) - X(t-) \in B\},$$

and let ν^p_X be its (dual predictable) compensator (satisfying the properties given in Theorem 4.7), so that $\tilde\nu_X := \nu_X - \nu^p_X$ is a local martingale measure, then

$$X(t) = X(0) + A(t) + X^c(t) + \int_{\mathbb{R}^d_*\times]0,t]} z\tilde\nu_X(\mathrm{d}z,\mathrm{d}s), \quad \forall t \geq 0, \tag{4.12}$$

where A is a predictable process with locally integrable variation and X^c is a continuous local martingale, both satisfying $A(0) = X^c(0) = 0$ and X^c is uniquely determined. Clearly, the integer measure ν depends only on the jump process δX, i.e., only the discontinuous part of X determines ν_X. If the semi-martingale X is quasi-left continuous (i.e., either $\mathcal{F}(\tau-) = \mathcal{F}(\tau)$ for every predictable stopping time τ or equivalently the predictable compensator ν^p_X satisfies $\nu^p_X(\mathbb{R}^d_*\times\{t\}) = 0$ almost surely), then the process A in (4.12) is continuous and uniquely determined.

Note the characteristic elements of a semi-martingale X, which are (1) the predictable process A with locally integrable variation (which is uniquely determined only when the semi-martingale is quasi-left continuous), (2) the predictable quadratic variation $\langle X^c\rangle$ and (3) the (dual predictable) compensator measure ν^p_X. If $X = M$ is a quasi-left continuous local martingale then $A = 0$ and there are only two characteristic elements to consider: (a) the predictable quadratic variation $\langle M^c\rangle$ (or the optional quadratic variation $[M]$) and (b) the predictable compensator ν^p (or the integer-valued measure ν). If the *special* character of the semi-martingale is removed, then the jumps may be not locally integrable and then the predictable compensator ν^p may be not integrable at infinity, i.e., only the function $z\mapsto |z|^2\wedge 1$ in ν^p-integrable, so that the predictable process

$$t \mapsto \int_{\mathbb{R}^d_*}\int_{]0,t]} (|z|^2\wedge 1)\,\nu^p(\mathrm{d}z,\mathrm{d}s)$$

is locally integrable. Thus the representation (4.12) becomes

$$X(t) = X(0) + A(t) + X^c(t) + \int_{\{|z|<1\}\times]0,t]} z\,\tilde\nu_X(\mathrm{d}z,\mathrm{d}s), \quad \forall t \geq 0, \tag{4.13}$$

where A contains a term of the form

$$\int_{|z|\geq 1}\int_{]0,t]} z\,\nu_X(\mathrm{d}z,\mathrm{d}t),$$

and $h(z) := z\mathbb{1}_{|z|<1}$ is used as the truncation function. However, our main interest is on processes with finite moments of all order, so that ν^p should integrate $z \mapsto |z|^n$ for all $n \geq 2$. The reader may consult He et al. [105, Section XI.2, pp. 305–311], after the stochastic integral is covered.

A fundamental example is the Poisson measures. We have

Definition 4.9 (Poisson-measure)**.** Let $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ be a filtered space. An integer-valued random measure ν on $\mathbb{R}^m_* \times [0,\infty)$ is called *Poisson measure* if

(a) the (nonnegative) measure $\Pi(B) := \mathbb{E}\{\nu(B)\}$ is a Radon measure on $\mathbb{R}^m_* \times [0,\infty)$, i.e., $\mathbb{E}\{\nu(K \times [0,t])\} < \infty$ for any compact subset K of $\mathbb{R}^m_*$ and for any $t \geq 0$,

(b) for any Borel measurable subset B of $\mathbb{R}^m_* \times (t,\infty)$ with $\Pi(B) < \infty$ the random variable $\nu(B)$ is independent of the σ-algebra $\mathcal{F}(t)$,

(c) Π satisfies $\Pi(\mathbb{R}^m_* \times \{t\}) = 0$ for every $t \geq 0$.

The measure Π is called *intensity measure* relative to the Poisson measure ν. If Π has the form $\Pi(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z) \times \mathrm{d}t$ for a (nonnegative) Radon measure π on $\mathbb{R}^m_*$ then ν is called a *homogeneous (or standard) Poisson measure*. If the condition (c) is not satisfied then ν is called *extended Poisson measure*. □

A standard Poisson measure ν on a Polish space $\mathcal{O} \times [0,\infty)$ (e.g., $\mathbb{R}^m_* \times [0,\infty)$ or even a non-locally compact separable metric space) relative to a σ-finite measure $\pi \times \mathrm{d}t$ on $\mathcal{B}\big(\mathcal{O} \times [0,\infty)\big)$ (called intensity) can be also defined as a random measure satisfying (a) for any Borel subset B of $\mathcal{O}$ with $\pi(B) < \infty$ and $t \geq 0$ the random variable $\nu(B \times]0,t]) = \nu(B,t)$ has a Poisson distribution with parameter $t\pi(B)$ and (b) for any $n \geq 1$ and any disjoint Borel sets $B_1, B_2, \ldots, B_n$ and $0 \leq t_0 < t_1 < \cdots t_n$ the random variables $\nu(B_1,t_1) - \nu(B_1,t_0), \nu(B_2,t_2) - \nu(B_2,t_1), \ldots, \nu(B_n,t_n) - \nu(B_n,t_{n-1})$ are independent. Given a σ-finite measure π on $\mathcal{B}(\mathcal{O})$, a standard Poisson measure ν can be constructed as follows. First, if π is a finite measure then we can find a sequence $\{\tau_1, \tau_2, \ldots\}$ of independent exponentially distributed random variables, with parameter $c = \pi(\mathcal{O})$ and a sequence $(\xi_1, \xi_2, \ldots)$ of $\mathcal{O}$-valued independent identically distributed random variables, with distribution $\pi/\pi(\mathcal{O})$ and independent of $\{\tau_1, \tau_2, \ldots\}$, in some (complete) probability space $(\Omega, \mathcal{F}, P)$. Thus the random integer measure on $\mathcal{O}$ defined by

$$\nu(B,t) := \sum_{i=1}^{\infty} \mathbb{1}_{\xi_i \in B}\mathbb{1}_{\tau_i \leq t}, \quad \forall B \in \mathcal{B}(\mathcal{O}),\ \forall t > 0$$

is the desired standard Poisson measure satisfying $\mathbb{E}\{\nu(B,t)\} = t\pi(B)$. Next, if ν is merely σ-finite then there exists a Borel partition of the whole space, $\mathcal{O} = \bigcup_n \mathcal{O}_n$, with $\pi(\mathcal{O}_n) < \infty$ and $\mathcal{O}_n \cap \mathcal{O}_k = \emptyset$ for $n \neq k$. For each $\mathcal{O}_n$ we

can find a Poisson measure ν_n as above, and make the sequence of integer-valued random measure $(\nu_1, \nu_2, \ldots)$ independent. Hence $\nu := \sum_n \nu_n$ provides a standard Poisson measure with intensity π. Remark that ν_n is a finite standard Poisson measure on $\mathcal{O}_n \times [0,\infty)$ considered on the whole $\mathcal{O} \times [0,\infty)$ with intensity π_n, $\pi_n(B) = \pi(B \cap \mathcal{O}_n)$.

Moreover, if $\mathcal{O} = \mathbb{R}^d_*$ then the jump random process corresponds to the measure π restricted to $\mathcal{O}_n$

$$\mathtt{j}_n(t,\omega) := \sum_{i=1}^{\infty} \xi_i^n \mathbb{1}_{\tau_i^n \le t}, \quad \forall t > 0$$

is properly defined, and if π integrates the function $z \mapsto |z|$ the jumps $\mathtt{j} = \sum_n \mathtt{j}_n$ (associated with ν_n) are defined almost surely. However, if π integrates only the function $z \mapsto |z|^2 \wedge 1$ then the stochastic integral is used to define the compensated jumps, formally $\mathtt{j} - \mathbb{E}\{\mathtt{j}\}$.

The same arguments apply to Poisson measures, if we start with an intensity measure defined on $\mathcal{O} \times [0,\infty)$. In this case, the (compensated) jumps is defined as a stochastic process, by integrating on $\mathcal{O} \times [0,t]$.

If the variable t is not explicitly differentiated, the construction of a Poisson (random is implicitly understood) measures ν on a Polish space Z, relative to a σ-finite measure Π can be simplified as follows: First, if Π is a finite measure then we can find a Poisson random variable η with parameter $c = \Pi(Z)$ and a sequence $(\zeta_1, \zeta_2, \ldots)$ of Z-valued independent identically distributed random variables, with distribution Π/c and independent of η in some (complete) probability space $(\Omega, \mathcal{F}, P)$. Then $\nu(B) = \sum_{k=1}^{\eta} \mathbb{1}_{\zeta_k \in B}$, for any B in $\mathcal{B}(Z)$, satisfies

$$\mathbb{E}\{\nu(B)\} = \sum_n \mathbb{E}\Big\{ \sum_{k=1}^{n} \mathbb{1}_{\zeta_k \in B} \mid \eta = n \Big\} = \sum_n \frac{n\Pi(B)}{c} P(\eta = n) = \Pi(B).$$

In particular, if $Z = \mathcal{O} \times [0,\infty)$ and $\Pi = \pi \times \mathrm{d}t$ then $\mathbb{E}\{\nu(B \times]0,t])\} = t\pi(B)$, for every B in $\mathcal{B}(\mathcal{O})$ and $t \ge 0$.

Thus, if Π is only σ-finite then partition the space $Z = \sum_n Z_n$ into sets with finite measure $\Pi(Z_n) < \infty$, and redo the construction with independent sequences $\{\eta^n\}$ and $\{\zeta_i^n\}$ to define $\nu(B) = \sum_n \sum_k \mathbb{1}_{k \le \eta^n} \mathbb{1}_{\zeta_k^n \in B}$.

As in Çınlar [46, Theorems 3.2 and 3.19, Chapter 6, pp. 264-270], we can now consider

Proposition 4.10. *Let $Z = \sum_n Z_n$ and $X = \sum_n X_n$ be partitions of the Polish spaces Z and X, and let $\mathtt{m}_n(z, \mathrm{d}y)$ be a transition kernel from Z_n into X_n, i.e., (a) for every B in $\mathcal{B}(X)$ the mapping $z \mapsto \mathtt{m}_n(z,B)$ is $\mathcal{B}(Z_n)$-measurable and (b) for every z in Z_n the set function $B \mapsto \mathtt{m}_n(z,B)$ is a probability on X_n. Suppose that $\{\xi_1^n, \xi_2^n, \ldots\}$ are X_n-valued random variables conditionally independent given $\{\eta^n, \zeta_i^n : i \ge 1\}$, for each $n \ge 1$, such that ξ_i^n has distribution $\mathtt{m}(\zeta_i^n, \cdot)$. Then $\mu(B) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{\xi_k^n \in B}$, for any B in $\mathcal{B}(X)$, is a Poisson measure with (marginal) intensity $\sum_n \int_{Z_n} \mathtt{m}_n(z, \cdot)\Pi(\mathrm{d}z)$ in X, and $\lambda(B) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{(\zeta_k^n, \xi_k^n) \in B}$, for any B in $\mathcal{B}(Z \times X)$, is a Poisson measure with (product) intensity $\sum_n \Pi_n \times \mathtt{m}_n = \sum_n \mathtt{m}_n(z, \mathrm{d}x)\mathbb{1}_{Z_n}\Pi(\mathrm{d}z)$ in $Z \times X$.*

Proof. Since the random variable $\{(\zeta_i^n, \xi_i^n) : i \geq 1\}$ is a sequence of independent identically distributed random variables with (product) distribution

$$P\big\{(\zeta_i^n, \xi_i^n) \in B\big\} = \int_{Z_n} \Pi(\mathrm{d}z) \int_{X_n} \mathbb{1}_B \mathtt{m}_n(z, \mathrm{d}x),$$

based on the above construction, we deduce that λ is a Poisson measure with (product) intensity $\sum_n \Pi_n \times \mathtt{m}_n$. Moreover, conditioning with respect to $\mathcal{B}(Z)$, we obtain the first assertion. Note that the marginal distribution is indeed

$$B \mapsto \mathtt{m}_n(\cdot, B)\Pi_n = \int_{Z_n} \mathtt{m}_n(z, B \cap X_n)\Pi(\mathrm{d}z),$$

for every B in $\mathcal{B}(X)$. □

If φ is a random transformation from Z into X, i.e., $(\omega, z) \mapsto \varphi(\omega, z)$ is a $\mathcal{F} \times \mathcal{B}(Z)$-measurable function from $\Omega \times Z$ into X. Then the marginal distributions

$$\mathtt{m}(z, B) = P\{\varphi(\omega, z) \in B\}, \quad \forall z \in Z, \quad \forall B \in \mathcal{B}(X)$$

defined a transition kernel as in Proposition 4.10. If ν is a Poisson measure with intensity Π on Z then

$$\mu(B) = \int_Z \mathbb{1}_{\{\varphi(\cdot,z) \in B\}} \nu(\mathrm{d}z) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{\varphi(\zeta_k^n, \cdot) \in B}, \quad \forall B \in \mathcal{B}(X)$$

and

$$\lambda(B) = \int_Z \mathbb{1}_{\{(z,\varphi(\cdot,z)) \in B\}} \nu(\mathrm{d}z) = \sum_n \sum_{k=1}^{\eta_n} \mathbb{1}_{(\zeta_k^n, \varphi(\xi_k^n, \cdot)) \in B},$$

for every $B \in \mathcal{B}(Z \times X)$, are Poisson measures with intensities $\int_Z \mathtt{m}(z, \cdot)\Pi(\mathrm{d}z)$ on X and $\Pi \times \mathtt{m} = \mathtt{m}(z, \mathrm{d}x)\Pi(\mathrm{d}z)$ on $Z \times X$.

It is clear that $Z = \mathbb{R}_*^m$ and $X = \mathbb{R}_*^d \times [0, \infty)$ are special cases. The (nonnegative) intensity measure can be written as sum of its continuous and discontinuous parts, i.e.,

$$\Pi = \Pi^c + \Pi^d, \quad \Pi^d(\mathrm{d}z, \mathrm{d}t) := \mathbb{1}_{\{t\,:\,\Pi(\mathbb{R}_*^m \times \{t\}) > 0\}} \Pi(\mathrm{d}z, \mathrm{d}t).$$

There is a characterization of Poisson measures as follows

Theorem 4.11. *An integer-valued random measure ν is a Poisson measure if an only if its compensator ν^p is deterministic and continuous, i.e., $\nu^p = \Pi$ and $\Pi(\mathbb{R}_*^m \times \{t\}) = 0$ for every $t \geq 0$. Moreover, for any Poisson measure ν and any pairwise disjoint measurable sets $(B_1, B_2, \ldots, B_n)$ with finite Π-measure, the set $\{\nu(B_1), \nu(B_2), \ldots, \nu(B_1)\}$ is a family of independent random variables and $\nu(B_i)$ has a Poisson distribution with mean $\Pi(B_i)$, for any i.* □

In view of the above characterization, $\nu^p = \Pi$ for a Poisson measure and because of the previous Theorem 4.7 we deduce that ν should integrate the function $|z|^2 \wedge 1$ when the jumps process A associated with the Poisson measure ν is a general semi-martingale. For an (special) semi-martingale the intensity Π should integrate $|z|^2 \wedge |z|$. Thus, we are only interested in Lévy measures ν which necessarily integrate the function $|z|^2 \wedge 1$.

It is clear that homogeneous (or standard) Poisson measures are associated with the jumps of Lévy processes via (4.7), and with Remark 4.8 in mind, the integer measures ν_i associated with each component of X_i in $\mathbb{R}_*$ may not reconstruct the measure ν associated with the X in $\mathbb{R}_*^m$, even if each component is independent of the others.

For a proof (including extended Poisson measure) we refer to Jacod and Shiryaev [117, Theorem II.4.8, pp. 104–106]. The reader may consult, for instance, Bremaud [32], where jump processes are discussed as point processes in the framework of the queue theory.

4.2 Stochastic Integrals

Let us fix a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$. A *simple predictable process* (or piecewise constant over stochastic intervals) is a stochastic process of the form $Y(t) = Y(\tau_{i-1})$ if $\tau_{i-1} < t \leq \tau_i$ with some $i = 1, \ldots, n$, where $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_n$ are stopping times and $Y(\tau_{i-1})$ is a $\mathcal{F}(\tau_{i-1})$ measurable random variable for any i, and $Y(t) = 0$ otherwise. It is called *bounded* if all $Y(\tau_{i-1})$ are bounded random variables. Note that any simple predictable process Y is left continuous with right-hand limits, so that $t \mapsto Y(t+)$ is a cad-lag process.

If X is an optional cad-lag process then we define the expression

$$Z(t) = \int_{(0,t]} Y(s)\mathrm{d}X(s) = \sum_{i=1}^{n} Y(\tau_{i-1})[X(t \wedge \tau_i) - X(t \wedge \tau_{i-1})], \tag{4.14}$$

as the integral of the simple predictable process (integrand) Y with respect to the optional cad-lag process (integrator) X. This integral process Z is cad-lag and optional, which is also continuous if X is so. On the other hand, the integration-by-part formula

$$\left\{ \begin{array}{l} X(b)Y(b) - X(a)Y(a) = \displaystyle\int_{(a,b]} X(t-)\mathrm{d}Y(t) + \int_{(a,b]} Y(t-)\mathrm{d}X(t) + \\ \qquad\qquad + \displaystyle\sum_{a<t\leq b} [X(t) - X(t-)]\,[Y(t+) - Y(t)], \end{array} \right. \tag{4.15}$$

yields the expression

$$\left\{ \begin{array}{l} Z(t) = \displaystyle\int_{(0,t]} Y(s)\mathrm{d}X(s) = \\ \qquad = X(t)Y(t) - \displaystyle\sum_{i=1}^{n} X(\tau_i)[Y(t \wedge \tau_i) - Y(t \wedge \tau_{i-1})], \end{array} \right. \tag{4.16}$$

which can be used to define the same integral process.

If $t \mapsto X(t,\omega)$ has also locally bounded variation for almost every ω then the measure theory can be used on (4.14) to extend the definition of the integral to a class of predictable processes, including all continuous adapted processes. On the other hand, we can use (4.16) to extend the definition of the integral to a class of predictable processes, including all continuous adapted processes with locally bounded variation. In either case, with this *pathwise analysis*, we are unable to see how two continuous processes of unbounded variation can be integrated, which is the case of a Wiener process as integrand and as integrator. In contrast with what follows, the fact that we use adapted processes is irrelevant in pathwise analysis. The reader may want to consult the classic reference McKean [167] for a comprehensive treatment.

4.2.1 Gaussian and Poisson Noises

The idea of a *noise* is the extension of a sequence of independent identically distributed random variables to the *continuous* context, where the two typical cases are Gaussian and Poisson noises. First, let us recall that we can build a (complete) probability space $(\Omega, \mathcal{F}, P)$, e.g., P is the Lebesgue measure on $(0,1)$, with the property that for any countable family of distributions $\{F_i\}$ on $\mathbb{R}^d$ there exists a family of independent random variables $\{\xi_i\}$ such that ξ_i is distributed accordingly to F_i, e.g., see Kallenberg [121, Theorem 3.19, pp. 55–57]. In particular, there exist two independent countable families of normally and exponentially distributed random variables, with parameters prescribed a priori, in some probability space $(\Omega, \mathcal{F}, P)$.

For instance, a simple Poisson noise with parameter $\lambda > 0$ can be regarded as a sequence $\dot{p} = \{\tau_n : n \geq 1\}$ of independent exponentially (with parameter $1/\lambda$) distributed random variables. Since $P(\lim_n \sum_{i=1}^n \tau_i = \infty) = 1$, the counting process $p_t = \sum_{n=1}^\infty \mathbb{1}_{\tau_1+\cdots+\tau_n \leq t}$, i.e.

$$p_t = \begin{cases} 0 & \text{if } t < \tau_1, \\ n & \text{if } \sum_{i=1}^n \tau_i \leq t < \sum_{i=1}^{n+1} \tau_i, \end{cases}$$

is defined almost surely and called a Poisson process, i.e., $p_0 = 0$, $p_t - p_s$ is Poisson distributed with mean $\lambda(t-s)$ and independent of p_s, for any $t > s \geq 0$. The paths are piecewise constant with jumps equal to 1. Moreover, if δ_n denotes the Dirac measure concentrated at n then

$$P\{p_t \in \mathrm{d}x\} = \mathrm{e}^{-\lambda t} \sum_{n=0}^\infty \delta_n(\mathrm{d}x) \frac{(\lambda t)^n}{n!} \quad \text{and} \quad \mathbb{E}\{\mathrm{e}^{\mathrm{i}\xi p_t}\} = \exp\left[t\lambda(\mathrm{e}^{\mathrm{i}\xi} - 1)\right]$$

are the transition density and the characteristic function. It is also clear that for $q_t = p_t - t\lambda$,

$$P\{q_t \in \mathrm{d}x\} = \mathrm{e}^{-\lambda t}\Big[\delta_0(\mathrm{d}x) + \sum_{n=1}^\infty \big(\delta_n(\mathrm{d}x) - 1\big) \frac{(\lambda t)^n}{n!}\Big]$$

is the transition function.

As seen below, the situation is much more complicate in the case of a white noise $\dot{w} = \{\xi_n : n \geq 1\}$, where the independent random variables ξ_n are standard normally distributed.

The White Noise

The simplest construction of a Wiener process with the L^2 orthogonal theory begins with an orthogonal basis $\{\varphi_n : n \geq 0\}$ in $L^2(]0,T[)$, and a sequence $\{\xi_n : n \geq 0\}$ of independent standard normally distributed random variables, which forms also an orthonormal system in $L^2 = L^2(\Omega, \mathcal{F}, P)$. Each function φ in $L^2(]0,T[)$ can be written as a converging orthogonal series

$$\varphi(s) = \sum_n (\varphi, \varphi_n)\varphi_n(s), \qquad \text{a.e. } s \in]0,T[,$$

where $(\cdot,\cdot)$ denotes the scalar product in $L^2(]0,T[)$, and $(\varphi,\varphi) = \sum_n |(\varphi,\varphi_n)|^2$. Thus the mapping $\varphi \mapsto w(\varphi) = \sum_n (\varphi,\varphi_n)\xi_n$ is an isometry from $L^2(]0,T[)$ into L^2 such that $w(\varphi)$ is a Gaussian random variable with $\mathbb{E}\{w(\varphi)\} = 0$ and $\mathbb{E}\{w(\varphi)w(\varphi')\} = (\varphi,\varphi')$, for every φ and φ' in $L^2(]0,T[)$. Hence, $\mathbb{1}_{]a,b]} \mapsto w(\mathbb{1}]a,b])$ could be regarded as a L^2-valued measure and $w(\varphi)$ is the integral. In particular, the orthogonal series

$$w(t) = w(\mathbb{1}_{]0,t[}) = \sum_n \xi_n \int_0^t \varphi_n(s)\mathrm{d}t, \qquad \forall t \geq 0$$

is converging in L^2, and

$$\mathbb{E}\{|w(t)|^2\} = \sum_n \Big| \int_0^t \varphi_n(s)\mathrm{d}s \Big|^2 = \int_0^T \mathbb{1}_{]0,t[}(s)\mathrm{d}s = t, \qquad \forall t \geq 0,$$

i.e., the above series yields a Gaussian process $t \mapsto w(t)$, which is continuous in L^2 and satisfies $\mathbb{E}\{w(t)\} = 0$ and $\mathbb{E}\{w(t)w(s)\} = t \wedge s$, for every t, s in $[0,T]$. Conversely, if a Wiener process $\{w(t) : t \geq 0\}$ is given then we can reconstruct the sequence $\{\xi_n : n \geq 0\}$ by means of the square-wave orthogonal basis $\{\varphi_n : n \geq 0\}$, where the integral $w(\varphi_n)$ reduces to a finite sum, namely,

$$\xi_n = w(\varphi_n) = \sum_{i=1}^{2^n} (-1)^{i-1} T^{-1/2} \big[w(t_{i,n}) - w(t_{i-1,n}) \big],$$

with $t_{i,n} = i2^{-n}T$, $i = 0, \ldots, 2^n$, $n \geq 0$. Finally, the almost surely continuity of the path requires either taking a particular version or using some martingale inequality. It is clear that in this construction there is not a precise to handle which random variables are involves in $w(s)$ when s belongs to $[0,t]$. However, a small change along this previous argument makes the trick, as we see in what follows.

The closed linear subspace H of $L^2(\Omega, \mathcal{F}, P)$ generated by the orthonormal sequence $\{\xi_n : n \geq 0\}$ is called a white noise (or Gaussian) space. If $\mathcal{F}_\xi$ is the

σ-algebra generated by the random variables $\{\xi_n : n \geq 0\}$ and the null sets in $\mathcal{F}$, then any random variable x in $L^2(\Omega, \mathcal{F}_\xi, P)$ which is independent of H is actually a constant, i.e., $x = \mathbb{E}\{x\}$. It is also clear that the Hilbert space H can be identified with the $L^2(]0,T[)$ via the above isometry. As discussed in the previous Section 4.1.1, based on the Hermit polynomials $h_n(x)$ we can construct an orthonormal basis

$$\Xi_n = \prod_i h_{n_i}(\xi_i)\sqrt{n_i!}, \qquad n = (n_i), \text{ only a finite number nonzero,} \tag{4.17}$$

for the space $L^2(\Omega, \mathcal{F}_\xi, P)$, which can be written as an infinite orthogonal sum of subspaces. It is clear that $\mathcal{F}_\xi$ is equal to $\mathcal{F}_w$, the σ-algebra generated by the random variables $\{w_t : t > 0\}$, it seems not obvious how to use the above construction to get an orthonormal basis corresponding to the σ-algebra $\mathcal{F}(t)$ generated by the random variables $\{w_s : 0 < s \leq t\}$.

Sometimes, another Hilbert space H is preferred instead of $L^2(]0,T[)$, i.e., we may begin with an orthogonal basis $\{e_n : n \geq 1\}$ in H and a sequence $\{\xi_n : n \geq 1\}$ of independent standard normally distributed random variables (after some adaptation, perhaps, with values in $\mathbb{R}^d$ or in some infinite dimension Banach space), which forms also an orthonormal system in $L^2 = L^2(\Omega, \mathcal{F}, P)$. Each function h in H can be written as a converging orthogonal series $h = \sum_n (h, e_n)_H e_n$, and $(h,h)_H = \sum_n |(h, e_n)_H|^2$. Thus the mapping $h \mapsto w(h) = \sum_n (h, e_n)_H \xi_n$ is an isometry from H into L^2 such that $w(h)$ is a Gaussian random variable with $\mathbb{E}\{w(h)\} = 0$ and $\mathbb{E}\{w(h)w(h')\} = (h, h')_H$, for every h and h' in H. Of particular interest in the case where $H = L^2(X, \mathcal{X}, \mu)$ for a σ-finite measure space $(X, \mathcal{X}, \mu)$. Choose a semi-ring $\mathcal{K}$ of $\mathcal{X}$ with finite measure, i.e., $\mu(K) < \infty$, for every K in $\mathcal{K}$ to consider the map $\mathbb{1}_K \mapsto w(\mathbb{1}_K)$ as a L^2-valued measure and $w(h)$ is the integral. Moreover, we may single-out a time variable, i.e., replace X and μ with $X \times]0,T[$ and $\mu \times \mathrm{d}t$, and get an orthonormal system of the form $e_i\varphi_n$. Hence, by relabeling the sequence $\{\xi_{i,n} : i, n \geq 1\}$ the orthogonal series

$$w_i(t) = w(e_i \mathbb{1}_{]0,t[}) = \sum_n \xi_{i,n} \int_0^t \varphi_n(s)\mathrm{d}t, \qquad \forall t \geq 0,\ i = 1, 2, \ldots,$$

is converging in L^2, and

$$\mathbb{E}\{|w_i(t)|^2\} = \sum_n \Big| \int_0^t \varphi_n(s)\mathrm{d}s \Big|^2 = \int_0^T \mathbb{1}_{]0,t[}(s)\mathrm{d}s = t, \qquad \forall t \geq 0,$$

i.e., the above series yields Gaussian processes $t \mapsto w_i(t)$, which are continuous in L^2 and satisfy $\mathbb{E}\{w_i(t)\} = 0$, $\mathbb{E}\{w_i(t)w_i(s)\} = t \wedge s$, for every t, s in $[0,T]$, and the process $(w_i(t) : t \geq 0)$ is independent of $(w_j(t) : t \geq 0)$ for every $i \neq j$. This construction is referred to as a general Wiener noise or white noise (random) measure.

The White Noise (details)

Formally, assume that the Hilbert space $L^2 = L^2(\Omega, \mathcal{F}, P)$ contains a sequence $\{e_{i,n} : i = 1, 2, \dots, 4^n,\ n \geq 1\}$ of independent standard normally distributed random variables, and set $\{e^n_r = e_{i,n} : r \in R_n\}$, indexed in r belonging to the diadic numbers $R = \bigcup_n R_n$, with $R_n = \{r = i2^{-n} : i = 1, 2, \dots, 4^n\}$. Because each $e_{i,n}$ has zero mean and are independent of each other, the sequence is orthogonal in L^2, actually, it is an orthonormal system since all variances are equal to 1. To simplify notation, assume that $\mathcal{F}$ is the sub σ-algebra generated by the sequence of random variables $\dot{w} = \{e^n_r : r \in R_n,\ n \geq 1\}$ (and all null sets). The closed linear subspace H of $L^2(\Omega, \mathcal{F}, P)$ generated by the elements in $\dot{w}$ is a which is called a white noise (or Gaussian) space. The system $\dot{w}$ is the ideal expression of the *white noise*, which is the formal derivative of the Wiener process w.

To given details of a simple construction a Wiener process $\{w_t : t > 0\}$ as the integral of (the function $s \mapsto \mathbb{1}_{s \leq t}$ with respect to) the system $\dot{w}$, we make use of the diadic property $t = \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t}$[1] to define the random variable

$$w_t = \sum_n 2^{-n} \sum_{i=1}^{4^n} e_{i,n} \mathbb{1}_{i2^{-n} \leq t}, \tag{4.18}$$

as a convergence series in $L^2(\Omega, \mathcal{F}, P)$, for every $t > 0$. Indeed, regard the expression as an orthogonal series expansion, and set $w_0 = 0$, for any $t \geq s \geq 0$, to have

$$\begin{aligned} \mathbb{E}\{|w_t - w_s|^2\} &= \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{E}\{|e_{i,n}|^2\} \mathbb{1}_{s < i2^{-n} \leq t} = \\ &= \sum_n 4^{-n} \sum_{i=1}^{4^n} \mathbb{1}_{s < i2^{-n} \leq t} = (t - s). \end{aligned}$$

Thus, $t \mapsto w_t$ provides a L^2-norm continuous random process satisfying (a) w_t is a Gaussian random variable with $\mathbb{E}\{w_t\} = 0$ and $\mathbb{E}\{|w_t|^2\} = t$, and (b) w_s is independent of $w_t - w_s$ for every $t > s$. If a parameter (variance) $a > 0$ is included then the Gaussian random variables $\{\xi_n : n \geq 0\}$ and $\{e^n_r : r \in R_n,\ n \geq 1\}$ have variance a and $\mathbb{E}\{|w_t|^2\} = ta$. Moreover

$$P\{w_t \in \mathrm{d}x\} = \mathrm{e}^{-|x|^2/(2ta)} \mathrm{d}x \quad \text{and} \quad \mathbb{E}\{\mathrm{e}^{\mathrm{i}\xi w_t}\} = \mathrm{e}^{-ta|\xi|^2/2}$$

are the transition density and the characteristic function.

Next, to check that the process $\{w_t : t \geq 0\}$ has a continuous version, we recall that $w_t - w_s$ is a Gaussian variable with zero mean and variance $|t - s|$, so that we deduce $\mathbb{E}\{|w_t - w_s|^4\} = 3|t - s|^2$, and therefore, we are allowed to select a continuous version.

[1] if $t = k2^{-m} = (k2^{n-m})2^{-n}$, $1 \leq k \leq 4^m$ then $k2^{n-m} \leq 4^n$, $\mathbb{1}_{i2^{-n} \leq t} = 1$ if and only if $i = 1, \dots, k2^{n-m}$, which yields $\sum_{i=1}^{4^n} \mathbb{1}_{i2^{-n} \leq t} = k2^{n-m} = t2^n$ if $k2^{n-m} = t2^n \geq 1$.

The concept of *stopping time* relative to a white noise $\dot{w}$ can be expressed as preserving orthogonality, i.e., a $[0,\infty]$-valued random variable τ is called $\dot{w}$-stopping time if $\{e_{i,n}\mathbb{1}_{i2^{-n}\le\tau}\}$ (or equivalently $\{e^n_r\mathbb{1}_{r\le\tau}\}$) remains an orthogonal system, for every $t>0$. For instance, if τ is a $\dot{w}$-stopping time then the formula (4.18) shows that $\mathbb{E}\{|w(t\wedge\tau)|^2\}=\mathbb{E}\{t\wedge\tau\}$ as expected.

Note that if x belongs to H then

$$\mathbb{E}\{xw_t\}=\sum_n 2^{-n}\sum_{i=1}^{4^n}\mathbb{E}\{xe_{i,n}\}\mathbb{1}_{i2^{-n}\le t},$$

and by taking $r=k2^{-m}$ with k some odd integer number between 1 and 4^m, we deduce $\mathbb{E}\{x(w_r-w_{r'})\}\to 2^{-m}\mathbb{E}\{xe_{k,m}\}$ as $r'\uparrow r$. This proves that any x in H which is orthogonal to any element in $\{w_t : t\ge 0\}$ is also orthogonal to any element in $\{e_{i,n} : i=1,\dots,4^n,\ n\ge 1\}$, i.e, the white noise subspace H is indeed the closed linear span of $\{w_t : t\ge 0\}$. Therefore the projection operator

$$\mathbb{E}\{x\mid w_s,\ s\le t\}=\sum_n\sum_{i=1}^{4^n}\mathbb{E}\{x\,e_{i,n}\}\,e_{i,n}\mathbb{1}_{i2^{-n}\le t},\tag{4.19}$$

is valid for every x in H. By means of the Hermit polynomials and $\{e_{i,n} : i2^{-n}=r\in R,\ r\le t\}$ we can construct an orthonormal basis for $L^2(\Omega,\mathcal{F}(t),P)$ as in (4.17), which yields an explicit expression for the conditional expectation with respect to $\mathcal{F}(t)$, for any square-integrable random variable x. In this context, remark that we have decomposed the Hilbert space H into an orthogonal series ($n\ge 1$) of finite dimensional subspaces generated by the orthonormal systems $\dot{w}^n=\{e^n_r : r\in R_n\}$.

The White Noise (converse)

Conversely, if a Wiener process $\{w_t : t\ge 0\}$ is given then the random variables $\bar{e}_{i,n}=2^{n/2}\big[w_{i2^{-n}}-w_{(i-1)2^{-n}}\big]$, are identically standard normally distributed, and the system $\{\bar{e}_{i,n} : i=1,\dots,4^n\}$ is independent, but $\{\bar{e}_{i,n} : i\ge 1,\ n\ge 1\}$ is not fully independent, i.e., $r=i2^{-n}=2i2^{-n-1}$ yields $\sqrt{2}\bar{e}_{i,n}=\bar{e}_{2i,n+1}+\bar{e}_{2i-1,n+1}$, which produces correlations. In this case, the representation (4.18) takes the form

$$w_t=\lim_n\Big[2^{-n/2}\sum_{i=1}^{4^n}\bar{e}_{i,n}\mathbb{1}_{i2^{-n}\le t}\Big],$$

or equivalently $w_t=\lim_n w_{k_n(t)2^{-n}}$, where $k_n(t)2^{-n}\le t<(k_n(t)+1)2^{-n}$, $1\le k_n(t)\le 4^n$. Moreover, the projection operator becomes

$$\mathbb{E}\{x\mid w_s,\ s\le t\}=\lim_n\sum_{i=1}^{4^n}\mathbb{E}\{x\,\bar{e}_{i,n}\}\,\bar{e}_{i,n}\mathbb{1}_{i2^{-n}\le t},$$

which can be proved to be convergent (as a particular case of a stochastic integral considered later on) in L^2.

To recover a white noise $\dot w = \{e^n_r : r \in R_n,\, n \geq 1\}$ as a diadic sequence of independent standard normally distributed random variables from a given Wiener process $w = \{w_t : t \geq 0\}$; we may use the square wave procedure, namely, for $i = 1, 2, \dots$ consider the Haar-type functions $f_i(s) = \mathbb{1}_{2i-1<2s\leq 2i} - \mathbb{1}_{2(i-1)<2s\leq 2i-1}$ and $f_{i,n}(s) = 2^{-n/2} f_i(s2^n)$, for $n \geq 0$. By construction, if $n \geq m$ then $f_{i,n} f_{j,m} = 0$ except for i within $(j-1)2^{n-m}+1$ and $j2^{n-m}$, and moreover, $\{f_{i,n}\}$ is an orthonormal system in $L^2(]0,\infty[)$. Therefore

$$e^n_{i2^{-n}} = w(f_{i,n}) = 2^{-n/2}\big[w_{(i-1)2^{-n}} - 2w_{(2i-1)2^{-n-1}} + w_{i2^{-n}}\big], \tag{4.20}$$

for $i = 1, \dots, 4^n$, $n \geq 1$, define a white noise which produces another Wiener process via (4.18), also given by the stochastic integral

$$\bar w_t = \sum_{n=1}^{\infty} 2^{-n} \sum_{i=1}^{4^n} w(f_{i,n}) \mathbb{1}_{i2^{-n}\leq t} = \int_0^t f_T(s)\mathrm{d}w_s \quad \forall T \geq t > 0,$$

where the real-valued function

$$f_t = \sum_{n=1}^{\infty} 2^{-n} \sum_{i=1}^{4^n} f_{i,n} \mathbb{1}_{i2^{-n}\leq t}, \qquad \int_0^{\infty} |f_t(s)|^2 \mathrm{d}s = t, \quad \forall t > 0,$$

is defined as an orthogonal series expansion in $L^2(]0,\infty[)$. Remark that $f_t(s) = f_T(s)$ a.e. s in $(0,t)$ for every $t \leq T$, and $f_t(s) = 0$ a.e. for s in (t,∞). Actually, for the probability measure $\mathrm{d}t/T$ defined on Borel σ-algebra on $]0,T[$, the family of random variables $\{\sqrt{T} f_t : t \in [0,T]\}$ is a Wiener process.

Furthermore, if a factor 2^{k-1} is added to the orthogonal series (4.18) then we may begin the sum with $n = k$ instead of $n = 1$. Comparing with the initial isometry given via orthonormal sequences, we note that the orthonormal system $\{f_{i,n}\}$ can be completed to be a basis by adding the functions $\tilde f_i(s) = \tilde f_{i,0}(s) = \mathbb{1}_{(i-1)<s\leq i}$, for $i = 1, 2 \dots$. Indeed, it suffices to check that $1/2\{\tilde f_{i,0}\} \pm 1/2\{f_{i,0}\}$ yields $\{\tilde{\tilde f}_{i,1}(s) = \mathbb{1}_{i-1<2s\leq i}\}$, and $1/2\{\tilde f_{i,1}\} \pm 1/2\{f_{i,1}\}$ yields $\{\tilde f_{i,2}(s) = \mathbb{1}_{i-1<4s\leq i-1}\}$ and so on. Thus, the isometry $w(f_{i,n}) = e_{i,n}$ and $w(\tilde f_i) = \tilde e_i$ mapping the basis $\{f_{i,n} : i = 1, \dots, 4^n,\, n \geq 0\} \cup \{\tilde f_i : i \geq 1\}$ in $L^2(]0,\infty[)$ into an orthornormal system $\{e_{i,n} : i = 1, \dots, 4^n,\, n \geq 0\} \cup \{\tilde e_i : i \geq 1\}$ in L^2 produces an expression very similar to (4.18), namely,

$$\begin{cases} \tilde w_t = \displaystyle\sum_{i=1}^{\infty} \tilde c_i(t)\tilde e_i + \sum_{n=0}^{\infty}\sum_{i=1}^{4^n} c_{i,n}(t) e_{i,n}, \\ \tilde c_i(t) = \displaystyle\int_0^t \tilde f_i(s)\mathrm{d}s, \qquad c_{i,n}(t) = \int_0^t f_{i,n}(s)\mathrm{d}s, \end{cases} \tag{4.21}$$

where the first series in i is a finite sum for each fixed $t > 0$, and the series in n reduces to a finite sum if $t = j2^{-m}$ for some $m \geq 0$ and $j \geq 1$. Essentially based

on Borel-Cantelli Lemma and the estimates

$$q_n = \max_{t\geq 0}\sum_{i=1}^{4^n}|c_{i,n}(t)e_{i,n}| = 2^{-n/2}\max_{i=1,\dots,4^n}|e_{i,n}|,$$
$$P\big(|e_{i,n}|>a\big)\leq\frac{2}{\sqrt{\pi}}\mathrm{e}^{-a^2/2},\qquad P\big(\max_{i=1,\dots,4^n}|e_{i,n}|>a\big)\leq 4^n\frac{2}{\sqrt{\pi}}\mathrm{e}^{-a^2/2},$$
$$P\big(q_n>\theta(2^{-n}\ln 8^n)^{1/2}\big)\leq\frac{2}{\sqrt{\pi}}4^{n(1-\theta^2)},\qquad \theta>1,$$

a more careful analysis shows the uniform convergence on any bounded time interval, almost surely. Actually, this is almost Ciesielski-Levy's construction as described in McKean [167, Section 1.2, pp. 5–8] or Karatzas and Shreve [124, Section 2.3, pp. 56–59]. Remark that with the expression (4.21), we cannot easily deduce a neat series expansion like (4.19) for the projection operator, i.e., since the functions $\{c_{i,n}\}$ have disjoint support only as i changes, for a fixed $t>0$, the orthogonal systems $\{\tilde{c}_i(s)\tilde{e}_i,\ c_{i,n}(s)e_{i,n}\ :\ s\leq t,\ \ i,n\}$ and $\{\tilde{c}_i(s)\tilde{e}_i,\ c_{i,n}(s)e_{i,n}\ :\ s>t,\ \ i,n\}$ are not orthogonal to each other, as in the case of the orthogonal series expansion (4.18). In the context of the orthogonal series expansion (4.21), the series

$$\langle\dot{\tilde{w}},\phi\rangle=\sum_{i=1}^{\infty}\tilde{e}_i\langle\tilde{f}_i,\phi\rangle+\sum_{n=0}^{\infty}\sum_{i=1}^{4^n}e_{i,n}\langle f_{i,n},\phi\rangle,\qquad\forall\phi\in\mathcal{S}(]0,\infty[),$$

could be referred to as white noise, the derivative in the sense of Schwartz distribution of a Wiener process, meaningful only as a generalized process.

On the other hand, note that we cannot take a fractional derivative to recover a white noise, i.e., the limit $(t-r)^{-1/2}\big[w_t-w_r\big]\to e_r$ as $t\downarrow r$ for a particular sequence of $\{t\}$. Indeed, if $r<t<s$ then w_t-w_r and w_s-w_t are independent, and hence

$$\mathbb{E}\Big\{\Big|\frac{w_s-w_r}{\sqrt{s-r}}-\frac{w_t-w_r}{\sqrt{t-r}}\Big|^2\Big\}=2-2\mathbb{E}\Big\{\Big(\frac{w_s-w_r}{\sqrt{s-r}}\Big)\Big(\frac{w_t-w_r}{\sqrt{t-r}}\Big)\Big\}=$$
$$=2\Big(1-\frac{\sqrt{t-r}}{\sqrt{s-r}}\Big).$$

Thus, if $\alpha_n\to 0$ then $(1-\sqrt{\alpha_{n+k}}/\sqrt{\alpha_n})$ does not converges to 0 as $n,k\to\infty$, which implies that the sequence $(w_{r+\alpha_n}-w_r)/\sqrt{\alpha_n}$ cannot be a Cauchy sequence in L^2. Therefore, we may have a subsequence such that $(w_{r+\alpha_n}-w_r)/\sqrt{\alpha_n}\to e_r$ weakly in L^2, but $E\{|e_r|^2\}\neq 1$, since otherwise, the sequence would converge in the L^2 norm.

The White Noise (another)

With the previous observation in mind, consider a countable family $\{e_r^n\}$ of standard normally distributed random variables, indexed for r in the diadic numbers $R=\bigcup_n R_n=\{r=i2^{-n}:i=1,\dots,4^n\}$ as early; but, we assume only that the finite family $\{e_r^n:r\in R_n\}$ is independent, for each fixed $n\geq 1$. Based

on the diadic property $2^{-n}\sum_{i=1}^{4^n}\mathbb{1}_{i2^{-n}\le t}=\max\{r:r\in R_n,\,r\le t\}\to t$, define the sequence of normally distributed random variables $w_0^n=0$ and

$$w_t^n=2^{-n/2}\sum_{i=1}^{4^n}e_{i2^{-n}}^n\,\mathbb{1}_{i2^{-n}\le t},\quad\forall t>0. \tag{4.22}$$

Note that $\mathbb{E}\{w_t^n\}=0$ and $\mathbb{E}\{|w_r^n|^2\}=r$, for every r in R. Thus, the classic Central Theorem shows that $\{w_r^n:n\ge1\}$ converges in law and $\lim_n\mathbb{E}\{|w_t^n-w_r^n|^2\}=t-r$, for any $t>r>0$. Since, for Gaussian variables with zero-mean we have the equality

$$\mathbb{E}\{|w_r^n-w_s^n|^4\}=3\big(\mathbb{E}\{|w_r^n-w_s^n|^2\}\big)^2=3|r-s|^2,\quad\forall r,s\in R,$$

this construction yields a Wiener measure W, i.e., a probability measure on $\Omega=C([0,\infty[)$ such that the coordinate functions $\omega:\Omega\mapsto\omega(t)=w_t(\omega)$ define a Wiener process.

Contrary to the previous series (4.18), the convergence in L^2 of the whole sequence $\{w_t^n:n\ge1\}$ is not automatically insured, we need to assume that the system $\{e_r^n:r\in R_n\}$ is compatible with the diadic numbers in the sense that without ambiguity we can remove the super-index n in e_r^n and use the notation $\{e_r:r\in R_n\}$. Indeed, e.g., by compactness, we can extract a subsequence $\{n_k\}$ such that $w_r^{n_k}\to w_r$ in L^2, for every r in R (i.e., only the random variables $\{e_r^{n_k}:r\in R_{n_k},\,k\ge1\}$ were used), but another convergent subsequence may have another limit (which uses another subset of random variables $\{e_r^n\}$). This previous argument can not used if we impose the condition $e_r^n=e_r^m$, for every n, m and r, i.e., compatibility with the diadic numbers. Moreover, under this assumption, we can single out all terms in the sum defining w_r^n using $\{e_r:r\in R_n\}\smallsetminus\{e_r:r\in R_{n-1}\}$ to obtain the relation $\sqrt{2}w_r^n=w_r^{n-1}+v_r^n$, for r in R_{n-1}, with v_r^n being a normally distributed random variable independent of w_r^n satisfying $\mathbb{E}\{v_r^n\}=0$ and $\mathbb{E}\{|v_r^n|^2\}=r$. By induction, we deduce

$$w_r^n=2^{(n(r)-n-1)/2}w_r^{n(r)-1}+\sum_{k=n(r)}^{n}2^{(k-n-1)/2}v_r^k, \tag{4.23}$$

where $n(r)=\inf\big\{n\ge1:r\in R_n\big\}$, $w_r^0=0$ and $\{w_r^{n(r)-1},v_r^{n(r)},\dots,v_r^n\}$ is an orthogonal system. This implies that the whole sequence $\{w_r^n:n\ge1\}$ converges in L^2, i.e., the limit

$$w_t=\lim_n\Big[2^{-n/2}\sum_{i=1}^{4^n}e_{i2^{-n}}\mathbb{1}_{i2^{-n}\le t}\Big],\quad\forall t>0 \tag{4.24}$$

exits in L^2, almost as an orthogonal series expansion if r belongs to R. Anyway, only the random variables $\{e_r^n:r\in R_n,\,n\ge1,\,r\le t\}$ intervene in getting $\{w_s:s\le t\}$, and the projector operator has the form

$$\mathbb{E}\{x\mid w_s,\,s\le t\}=\lim_n\sum_{i=1}^{4^n}\mathbb{E}\{x\,e_{i2^{-n}}\}\,e_{i2^{-n}}\mathbb{1}_{i2^{-n}\le t}, \tag{4.25}$$

as a limit in L^2 of almost an orthogonal series expansion, valid for any x in the closed linear span of $\{w_t : t \geq 0\}$. Hermit polynomials are needed to get a series expansion for any x in $L^2(\Omega, \mathcal{F}, P)$.

Let us summarize the main points proved above:

Proposition 4.12. *Let $\{e_{i,n} : i = 1, \dots, 4^n, n \geq 1\}$ be a countable family of identically distributed random variables with $\mathbb{E}\{e_{i,n}\} = 0$ and $\mathbb{E}\{|e_{i,n}|^2\} = 1$, and such that each $\dot{w}^n = \{e_{i,n} : i = 1, \dots, 4^n\}$ is a sub-family of independent random variables. For convenience we may take all $e_{i,n}$ normally distributed, but this is not necessary.*

(a) *If $\dot{w}^1, \dots, \dot{w}^n, \dots$ are independent then the orthogonal series expansion (4.18) yields a Wiener process, and the conditional expectation operator (4.19).*

(b) *If $e_r^n = e_{i,n}$, with $r = i2^{-n}$, $i = 1, \dots, 4^n$, then the limit of the expression (4.22) exists in law and defines a Wiener measure.*

(c) *If, besides the conditions of (b), also we suppose the diadic compatibility assumption, i.e., $\{e_r^n = e_r : r = i2^{-n}, i = 1, \dots, 4^n, n \geq 1\}$ is an independent family, then the limit (4.24) exits in L^2, almost as an orthogonal series expansion, and the conditional expectation operator takes the form (4.25).* □

• *Remark* 4.13. If $\dot{w} = \{e_r^n : r \in R_n, n \geq 1\}$ is a diadic family of identically distributed random variables with $\mathbb{E}\{e_r^n\} = 0$ and $\mathbb{E}\{|e_r^n|^2\} = 1$ then the diadic compatibility assumption as in (c) of Proposition 4.12 is satisfied for the diadic family $\dot{w}' = \{e_r : r \in R_n, n \geq 1\}$ defined by $e_r = \sum_{k=1}^{\infty} 2^{k/2} e_r^{k+n(r)-1}$, with $n(r) = \min\{n \geq 1 : r \in R_n\}$. In this case, note that the orthogonal series expansion (4.18) for the white noise $\dot{w}$ is very similar to the expression (4.24) for the white noise $\dot{w}'$. It is clear that there are infinite many choices to obtain a white noise $\dot{w}'$ from the initial $\dot{w}$, e.g., any sequence $\{k_n : n \geq 1\}$ with $k_n \geq n(r)$ will produce a suitable $\dot{w}'$, where $e_r = e_r^{k(n)}$, for r in $R_n \setminus R_{n-1}$, with $R_0 = \emptyset$. □

• *Remark* 4.14. Under the compatibility assumption as in (c) of Proposition 4.12, we may use the equality (4.23) to obtain

$$\sum_{n=n(r)}^{\infty} 2^{-n/2} w_r^n = 2^{(n(r)-1)/2} w_r^{n(r)-1} + \sum_{n=n(r)}^{\infty} 2^{-n} \sum_{k=n(r)}^{n} 2^{(k-1)/2} v_r^k,$$

and exchanging the double sum

$$\sum_{n=n(r)}^{\infty} 2^{-n} \sum_{k=n(r)}^{n} 2^{(k-1)/2} v_r^k = \sum_{k=n(r)}^{\infty} 2^{(k-1)/2} v_r^k \sum_{n=k}^{\infty} 2^{-n} =$$

$$= \sum_{k=n(r)}^{\infty} 2^{-(k-1)/2} v_r^k.$$

This shows that the series (4.18), with $e_{i,n} = e_{i2^{-n}}$, converges in L^2, as an orthogonal series expansion relative to $\{w_r^{n(r)-1}, v_r^{n(r)}, \dots, v_r^n, \dots\}$, with $t = r$ in R. For a non-diadic t, we have an almost orthogonal series expansion. □

• *Remark* 4.15. The above arguments can be used to construct the integral of a function φ belonging to $L^2(]0,\infty[)$. For instant, if $\varphi_n(s) = \varphi(i2^{-n})$ for s in $](i-1)2^{-n}, i2^{-n}]$, $i = 1,\dots,4^n$, then

$$2^{-n}\sum_{i=1}^{4^n}|\varphi(i2^{-n})|^2 = \int_0^{4^n}|\varphi_n(s)|^2\mathrm{d}s.$$

Therefore, e.g., we may replace (4.18) and (4.24) with

$$w(\varphi) = \sum_n 2^{-n}\sum_{i=1}^{4^n}\varphi(i2^{-n})e_{i,n} \quad \text{and} \quad 2^{-n/2}\sum_{i=1}^{4^n}\varphi(i2^{-n})e_{i2^{-n}} \to w(\varphi),$$

to obtain the integral of φ with respect to the Wiener noise $\dot{w}$. Essentially, this is like using the diadic system of functions $\phi_{i,n} = (-1)^{i-1}\mathbb{1}_{](i-1)2^{-n},i2^{-n}]}$, $i = 1,\dots,4^n$, and $\{(-1)^{i-1}e_{i,n}\}$ to establish a mapping similar (after a proper scaling) to the initial isometry in the beginning of this subsection. Note that $\{\phi_{i,n}\}$ is not a fully orthogonal system, but $\phi_n = \sum_{i=1}^{4^n}\phi_{i,n}$ is a finite sum of functions with disjoint supports and ϕ_n is orthogonal to ϕ_m if $n \neq m$. It is clear that behind is the orthonormal system obtained from (4.20). □

The Poisson Measure

The construction of Poisson (random) measure and some of its properties are necessary to discuss general Poisson noises. One way is to follow the construction of the general Wiener noise or white noise (random) measure, but using Poisson (random) variables instead of Gaussian (random) variables.

If $\{\tau_{i,n} : i \geq 1\}$ is a sequence of independent exponentially (with parameter 1) distributed random variables then random variables $\zeta_n(\lambda) = \sum_k \mathbb{1}_{\theta_{k,n}\leq\lambda}$, with $\theta_{k,n} = \tau_{1,n} + \cdots + \tau_{n,n}$, is a sequence of independent identically distributed random variables having a Poisson distribution with parameter λ. Hence, $\tilde{\zeta}_n(\lambda) = \zeta_n(\lambda) - \lambda$ has mean zero and variance $\mathbb{E}\{|\tilde{\zeta}_n(\lambda)|^2\} = \lambda$. If $\{h_n : n \geq 1\}$ is a complete orthogonal system in a Hilbert space H with $(h_n,h_n)_H = 1/k_n$, then any function h in H can be written as a converging orthogonal series $h = \sum_n (h,h_n)_H k_n h_n$, and $(h,h)_H = \sum_n |(h,h_n)_H|^2 k_n$. Thus the mapping $h \mapsto q(h) = \sum_n (h,h_n)_H \tilde{\xi}_n(k_n)$ is a linear isometry from H into $L^2 = L^2(\Omega,\mathcal{F},P)$, and if $(h,h_n)_H = 1$ for any n in a finite subset of indexes N_h and $(h,h_n)_H = 0$ otherwise then $p(h) = \sum_n (h,h_n)_H \xi_n(k_n)$ is Poisson random variable with parameter $\sum_{n\in N_h} k_n$. In any case, if the series $m(|h|) = \sum_n |(h,h_n)_H| k_n < \infty$ then $p(h) = \sum_n (h,h_n)_H \xi_n(k_n)$ is convergent, and $p(h) = q(h) + m(h)$, with $\sum_n (h,h_n)_H k_n$.

Another construction is developed for a more specific Hilbert space, namely, $H = L^2(Y,\mathcal{Y},\mu)$ with a σ-finite measure space $(Y,\mathcal{Y},\mu)$, where the Poisson character is imposed on the image of $\mathbb{1}_K$ for any K in $\mathcal{Y}$ with $\mu(K) < \infty$.

Two steps are needed, first assume $\mu(Y) < \infty$ and choose a sequence $\{\zeta_n : n \geq 1\}$ of independent identically distributed following the probability law given by $\mu/\mu(Y)$ and also choose an independent Poisson distributed variable η with

parameter $\lambda = \mu(Y)$. Define $p(A) = 0$ when $\eta = 0$ and $p(A) = \sum_{n=1}^{\eta} \mathbb{1}_{\zeta_n \in A}$ otherwise, for every A in $\mathcal{Y}$. The random variable $p(A)$ takes only nonnegative integer values, $p(Y) = \eta$, and if $A_1, \dots, A_k$ is a finite partition of Y, i.e., $Y = \sum_i A_i$, and $n_1 + \cdots + n_k = n$ then

$$P\big(p(A_1) = n_1, \dots, p(A_k) = n_k\big) = \\ = P\big(p(A_1) = n_1, \dots, p(A_k) = n_k \,:\, p(Y) = n\big) P\big(p(Y) = n\big),$$

which are multinomial and Poisson distribution, and so

$$P\big(p(A_1) = n_1, \dots, p(A_k) = n_k\big) = \\ = n! \frac{\big(\mu(A_1)\big)^{n_1}}{\big(\mu(Y)\big)^{n_1} n_1!} \cdots \frac{\big(\mu(A_k)\big)^{n_k}}{\big(\mu(Y)\big)^{n_k} n_k!} \, \mathrm{e}^{-\mu(Y)} \frac{\big(\mu(Y)\big)^n}{n!},$$

and summing over $n_1, \dots, n_k$ except in n_i , we obtain

$$P\big(p(A_i) = n_i\big) = \mathrm{e}^{-m(A_i)} \frac{\big(\mu(A_i)\big)^{n_i}}{n_i!}.$$

Thus the mapping $A \mapsto p(A)$ satisfies:

(1) for every ω, $A \mapsto p(A, \omega)$ is measure on Y;

(2) for every measurable set A, the random variable $p(A)$ has a Poisson distribution with parameter (or mean) $\mu(A)$;

(3) if $A_1, \dots, A_k$ are disjoint then $p(A_1), \dots, p(A_k)$ are independent.

In the previous statements, note that if $\mu(A) = 0$ then the random variable $p(A) = 0$, which is (by convention) also referred to as having a Poisson distribution with parameter (or intensity) zero.

For the second step, because μ is σ-finite, there exists a countable partition $\{Y_k : k \geq 1\}$ of Y with finite measure, i.e., $Y = \sum_k Y_k$ and $\mu(Y_k) < \infty$. Now, for each k with construct p_k (as above) corresponding to the finite measure μ_k, with $\mu_k(A) = \mu(A \cap Y_k)$, in a way that the random variable involved $\zeta_{k,n}$ and η_k are all independent in k. Hence the mapping $A \mapsto p_k(A)$ satisfies (1), (2) and (3) above, and also:

(4) for every choice of $k_1, \dots, k_n$ and $A_1, \dots, A_n$ in $\mathcal{A}$, the random variables $p_{k_1}(A_1), \dots, p_{k_n}(A_n)$ are independent.

Since a sum of independent Poisson (random) variables is again a Poisson variable, the series $p(A) = \sum_k p_k(A)$ defines a Poisson (random) variable with parameter (or mean) $\mu(A)$ whenever $\mu(A) < \infty$. If $\mu(A) = \infty$ then

$$\sum_k P\big(p_k(A) \geq 1\big) = \sum_n \big(1 - \mathrm{e}^{-\mu(A \cap Y_k)}\big) = \infty,$$

since $\lim_n (1 - \mathrm{e}^{-\mu(A \cap Y_k)}\big) = 1$, i.e., if $\mu(A) = \infty$ then $p(A) = \infty$ almost surely. Hence, the mapping $A \mapsto p(A)$ satisfies (1), (2) and (3), as long as a random variable which is equal to infinite (or to zero) is considered a Poisson variable with parameter $\lambda = \infty$ (or $\lambda = 0$). In this case, a Poisson variable with $\lambda = \infty$

(or $\lambda = 0$) means a (degenerate) random variable which is equal to $+\infty$ (or to 0) almost surely. Remark that contrary to the Wiener process, to define the Poisson measure, the previous construction uses *independence* instead of orthogonality.

Recall that for convenience, if μ is a measure and f a function then $\mu(f)$ denotes the integral of f with respect to μ. In particular, $\mu(A) = \mu(\mathbb{1}_A)$ and $p(A) = p(\mathbb{1}_A)$.

Proposition 4.16. *If $(Y, \mathcal{Y}, \mu)$ is a σ-finite measure space then the previous construction yields a Poisson random measure p with intensity measure μ, i.e., (1), (2), (3) above are satisfied. Moreover, if $\mu(Y) < \infty$ and $\varphi : Y \to \mathbb{R}^m$ is a measurable function then*

$$p(\varphi) = \int_Y \varphi(y)\, p(\mathrm{d}y)$$

defines a random variable on $\mathbb{R}^m$ with compound Poisson distribution, namely,

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}r\cdot p(\varphi)}\} = \exp\Big[\int_Y \big(\mathrm{e}^{\mathrm{i}r\cdot\varphi(y)} - 1\big)\mu(\mathrm{d}y)\Big], \quad \forall r \in \mathbb{R}^m.$$

Moreover, if $\mu(|\varphi|^2) < \infty$ then $\mathbb{E}\{p(\varphi)\} = \mu(\varphi)$ and $\mathbb{E}\{|q(\varphi)|^2\} = \mu(|\varphi|^2)$, where $q = p - \mu$. Furthermore, if $A_1, \dots, A_n$ are disjoint measurable sets then the random variables $p(\varphi\mathbb{1}_{A_1}), \dots, p(\varphi\mathbb{1}_{A_n})$ are independent.

Proof. From the construction we check that for every ω, the measure $A \mapsto p(A, \omega)$is supported in a finite number of points, namely, $\zeta_i(\omega)$ for $i = 1, \dots, \eta(\omega)$. Thus, the expression of the random variable $p(A)$ is finite. Using a diadic approximation of φ, i.e., we partition $\mathbb{R}^m$ into diadic cubes of the form $C_{j,n} =](j_1 - 1)2^{-n}, j_1 2^{-n}] \times \cdots \times](j_m - 1)2^{-n}, j_m 2^{-n}]$, with $j = (j_1, \dots, j_m)$ and set $\varphi_n(x) = j2^{-n}$ for every x in $C_{j,n}$, we have $|\varphi(x) - \varphi_n| \le 2^{-n}\sqrt{m}$. Since

$$p(\varphi_n, \omega) = \sum_j (j2^{-n}) p(\varphi^{-1}(C_{j,n}, \omega),$$

from the definition of the Poisson measure p we deduce

$$\begin{aligned}\mathbb{E}\{\mathrm{e}^{\mathrm{i}r\cdot p(\varphi_n)}\} &= \prod_j \mathbb{E}\{\mathrm{e}^{\mathrm{i}r\cdot j2^{-n} p(\varphi^{-1}(C_{j,n}))}\} = \\ &= \prod_j \exp\big[(\mathrm{e}^{\mathrm{i}r\cdot j2^{-n}} - 1)\mu(\varphi^{-1}(C_{j,n}))\big] = \exp\Big[\int_Y (\mathrm{e}^{\mathrm{i}r\cdot\varphi_n(y)} - 1)\mu(\mathrm{d}y)\Big]\end{aligned}$$

and the first part follows as $n \to \infty$.

Once the expression of the characteristic function have been proved, the mean and the variance are calculated by taking derivative with respect to the parameter r, and the last part, regarding the independence, is deduced by the convergence of $p(\varphi_n)$ to $p(\varphi)$ and the property (3) of Poisson measure discussed above. □

Remark that as it was defined, for each ω fixed, the Poisson random measure $p(\cdot,\omega)$ is a finite sum of Dirac measures. Hence, p is also called Poisson point measure. The companion measure $q = p - \mu$ is referred to as a centered (or martingale) Poisson (random or point) measure.

Sometimes, we may single-out a time variable, i.e., replace Y and μ with $Y\times]0,\infty[$ and $\mu\times \mathrm{d}t$. In this case, the variable η can be specifically constructed as a Poisson process with parameter $\lambda = \mu(Y) < \infty$, i.e.,

$$\eta(t) = \sum_n \mathbb{1}_{t\geq\theta_n}, \quad \forall t > 0,$$

where $\theta_n = \tau_1 + \cdots + \tau_n$ and $\{\tau_n : n \geq 1\}$ is a sequence of independent exponentially distributed (with parameter λ) random variable. In this case

$$p(A\times]a,b]) = \sum_{n=1}^{\eta(b)} \mathbb{1}_{\zeta_n\in A} - \sum_{n=1}^{\eta(a)} \mathbb{1}_{\zeta_n\in A} = \sum_n \mathbb{1}_{\zeta_n\in A}\mathbb{1}_{a<\theta_n\leq b}, \quad \forall a \leq b.$$

If $\mu(Y) = \infty$ then express the space Y as countable number of disjoint sets with finite measure (i.e., $Y = \sum_k Y_k$ with $\mu(Y_k) < \infty$), and find sequences of independent variables $\zeta_{n,k}$ with distribution $\mu(\cdot\cap Y_k)/\mu(Y_k)$ and $\tau_{n,k}$ exponentially distributed with parameter $\mu(Y_k)$, for any $n,k \geq 1$. The Poisson measure is given by

$$p(A\times]a,b]) = \sum_{n,k} \mathbb{1}_{\zeta_{n,k}\in A}\mathbb{1}_{a<\theta_{n,k}\leq b}, \quad \forall a \leq b,$$

where $\theta_{k,n} = \tau_{1,k} + \cdots + \tau_{n,k}$. Our interest is the case where $Y = \mathbb{R}^d_*$ and $\zeta_{n,k}$ is interpreted as the jumps of a Lévy process.

The Poisson Noise I

Another type of complications appear in the case of the compound Poisson noise, i.e., like a Poisson process with jumps following some prescribed distribution, so that the paths remain piecewise constant.

Consider $\mathbb{R}^d_* = \mathbb{R}^d \setminus \{0\}$ and $\mathcal{B}_* = \mathcal{B}(\mathbb{R}^d_*)$, the Borel σ-algebra, which is generated by a countable semi-ring $\mathcal{K}$. (e.g., the family of d-intervals $]a,b]$ with closure in $\mathbb{R}^d_*$ and with rational end points). Now, beginning with a given (non-zero) finite measure m in $(\mathbb{R}^d_*, \mathcal{B}_*)$, we construct a sequence $\dot{q} = \{(z_n, \tau_n : n \geq 1\}$ of independent random variables such that each τ_n is exponentially distributed with parameter $m(\mathbb{R}^d_*)$ and z_n has the distribution law $A \mapsto m(A)/m(\mathbb{R}^d_*)$, thus, the random variables $\theta_n = \tau_1 + \cdots + \tau_n$ have $\Gamma(m(\mathbb{R}^d_*), n)$ distribution. The series $\eta_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is almost surely a finite sum and defines a Poisson process with parameter $m(\mathbb{R}^d_*)$, satisfying $\mathbb{E}\{\eta_t\} = tm(\mathbb{R}^d_*)$ and $\mathbb{E}\{|\eta_t - tm(\mathbb{R}^d_*)|^2\} = tm(\mathbb{R}^d_*)$. Moreover, we may just suppose given a $\mathbb{R}^d$-valued compound Poisson process $\{N_t : t \geq 0\}$ with parameter $(\lambda, m/\lambda)$ or m, i.e., with the following characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}\zeta\cdot N_t(z)}\} = \exp\Big\{t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathrm{i}\zeta\cdot z} - 1\big]m(\mathrm{d}z)\Big\}, \qquad \forall \xi \in \mathbb{R}^d,$$

as a Lévy process, with $N_t = \sum_n z_n \mathbb{1}_{t\geq\theta_n}$.

In any case, the counting measure either

$$p_t(K) = \sum_n \mathbb{1}_{z_n\in K}\mathbb{1}_{t\geq\theta_n}, \quad \forall K\in\mathcal{K},\ t\geq 0,$$

or equivalently

$$p_t(K) = \sum_{n=1}^{\eta(t)} \mathbb{1}_{z_n\in K}, \qquad \eta(t) = \sum_n \mathbb{1}_{t\geq\theta_n}, \quad \forall K\in\mathcal{K},\ t\geq 0,$$

is a Poisson process with parameter $m(K)$, $\eta(t)$ is also a Poisson process with parameter $tm(\mathbb{R}^d_*)$. Moreover, if $K_1,\dots,K_k$ are any disjoint sets in $\mathcal{K}$ then $p_t(K_1),\dots,p_t(K_k)$ are independent processes. Indeed, if $n = n_1+\cdots+n_k$ and $\mathbb{R}^d_* = K_1\cup\cdots\cup K_k$ then

$$\begin{aligned}
&P\big(p_t(K_1)=n_1,\dots p_t(K_k)=n_k\big) = \\
&\quad = P\big(p_t(K_1)=n_1,\dots p_t(K_k)=n_k \mid p_t(K)=n\big)P\big(p_t(K)=n\big) = \\
&\quad = P\big(\sum_{i=1}^n \mathbb{1}_{z_i\in K_1}=n_1,\dots\sum_{i=1}^n \mathbb{1}_{z_i\in K_k}=n_k \mid p_t(\mathbb{R}^d_*)=n\big)P\big(\eta(t)=n\big),
\end{aligned}$$

which are multinomial and Poisson distribution, and so

$$\begin{aligned}
&P\big(p_t(K_1)=n_1,\dots p_t(K_k)=n_k\big) = \\
&\quad = \frac{\big(m(K_1)\big)^{n_1}}{\big(m(\mathbb{R}^d_*)\big)^{n_1}n_1!}\cdots\frac{\big(m(K_k)\big)^{n_k}}{\big(m(\mathbb{R}^d_*)\big)^{n_k}n_k!}\,\mathrm{e}^{-m(\mathbb{R}^d_*)}\frac{\big(m(\mathbb{R}^d_*)\big)^n}{n!},
\end{aligned}$$

and summing over $n_1,\dots,n_k$ except in n_j , we obtain

$$P\big(p_t(K_j)=n_j\big) = \mathrm{e}^{-m(K_j)}\frac{\big(m(K_j)\big)^{n_j}}{n_j!},$$

which proves that $p_t(K_j)$ are independent Poisson processes. This implies that

$$\mathbb{E}\{p_t(K)\} = tm(K), \qquad \mathbb{E}\big\{|p_t(K)-tm(K)|^2\big\} = tm(K),$$

for every K in $\mathcal{K}$ and $t\geq 0$. Hence, the (martingale or centered) measure

$$q_t(K) = \sum_n \mathbb{1}_{z_n\in K}\mathbb{1}_{t\geq\theta_n} - tm(K), \qquad \mathbb{E}\{q_t(K)\} = 0, \quad \forall K\in\mathcal{K}$$

satisfies $\mathbb{E}\{q_t^2(K)\} = tm(K)$, and if $K\cap K' = \emptyset$ then $q_t(K)$ and $q_t(K')$ are orthogonal and, in general, $\mathbb{E}\{q_t(K)q_t(K')\} = tm(K\cap K')$.

• *Remark* 4.17. Sometime it is more convenient not to distinguish the time t in the Poisson measure, i.e., to consider p as a random integer measure on $\mathbb{R}^d_*\times(0,\infty)$. In this case, either two steps are necessary or only a construction on $\mathbb{R}^d_*\times(0,b]$ $(b<\infty)$ is achieved. For instance, given a bounded measure Π on $\mathbb{R}^d_*\times(0,b]$ proceed as follows: (1) find a sequence $\{z_n : n\geq\}$ of independent

random variables with identical distribution Π/c, $c = \Pi(\mathbb{R}^d_* \times (0,b])$, and (2) find an independent Poisson distributed (with parameter c) random variable η, and then define $p(B) = \sum_{n=1^\eta} \mathbb{1}_{z_n \in B}$. By using independent copies of p, we can patch the definition of p from $\mathbb{R}^d_* \times (0,b]$ into $\mathbb{R}^d_* \times (b,2b]$ and so on, to get p defined on the whole $\mathbb{R}^d_* \times (0,\infty)$, and clearly $\Pi(\mathrm{d}z,\mathrm{d}t) = m(\mathrm{d}z)\mathrm{d}t$. In this construction, the origin $\{0\}$ plays not particular role, so that the intensity Π needs only to be a σ-finite Borel measure on some Polish space. Later, to integrate the function z to reproduce the jumps, the Lévy measure condition appears. □

Now, a (real-valued) simple function relative to the semi-ring $\mathcal{K}$ is a finite sum of terms (with disjoint K's) of the form $\alpha\mathbb{1}_K(z)$ (which is equal to α when z belongs to K and 0 otherwise). Each term integrates with respect to $p_t(\mathrm{d}z)$ and $q_t(\mathrm{d}z)$ as follows

$$\int_{\mathbb{R}^d_*} \alpha\mathbb{1}_K(z)\, q_t(\mathrm{d}z) = \alpha\, q_t(K), \qquad \mathbb{E}\Big\{\Big|\int_{\mathbb{R}^d_*} \alpha\mathbb{1}_K(z)\, q_t(\mathrm{d}z)\Big|^2\Big\} = \alpha^2 t m(K).$$

This definition is extended by linearity (uniquely) to any simple function, ψ and because each $\{q_t(K)\}$ are independent when the K's are disjoint, we preserve the relation

$$\mathbb{E}\{|q_t(\psi)|^2\} = \mathbb{E}\Big\{\Big|\int_{\mathbb{R}^d_*} \psi(z)\, q_t(\mathrm{d}z)\Big|^2\Big\} = t\int_{\mathbb{R}^d_*} |\psi(z)|^2 m(\mathrm{d}z) = tm(|\psi|^2).$$

Remark that to simplify the notation, we write $q_t(\psi)$ and $m(\psi)$ to symbolize the integral of a function ψ, e.g., $m(K) = m(\mathbb{1}_K) = m(|\mathbb{1}_K|^2)$. Moreover, because m is a finite measure, if $m(|\psi|^2) < \infty$ then $m(|\psi|) < \infty$.

Again, this integral $\psi \mapsto q_t(\psi)$ is extended as a linear isometry between Hilbert spaces, from $L^2(m) = L^2(\mathbb{R}^d_*, \mathcal{B}_*, tm)$ into $L^2(\Omega, \mathcal{F}, P)$, and

$$q_t(\psi) = \sum_n \psi(z_n)\mathbb{1}_{t\geq\theta_n} - tm(\psi), \qquad \text{with} \quad \mathbb{E}\{q_t(\psi)\} = 0, \tag{4.26}$$

reduces to a finite sum almost surely. This is the same argument as the case of random orthogonal measures, but in this case, this is also a pathwise argument. Indeed, we could use the almost surely finite sum (4.26) as definition.

A priori, the above expression of $q_t(\psi)$ seems to depend on the pointwise definition of ψ, however, if $\psi = \psi'$ m-almost everywhere then $q_t(\psi) = q_t(\psi')$ almost surely. Moreover, $\mathbb{E}\{q_t(\psi)q_s(\psi')\} = (t\wedge s)m(\psi\psi')$ and the process $t \mapsto q_t(\psi)$ is continuous in the L^2-norm.

As mentioned early, $N_t = \sum_n z_n \mathbb{1}_{t\geq\theta_n}$ is a $\mathbb{R}^d$-valued compound Poisson process, and therefore, the expression

$$t \mapsto p_t(\psi) = \sum_n \psi(z_n)\mathbb{1}_{t\geq\theta_n}, \qquad \forall \psi \in L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$$

defines a real-valued compound Poisson process with characteristic function

$$\mathbb{E}\{\mathrm{e}^{\mathrm{i}p_t(\psi)}\} = \exp\Big\{t\int_{\mathbb{R}^d_*} \big[\mathrm{e}^{\mathrm{i}\psi(z)} - 1\big] m(\mathrm{d}z)\Big\}.$$

This yields

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}q_t(\psi)}\} = \exp\Big\{t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\psi(z)} - 1 - \mathtt{i}\psi(z)\big]m(\mathrm{d}z)\Big\}, \tag{4.27}$$

for every ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$.

If $m(|z|) < \infty$ then $\mathbb{E}\{|z_n|\} < \infty$ and $\mathbb{E}\{|N_t|\} = tm(|z|)$. Moreover, we can define the $\mathbb{R}^d$-valued Lévy process $q_t(z) = N_t - tm(z)$ with characteristic $(0, 0, m)$, i.e.,

$$\mathbb{E}\{\mathrm{e}^{\mathtt{i}\zeta\cdot q_t(z)}\} = \exp\Big\{t\int_{\mathbb{R}^d_*}\big[\mathrm{e}^{\mathtt{i}\zeta\cdot z} - 1 - \mathtt{i}\zeta\cdot z\big]m(\mathrm{d}z)\Big\} \tag{4.28}$$

and transition density

$$\begin{cases} P(q_t(z)\in \mathrm{d}x) = \mathrm{e}^{-m(\mathbb{R}^d_*)t}\Big[\delta_0(\mathrm{d}x) + \displaystyle\sum_{n=1}^{\infty}\big(m^{\star n}(\mathrm{d}x) - m^{\star n}(\mathbb{R}^d_*)\big)\frac{t^n}{n!}\Big], \\ m^{\star(n+1)}(B) = (m^{\star n}\star m)(B) = \displaystyle\int_{\mathbb{R}^d_*\times\mathbb{R}^d_*}\mathbb{1}_B(x+y)\,m^{\star n}(\mathrm{d}x)\,m(\mathrm{d}y), \end{cases} \tag{4.29}$$

where $m^{\star 1} = m$ and $m^{\star n}(\mathbb{R}^d_*) = (m(\mathbb{R}^d_*))^n = \lambda^n$. Next, remarking that $t\mapsto q_t(z)$ is continuous except for $t = \theta_n$ that $q_t(z) - q_{t-}(z) = N_t - N_{t-} = z_n$, the expression

$$q_t(K) = \sum_{s\le t}\mathbb{1}_{q_t(z)-q_{t-}(z)\in K} - tm(K) \tag{4.30}$$

is a finite sum almost surely, and can be used to reconstruct the counting measure $\{q_t(K) : K\in\mathcal{K}\}$ from the $\{q_t(z) : t\ge 0\}$. Indeed, just the knowledge that the paths $t\mapsto q_t(z)$ are cad-lag, implies that the series (4.30) reduces to a finite sum almost surely.

The terms $\psi(z_n)\mathbb{1}_{t\ge\theta_n}$ in the series (4.26) are not independent, but setting $\lambda = m(\mathbb{R}^d_*)$ and $m' = m/\lambda$ we compute

$$\begin{aligned} &\mathbb{E}\big\{|\psi(z_n)\mathbb{1}_{t\ge\theta_n}|^2\big\} = m'(|\psi|^2)\,r_n(t), \\ &\mathbb{E}\big\{\psi(z_n)\mathbb{1}_{t\ge\theta_n}\psi(z_k)\mathbb{1}_{t\ge\theta_k}\big\} = |m'(\psi)|^2 r_n(t), \quad \forall k > n\ge 1, \end{aligned}$$

where

$$\mathbb{E}\{\mathbb{1}_{t\ge\theta_n}\} = \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!}\mathrm{e}^{-\lambda s}\mathrm{d}s = r_n(t)$$

with $\sum_n r_n(t) = tm(\mathbb{R}^d_*)$. Thus, the Gram-Schmidt orthogonalization procedure can be used to define $e_0(\psi,t) = -tm(\psi)$, $e_1(\psi,t) = \psi(z_1)\mathbb{1}_{t\ge\theta_1} - m'(\psi)r_1(t)$, $\mathbb{E}\{|e_1(\psi,t)|^2\} = \big(m'(|\psi|^2) - |m'(\psi)|^2 r_1(t)\big)r_1(t)$, and

$$e_2(\psi,t) = \psi(z_2)\mathbb{1}_{t\ge\theta_2} - m'(\psi)r_2(t) - \frac{|m'(\psi)|^2}{m'(|\psi|^2) - |m'(\psi)|^2 r_1(t)}\,e_1(\psi,t),$$

and a more complicate expression for $n \geq 2$. Actually, this is equivalent to

$$e_n(\psi, t) = \psi(z_n)\mathbb{1}_{t\geq\theta_n} - \mathbb{E}\{\psi(z_n)\mathbb{1}_{t\geq\theta_n} \,|\, \psi(z_i)\mathbb{1}_{t\geq\theta_i},\, i = 1, \dots, n-1\}$$

the conditional expectation, and $q_t(\psi) = \sum_{n\geq 1} e_n(\psi, t)$.

Alternatively, if $e'_n(\psi, t) = \big(\psi(z_n) - m'(\psi)\big)\mathbb{1}_{t\geq\theta_n}$ then $\mathbb{E}\{e'_n(\psi, t)\} = 0$, $\mathbb{E}\{|e'_n(\psi, t)|^2\} = \big(m'(|\psi|^2) - |m'(\psi)|^2\big)r_n(t)$, and for $k > n \geq 1$,

$$\mathbb{E}\{e'_n(\psi, t)e'_k(\psi, t)\} = \mathbb{E}\{(\psi(z_n) - m'(\psi))(\psi(z_k) - m'(\psi))\mathbb{1}_{t\geq\theta_n}\} = 0.$$

Also, define $e''_n(\psi, t) = m'(\psi)\big(\mathbb{1}_{t\geq\theta_n} - r_n(t)\big)$, which satisfies $\mathbb{E}\{e''_n(\psi, t)\} = 0$, $\mathbb{E}\{|e''_n(\psi, t)|^2\} = |m'(\psi)|^2 r_n(t)\big(1 - r_n(t)\big)$, $\mathbb{E}\{e''_n(\psi, t)e''_k(\psi, t)\} = 0$ if $n \neq k$, and $\mathbb{E}\{e''_n(\psi, t)e'_k(\psi, t)\} = 0$ for any n, k. Therefore $\{e'_n(\psi, t), e''_k(\psi, t) : n, k \geq 1\}$ is an orthogonal system such that $q_t(\psi) = \sum_n e'_n(\psi, t) + \sum_k e''_n(\psi, t)$ or $q_t(\psi) = \sum_n e'_n(\psi, t) + m'(\psi)N_t$, where $N_t = \sum_n \mathbb{1}_{t\geq\theta_n}$ is a Poisson process with parameter $m(\mathbb{R}^d_*)$. Comparing with the white noise, the orthogonality is not necessary since the series defining (4.26) is finite almost surely.

If $\mathcal{F}_\psi$ is σ-algebra generated by $\{q_s(\psi) : s \leq t\}$ (or equivalently by the countable family $\{e_r(\psi) : r \leq t,\, r \in R\}$), then the closure linear subspace H_ψ of $L^2(\Omega, \mathcal{F}_\psi, P)$ spanned by $\{q_t(\psi) : t \geq 0\}$ could be called the Poisson noise relative to any nonzero ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}, m)$. If we normalize the orthogonal system then the projection operator

$$\mathbb{E}\{x \,|\, q_s(\psi),\, s \leq t\} = \sum_n \frac{\mathbb{E}\{x\, e_n(\psi, t)\}}{\mathbb{E}\{|e_n(\psi, t)|^2\}}\, e_n(\psi, t), \tag{4.31}$$

valid only for x in H_ψ. Contrary to the white noise, there is not an equivalent to the Hermit polynomials (in general), and we do not have an easy construction of an orthonormal basis for $L^2(\Omega, \mathcal{F}_\psi, P)$.

• *Remark* 4.18. The above argument used to construct $q_t(\psi)$ for every ψ in $L^2(m)$ can be adapted to define $q_t(\Psi) = q(\Psi\,\mathbb{1}_{]0,t]})$ as the double integral of functions $\Psi = \Psi(t, z)$ belonging to $L^2(]0, \infty[\times\mathbb{R}^d_*, \mathrm{d}t \times \mathrm{d}m)$, where

$$\mathbb{E}\{|q(\Psi)|^2\} = \int_0^\infty \mathrm{d}t \int_{\mathbb{R}^d_*} |\Psi(t, z)|^2 m(\mathrm{d}z),$$

and $\mathbb{E}\{q(\Psi)\} = 0$. Even $\mathbb{R}^n$-valued functions Ψ can be handled with the same argument. □

The Poisson Noise II

Even more complicate is the case of the general Poisson noise, which is regarded as Poisson point process or Poisson measure, i.e., the paths are cad-lag functions, non necessary piecewise constant.

Let m be a σ-finite measure in $(\mathbb{R}^d_*, \mathcal{B}_*)$, with the Borel σ-algebra being generated by a countable semi-ring $\mathcal{K}$. We partition the space $\mathbb{R}^d_*$ is a disjoint union $\mathbb{R}^d_* = \sum_k \mathtt{R}_k$ with $0 < m(\mathtt{R}_k) < \infty$ to apply the previous construction

for the finite measures $m_k = m(\cdot \cap \mathtt{R}_k)$ in such a way that the sequences $\dot{q}_k = \{(z_{n,k}, \tau_{n,k}) : n \geq 1\}$ are independent for $k \geq 1$. Therefore, the sequence of counting measures $\{q_{t,k}(K) : k \geq 1\}$ is orthogonal, with $\mathbb{E}\{|q_{t,k}(K)|^2\} = tm(K \cap \mathtt{R}_k)$, and the series $q_t(K) = \sum_k q_{t,k}(K)$ is now defined as a limit in $L^2(\Omega, \mathcal{F}, P)$ satisfying $\mathbb{E}\{q_t(K)\} = 0$ and $\mathbb{E}\{|q_t(K)|^2\} = tm(K)$. Remark that we could assume given a sequence $\{N_{t,k} : k \geq 1\}$ of independent $\mathbb{R}^d$-valued compound Poisson processes with parameter m_k, but the series $\sum_k N_{t,k}$ may not be convergent.

Next, the same argument applies for the integrals, i.e., $q_t(\psi) = \sum_k q_t(\psi)$ makes sense (as a limit in the L^2-norm) for every ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$, and $\mathbb{E}\{q_t(\psi)\} = 0$, $\mathbb{E}\{|q_t(\psi)|^2\} = tm(|\psi|^2)$. However the (double) series

$$q_t(\psi) = \sum_k \Big[\sum_n \psi(z_{n,k}) \mathbb{1}_{t \geq \theta_{n,k}} - tm_k(\psi) \Big], \quad \forall \psi \in L^2(\mathbb{R}^d_*, \mathcal{B}_*, m), \quad (4.32)$$

does not necessarily reduces to a finite sum almost surely, $m(|\psi|)$ may not be finite and the pathwise analysis can not be used anymore.

Nevertheless, if we add the condition that any K in $\mathcal{K}$ is contained in a finite union of $\mathtt{R}_k$, then $q_t(K) = \sum_k q_{t,k}(K)$ does reduce to a finite sum almost surely, and we can construct the integral almost as in the case of the composed Poisson noise. This is to say that, for any K in $\mathcal{K}$, the path $t \mapsto q_t(K)$ is a piecewise constant function almost surely. Similarly, if ψ vanishes outside of a finite number of $\mathtt{R}_k$ then the series (4.32) reduces to a finite sum almost surely.

The martingale estimate[2]

$$\mathbb{E}\{\sup_{t \leq T} |q_t(\psi)|^2\} \leq 3m(|\psi|^2)\, T, \qquad \forall T > 0,$$

shows that the limit defining $q_t(\psi)$ converges uniformly on any bounded time interval $[0, T]$, and so, it is a cad-lag process. Another way is to make use of the estimate $\mathbb{E}\{|q_t(\psi) - q_s(\psi)|^2\} = m(\psi)|t - s|$ (and the property of independent increments) to select a cad-lag version.

Therefore, the (double) integral $q_t(\psi)$ is defined above as a L^2-continuous random process by means of a L^2 converging limit as in (4.32).

Actually, the random measure $\{q_t(\mathrm{d}z) : t \geq 0,\, z \in \mathbb{R}^d_*\}$ is a *centered Poisson measure* Lévy measure m, namely, for every ψ in $L^2(\mathbb{R}^d_*, \mathcal{B}_*, m)$, the integral $\{q_t(\psi) : t \geq 0\}$ is a Lévy process with characteristic $(0, 0, m_\psi)$, where m_ψ is pre-image measure of m, i.e., $m_\psi(B) = m(\psi^{-1}(B))$, for every Borel set B in $\mathbb{R}$, and the expression (4.27) of the characteristic function of $q_t(\psi)$ is valid.

Since the measure m is not necessarily finite, only if $m(\psi) < \infty$ we can add the counting process to define the integral $p_t(\psi)$ as in the case of a compound Poisson process, i.e., the (double) series

$$p_t(\psi) = \sum_{n,k} \varphi(z_{n,k}) \mathbb{1}_{t \leq \theta_{n,k}}$$

[2]Note that $\{q_t(\psi) : t \geq 0\}$ is a separable martingale, so that Doob's inequality or regularization suffices to get a cad-lag version

converges in $L^1(\Omega,\mathcal{F},P)$, but does not necessarily reduces to a finite sum almost surely. Any way, we have the equality $\mathbb{E}\{q_t(\psi)q_s(\psi')\}=(t\wedge s)m(\psi\psi')$, for every ψ and ψ' in $L^2(\mathbb{R}^d_*,\mathcal{B}_*,m)$ and any $t,s>0$.

Thus, if $m(|z|)<\infty$ then the series $\sum_k\sum_n|z_n|\mathbb{1}_{t\geq\theta_{n,k}}=\sum_k m_k(|z|)=m(|z|)$ converges, and therefore, the $\mathbb{R}^d$-valued Lévy process

$$N_t=\sum_k\sum_n z_{n,k}\mathbb{1}_{t\geq\theta_{n,k}}$$

is meaningful and $N_t=q_t(z)+tm(z)$. In general, if only $m(|z|^2\wedge|z|)<\infty$ then the $\mathbb{R}^d$-valued Lévy process $\{q_t(z):t\geq0\}$ with characteristic function (4.28) remains meaningful, and the expression (4.30) allows us to reconstruct the counting measure $\{q_t(K):K\in\mathcal{K}\}$ from the $\{q_t(z):t\geq0\}$. However, the expression of the transition density is not so immediately, for each finite measure $m_k=m(\cdot\cap\mathtt{R}_k)$ we have an explicit series but the limit in k is not so clear. Any way, for a bounded set B with $m(B)<\infty$, the transition density of $\{q_t(z\mathbb{1}_B):t\geq0\}$ is given by a series similar to (4.29).

Observe that if the measure m integrates the function $z\mapsto|z|^2$ then

$$q_t(z)=\sum_k\Big[\sum_n z_{n,k}\mathbb{1}_{t\geq\theta_{n,k}}-tm_k(z)\Big]$$

converges in L^2, and because $P(\lim_{n,k}\theta_{n,k}=\infty)=1$ and $m(\mathbb{1}_{|z|\geq\varepsilon})<\infty$, $\varepsilon>0$, the series $\sum_k\sum_n\mathbb{1}_{|z_{n,k}|\geq\varepsilon}\mathbb{1}_{t\geq\theta_{n,k}}$ is a finite sum almost surely, for every $\varepsilon>0$. Therefore, a convenient semi-ring $\mathcal{K}$ is the countable class of d-intervals $]a,b]$ with closure in $\mathbb{R}^d_*$ and with rational end points, in this way, if $m(|z|^2\wedge1)<\infty$ then $q_t(K)$, given by either (4.32) or (4.30), reduces to a finite sum almost surely, for every K in $\mathcal{K}$. Usually, an *intensity measure* m (not necessarily in $\mathbb{R}^d_*$) is associated with $\{q_t(\mathrm{d}z)\}$ (regarded as a Poisson martingale measure), whist a *Lévy measure* m (on $\mathbb{R}^d_*$), which necessary satisfies $m(|z|^2\wedge1)<\infty$, is associated with $\{q_t(z)\}$ (regarded as a $\mathbb{R}^d$-valued centered Poisson process). However, we prefer to assume $m(|z|^2)<\infty$ to obtain a $\mathbb{R}^d$-valued Lévy process $\{q_t(z):t\geq0\}$ with finite second-order moments.

If $\mathcal{K}$ is a countable semi-ring (with each K separated from $\{0\}$) generating the Borel σ-algebra in $\mathbb{R}^d_*$ then, perhaps, the system $\dot{q}_t=\{e_{n,k}(K,t):n,k\geq1,\ K\in\mathcal{K}\}$, with

$$e_{n,k}(K,t)=\big[\mathbb{1}_{z_{n,k}\in K}\mathbb{1}_{\theta_{n,k}\leq t}-\mathbb{E}\{\mathbb{1}_{z_{n,k}\in K}\}\mathbb{E}\{\mathbb{1}_{\theta_{n,k}\leq t}\}\big],$$

is the ideal expression of a Poisson noise with Lévy measure m. Similarly, if ψ in $L^2(\mathbb{R}^d_*,\mathcal{B}_*,m)$ then for every $n\geq1$, $\{e_{n,k}(\psi,t):k\geq1\}$, with

$$e_{n,k}(\psi,t)=\big[\psi(z_{n,k})\mathbb{1}_{\theta_{n,k}\leq t}-\mathbb{E}\{\psi(z_{n,k})\}\mathbb{E}\{\mathbb{1}_{\theta_{n,k}\leq t}\}\big],$$

is an orthogonal system in $L^2(\Omega,\mathcal{F},P)$, with $\mathbb{E}\{e_{n,k}(\psi,t)\}=0$ and repeating the orthogonalization of the case with finite measure m, an orthogonal system $\{\tilde{e}_{n,k}(\psi,t):n,k\geq1\}$ can be constructed. Hence, the projection operator has a form similar to (4.31). It is also clear that we can extend Remark 4.18 to this general Poisson noise.

To conclude this long introduction (of Wiener and Poisson noises or processes), let us mention that the previous arguments could be used to define a Wiener process $\{w_t : t \geq 0\}$ and a $\mathbb{R}^d$-valued (centered) Poisson process $\{q_t(z) : t \geq 0\}$ or martingale Poisson measure $\{q_t(\mathrm{d}z) : t \geq 0\}$ with Lévy measure m on $\mathbb{R}^d_*$, independent one of each other. Essentially, the arguments go as follows: a convergent orthogonal (better, independent identically distributed random variables with zero mean) series is used for the Wiener process, and a two procedure is needed for the Poisson measure, namely, an almost surely finite series (when the Lévy measure is finite) and next a convergent orthogonal series. As mentioned early, the above constructions can handle real-valued functions in $L^2(]0,\infty[)$ or $L^2(m(\mathrm{d}z) \times \mathrm{d}t)$ instead of just integrating functions constant in time (1 and ψ), and eventually random functions which are appropriate limits of a linear combination of terms like $\mathbb{1}_{]0,\tau]}$, with a bounded stopping time τ.

Summing-up, these constructions, specially the extension to random functions, are called *stochastic integrals*. The class of random functions that are *integrable* with respect to either a $\mathbb{R}^d$-valued Wiener process w or a Poisson martingale measure q with Lévy measure m in $\mathbb{R}^d_*$ are processes either $(f(t) : t \geq 0)$ or $\{g(z,t) : z \in \mathbb{R}^d_*,\, t \geq 0\}$ satisfying almost surely the integrability condition

$$\text{either} \quad \int_0^T |f(t)|^2 \mathrm{d}t < \infty \quad \text{or} \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}^d_*} |g(z,t)|^2 \pi(\mathrm{d}z) < \infty,$$

and the non-anticipative assumption, i.e., for every $t \geq 0$, either $f(t)$ or $g(z,t)$ is independent of the increments, either $\{w(s) - w(t) : s > t\}$ or $\{q_s(K) - q_t(K) : s > t,\, K \in \mathcal{K}\}$, with $\mathcal{K}$ a countable semi-ring (each K separated from $\{0\}$) generating the Borel σ-algebra in $\mathbb{R}^d_*$. This non-anticipating property with respect to the previous constructions translates into an independent condition of either $f(t)$ or $g(z,t)$ with respect to the sequence od random variables

$$\begin{aligned} &\text{either} \quad \{e_{i,n} : i = 1, \ldots, 4^n,\, n \geq 1,\, i2^{-n} > t\} \\ &\text{or} \quad \{\mathbb{1}_{z_{n,k} \in K} \mathbb{1}_{s \geq \theta_{n,k} > t} : n, k \geq 1,\, K \in \mathcal{K},\, s > t\}, \end{aligned}$$

with the notation (4.18) and (4.32). The main point of these constructions is to note that the stochastic integrals are intrinsically connected with the construction of Lévy processes. However, in what follows, the focus is on the integrands (i.e., processes that are integrable) with respect to a Lévy process.

4.2.2 Relative to Wiener processes

Let $(w(t) : t \geq 0)$ be a real-valued standard Wiener process in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, i.e., $w(t)$ and $w^2(t) - t$ are continuous martingales relative to the filtration $(\mathcal{F}_t : t \geq 0)$, with $w(0) = 0$. Denote by $\mathcal{E}$ the vector space of all processes of the form $f(t,\omega) = f_{i-1}(\omega)$ if $t_{i-1} < t \leq t_i$ with some $i = 1, \ldots, n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. it is clear

what the integral should be for any integrand in $\mathcal{E}$, namely

$$\int f(s)\mathrm{d}w(s) := \sum_{i=1}^{n} f_{i-1}[w(t_i) - w(t_{i-1})], \tag{4.33}$$

and

$$\int_{(0,t]} f(s)\mathrm{d}w(s) := \sum_{i=1}^{n} f_{i-1}[w(t \wedge t_i) - w(t \wedge t_{i-1})], \quad \forall t \geq 0,$$
$$\int_{(a,b]} f(s)\mathrm{d}w(s) := \int_{(0,b]} f(s)\mathrm{d}w(s) - \int_{(0,a]} f(s)\mathrm{d}w(s),$$

for every $b > a \geq 0$. Notice that

$$\int_{(a,b]} f(s)\mathrm{d}w(s) = \int f(s)\, \mathbb{1}_{(a,b]}(s)\mathrm{d}w(s),$$

for every $b > a \geq 0$. This definition (4.33) is independent of the particular representation used and because f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable we obtain

$$\mathbb{E}\Big\{\Big|\int f(s)\mathrm{d}w(s)\Big|^2\Big\} = \sum_{i=1}^{n} \mathbb{E}\{|f_{i-1}|^2(t_i - t_{i-1})\} = \mathbb{E}\{\int |f(s)|^2\mathrm{d}s\}, \tag{4.34}$$

for every f in $\mathcal{E}$. Moreover the processes

$$\int_{(0,t]} f(s)\mathrm{d}w(s) \quad \text{and} \quad \Big|\int_{(0,t]} f(s)\mathrm{d}w(s)\Big|^2 - \int_0^t |f(s)|^2\mathrm{d}s, \tag{4.35}$$

for every $\forall t \geq 0$, are continuous martingales, and

$$\mathbb{E}\Big\{\Big[\int f(s)\mathrm{d}w(s)\Big]\,\Big[\int g(s)\mathrm{d}w(s)\Big]\Big\} = \mathbb{E}\Big\{\int f(s)\, g(s)\mathrm{d}s\Big\}, \tag{4.36}$$

for any two stochastic processes f and g in $\mathcal{E}$. Denote by $\bar{\mathcal{E}}$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes f for which there exists a sequence $(f_1, f_2, \ldots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big\{\int |f_n(s) - f(s)|^2\mathrm{d}s\Big\} = 0.$$

Based on the martingale inequality

$$\mathbb{E}\Big\{\sup_{0\leq t\leq T}\Big|\int_{(0,t]} f(s)\mathrm{d}w(s)\Big|^2\Big\} \leq 4\,\mathbb{E}\Big\{\int_0^T |f(s)|^2\mathrm{d}s\Big\}, \tag{4.37}$$

for every $T \geq 0$, and the isometry identity (4.34), this linear operation can be extended to the closure $\bar{\mathcal{E}}$, preserving linearity and the properties (4.34), (4.35), (4.36). This is called Itô integral or generally *stochastic integral*. Besides a density argument, the estimate (4.37) is used to show that the stochastic integral on $(0, t]$ is a continuous process as a function of $t \geq 0$, for any f in $\bar{\mathcal{E}}$.

If τ and θ are stopping times with $\theta \le \tau \le T$ (with T a constant) then the process

$$\mathbb{1}_{]\!]\theta,\tau]\!]} : (\omega, t) \mapsto \mathbb{1}_{\theta(\omega)<t\le\tau(\omega)}$$

is elementary predictable process, indeed, for any partition $0 = t_0 < t_1 < \cdots < t_n$, with $t_n \ge T$ we have

$$\mathbb{1}_{]\!]\theta,\tau]\!]} = \sum_{i=1}^{n} \mathbb{1}_{[\theta\le t_{i-1}]\setminus[\tau\le t_i]}\ \mathbb{1}_{]t_{i-1},t_i]},$$

so that

$$\int \mathbb{1}_{]\!]\theta,\tau]\!]}(s)\mathrm{d}w(s) = \sum_{i=1}^{n} \mathbb{1}_{[\theta\le t_{i-1}]}\ \mathbb{1}_{[\tau\le t_i]}\ [w(t_i) - w(t_{i-1}] =$$
$$= \sum_{0\le i<j\le n} \mathbb{1}_{[\theta=t_i]}\ \mathbb{1}_{[\tau=t_j]}\ [w(\tau) - w(\theta)] = w(\tau) - w(\theta),$$

Even more general, we have the equality

$$\int_{(\theta,\tau]} c\, f(s)\mathrm{d}w(s) = c \int_{(\theta,\tau]} f(s)\mathrm{d}w(s), \tag{4.38}$$

for every bounded random variable c which is $\mathcal{F}_\theta$-measurable and any f in $\bar{\mathcal{E}}$. A way of proving (4.38) is to approximate the stopping times by finite-valued stopping times, which also show that in (4.33) we may replace the deterministic times t_i by stopping times τ_i, i.e.,

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \sum_{i=1}^{n} f_{i-1}[w(t\wedge\tau_i) - w(t\wedge\tau_{i-1})], \tag{4.39}$$

for every $t \ge 0$ and any processes of the form $f(t,\omega) = f_{i-1}(\omega)$ if $\tau_{i-1} < t \le \tau_i$ with some $i = 1, \ldots, n$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_n \le T$, with T a real number, and f_i are $\mathcal{F}(\tau_i)$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise.

Now, we may extend this stochastic integral by localizing the integrand, i.e., denote by $\bar{\mathcal{E}}_{\mathrm{loc}}$ the space of all processes f for which there is a sequence $(\tau_1 \le \tau_2 \le \cdots)$ of stopping times such that $P(\tau_n < \infty)$ converges to zero and the processes $f_k(t,\omega) := f(t,\omega)$ for $t \le \tau_k$ (with $f_k(t,\omega) := 0$ otherwise) belong to $\bar{\mathcal{E}}$. Since, almost surely we have

$$\int_{(0,t]} f_k(s)\mathrm{d}w(s) = \int_{(0,t]} f_n(s)\, dw(s), \quad \forall t \le \tau_k,\ k \le n,$$

and both processes are continuous, we can define

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \lim_n \int_{(0,t]} f_n(s)\, dw(s), \quad \forall t \ge 0,$$

in a unique way and independent of the *localizing* sequence $(\tau_1 \le \tau_2 \le \cdots)$ used. For processes in $\bar{\mathcal{E}}_{\mathrm{loc}}$ the equalities (4.34) and (4.36) are no longer meaningful,

but the processes (4.35) become continuous local martingales. A very useful estimate, similar to the martingale inequality (4.37) but adapted to the local case is the following inequality

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_{(0,t]} f(s)\mathrm{d}w(s)\Big| \ge \varepsilon\Big\} \le \frac{\delta}{\varepsilon^2} + P\Big\{ \int_0^T |f(s)|^2\mathrm{d}s \ge \delta\Big\}, \qquad (4.40)$$

for any positive numbers T, ε and δ.

It is important to remark that the stochastic integral is initially defined in a L^2 space, where an element is an equivalence class relative to the product measure $P\times \mathrm{d}\ell$, with $\mathrm{d}\ell$ the Lebesgue measure on the semi-line $[0,\infty)$. For the sake of simplicity, we write $\Omega\times[0,\infty)$ or $[0,\infty)\times\Omega$ indistinctly as long as no confusion may arrive, i.e., processes are written $f(t,\omega)$ or $f(\omega,t)$. Next, by means of martingale properties we can select a *good* version to make the processes (4.35) continuous (local) martingales. By a simple argument of monotone classes, we deduce that $\bar{\mathcal{E}}$ contains the Hilbert space $L^2(\Omega\times[0,\infty),\mathcal{P},P\times\mathrm{d}\ell)$. On the other hand, it is also clear that any stochastic process in $\bar{\mathcal{E}}_{\rm loc}$ is measurable relative to the σ-algebra $\bar{\mathcal{P}}$, generated by $\mathcal{P}$ and all $P\times\mathrm{d}\ell$-null subsets of $\Omega\times[0,\infty)$. As mentioned above, all concepts (in particular the stochastic integral) are up to or except to an evanescent set. However, the stochastic integral is defined up to a $P\times\mathrm{d}\ell$-null subset of $\Omega\times[0,\infty)$, and then a good version is chosen. Thus, the next question is what processes are in $\bar{\mathcal{E}}$ or $\bar{\mathcal{E}}_{\rm loc}$ besides those that are predictable, i.e., what can be said about completion σ-algebra $\bar{\mathcal{P}}$ of the predictable σ-algebra $\mathcal{P}$.

Adapted, Predictable and Other Properties

Following Doob [59] we can prove that any adapted square integrable process $f(t,\omega)$ is in $\bar{\mathcal{E}}$. First assume f is bounded and vanishes for t outside of a bounded interval, then we partition the real line $\mathbb{R}$ into intervals $(k\varepsilon,(k+1)\varepsilon]$ with $k=0,\pm1,\pm2,\dots$, $\varepsilon>0$, and we define $f_{\varepsilon,s}(t,\omega) := f(\alpha_\varepsilon(t-s)+s,\omega)$, where $\alpha_\varepsilon(r) := k\varepsilon$ for any r in the subinterval $(k\varepsilon,(k+1)\varepsilon]$, where f has been extended for $t\le 0$. The restriction to $[0,\infty)$ of the process $f_{\varepsilon,s}$ belongs to $\mathcal{E}$ for any $\varepsilon>0$ and s in $\mathbb{R}$. We claim that there exist a sequence $(\varepsilon_1>\varepsilon_2>\cdots)$ and some s such that

$$\lim_{n\to\infty} \mathbb{E}\Big\{ \int |f_{\varepsilon_n,s}(t,\omega)-f(t,\omega)|^2\mathrm{d}t\Big\} = 0.$$

Indeed, the continuity of the translation in $\mathbb{R}$ with respect to the Lebesgue measure and the fact that $\alpha_\varepsilon(r)-r\to 0$ as $\varepsilon\to 0$ show that

$$\lim_{\varepsilon\to 0} \int |f(\alpha_\varepsilon(t)+s,\omega)-f(t+s,\omega)|^2\mathrm{d}s = 0, \quad \forall t,\omega.$$

Since all processes considered are bounded and vanish outside of a fixed finite interval, we have

$$\lim_{\varepsilon\to 0} \int \mathbb{E}\Big\{ \int \big|f(\alpha_\varepsilon(t)+s,\omega)-f(t+s,\omega)\big|^2\mathrm{d}s\Big\}\mathrm{d}t = 0.$$

Fubini's Theorem allows us to exchange the integration order of the variables s and t, proving the claim. Finally, for the general case, we define $f_n(t,\omega) := f(t,\omega)$ if $0 \le t \le n$ and $|f(t,\omega)| \le n$, and $f_n(t,\omega) := 0$ otherwise. Applying the previous approximation to f_n we complete the proof. Thus any measurable adapted process is $\bar{\mathcal{P}}$ measurable.

It is easy to check that any progressively measurable process such that

$$P\{\int_0^t |f(s,\omega)|^2 \mathrm{d}s < \infty\} = 1, \quad \forall t \ge 0 \tag{4.41}$$

belongs to $\bar{\mathcal{E}}_{\mathrm{loc}}$. Indeed, the expression

$$\tau_n := \inf\{t \ge 0 : \int_0^t |f(t,\omega)|^2 \mathrm{d}s \ge n\}$$

define a localizing sequence of stopping times for the process f, which proves the claim. However, when f is only a measurable adapted process, τ_n may not be a stopping time. In this case, we can always approximate f by truncation, i.e, $f_n(t,\omega) := f(t,\omega)$ if $|f(t,\omega)| \le n$ and $f_n(t,\omega) := 0$ otherwise, so that

$$\lim_n P\{\int_0^T |f_n(t,\omega) - f(t,\omega)|^2 \mathrm{d}s \ge \delta\} = 0, \quad \forall T, \delta \ge 0.$$

Since f_n belongs to $\bar{\mathcal{E}}$, for every $n \ge 1$, the estimate (4.40) proves also that $\bar{\mathcal{E}}_{\mathrm{loc}}$ contains all measurable adapted processes satisfying (4.41). If f is a cad-lag adapted process then $t \mapsto f(t-)$ and f are progressively measurable, condition (4.41) is satisfied and

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \int_{(0,t]} f(s-)\mathrm{d}w(s), \quad \forall t > 0.$$

Moreover, let $0 = \tau_0^n \le \tau_1^n < \cdots < \tau_k^n < \cdots$ be a sequence of stopping times such that $P\{\sup_k \tau_k^n < \infty\} \to 0$ and $P\{\sup_k(\tau_k^n - \tau_{k-1}^n > \delta\} \to 0$, for any $\delta > 0$, as $n \to \infty$. Then, for any given adapted stochastically left continuous process f, we define the sequence of simple predictable processes $f_{n,m}(t,\omega) := f(\tau_k^n,\omega)$ if $|f(\tau_k^n,\omega)| \le m$ and $\tau_k^n < t \le \tau_{k+1}^n$, $k = 0,1,2,\ldots$, and $f_{n,m}(t,\omega) := 0$ otherwise. It is clear that

$$\lim_n P\{|f_{n,m}(t,\omega) - f_m(t,\omega)| \ge \delta\} = 0, \quad \forall t, \delta, m > 0,$$

where $f_m(t,\omega) := f(t,\omega)$ if $|f(t,\omega)| \le m$ and $f_m(t,\omega) := 0$ otherwise. Because $|f_{m,n}|$ is bounded by m, this yields

$$\lim_n P\Big\{\int_0^T |f_{n,m}(t,\omega) - f_m(t,\omega)|^2 \mathrm{d}t \ge \delta\Big\} = 0, \quad \forall T, \delta, m > 0.$$

Hence, by means of (4.40)

$$\lim_n P\Big\{\sup_{0\le t\le T} \Big|\int_{]0,t]} [f_{n,m}(t) - f_m(t)]\mathrm{d}w(t)\Big| \ge \varepsilon\Big\} = 0,$$

for every $T, \varepsilon, m > 0$. Thus, for each $t, m > 0$, the expression

$$\int_{]0,t]} f_{n,m}(s)\mathrm{d}w(s) = \sum_{k=0}^{\infty} f_m(\tau_k^n, \omega)\,[w(t \wedge \tau_{k+1}^n, \omega) - w(t \wedge \tau_k^n, \omega)],$$

for every $t > 0$, is an approximation of the stochastic integral provided f satisfies (4.41). Recall that $f_m(t,\omega) = f(t,\omega)$ if $|f(t,\omega)| \le m$, so that f_m converges to f almost surely in L^2. It is clear that condition (4.41) is satisfied when f is cad-lag. A typical case is when $\tau_k^n := k2^{-n}$. Thus, if the equivalence class, containing an element f in $\bar{\mathcal{E}}_{\mathrm{loc}}$, contains a predictable element (in the previous case $f(t-)$ for any $t > 0$) then we write the stochastic integral with the predictable representative of its equivalence class.

It can be proved, see Dellacherie and Meyer [58, Theorem VIII.1.23, pp. 346-346] that for any f in $\bar{\mathcal{E}}_{\mathrm{loc}}$ we have

$$\begin{cases} \text{if } \; f(s,\omega) = 0, \; \forall (s,\omega) \in]a,b] \times F, \; F \in \mathcal{F} \\ \text{then } \; \displaystyle\int_{(a,b]} f(s)\mathrm{d}w(s) = 0 \;\; \text{a.s. on } \; F. \end{cases} \tag{4.42}$$

This expresses the fact that even if the construction of the stochastic integral is not *pathwise*, it retains some local character in Ω.

From the definition it follows that if f is a cad-lag adapted process with locally bounded variation then

$$\int_{(0,t]} f(s)\mathrm{d}w(s) = \int_{(0,t]} f(s-)\mathrm{d}w(s) = f(t)w(t) - \int_{(0,t]} w(s)\mathrm{d}f(s),$$

where the last integral is in the Riemann-Stieltjes or Lebesgue-Stieltjes sense. However, the Wiener process w has unbounded local variation. Let $\varpi := (0 = t_0 < t_1 < \cdots < t_n = t)$ be a partition of $[0,t]$, with $\|\varpi\| := \max_i(t_i - t_{i-1})$ and consider the Riemann sums

$$S_\varpi := \sum_{i=1}^{n} w(t_i^*)[w(t_i) - w(t_{i-1})], \quad \text{with} \quad t_{i-1} \le t_i^* \le t_i,$$

which can be rewritten as

$$S_\varpi = \frac{w^2(t)}{2} - \frac{1}{2}\sum_{i=1}^{n}[w(t_i) - w(t_{i-1})]^2 + \sum_{i=1}^{n}[w(t_i^*) - w(t_{i-1})]^2 + \\ + \sum_{i=1}^{n}[w(t_i) - w(t_i^*)][w(t_i^*) - w(t_{i-1})].$$

Since

$$\mathbb{E}\Big\{\sum_{i=1}^{n}[w(t_i) - w(t_{i-1})]^2\Big\} = \sum_{i=1}^{n}[t_i - t_{i-1}] = t,$$

$$\mathbb{E}\big\{\big[\sum_{i=1}^{n}[w(t_i^*)-w(t_{i-1})]^2-\sum_{i=1}^{n}(t_i^*-t_{i-1})\big]^2\big\}=$$
$$=\mathbb{E}\big\{\sum_{i=1}^{n}[w(t_i^*)-w(t_{i-1})]^4\big\}-\sum_{i=1}^{n}(t_i^*-t_{i-1})^2\le 2t\|\varpi\|,$$

and

$$\mathbb{E}\big\{\big[\sum_{i=1}^{n}[w(t_i)-w(t_i^*)][w(t_i^*)-w(t_{i-1})]\big]^2\big\}=$$
$$=\sum_{i=1}^{n}(t_i-t_i^*)(t_i^*-t_{i-1})\le t\|\varpi\|,$$

we deduce that

$$\lim_{\|\pi\|\to 0}\big[S_\varpi-\sum_{i=1}^{n}(t_i^*-t_{i-1})\big]=\frac{w^2(t)}{2}-\frac{t}{2},$$

in the L^2-sense. In the Itô integral, $t_i^*=t_{i-1}$ so that

$$\int_{(0,t]}w(s)\mathrm{d}w(s)=\frac{w^2(t)}{2}-\frac{t}{2},\quad\forall t\ge 0.$$

However, any choice $t_i^*=(1-r)t_{i-1}+rt_i$, with $0\le r\le 1$, could be possible. In particular Fisk-Stratonovich integral, where $r=1/2$, $t_i^*=(t_{i-1}+t_i)/2$, yields a symmetric calculus, very useful in some physical and mechanical models. However, Itô integral, i.e., the choice $r=1$, $t_i^*=t_{i-1}$, is more oriented to control models, where the adapted (or predictable) character (i.e., non-interaction with the future) is an essential property. Moreover, the martingale property is preserved.

Working by coordinates, this stochastic integral can be extended to a $\mathbb{R}^d$-valued Wiener process and $n\times d$ matrix-valued predictable processes.

4.2.3 Relative to Poisson measures

Let $\{p(t):t\ge 0\}$ be a real-valued compound Poisson process with parameters (c,ν), where $c>0$ and γ is a distribution in $\mathbb{R}_*^d:=\mathbb{R}^d\setminus\{0\}$, in a given filtered space $(\Omega,\mathcal{F},P,\mathcal{F}_t:t\ge 0)$. This means that

$$p(t,\omega)=\begin{cases}0 & \text{if } t<\theta_1(\omega),\\ Z_n(\omega) & \text{if } \theta_n(\omega)\le t<\theta_{n+1}(\omega),\end{cases}$$

where $\theta_n:=\tau_1+\tau_2+\cdots+\tau_n$, $\{\tau_n:n=1,2,\dots\}$ is a sequence of independent exponentially distributed (with parameter c) random variables, $Z_n:=\zeta_1+\zeta_2+\cdots+\zeta_n$, $\{\zeta_n:n=1,2,\dots\}$ is a sequence of independent identically distributed (with distribution law γ) random variables, independent of the sequence $\tau_1,\tau_2\dots$. In particular, if γ is δ_1, the Dirac measure at $z=1$ then

$Z_n = n$, the case of a standard Poisson process. Notice that $p(t) - ct$ and $p^2(t) - ct$ are martingales relative to the filtration $(\mathcal{F}_t : t \geq 0)$, with $p(0) = 0$. Since the function $t \mapsto p(t, \omega)$ is cad-lag, piecewise constant and with bounded variation for any ω, the integral with respect to $p(t)$ is covered by the measure theory, i.e., a pathwise integration. For a bounded left-continuous process $f(t, \omega)$ we can define

$$\left\{ \int_{(0,t]} f(s,\omega)\mathrm{d}p(s,\omega) := \sum_{n=1}^{\infty} f(\theta_n(\omega),\omega)\, \mathbb{1}_{\theta_n(\omega)\leq t} = \right. \\ \left. = \sum_{n=1}^{N(t,\omega)} f(\theta_n(\omega),\omega), \right. \tag{4.43}$$

for each ω, where $N(t, \omega) := n$ if $\theta_n(\omega) \leq t < \theta_{n+1}(\omega)$, i.e., a standard Poisson process. Because $t \mapsto \mathbb{E}\{p(t)\}$ is continuous, we have

$$\int_{(0,t]} f(s,\omega)\mathrm{d}\mathbb{E}\{p(s,\cdot)\} = \int_{(0,t]} f(s+,\omega)\mathrm{d}\mathbb{E}\{p(s,\cdot)\},$$

but

$$\int_{(0,t]} p(s-,\omega)\mathrm{d}p(s,\omega) = \sum_{n=1}^{\infty} p(\theta_n(\omega)-,\omega)\, \mathbb{1}_{\theta_n(\omega)\leq t} = \\ = \sum_{k=1}^{N(t,\omega)} Z_{k-1}(\omega)\, \zeta_k(\omega),$$

and

$$\int_{(0,t]} p(s,\omega)\mathrm{d}p(s,\omega) = \sum_{n=1}^{\infty} p(\theta_n(\omega),\omega)\, \mathbb{1}_{\theta_n(\omega)\leq t} = \sum_{k=1}^{N(t,\omega)} Z_k(\omega)\, \zeta_k(\omega).$$

Thus, for a given compound Poisson process $p(t)$ as above and a left-continuous (or only predictable) process $f(t)$ (without begin locally integrable), we can use (4.43) to define the *stochastic integral*, which is just a *pathwise* sum (integral) in this case, with is a jump process similar to the compound Poisson process. Similar arguments apply to the centered compound Poisson process $t \mapsto \big(p(t) - \mathbb{E}\{p(t)\}\big)$, and the integral is the difference of random pathwise integral and a deterministic integral.

Next step is to consider a standard Poisson measure $\{p(\cdot, t) : t \geq 0\}$ with Lévy (intensity) measure $\pi(\cdot)$ in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, i.e., (a) $\pi(\cdot)$ is a Radon measure on $\mathbb{R}^m_* := \mathbb{R}^m \setminus \{0\}$, i.e., $\pi(K) < \infty$ for any compact subset K of $\mathbb{R}^m_*$; (b) $\{p(B, t) : t \geq 0\}$ is a Poisson process with parameter $\pi(B)$, for any Borel subset B in $\mathbb{R}^d_*$ with $\pi(B) < \infty$ (here $p(B, t) := 0$ if $\pi(B) = 0$); (c) the Poisson processes $p(\cdot, B_1), p(\cdot, B_2), \ldots, p(\cdot, B_n)$ are independent if $B_1, B_2, \ldots, B_n$ are disjoint Borel set in $\mathbb{R}^m_*$ with $\pi(B_i) < \infty$, $i = 1, \ldots, n$.

Given a Radon measure π in $\mathbb{R}^m_*$ (which integrates the function $|z|^2 \wedge 1$, so that it can be called a Lévy measure), we write $\pi = \sum_k \pi_k$, where $\pi_k(B) :=$

$\pi(B \cap R_k)$, $\mathbb{R}^m_* = \bigcup_k R_k$, $\pi(R_k) < \infty$ and $R_k \cap R_\ell = \emptyset$ if $k \neq \ell$. To each π_k we may associate a compound Poisson process and a point process by the expressions

$$Y_k(t) := \sum_{n=1}^{\infty} \zeta_{n,k} \mathbb{1}_{t \geq \theta_{n,k}},$$
$$\delta Y_k(t) := Y_k(t) - Y_k(t-) = \zeta_{n,k} \mathbb{1}_{t=\theta_{n,k}}, \quad \forall t \geq 0,$$

where $\theta_{n,k} := \tau_{1,k} + \tau_{2,k} + \cdots + \tau_{n,k}$, $\{\tau_{n,k} : n = 1, 2, \dots\}$ is a sequence of independent exponentially distributed (with parameter $\pi(R_k) = c_k$) random variables, and $\{\zeta_{n,k} : n = 1, 2, \dots\}$ is another sequence of independent identically distributed (with distribution law π_k/c_k) random variables, and the two sequences $\{\tau_{n,h} : n, k \geq 1\}$, $\{\zeta_{n,k} : n, k \geq 1\}$ are independent. The jump process $\delta Y = \sum_k \delta Y_k$ is indeed a Poisson point process with characteristic measure π, i.e., with $Z_{n,k} := \zeta_{1,k} + \zeta_{2,k} + \cdots + \zeta_{n,k}$,

$$p(B \times]s,t]) := \sum_{n,k=1}^{\infty} \mathbb{1}_{s < \theta_{n,k} \leq t} \, \mathbb{1}_{Z_{n,k} \in B}, \quad \forall t > s \geq 0, \; B \in \mathcal{B}(\mathbb{R}^m_*),$$

is a standard Poisson random measure with intensity measure

$$\mathbb{E}\{p(B \times]s,t])\} = (t-s)\pi(B).$$

In general, we cannot arrange the jumps in the increasing order like the case of a compound Poisson process, because there may occur accumulation of small jumps. With any of the notation $p(B,t)$ or $p(B \times]0,t])$ or $p(B,]0,t])$ the integer-valued random measure p (see Section 4.1.3) is also called a standard Poisson random measure. From the process viewpoint, $p(B,]s,t])$ is defined as the (finite) number of jumps (of a cad-lag process Y) belonging to B within the interval $]s,t]$. Note that the predictable compensator of the optional random measure $p(\cdot,t)$ is the deterministic process πt. Thus, for a predictable process of the form $F(z,t,\omega) = f(t,\omega)\,\mathbb{1}_{z \in B}$ the expression

$$\int_{\mathbb{R}_k \times]0,t]} F(z,s,\omega)\, p(\mathrm{d}z, \mathrm{d}s) = \sum_{n=1}^{\infty} f(\theta_{n,k}(\omega), \omega) \mathbb{1}_{0 < \theta_{n,k}(\omega) \leq t} \, \mathbb{1}_{Z_{n,k}(\omega) \in B}$$

is indeed a finite stochastic pathwise sum (as previously). However, the passage to the limit in k is far more delicate and requires more details.

With the above introduction, let ν be an integer-valued random measure, which is a Poisson measure as in Definition 4.9, with Lévy measure $\Pi(B \times]s,t]) = \mathbb{E}\{\nu(B \times]s,t])\}$, $\Pi(\mathbb{R}^m_* \times \{t\}) = 0$ for every $t \geq 0$, and local martingale measure $\tilde{\nu} := \nu - \Pi$, in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$. In particular, a standard Poisson measure $\{p(\cdot,t) : t \geq 0\}$ with Lévy (characteristic or intensity) measure $\pi(\cdot)$, and $\Pi(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z) \times \mathrm{d}t$. Note that we reserve the notation p for a standard Poisson measure. Denote by $\mathcal{E}$ the vector space of all processes of the form $f(z,t,\omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \leq t_i$ and z belongs to K_j with some $i = 1, \dots, n$, and $j = 1, \dots, m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real

numbers, K_j are disjoint sets with compact closure in $\mathbb{R}^m_*$ and $f_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. It is clear what the integral should be for any integrand in $\mathcal{E}$, namely

$$\begin{cases} \displaystyle\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) := \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i-1,j}\,\tilde\nu(K_j\times]t_{i-1},t_i]), \\ \displaystyle\int_{\mathbb{R}^m_*\times(a,b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) := \int f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s), \end{cases} \tag{4.44}$$

for every $b > a \geq 0$. Notice that

$$\int_{\mathbb{R}^m_*\times(0,\infty)} f(z,s)\,\mathbb{1}_{(0,t]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i-1,j}\,\tilde\nu(K_j\times]t\wedge t_{i-1}, t\wedge t_i])$$

and

$$\int f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) = \\ = \int_{\mathbb{R}^m_*\times(0,b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) - \int_{\mathbb{R}^m_*\times(0,a]} f(z,s)\,\tilde\nu(\mathrm{d}s,\mathrm{d}z),$$

for every $t > 0$.

If ν is a standard (or homogeneous) Poisson measure, i.e., $\mathbb{E}\{\nu(B\times(]s,t])\} = (t-s)\pi(B)$, then $p(K,t) := \nu(K\times]0,t])$ is a Poisson process with parameter $\pi(K)$, then for any left-continuous adapted process of the form $f(z,t,\omega) = f_j(t,\omega)$ when z belongs to K_j, we can calculate the stochastic integral, namely,

$$\int_{\mathbb{R}^m_*\times(0,t]} \sum_{j=1}^{m} f_i(s)\,\mathbb{1}_{K_j}(z)\,\nu(\mathrm{d}z,\mathrm{d}s) := \sum_{j=1}^{m}\sum_{k=1}^{p(t,K_j,\omega)} f_j(\theta_k(\omega,K_j),\omega),$$

for every $t \geq 0$, where $\theta_k(\omega,K_j)$ is the time of the k jumps of the Poisson process $t \mapsto p(K_j,t)$. In the case of a compound-Poisson process as above, we may forget about the K dependency, and make the previous pathwise definition, both concepts agree. In general, from $\nu = \tilde\nu + \Pi$, with $\Pi = \pi\times\mathrm{d}t$, we can define the stochastic integral relative to an integer-valued random measure ν.

This definition is independent of the particular representation used. Since for any K_1 disjoint of K_2 and any $t \geq 0$ the random variables $p(K_1,t)$ and $p(K_2,t)$ are orthogonal, and because f_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable we obtain

$$\mathbb{E}\Big\{\Big|\int_{\mathbb{R}^m_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} = \mathbb{E}\Big\{\int_{\mathbb{R}^m_*\times(0,t]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\Big\}, \tag{4.45}$$

for every f in $\mathcal{E}$. Moreover the processes

$$\begin{cases} \displaystyle\int_{\mathbb{R}^m_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) \quad \text{and} \\ \displaystyle\Big|\int_{\mathbb{R}^d_*\times(0,t]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s)\Big|^2 - \int_{\mathbb{R}^m_*\times(0,t]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s), \end{cases} \tag{4.46}$$

with $t \geq 0$ are cad-lag (quasi-left continuous) martingales, and

$$\left\{ \begin{aligned} &\mathbb{E}\Big\{\Big[\int_{\mathbb{R}^m_* \times (0,\infty)} f(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big] \times \\ &\qquad \times \Big[\int_{\mathbb{R}^m_* \times (0,\infty)} g(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big]\Big\} = \\ &\qquad = \mathbb{E}\Big\{\int_{\mathbb{R}^m_* \times (0,\infty)} f(z,s)\,g(z,s)\,\Pi(\mathrm{d}z,\mathrm{d}s)\Big\}, \end{aligned} \right. \tag{4.47}$$

for any two stochastic processes f and g in $\mathcal{E}$. Denote by $\bar{\mathcal{E}}_\Pi$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes f for which there exists a sequence $(f_1, f_2, \dots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty} \mathbb{E}\Big\{\int_{(0,\infty)\times\mathbb{R}^m_*} |f_n(z,s) - f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\Big\} = 0.$$

As in the previous section, the martingale inequality

$$\left\{ \begin{aligned} &\mathbb{E}\Big\{\sup_{0\leq t\leq T}\Big|\int_{\mathbb{R}^d_* \times (0,t]} f(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} \leq \\ &\qquad \leq 4\,\mathbb{E}\Big\{\int_{\mathbb{R}^d_* \times (0,T]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s)\Big\}, \end{aligned} \right. \tag{4.48}$$

holds for every $T \geq 0$, and also the isometry identity (4.45). Hence, this linear operation can be extended to the closure $\bar{\mathcal{E}}_\Pi$, preserving linearity and the properties (4.45), (4.46), (4.47). This is called Itô integral or generally *stochastic integral*, with respect to a Poisson measure. Next, by localizing the integrand, this definition is extended to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$, the space of all processes f for which there is a sequence $(\tau_1 \leq \tau_2 \leq \cdots)$ of stopping times such that $P(\tau_n < \infty)$ converges to zero and the processes $f_k(t,\omega) := f(t,\omega)$ for $t \leq \tau_k$ (with $f_k(t,\omega) := 0$ otherwise) belong to $\bar{\mathcal{E}}_\Pi$. As in the case of the Wiener process, a key role is played by the following inequality

$$\left\{ \begin{aligned} &P\Big\{\sup_{0\leq t\leq T}\Big|\int_{\mathbb{R}^m_* \times (0,t]} f(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big| \geq \varepsilon\Big\} \leq \frac{\delta}{\varepsilon^2} + \\ &\qquad + P\Big\{\int_{\mathbb{R}^m_* \times (0,T]} |f(z,s)|^2\,\Pi(\mathrm{d}z,\mathrm{d}s) \geq \delta\Big\}, \end{aligned} \right. \tag{4.49}$$

for any positive numbers T, ε and δ.

The class of processes that we can integrate are those in $\bar{\mathcal{E}}_\Pi$ or more general in $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$, but the stochastic integral is initially defined in a L^2 space, where an element is an equivalence class relative to the product measure $P \times \Pi$, with $\Pi = \Pi(\mathrm{d}z,\mathrm{d}s)$ the Lévy measure on $\mathbb{R}^m_* \times [0,\infty)$. Again, for the sake of simplicity, we write $\Omega \times \mathbb{R}^m_* \times [0,\infty)$ or $\mathbb{R}^m_* \times]0,\infty) \times \Omega$ or $]0,\infty) \times \mathbb{R}^m_* \times \Omega$ indistinctly as long as no confusion may arrive, i.e., processes are written $f(\omega,t,z)$ or $f(z,\omega,t)$ or $f(t,z,\omega)$. Next, by means of martingale properties we can select a *good* version to make the process (4.46) a cad-lag (local) martingale. By a simple argument of monotone classes, we deduce that (as in the case of the Wiener process) the

closure $\bar{\mathcal{E}}_\Pi$ (of all elementary processes in $\mathbb{R}^d_* \times [0,\infty)$) contains the Hilbert space $L^2(\mathbb{R}^d_* \times [0,\infty) \times \Omega, \mathcal{B} \times \mathcal{P}, \Pi \times P)$, $\Pi = \Pi(\mathrm{d}z, \mathrm{d}s)$.

On the other hand, it is also clear that any stochastic process in $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$ is measurable relative to the σ-algebra $\overline{\mathcal{B} \times \mathcal{P}}$, generated by $\mathcal{B} \times \mathcal{P}$ and all $\Pi \times P$-null subsets of $\mathbb{R}^d_* \times [0,\infty) \times \Omega$. Again, we notice that the value at time 0 is irrelevant.

Lévy and Point Processes Comments

If the Lévy measure is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}\ell$ on $[0,\infty)$, i.e., $\Pi(\mathrm{d}z, \mathrm{d}s) = \pi(\mathrm{d}z) \times \mathrm{d}\ell$, then (as in the case of the Wiener process) any measurable adapted process $f(z,s,\omega)$ is equivalent to a $\mathcal{B} \times \mathcal{P}$-measurable process, so it belongs to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$ whenever

$$P\Big\{ \int_{\mathbb{R}^d_* \times (0,T]} |f(z,s)|^2 \, \Pi(\mathrm{d}z, \mathrm{d}s) < \infty \Big\} = 1, \quad \forall\, T > 0 \tag{4.50}$$

is satisfied. This holds for standard Poisson measures.

Because the Lévy measure does not charge on $\mathbb{R}^d_* \times \{t\}$, for every $t \geq 0$, see Theorem 4.11, the stochastic integral is a cad-lag quasi-left continuous and the argument developed for Wiener processes applies proving that any progressively measurable process satisfying (4.50) belongs to $\bar{\mathcal{E}}_{\Pi,\mathrm{loc}}$.

The above stochastic integral can be constructed also for an *extended Poisson measure* (see Jacod and Shirayaev [117, Definition 1.20, Chapter 2, p. 70]), where $\Pi(\mathbb{R}^d_* \times \{t\})$ may not vanish for some $t > 0$. Actually, the stochastic integral can be constructed for any *orthogonal* measures, see Definition 3.33 in Chapter 3.

On the other hand, a (homogeneous) Poisson measure $p(\mathrm{d}z, \mathrm{d}s)$ with Lévy measure π always satisfies $p(\mathbb{R}^m_*, \{0\}) = 0$ and can be approximated by another Poisson measure $p_\varepsilon(\mathrm{d}z, \mathrm{d}s)$ with Lévy measure $\pi_\varepsilon = \mathbb{1}_{K_\varepsilon}\pi$, where the support $K_\varepsilon = \{0 < \varepsilon \leq |z| \leq 1/\varepsilon\}$ of π_ε is a compact on $\mathbb{R}^m_*$, i.e., all jumps smaller than ε or larger than $1/\varepsilon$ have been eliminated. The integer measure p_ε is associated with a compound Poisson process and has a finite (random) number of jumps, i.e., for any $t > 0$ there is an integer $N = N(t,\omega)$, points $z_i = z_i(t,\omega)$ in K_ε for $i = 1, \ldots, N$ and positive reals $\theta_i = \theta_i(t,\omega)$, $i = 1, \ldots, N$ such that $p(B,]a,b], \omega) = \sum_{n=1}^N \mathbb{1}_{z_i \in B} \mathbb{1}_{a < \theta_i \leq b}$, for every $B \in \mathcal{B}(\mathbb{R}^m_*)$, $0 \leq a < b \leq t$. In this case, the forward stochastic integral can be written as

$$\int_{\mathbb{R}^m_* \times (0,t]} f(z,s)\, \tilde{p}(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^N f(z_i, \theta_i -) - \int_0^t \mathrm{d}s \int_K f(z,s) \pi(\mathrm{d}z), \tag{4.51}$$

for any adapted cad-lag process $f(z,s)$, continuous in z.

Alternatively, we may regard the integer measure ν as a point process, i.e.,

$$\nu(B,]a,b]) = \sum_{i=1}^\infty \mathbb{1}_{\{p_i \in B\}} \mathbb{1}_{\{a < \tau_i \leq b\}}$$

to consider the pathwise integrals

$$\int_{\mathbb{R}^m\times]0,t]} f(z,s)\,\nu(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^{\infty} f(p_i,\tau_i)\mathbb{1}_{0<\tau_i\le t}$$

and $$\int_{\mathbb{R}^m\times]0,t]} f(z,s)\,\Pi(\mathrm{d}z,\mathrm{d}s),$$

defined for integrable (with respect to ν and Π) processes f. Later, a martingale argument allows the extension to square-integrable with respect Π, e.g., see Ikeda and Watanabe [110, Chapter 2]. Both approaches are equivalent and the expression (4.51) remains valid for f integrable with respect to ν and square-integrable with respect to Π.

It should be clear that the starting point is an integer-valued random measure ν (see Definition 4.6) which yields a compensated local martingale measure $\tilde{\nu} := \nu - \nu^p$, where ν^p is the (unique dual) predictable projection of ν (see Theorem 4.7 and Definition 3.3.5 of the previous chapter). Recall that a local martingale M is called *purely discontinuous* if $M(0) = 0$ and the product $M\,N$ is a local martingale for any continuous local martingale N. Stochastic integrals with respect to a compensated local martingale measure $\tilde{\nu}$ are purely discontinuous local martingales. Also, given an optional locally integrable process X with $X(0) = 0$ there exists a unique predictable projection ${}^pX$, i.e. a predictable locally integrable process such that $\mathbb{E}\{{}^pX\,\mathbb{1}_{\tau<\infty}\} = \mathbb{E}\{X\,\mathbb{1}_{\tau<\infty}\}$ for any predictable stopping time τ, such that $t \mapsto \mathbb{1}_{\tau\le t}$ is a predictable process. In particular (e.g., Jacod and Shirayaev [117, Theorem 2.28, Corallary 2.31, Chapter 1, p. 23–24]) for a local martingale M we have ${}^pM(t) = M(t-)$ and $\delta M(t) = 0$ for every $t > 0$.

• *Remark* 4.19. Let $p(\mathrm{d}z,\mathrm{d}s)$ be a Poisson measure with Lévy measure given by $\Pi(\mathrm{d}z,\mathrm{d}s) = \pi(\mathrm{d}z,s)\mathrm{d}s$ in $\mathbb{R}^m_* \times [0,\infty)$ with $\Pi(\mathbb{R}^m_*,\{0\}) = 0$ and let γ be a Borel function from $\mathbb{R}^m_* \times [0,\infty)$ into $\mathbb{R}^d$ square-integrable with respect to Π on any set of the form $\mathbb{R}^m_* \times (0,T]$, for any constant $T > 0$, and cad-lag in $[0,\infty)$. The Poisson measure p can be viewed as a Poisson point process in $\mathbb{R}^m_*$, i.e.,

$$p(B,]a,b]) = \sum_{i=1}^{\infty} \mathbb{1}_{\{p_i\in B\}}\,\mathbb{1}_{\{a<\tau_i\le b\}},$$

where the masses $\{p_i\}$ are in $\mathbb{R}^m_*$ and $\{\tau_i\}$ are stopping times (non necessary non-decreasing in i). Then we may define the stochastic integral

$$I(t,\tilde{p}) = \int_{\mathbb{R}^m_*\times(0,t]} \gamma(z,s)\,\tilde{p}(\mathrm{d}z,\mathrm{d}s),$$

which has a jump only at $t = \tau_i$ if $\gamma(p_i,\tau_i-) \neq 0$ for some i. If $z \mapsto \gamma(z,\cdot)$ is integrable with respect to p and Π (e.g., bounded, continuous in z and vanishing near $z = 0$) then

$$I(t,\gamma,\tilde{p}) = \sum_{i=1}^{\infty} \gamma(p_i,\tau_i-)\,\mathbb{1}_{\{0<\tau_i\le t\}} - \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma(z,s)\,\pi(\mathrm{d}z,s)\mathrm{d}s,$$

which is a pathwise integral. The integer measure p_γ associate with the martingale $t \mapsto I(t, \gamma, \tilde{p})$ satisfies

$$p_\gamma(B,]a, b]) = \sum_{i=1}^{\infty} \mathbb{1}_{\{\gamma(p_i, \tau_i -) \in B\}} \mathbb{1}_{\{a < \tau_i \leq b\}},$$

which is a Poisson measure with

$$\pi_\gamma(B, s) = \pi\big(\{(z, s) \in \mathbb{R}^m_* \times [0, \infty) : \gamma(z, s) \in B\}, s\big)$$

and $\Pi_\gamma(\mathrm{d}z, \mathrm{d}s) = \pi_\gamma(\mathrm{d}z, s)\mathrm{d}s$ as its Lévy measure on $\mathbb{R}^d_0 \times [0, \infty)$. □

Recall that δ denotes the jumps operator $\delta X(t) = X(t) - X(t-)$, the jumps of a local martingale have the following structure,

Theorem 4.20 (jump structure)**.** *Let X be an optional locally integrable process with $X(0) = 0$. Then there exists a (unique purely discontinuous) local martingale M such that δM and X are indistinguishable (i.e., except on a set of measure zero we have $\delta M(t) = X(t)$, for every $t \geq 0$) if and only if the predictable projection ${}^pX = 0$ and the increasing process $t \mapsto \sqrt{\sum_{s \leq t} |X(s)|^2}$ is (locally) integrable. Moreover, M is a (locally) square integrable martingale if and only if $t \mapsto \sum_{s \leq t} |X(s)|^2$ is (locally) integrable and M is a local martingale with (locally) bounded variation paths if and only if $t \mapsto \sum_{s \leq t} |X(s)|$ is (locally) integrable.*

Proof. One part of the argument goes as follows. (1) First, if X is locally square integrable predictable process with ${}^pX = 0$ then a local martingale M satisfying $\delta M(t) = X(t)$, for every $t \geq 0$, can be constructed, essentially the case of the stochastic integral. (2) Second, if X is locally integrable predictable process with ${}^pX = 0$ then $A(t) := \sum_{s \leq t} X(s)$ and $A - A^p$ have locally integrable bounded variation paths, where A^p is its compensator. Since $\delta(A^p) = {}^p(\delta A) = {}^pX = 0$, we can set $M := A - A^p$ to obtain $\delta M = X$, which is a local martingale with locally integral bounded variation paths. Finally, the general case is a superposition of the above two arguments. Indeed, let X be an optional process with ${}^pX = 0$ and $\sqrt{A}$ locally integrable, where $A := \sum_{s \leq t} |X(s)|^2$. Set $Y := X\mathbb{1}_{|X|>1}$, $X'' := Y - {}^pY$ and $X' := X - X''$, so ${}^pX' = {}^pX'' = 0$. The increasing process $B(t) := \sum_{s \leq t} |Y(s)|$ satisfies $|\delta B| \leq \sqrt{|\delta A|}$ so that B is locally integrable. Because ${}^p(\delta B) = \delta(B^p)$ we have $\sum_{s \leq t} |{}^pY(s)| \leq B^p(t)$, so that $\alpha(t) := \sum_{s \leq t} |X''(s)|$ is also locally integrable. In view of the previous argument (2), there is a local martingale M'' with locally integrable bounded paths such that $\delta M'' = X''$. Next, because $|X'|^2 \leq 2|X|^2 + 2|X''|^2$ the process $\beta(t) := \sum_{s \leq t} |X'(s)|^2$ takes finite values. Since ${}^pX = 0$ we have ${}^pY = -{}^p(X\mathbb{1}_{|X| \leq 1})$, $|{}^pY| \leq 1$ and $|X'| \leq 2$, which yields $\delta\beta(t) \leq 4$, proving that the increasing process β is locally integrable. Again, in view of the previous argument (1), there is a local martingale M' such that $\delta M' = X'$. The proof is ended by setting $M = M' + M''$.

Since any local martingale M can (uniquely) expressed as the sum $M = M^c + M^d$, where M^c is a continuous local martingale and M^d is a purely discontinuous

local martingale (with $M^d(0) = 0$), the purely discontinuous part M^d is uniquely determined by the jumps δM. So adding the property *purely discontinuous* to the above martingale, we have the uniqueness. Full details can be found in Jacod and Shirayaev [117, Theorem 4.56, Chapter 1, p. 56–57]. □

Let ν be a quasi-left continuous integer-valued random measure (in particular, a Poisson measure), i.e,

$$\nu(B\times]a,b],\omega) := \sum_{n=1}^{\infty} \mathbb{1}_{a_n(\omega)\in B}\mathbb{1}_{\tau_n(\omega)\in]a,b]},$$
$$\mathbb{E}\{\nu(\mathbb{R}_*^m \times \{t\})\} = \mathbb{E}\Big\{\sum_{n=1}^{\infty} \mathbb{1}_{\tau_n(\omega)=t}\Big\} = 0,$$

for every B in $\mathcal{B}(\mathbb{R}_*^m)$, $b > a \geq 0$ and $t \geq 0$, where $\{a_n : n \geq 1\}$ is a sequence of points in $\mathbb{R}_*^m$ such that a_n is $\mathcal{F}(n)$-measurable, and $\{\tau_n : n \geq 1\}$ is a (unordered) sequence of predictable stopping times. Then, the stochastic integral with respect to ν is (uniquely) defined for any predictable process $f(z,s,\omega)$ such that $F : t \mapsto \sqrt{\sum_n |f(a_n,\tau_n)|^2}$ is locally integrable, in particular if $\mathbb{E}\{|f(a_n,\tau_n)|^2\} < \infty$ for every $n \geq 1$. If ν is not quasi-left continuous (e.g., an extended Poisson measure) then the predictable projection of F may not vanish, i.e., ${}^pF(t) = \sum_n f(a_n,t)\,\mathbb{1}_{\tau_n=t}$, when every the (pathwise) series converges absolutely. Thus f is integrable with respect to ν if the (optional) process $F(t) - {}^pF(t)$ is locally integrable, see Jacod and Shirayaev [117, Definition 1.27, Chapter 2, p. 72].

For future reference, we conclude this subsection with the following summery of key properties and relations.

Let us go back to the case of a Poisson measure ν with Lévy measure (properly saying, intensity or characteristic measure) Π, i.e., $\Pi(B\times]s,t]) = \mathbb{E}\{\nu(B\times]s,t])\}$, $\Pi(B\times\{t\}) = 0$, for every $t > s > 0$ and Borel subset B of $\mathbb{R}_*^m$, and Π integrates the function $z \mapsto |z|^2 \wedge |z|$ on $\mathbb{R}_*^m\times]0,T]$, for every $T > 0$.

Next, we construct a local martingale measure $\tilde{\nu} = \nu - \Pi$, and its associated purely jumps (which is quasi-continuous from the left, i.e., with no deterministic jumps) local martingale process $L = (L_i)$ with values in $\mathbb{R}^m$,

$$L_i(t) = \int_{\mathbb{R}_*^m\times]0,t]} z_i\tilde{\nu}(\mathrm{d}z,\mathrm{d}s), \quad \forall\, t \geq 0,\ i = 1,\dots,m,$$

with predictable compensator

$$L_i^p(t) = \int_{\mathbb{R}_*^m\times]0,t]} z_i\Pi(\mathrm{d}z,\mathrm{d}s), \quad \forall\, t \geq 0,\ i = 1,\dots,m.$$

Usually, L is referred to as the *canonical* compensated Poisson (jump) process associated with the Poisson measure ν, and reciprocally, ν is referred to as the *canonical* Poisson measure associated with the compensated Poisson (jump) process L.

For a predictable process $f(x,s,\omega)$ satisfying the integrability condition (4.50) we can define the stochastic integral (a real-valued local martingale)

$$I(t) = \int_{\mathbb{R}_*^m \times]0,t]} f(z,s)\tilde{\nu}(\mathrm{d}z,\mathrm{d}s) \quad \forall\, t \geq 0,$$

and $I(0) = 0$, as a cad-lag process (and quasi-continuous from the left). If the integrand takes the form $f(z,s,\omega) = \sum_{i=1}^m g_i(t,\omega)z_i$ then we can write

$$I(t) = \sum_{i=1}^m \int_{]0,t]} g_i(s)\mathrm{d}L_i(s) \quad t \geq 0.$$

Always, we have the following properties on their jumps:

$$I(t) - I(t-) := \delta I(t) = f(\delta L(t),t)\mathbb{1}_{\{|\delta L(t)|>0\}}, \quad \forall\, t > 0.$$

The stochastic integral process $I(t)$ is a locally integrable bounded variation process if and only if

$$P\Big\{ \int_{\mathbb{R}_*^d \times (0,t]} |f(z,s)|\, \Pi(\mathrm{d}z,\mathrm{d}s) < \infty \Big\} = 1, \quad \forall\, t > 0$$

or equivalently

$$P\Big\{ \sum_{0<s\leq t} |\delta I(s)| < \infty \Big\} = 1, \quad \forall\, t > 0,$$

and in this case we have

$$I(t) = \sum_{0<s\leq t} f(\delta L(s),s)\mathbb{1}_{\{|\delta L(s)|>0\}} - \int_0^t f(z,s)\Pi(\mathrm{d}z,\mathrm{d}s), \quad \forall\, t > 0,$$

where the series converges absolutely almost surely. It is clear that the separation of the stochastic integral into a series of jumps and Lebesgue-type integral is not possible in general. However, the definition allows a suitable limit $I(t) = \lim_{\varepsilon\to 0} I_\varepsilon(t)$, where $I_\varepsilon(t)$ is the stochastic integral (of finite jumps almost surely) associated with the Lévy measure $\Pi_\varepsilon(B\times]s,t]) = \Pi\big((B\cup\{|z|\geq\varepsilon\}\times]s,t]\big)$, which can be written as previously (actually the series of jumps becomes a stochastic finite sum). In any case, the series of the jumps squared is absolutely convergent almost surely, and the process

$$t \mapsto \sum_{0<s\leq t} [I(s) - I(s-)]^2 - \int_0^t |f(z,s)|^2 \Pi(\mathrm{d}z,\mathrm{d}s)$$

is a local martingale.

Note that the integer measure ν_I on $\mathbb{R}_*$ induced by the jumps of $I(t)$, namely,

$$\nu_I(K\times]0,t]) = \sum_{0<s\leq t} \mathbb{1}_{\{f(\delta L(s),s)\in K\}}, \quad \forall t > 0,\ K \subset \mathbb{R}_*,\ \text{ compact},$$

with predictable compensator

$$\nu_I^p(K\times]0,t]) = \int_0^t \Pi\big(\{z\in\mathbb{R}_*^m : f(z,s)\in K\}, \mathrm{d}s\big),$$

yield the martingale measure $\tilde{\nu}_I = \nu - \nu^p$.

If we take an integrand $f(z,t,\omega)$ with values in $\mathbb{R}^n$ then the stochastic integral I will take values in $\mathbb{R}^n$ and its associated integer measure ν_I would be defined in $\mathbb{R}_*^n$.

Certainly, if we begin with a Lévy measure Π that integrates only $|z|^2\wedge 1$ then we need to split the jumps into two classes (small and large) to express the above properties. Also, recall that if we begin with Lévy processes $\ell_i(t)$, $i=1,\dots,m$ we may construct the integer measure ν (which is actually a standard Poisson measure) associated with the *jumps* of the $\mathbb{R}^m$-valued process $\ell=(\ell_1,\dots,\ell_m)$. The Lévy measure associated with (standard) Poisson measure ν or the Lévy m-dimensional process ℓ is the same (of the form $\pi(\mathrm{d}z)\mathrm{d}s$), and the canonical compensated Poisson process L has exactly the same jumps as ℓ, i.e., $\delta\ell(t) = \delta L(t)$, for every $t>0$. Note that the Lévy measure $\pi(\mathrm{d}z)$ in $\mathbb{R}_*^m$ is not necessarily the product measure of the individual Lévy measures $\pi_i(\mathrm{d}z_i)$ in $\mathbb{R}_*$ of each ℓ_i, even if the ℓ_i are independent, one needs also to assume no simultaneous jumps. Actually, if ℓ_i are independent then $\pi(\mathrm{d}z) = \sum_i \pi_i(\mathrm{d}z_i)$, after identifying the measure $\pi_i(\mathrm{d}z_i)$ in $\mathbb{R}_*^1$ with the measure $\pi_i(\mathrm{d}z_i)\times 0_i$ in $\mathbb{R}_*^m$ where 0_i is the zero-measure in $(\mathrm{d}z_1,\dots,\mathrm{d}z_{i-1},\mathrm{d}z_{i+1},\dots,\mathrm{d}z_m)$.

4.2.4 Extension to Semi-martingales

Remark that the initial intension is to integrate a process $f(s)$ or $f(z,t)$ which is adapted (predictable) with respect to a Wiener process $w(s)$ or centered Poisson measure $\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)$. This is to say that in most of the cases, the filtration $\{\mathcal{F}(t): t\geq 0\}$ is generated by the Wiener process or the Poisson measure, which is completed for convenience. However, what is mainly used in the construction of the stochastic integral are the following conditions:

(1) the filtration $\mathbb{F}=\{\mathcal{F}(t): t\geq 0\}$ is complete and right-continuous,

(2) the integrand f is predictable with respect to filtration $\mathbb{F}$,

(3) the integrator w (or $\tilde{\nu}$) is a (semi-)martingale with respect to filtration $\mathbb{F}$.

Thus we are interested in choosing the filtration $\mathbb{F}$ as large as possible, but preserving the (semi-)martingale character. e.g., the non-anticipative filtration $\mathbb{A}$, where $\mathcal{A}(t)$ is defined as the σ-algebra of all sets in $\mathcal{F}$ which are independent of either $w(t_1)-w(t_0),\dots,w(t_n)-w(t_{n-1})$ or $\tilde{\nu}(K_j\times]t_{i-1},t_i])$, for any $j=1,\dots,m$ and $t\leq t_0<t_1<\cdots<t_n$. Note that $\mathcal{A}(t)$ contains all null sets in $\mathcal{F}$ and the cad-lag property of w (or $\tilde{\nu}$) shows that $\mathcal{A}(t)=\bigcap_{s>t}\mathcal{A}(s)$. Because $w(t)$ (or $\tilde{\nu}(K\times]s,t]$)) is independent of any future increment, the σ-algebra $\mathcal{F}(t)$ generated by $\{w(s): s\leq t\}$ (or by $\{\tilde{\nu}(K\times]0,s]): s\leq t\}$) is included in $\mathcal{A}(t)$.

Moreover, since

$$\mathbb{E}\{w(t) \mid \mathcal{A}(s)\} = \mathbb{E}\{w(t) - w(s) \mid \mathcal{A}(s)\} + \mathbb{E}\{w(s) \mid \mathcal{A}(s)\} = \\ = \mathbb{E}\{w(t) - w(s)\} + w(s) = w(s),$$

the martingale character is preserved.

Actually, the cancelation is produced when the integrator is independent and has increment of zero-mean, even least, when the increments of the integrator are orthogonal to the integrand, e.g., $\mathbb{E}\{f(s)[w(t) - w(s)]\} = \mathbb{E}\{f(s)\}\mathbb{E}\{w(t) - w(s)\} = 0$ for $t > s$. Thus, define the class $\mathcal{E}^*$ of processes of the form $f(z, t, \omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \leq t_i$ and z belongs to K_j with some $i = 1, \ldots, n$, and $j = 1, \ldots, m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint sets with compact closure in $\mathbb{R}^m_*$ and $f_{i-1,j}$ is a bounded random variable which is *orthogonal* to $\tilde{\nu}(K_j \times]t_{i-1}, t_i])$ (in particular $\mathcal{F}(t_{i-1})$-measurable) for any i, and $f(t, \omega) = 0$ otherwise, and an analogous definition for the Wiener process case. The stochastic integral is then initially defined on the class $\mathcal{E}^*$ and the extension procedure can be carried out successfully, we refer to Section 3.13 of the the previous chapter on *Random Orthogonal Measures*. In any case, remark that if f is a deterministic function then to define the stochastic integral we need the local L^2-integrability in time, e.g., an expression of the form $s \mapsto s^\alpha$ or $(z, s) \mapsto (z \wedge 1)s^\alpha$ is integrable as long as $\alpha > -1/2$.

Space of Semi-martingales

Let us now consider the space $\mathcal{S}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, $1 \leq p \leq \infty$ of p-integrable semi-martingale on $[0, \infty]$ is defined as the cad-lag processes X with a decomposition of the form $X = M + A^+ - A^-$ where M is a local martingale and A^+, A^- are adapted monotone increasing processes with $A^+(0) = A^-(0) = 0$, both relative to $(\mathcal{F}_t : t \geq 0)$ and such that the quantity

$$\|X\|_{S^p} := \inf_{X = M + A^+ - A^-} \|M, A^+, A^-\|_{S^p},$$

where

$$\|M, A^+, A^-\|_{\mathbb{S}^p} := \mathbb{E}\Big\{ \big[\sqrt{[M](\infty)} + |A^+(\infty)| + |A^-(\infty)|\big]^p \Big\}^{1/p},$$

is finite. This is a semi-norm and by means a of equivalence classes we define the *non-separable* Banach space $\mathbb{S}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$.

Going back to the above definition of the semi-norm $\|X\|_{\mathbb{S}^p}$, if the square bracket process $\sqrt{[M](\infty, \omega)}$ is replaced with maximal process $M^*(\infty, \omega) = \sup_{t \geq 0} |M(t, \omega)|$ then we obtain an equivalent semi-norm.

This procedure can be localized, i.e., define $\mathcal{S}^p_{\mathrm{loc}}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ and the space of equivalence classes $\mathbb{S}^p_{\mathrm{loc}}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ as the spaces of semi-martingales X such that there is a sequence of stopping times $\tau_k \to \infty$ as $k \to \infty$ satisfying $X_k(\cdot) := X(\cdot \wedge \tau_k)$ belongs to $\mathcal{S}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, for any $k \geq 1$. Thus $\mathcal{S}^1_{\mathrm{loc}}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ is the space of special semi-martingales.

A further step is to consider $\mathcal{S}^0(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ the space of all semi-martingales (including non-special) X on the closed real semi-line $[0,\infty]$, i.e., $X=M+A^+-A^-$ where M is a local martingale in $[0,\infty]$ and A^+, A^- are adapted monotone increasing processes with $A^+(0)=A^-(0)=0$ and $A^+(\infty)$, $A^-(\infty)$ are almost surely finite. With the topology induced by the semi-distance

$$\begin{aligned}\llbracket X\rrbracket_{\mathbb{S}^0}&:=\inf_{X=M+A^+-A^-}\llbracket M,A^+,A^-\rrbracket_{\mathbb{S}^0},\\ \llbracket M,A^+,A^-\rrbracket_{\mathbb{S}^0}&:=\mathbb{E}\{1\wedge\big(\sqrt{[M](\infty)}+|A^+(\infty)|+\\ &\quad+|A^-(\infty)|\big)\}+\sup_\tau\mathbb{E}\{|M(\tau)-M(\tau-)|\},\end{aligned}$$

for any stopping time τ. Thus $\mathbb{S}^0(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$, after passing to equivalence classes, is a non-separable complete vector space. A closed non-separable subspace is the set $\mathbb{S}^p_{\rm c}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ of all continuous p-integrable semi-martingales, which admits a localized space denoted by $\mathbb{S}^p_{\rm c,loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$. The reader may take a look at Protter [206, Section V.2, pp. 138–193] for others similar spaces of semi-martingales.

A companion (dual) space is the set $\mathcal{P}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ of p-integrable predictable processes X, i.e., besides being predictable we have

$$||X||_{\mathbb{P}^p}:=\Big\{\int_0^\infty \mathrm{d}t\int_\Omega |X(t,\omega)|^p\,P(d\omega)\Big\}^{1/p},$$

which yields the non-separable Banach space $\mathbb{P}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$. Its localized spaces $\mathcal{P}^p_{\rm loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ and $\mathbb{P}^p_{\rm loc}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$, $p\geq 1$, are defined by the conditions (1) X is a predictable process and (2) such that there is an increasing sequence of stopping times $\tau_k\to\infty$ as $k\to\infty$ such that the processes $X_k:=\mathbb{1}_{]0,\tau_k]}X$ belong to $\mathcal{P}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$, for any $k\geq 1$.

Notice that the uncountable set of bounded and adapted left-continuous (having right-hand limit) processes is a dense subspace of $\mathbb{P}^p(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$. However, the set $\mathbb{P}^p_{\rm c}(\Omega,\mathcal{F},P,\mathcal{F}_t,t\geq 0)$ of bounded and continuous (adapted, p-integrable) processes is neither dense nor closed. We refer to Dellacherie and Meyer [58, Sections VII.3.96–105, pp. 308–324].

It is clear by now that semi-martingales are *desirable integrators* while predictable processes are *desirable integrands*. Semi-martingales contain two type of (localized) processes, (1) a bounded variation process which is integrated following the classic measure theory and (2) a local martingale which is the main study of stochastic integrals. To focus in the stochastic integral itself, the natural integrators (without localizing) are the so-called *quasi-martingales*, defined as an adapted cad-lag process X satisfying $\mathbb{V}\mathrm{ar}(X):=\sup\{\mathbb{V}\mathrm{ar}(X,\pi):\pi\}<\infty$, where $\pi=\{t_0,t_1,\dots,t_n\}$, $0=t_0<t_i<t_{i+1}$,

$$\mathbb{V}\mathrm{ar}(X,\pi):=\sum_{i=1}^n\big|\mathbb{E}\{X(t_i)-X(t_{i-1})\,|\,\mathcal{F}(t_{i-1})\}\big|+|X(t_n)|. \tag{4.52}$$

It can be proved, e.g. see Rogers and Williams [214, Section VI.41, pp. 396–398]), that any quasi-martingale admits a representation $X=Y-Z$, where Y

and Z are two nonnegative super-martingales such that $\mathbb{V}\mathrm{ar}(X) = \mathbb{V}\mathrm{ar}(Y) + \mathbb{V}\mathrm{ar}(Z)$ and that if $X = \bar{Y} - \bar{Z}$ are two other nonnegative super-martingales then $\bar{Y} - Y = \bar{Z} - Z$ is also a nonnegative super-martingale.

Given a filtered probability space $(\Omega, P, \mathcal{F}, \mathcal{F}_t : t \geq 0)$, let $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ be the *measurable*, *optional* and *predictable* σ-algebras on $[0,\infty)\times\Omega$. Now, a subset N of $[0,\infty)\times\Omega$ is called *evanescent* if $P\{\omega \in \Omega : (t,\omega) \in N\} = 0$ for every $t \geq 0$. We suppose that $\mathcal{M}$, $\mathcal{O}$ and $\mathcal{P}$ have been augmented with all evanescent sets.

For a given integrable monotone increasing (bounded variation) cad-lag process A, with its associated continuous and jump parts $A(t) = A^c(t) + [A(t+) - A(t-)]$, we may define a (signed) measure μ by the expression

$$\begin{aligned}\mu(X) &:= \mathbb{E}\Big\{\int_{[0,\infty)} X(t)\mathrm{d}A(t)\Big\} = \\ &= \mathbb{E}\Big\{\int_0^\infty X(t)\mathrm{d}A^c(t) + \sum_{t\geq 0} X(t)\,[A(t+) - A(t-)]\Big\}\end{aligned}$$

for any nonnegative $\mathcal{M}$ measurable process X. This measure vanishes on evanescent sets. Conversely, it can be proved (Doléans' Theorem, e.g., Rogers and Williams [214, Section VI.20, pp. 249–351]) that any bounded measure μ on $\mathcal{M}$, which vanishes on evanescent sets, can be represented (or disintegrated) as above for some process A as above. Furthermore, if μ satisfies

$$\mu(X) = \mu({}^oX) \quad \text{or} \quad \mu(X) = \mu({}^pX)$$

then A is optional or predictable.

Denote by $\mathcal{D}_0$ the vector space either (1) of all adapted cad-lag and bounded processes or (2) of all processes X of the form

$$X = \sum_{i=0}^n X_i\,\mathbb{1}_{[\tau_i,\tau_{i+1}[}, \quad 0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_n \leq \tau_{n+1} = \infty,$$

for any n and stopping times τ_i. Now, if $A[\cdot]$ is a linear and positive functional on $\mathcal{D}_0$ satisfying the condition

$$P\{\lim_n \sup_{0\leq s\leq t} |X_n(s)|\} = 0,\ \forall t \geq 0 \ \text{ implies } \ \lim_n A(X_n) = 0, \tag{4.53}$$

then there should exist two integrable monotone increasing cad-lag processes A^o, A^p, with $A^o(-0) = 0$, A^o optional and purely jumps, and with A^p predictable, such that

$$A[X] = \mathbb{E}\Big\{\int_{(0,\infty]} X(t-)\,dA^p(t) + \sum_{t\geq 0} X(t)\,[A^o(t) - A^o(t-)]\Big\},$$

for any X in $\mathcal{D}_0$, and the above representation is unique up to an evanescent set. Indeed, by means of condition (4.53) the functional $A[\cdot]$ can be extended to a bounded positive measure which vanishes on evanescent sets and the result follows from the previous representation.

Similarly, an adapted process A, which is right-continuous in probability (not necessarily cad-lag), is a suitable integrator if and only if the set of random variables

$$\int X \mathrm{d}A := \sum_{i=0}^{n} X_i \left[A(\tau_{i+1}) - A(\tau_i)\right]$$

remains bounded (e.g., in probability or in L^2) for every elementary predictable process X satisfying

$$X = \sum_{i=0}^{n} X_i \, \mathbb{1}_{[\tau_i, \tau_{i+1}[}, \qquad |X| \leq 1.$$

For instance, the reader is referred to the book Bichteler [25, Chapter 2, pp. 43–86] for a carefully analysis on this direction.

Then, a desirable property for a linear positive function $M[\cdot]$ defined on $\mathcal{D}_0$ to be called stochastic integral is the following condition

$$\begin{cases} \text{if } \; P\{\lim\limits_{n} \sup\limits_{0\leq s\leq t} |X_n(s)| \geq \delta\} = 0, \; \forall\, t \geq 0, \; \delta > 0 \\ \text{then } \; P\{\lim\limits_{n} \sup\limits_{0\leq t\leq T} |M[X_n \, \mathbb{1}_{]0,t]}]| \geq \varepsilon\} = 0, \end{cases} \tag{4.54}$$

for every $T \geq 0$ and $\varepsilon > 0$, or even a weaker version of it.

For a given adapted cad-lag integrable real-valued process $\{Z(t) : t \geq 0\}$ we can define a functional $Z[\cdot]$ on $\mathcal{D}_0$ as follows:

$$Z[\sum_{i=0}^{n} X_i \, \mathbb{1}_{[\tau_i, \tau_{i+1}[}] := \sum_{i=0}^{n} X_i \left(Z(\tau_{i+1}) - Z(\tau_i)\right), \tag{4.55}$$

which can be initially defined on predictable rectangles $F\times]a, b]$, F in $\mathcal{F}(a)$ by means of

$$\begin{cases} \lambda_Z(]a, b] \times F) := \mathbb{E}\{\mathbb{1}_F[Z(b) - Z(a)]\}, \\ \lambda_Z(\{0\} \times F_0) := 0, \quad \forall\, F_0 \in \mathcal{F}(0), \end{cases} \tag{4.56}$$

and then extended by additivity. If the process Z is only locally integrable, we may suppress the last term with $\tau_{n+1} = +\infty$ or consider only (deterministic) times t_i instead of stopping times τ_i. If the functional $Z[\cdot]$ or equivalent the additive set function λ_Z is nonnegative, then λ_Z is called a *content*.

It is clear that $\lambda_Z \geq 0$ if Z is monotone increasing. However, $\lambda_Z = 0$ if Z is a martingale and $\lambda_Z \geq 0$ if Z is a sub-martingale. If $\{M(t) : t \geq 0\}$ is a square integrable then $\{M^2(t) : t \geq 0\}$ is a sub-martingale and hence $\lambda_{M^2} \geq 0$, moreover

$$\begin{cases} \lambda_{M^2}(]a, b] \times F) := \mathbb{E}\{\mathbb{1}_F[M^2(b) - M^2(a)]\} = \\ \qquad = \mathbb{E}\{\mathbb{1}_F[M(b) - M(a)]^2\}, \quad \forall\, b > a \geq 0, \; F \in \mathcal{F}(a). \end{cases} \tag{4.57}$$

The extension of λ_{M^2} to a measure on $(\mathbb{R}^+ \times \Omega, \mathcal{P})$ is called Doléans measure. It can be proved (e.g. Chung and R.J. Williams [45, Theorem 2.16, Chapter 2,

pp. 52–53]) that if the process Z is a positive sub-martingale then the content λ_Z can be uniquely extended to a σ-finite measure on $\mathcal{P}$. In particular this applies to λ_{M^2}.

Extension Argument

Denote by $\mathcal{E}$ the vector space of all processes of the form $X(t,\omega) = X_{i-1}(\omega)$ if $t_{i-1} < t \leq t_i$ with $i = 1, \dots, n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and X_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $X(t,\omega) = 0$ otherwise. Elements in $\mathcal{E}$ are called elementary predictable processes. Given a square integrable $\{M(t) : t \geq 0\}$ we denote by μ_M its Doléans measure and define the stochastic integral as follows:

$$\begin{cases} \int X(s)\mathrm{d}M(s) := \sum_{i=1}^{n} X_{i-1}[M(t_i) - M(t_{i-1})], \\ \int_{(0,t]} X(s)\mathrm{d}M(s) := \sum_{i=1}^{n} X_{i-1}[M(t \wedge t_i) - M(t \wedge t_{i-1})], \\ \int_{(a,b]} X(s)\mathrm{d}M(s) := \int_{(0,b]} f(s)\mathrm{d}M(s) - \int_{(0,a]} X(s)\mathrm{d}M(s), \end{cases} \tag{4.58}$$

for every $t \geq 0$ and $b > a \geq 0$. Notice that

$$\int_{(a,b]} X(s)\mathrm{d}M(s) = \int X(s)\, \mathbb{1}_{(a,b]}(s)\mathrm{d}M(s),$$

for every $b > a \geq 0$. This definition (4.58) (defined up to an evanescent set) is independent of the particular representation used and because X_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable we obtain

$$\begin{cases} \mathbb{E}\{\big|\int X(s)\mathrm{d}M(s)\big|^2\} = \sum_{i=1}^{n} \mathbb{E}\{|X_{i-1}|^2[M^2(t_i) - M^2(t_{i-1})]\} = \\ \qquad\qquad = \int_{\mathbb{R}^+\times\Omega} |X|^2 \mathrm{d}\mu_M, \end{cases} \tag{4.59}$$

for every X in $\mathcal{E}$, and

$$\mathbb{E}\big\{\big[\int X(s)\mathrm{d}M(s)\big]\,\big[\int Y(s)\mathrm{d}M(s)\big]\big\} = \int_{\mathbb{R}^+\times\Omega} XY\mathrm{d}\mu_M, \tag{4.60}$$

for any two stochastic processes X and Y in $\mathcal{E}$.

Moreover the process $Z(t) = (X \diamond M)(t)$,

$$(X \diamond M)(t) := \int_{(0,t]} X(s)\mathrm{d}M(s), \quad \forall t \geq 0, \tag{4.61}$$

is (cad-lag) square integrable martingale, which is continuous if M is so. Since,

$$\begin{aligned} \mu_Z(]a,b] \times F) &= \mathbb{1}_F[Z(b) - Z(a)]^2) = \\ &= \mathbb{E}\big\{\mathbb{1}_F\big[\int_{(a,b]} X(s)\mathrm{d}M(s)\big]^2\big\} = \int_{(a,b]\times F} |X|^2 \mathrm{d}\mu_M, \end{aligned}$$

we deduce that

$$\mu_{X\diamond M}(B) = \int_B |X|^2 \mathrm{d}\mu_M, \quad \forall\, B \in \mathcal{P}. \tag{4.62}$$

If X belongs to $\mathcal{E}$, F is a $\mathcal{F}(a)$-measurable set and τ a stopping time which takes only finitely many values then $\mathbb{1}_F X$ and $\mathbb{1}_{]0,\tau]}X$ belong to $\mathcal{E}$ and

$$\begin{cases} \displaystyle\int_{]a,b]} \mathbb{1}_F X(s)\mathrm{d}M(s) = \mathbb{1}_F \int_{]a,b]} X(s)\mathrm{d}M(s), \\ [X \diamond M](\tau) = \displaystyle\int \mathbb{1}_{]0,\tau]}(s)X(s)\mathrm{d}M(s). \end{cases} \tag{4.63}$$

It is also clear from the expression (4.58) that the jumps of $(X\diamond M)$ are produced only by jumps of the integrator M, i.e.,

$$(X \diamond M)(t) - (X \diamond M)(t-) = X(t)[M(t) - M(t-)], \quad \forall\, t > 0, \tag{4.64}$$

except for a set of measure zero.

Denote by $\bar{\mathcal{E}}_M$ the L^2-closure of $\mathcal{E}$, i.e., the Hilbert space of all processes X for which there exists a sequence $(X_1, X_2, \ldots)$ of processes in $\mathcal{E}$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}^+\times\Omega} |X_n - X|^2 \mathrm{d}\mu_M = 0.$$

Based on the isometry identity (4.59), and the maximal martingale inequality, for every $T \geq 0$,

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T} \Big| \int_{(0,t]} X(s)\mathrm{d}M(s)\Big|^2 \Big\} \leq 4\,\mathbb{E}\Big\{ \Big| \int_0^T X(s)\mathrm{d}M(s)\Big|^2 \Big\}, \tag{4.65}$$

this linear operation (called *stochastic integral*) can be extended to the closure $\bar{\mathcal{E}}_M$, preserving linearity and the properties (4.59), ..., (4.64). Moreover, (4.63) holds for any bounded $\mathcal{F}(a)$-measurable function f replacing $\mathbb{1}_F$ (even if a is a bounded stopping times) and any bounded stopping time τ.

In general, it is proved in Doob [59, Section IX.5, pp. 436–451] that any martingale M with orthogonal increments (i.e., a square-integrable martingale), the Hilbert space $\bar{\mathcal{E}}_M$ contains all adapted process X and square-integrable respect to the product measure $P(\mathrm{d}\omega)$ times the Lebesgue-Stieltjes measure $\mathrm{d}\mathbb{E}\{|M(t) - M(0)|^2\}$.

It is convenient to localize the above processes, i.e., we say that a measurable process X belongs to $\bar{\mathcal{E}}_{M,\mathrm{loc}}$ if and only if there exists a sequence of stopping times $\{\tau_k : k \geq 1\}$ such that $\tau_k \to \infty$ almost sure and $\mathbb{1}_{]0,t\wedge\tau_k]}X$ belongs to $\bar{\mathcal{E}}_{M_k}$, for every $t > 0$, where $M_k := \{M(s \wedge \tau_k) : s \geq 0\}$. Therefore, the stochastic integral $X \diamond M$ is defined as the almost sure limit of the sequence $\{X_k \diamond M_k : k \geq 1\}$, with $X_k := \mathbb{1}_{]0,\tau_k]}X$. This should be validated by a suitable condition to make this definition independent of the choice of a localizing sequence, see Chung and Williams [45, Theorem 2.16, Chapter 2, pp. 23–48].

The use of the quadratic variation process is simple when dealing with a continuous square integrable martingale. The general case is rather technical.

Anyway, a key point is the following: If $M = \{M(t) : t \geq 0\}$ is a locally square integrable martingale then there exists an increasing predictable process $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a local martingale, which is continuous if and only if M is quasi-left continuous (e.g., Jacod and Shiryaev [117, Theorem 4.2, Chapter 1, pp. 38–39]). It is clear that we have, first for X in $\mathcal{E}$ and then for every X in $\mathcal{E}_M$, the relation

$$\langle X \diamond M \rangle(t) = \int_0^t |X(s)|^2 \mathrm{d}\langle M \rangle(s), \quad \forall t \geq 0, \tag{4.66}$$

so that the process

$$\Big[\int_{(0,t]} X(s)\mathrm{d}M(s)\Big]^2 - \int_0^t |X(s)|^2 \mathrm{d}\langle M \rangle(s), \quad \forall t \geq 0, \tag{4.67}$$

is a (cad-lag) local martingale.

Lenglart's domination property (see inequality (3.32) in Chapter 3 or more details in Jacod and Shiryaev [117, Section 1.3c, pp. 35–36]) yields the useful estimate

$$\left\{ \begin{array}{l} P\big\{ \sup\limits_{0\leq t\leq T} \big| \int_{(0,t]} X(s)\mathrm{d}M(s)\big| \geq \varepsilon \big\} \leq \dfrac{\delta}{\varepsilon^2} + \\ \qquad\qquad + P\big\{ \int_0^T |X(s)|^2 \mathrm{d}\langle M \rangle(s) \geq \delta \big\}, \end{array} \right. \tag{4.68}$$

for any positive numbers T, ε and δ. By means of this estimate, all properties (4.59), ..., (4.64), (4.67), (4.68) hold, except that the process (4.61) is now a (cad-lag, continuous whenever M is such) local square martingale. Moreover, the continuity property (4.54) is now verified.

Since any continuous local martingale is a local square integral martingale, the stochastic integral is well defined. To go one step further and define the stochastic integral for any (cad-lag, not necessarily continuous and not necessarily local square integrable) local martingale M, we need to define the (optional) quadratic variation, see (3.28) in Chapter 3 or for more detail see for instance Dellacherie and Meyer [58, Chapters V–VIII] or Liptser and Shiryayev [158],

$$\left\{ \begin{array}{l} [M](t) := \langle M^c \rangle(t) + A_M(t), \quad \text{with} \\ A_M(t) := \sum\limits_{s\leq t} [M(s) - M(s-)]^2, \quad \forall t \geq 0, \end{array} \right. \tag{4.69}$$

where M^c is the continuous part of the (local) martingale M and the second term in the right-hand side A_M is an optional monotone increasing process null at time zero, not necessarily locally integrable, but such that $\sqrt{A_M}$ is locally integrable. It can be proved (see Rogers and Williams [214, Theorem 37.8, Section VI.7, pp. 389–391]) that the process $[M]$ given by (4.69) is the unique optional monotone increasing process null at time zero such that $M^2 - [M]$ is a local martingale and $[M](t) - [M](t-) = [M(t) - M(t-)]^2$ for every $t > 0$.

On the other hand, a local martingale admits a unique decomposition $M = M_0 + M^c + M^d$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, M^c is a continuous local martingale (null at $t = 0$) and M^d is a purely discontinuous local martingale, i.e., $M^d(0) = 0$ and for every continuous local martingale N the product $M^d N$ is a local martingale. Let us show that for a given $\varepsilon > 0$, any local martingale M admits a (non unique) decomposition $M = M_0 + M_\varepsilon^{'} + M_\varepsilon^{''}$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, $M_\varepsilon^{'}$ is a (cad-lag, only the small jumps) local martingale (null at $t = 0$) satisfying $|M_\varepsilon^{'}(t) - M_\varepsilon^{'}(t-)| \leq \varepsilon$ for every $t > 0$, and $M_\varepsilon^{''}$ is a (cad-lag, only the large jumps) local martingale (null at $t = 0$) which have local bounded variation. Indeed, set $\delta M(t) := M(t) - M(t-)$ and because M is a cad-lag process we can define $A(t) := \sum_{s \leq t} \delta M(s) \mathbb{1}_{|\delta M(s)| > \varepsilon/2}$, whose variation process $\mathrm{var}(A, t) := \sum_{s \leq t} |\delta M(s)| \mathbb{1}_{|\delta M(s)| > \varepsilon/2}$ is finite for almost every path. Setting $\tau_k := \inf\{t > 0 : \mathrm{var}(A, t) > k \text{ or } |M(t)| > k\}$ we obtain $\mathrm{var}(A, \tau_k) \leq k + |\delta M(\tau_k)|$, i.e, $\mathrm{var}(A, \tau_k) \leq 2k + |M(\tau_k)|$ so that the sequence of stopping times $\{\tau_k : k \geq 1\}$ is a reducing sequence for $\mathrm{var}(A, \cdot)$, proving that the process $\mathrm{var}(A, \cdot)$ is local integrable. Therefore A admits a dual predictable compensator A^p, see Definition 3.5 in Chapter 3. It is clear that $M_\varepsilon^{''} := A - A^p$ is a local martingale with local bounded variation. A simple calculation show that $M_\varepsilon^{'} := M - A + A^p$ satisfies $|\delta M(t)| \leq \varepsilon$, for every $t > 0$. Moreover, since $M_\varepsilon^{''}$ is also a purely discontinuous martingale, i.e., $M_\varepsilon^{''}$ is orthogonal to any continuous local martingale N, namely $M_\varepsilon^{''} N$ is a local martingale, see Jacod and Shiryaev [117, Section 1.4b, pp. 40–43]).

Thus, an essential fact needed to complete the definition of stochastic integral is that either a local martingale or semi-martingale M admits a (non-unique) decomposition $M = M_0 + M_1 + M_2$, where M_0 is a $\mathcal{F}(0)$-measurable random variable, M_1 is a cad-lag process with locally bounded variation paths and M_2 is a local square integrable martingale, both null at time zero. Therefore, integration with respect to M_1 is pathwise (as in the classic measure theory) and integration with respect to M_2 is as above, via the martingale argument.

The only continuous local martingale which has bounded variation paths is the constant process. However, there are (e.g., the Poisson process, after subtracting its compensator) (cad-lag, non-continuous) local martingale with bounded variation paths. Therefore there are two possible interpretations when taking those processes as integrators. This is resolved by using the predictable version representing the equivalence class of the integrand process. For instance, if X is a cad-lag process and M is a local martingale with bounded variation paths, e.g., for a sequence $\{a_i, \tau_i : i \geq 1\}$ of stopping times $\tau_1 \leq \tau_2 \leq \cdots$, with $\tau_i \to \infty$, and $\mathcal{F}(\tau_i)$-measurable random variables a_i we have

$$A(t) := \sum_{i=1}^{\infty} a_i \mathbb{1}_{\tau_i \leq t}, \quad \forall\, t \geq 0, \qquad M := A - A^p,$$

where A^p is the dual compensator of A. The expression

$$\int_{]0,t]} X(t) \mathrm{d}A(t) = \sum_{i=1}^{\infty} X(\tau_i) a_i \mathbb{1}_{\tau_i \leq t}, \quad \forall\, t \geq 0,$$

is pathwise interpreted (and well defined) in the Riemann-Stieltjes sense if and only if the process X is left-continuous at each jump time, i.e., $X(\tau_i) = X(\tau_i-)$, for every $i \geq 1$. On the other hand, the measure induced by A or by $A_- : t \mapsto A(t-)$ (its left-continuous version) is the same sum of Dirac measures so that the expression

$$\int_{]0,t]} X(t)\mathrm{d}A_-(t) = \sum_{i=1}^{\infty} X(\tau_i)a_i \mathbb{1}_{\tau_i \leq t}, \quad \forall\, t \geq 0,$$

is pathwise interpreted (and well defined) in the Riemann-Stieltjes sense if and only if the process X is right-continuous at each jump time, i.e., $X(\tau_i) = X(\tau_i+)$, for every $i \geq 1$. In the Lebesgue-Stieltjes sense, it does not matter which version A or A_- is used to derived the measure, proving that a bounded process X is integrable if it is right (or left) continuous at τ_i for every $i \geq 0$.

The dual compensator A^p of a (cad-lag) process A with locally integrable bounded variation satisfied, see Definition 3.5 in Chapter 3,

$$\mathbb{E}\{\int_{[0,T_k)} X(t,\omega)\mathrm{d}A^p(t,\omega) = \mathbb{E}\{\int_{[0,T_k)} {}^pX(t,\omega)\mathrm{d}A(t,\omega),$$

for every $k \geq 1$ and for any bounded measurable process X, where the predictable projection ${}^pX$, is such that for any predictable stopping time τ we have $\mathbb{E}\{{}^pX\mathbb{1}_{\tau<\infty}\} = \mathbb{E}\{X\mathbb{1}_{\tau<\infty}\}$. The sequence of stopping times $\{T_k : k \geq 1\}$ localizes A, i.e., the process $t \mapsto A(t\wedge T_k)$ has integrable bounded variation (meaning in this case $\mathbb{E}\{A(T_k)\} < \infty$) and $T_k \to \infty$ almost surely. We deduce that the stochastic integral with respect to an integrator $A - A^p$ is always zero for any predictable process X. Recall that the stochastic integral is meaningful only for the predictable member representing a given equivalence class of processes used as integrand.

Therefore, we conclude that as long as the predictable (in particular any adapted left-hand continuous) version of the integrand (equivalence class) process is used, the pathwise and stochastic integral coincide.

Back to Integer Random Measures

Let ν be an integer-valued (random) measure, see Definition 4.6, and let ν^p be a good version of its compensator, see Theorem 4.7. For instance, if ν is a extended Poisson measure then ν^p is a deterministic Radon measure on $\mathbb{R}^m_* \times [0,\infty)$ with $\nu^p(\mathbb{R}^m_* \times \{0\}) = 0$. Denote by ν_{qc} the quasi-continuous part of ν, i.e.,

$$\begin{aligned}
&\nu_{qc}(B\times]a,b]) := \nu(B\times]a,b]) - \nu^p_d(B\times]a,b]),\\
&\nu^p_d(B\times]a,b]) := \sum_{a<s\leq b} \nu^p(\{s\}\times B),
\end{aligned}$$

with $\nu^p_c = (\nu_{qc})^p$, where

$$\nu^p_c := \nu^p(B\times]a,b]) - \nu^p_d(B\times]a,b]),$$

is a good version of the compensator of ν_{qc}. The measure ν_d^p contains all non-predictable discontinuities, which are not handled with the stochastic integral, they must be treated pathwise, by means of the classic measure theory. For instance, if $\nu = \nu_X$ defined as the number of jumps associated to a (cad-lag) local martingale (or semi-martingale) X, see (4.7) then ν_d^p is locally integrable. The integral with respect to the predictable discontinuous part $\nu_d := \nu - \nu_{qc}$ is part of the stochastic integral. Thus, using the (cad-lag and quasi-left continuous, purely discontinuous) local martingale measure $\tilde\nu_{qc} := \nu_{qc} - \nu_c^p = \nu - \nu^p$, we proceed as in Section 4.2.3 to define the stochastic integral, essentially replacing the Lévy measure $m(\mathrm{d}s, \mathrm{d}z)$ by (continuous part of) the compensator ν_c^p. Thus, for a elementary predictable process f of the form $f(t,z,\omega) = f_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint compact subsets of $\mathbb{R}_*^m$ and $f_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise, we set

$$\int_{\mathbb{R}_*^m\times(0,\infty)} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) := \sum_{i=1}^{n}\sum_{j=1}^{m} f_{i-1,j}\,\tilde\nu_{qc}(K_j\times]t_{i-1},t_i]),$$

and

$$\int_{\mathbb{R}_*^m\times(a,b]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) := \int_{\mathbb{R}_*^m\times(0,\infty)} f(z,s)\,\mathbb{1}_{(a,b]}(s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s),$$

for every $b > a \ge 0$. The L^2-closure of all elementary predictable processes $\mathcal{E}$ is denoted by $\bar{\mathcal{E}}_\nu$, i.e., processes f such that there is a sequence $(f_1, f_2, \dots)$ of processes in $\mathcal{E}$ such that

$$\lim_{k\to\infty} \mathbb{E}\Big\{ \int_{\mathbb{R}_*^m\times(0,\infty)} |f_k(z,s) - f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s)\Big\} = 0.$$

Notice that we may use (indistinctly), ν_c^p or ν_{qc} in the above condition, both are random measure. Based on the isometry and estimate

$$\begin{aligned}
&\mathbb{E}\Big\{\Big|\int_{\mathbb{R}_*^m\times(0,T]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} = \mathbb{E}\Big\{\int_{\mathbb{R}_*^m\times(0,T]} |f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s)\Big\},\\
&\mathbb{E}\Big\{\sup_{0\le t\le T}\Big|\int_{\mathbb{R}_*^m\times(0,t]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s)\Big|^2\Big\} \le\\
&\qquad\qquad \le 4\,\mathbb{E}\Big\{\int_{\mathbb{R}_*^m\times(0,T]} |f(z,s)|^2\,\nu_c^p(\mathrm{d}z,\mathrm{d}s)\Big\},
\end{aligned}$$

for every $T \ge 0$, the stochastic integral is defined in the Hilbert space $\bar{\mathcal{E}}_\nu$, which can be also extended to the localized space $\bar{\mathcal{E}}_{\nu,\mathrm{loc}}$. Therefore, the integral with respect to $\tilde\nu$ when it is not quasi-left continuous is defined by

$$\left\{\begin{aligned}
\int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\tilde\nu(\mathrm{d}z,\mathrm{d}s) &= \int_{\mathbb{R}_*^m\times]a,b]} f(z,s)\,\tilde\nu_{qc}(\mathrm{d}z,\mathrm{d}s) +\\
&+ \int_{\mathbb{R}_*^m\times]a.b]} f(z,s)\,\nu_d^p(\mathrm{d}z,\mathrm{d}s),
\end{aligned}\right. \tag{4.70}$$

where the second term is a pathwise Lebesgue-Stieltjes integral.

Taking the quasi-left continuous part $\tilde{\nu}_{qc}$, the process

$$f \diamond \tilde{\nu}_{qc} : t \mapsto \int_{\mathbb{R}_*^m \times (0,t]} f(z,s)\, \tilde{\nu}_{qc}(\mathrm{d}z, \mathrm{d}s),$$

is a (local) martingale with predictable quadratic variation process

$$\langle f \diamond \tilde{\nu}_{qc} \rangle = \int_{\mathbb{R}_*^m \times (0,t]} |f(z,s)|^2\, \nu_c^p(\mathrm{d}z, \mathrm{d}s),$$

so that denoting by $\nu_{qc}(f)$ its associate integer-valued measure with (good predictable) compensator $\nu_{qc}^p(f)$ and local martingale measure $\tilde{\nu}_{qc}(f) := \nu_{qc}(f) - \nu_{qc}^p(f)$ we have the substitution formula

$$\int_{\mathbb{R}_*^m \times (0,t]} g(z,s)\, \tilde{\nu}_{qc}^f(\mathrm{d}z, \mathrm{d}s) = \int_{\mathbb{R}_*^m \times (0,t]} g(z,s)\, f(z,s)\, \tilde{\nu}_{qc}(\mathrm{d}z, \mathrm{d}s), \tag{4.71}$$

first for elementary predictable processes g, which is extended by continuity to any integrable processes f and g.

When the stochastic integral is defined for random measures associated to a semi-martingale, i.e., the integer-valued measure ν_M associated with a (cad-lag) local martingale (or semi-martingale) M is the same as the one associated with its jumps part, $M^d := M - M^c$, i.e., $\nu_M = \nu_{M^c}$, a general form of the stochastic integral takes the form

$$\int_{]a,b]} X(s)\, \mathrm{d}M^c(s) + \int_{\mathbb{R}_*^m \times]a,b]} f(z,s)\, \tilde{\nu}_{qc}(\mathrm{d}z, \mathrm{d}s) + \int_{\mathbb{R}_*^m \times]a,b]} f(z,s)\, \nu_d^p(\mathrm{d}z, \mathrm{d}s),$$

where the first (stochastic) integral is a continuous local martingale, the second (stochastic) integral is a purely discontinuous local martingale and the last term makes sense as a Lebesgue-Stieltjes pathwise integral. Notice that integral with respect to ν_c^p or ν^p is part of the stochastic integral with respect to $\tilde{\nu}_{qc}$ or μ, respectively, i.e., if

$$P\{\int_{\mathbb{R}_*^m \times]a.b]} |f(z,s)|\, \nu(\mathrm{d}z, \mathrm{d}s) < \infty\} = 1$$

then we have

$$\begin{aligned}
&\int_{\mathbb{R}_*^m \times]a.b]} f(z,s)\, \nu(\mathrm{d}z, \mathrm{d}s) = \\
&\quad = \int_{\mathbb{R}_*^m \times]a,b]} f(z,s)\, \tilde{\nu}(\mathrm{d}z, \mathrm{d}s) + \int_{\mathbb{R}_*^m \times]a.b]} f(z,s)\, \nu^p(\mathrm{d}z, \mathrm{d}s) = \\
&\quad = \int_{]a,b] \times \mathbb{R}_*^m} f(z,s)\, \tilde{\nu}_{qc}(\mathrm{d}z, \mathrm{d}s) + \int_{\mathbb{R}_*^m \times]a.b]} f(z,s)\, \nu_c^p(\mathrm{d}z, \mathrm{d}s),
\end{aligned}$$

almost surely. Moreover, any integer-valued measure ν has the form

$$\nu(B \times]a,b]) = \sum_{i=1}^{\infty} \mathbb{1}_{a < \tau_i \le b}\, \mathbb{1}_{\zeta_i \in B}, \quad \forall\, b > a \ge 0,\ B \in \mathcal{B}(\mathbb{R}_*^m),$$

for some sequence $\{\tau_i, \zeta_i : i \geq 1\}$, where the stopping times τ_i cannot be ordered, i.e., it is not necessarily true that $\tau_i \leq \tau_{i+1}$, and the $\mathbb{R}^m_*$-valued random variables ζ_i are $\mathcal{F}(\tau_i)$-measurable, but $\nu(\mathbb{R}^m_* \times \{0\}) = 0$ and $\nu(K \times]a, b]) < \infty$ for any $b > a \geq 0$ and any compact subset K of $\mathbb{R}^m_*$. Thus, we expect

$$\int_{\mathbb{R}^m_* \times]a.b]} f(z,s)\nu(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i \leq b}\, f(\zeta_i, \tau_i),$$

whenever the above series converges absolutely and f is a continuous process.

To integrate a general predictable process $f = f(s, z, \omega)$, we may proceed as follows: first we separate the integrable jumps (jumps of order 1) from the square integrable jumps (jumps of order 2), namely, first we define

$$f_1(s) := \sum_{i=1}^{\infty} \mathbb{1}_{\tau_i = s}\, f(\zeta_i, \tau_i),$$

whenever sum is absolutely convergent, i.e.,

$$\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i = s}\, |f(\zeta_i, \tau_i)| < \infty,$$

and $f_1(s) = 0$ otherwise. The particular case where $f(z, t, \omega) = 0$ for any z such that $|z| < \varepsilon$, for some $\varepsilon = \varepsilon(\omega) > 0$ is the leading example, since the above series becomes a finite sum. Recalling that the jump process $t \mapsto \sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \leq t}\, f_1(\tau_i)$ is a cad-lag process, so it has only a finite number of jumps greater than $\varepsilon > 0$ on any bounded time interval $[0, T]$, $T > 0$, we can set, for any $b > a \geq 0$

$$\int_{\mathbb{R}^m_* \times]a.b]} f(z,s)\, \nu^p(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i \leq b}\, f_1(\tau_i),$$

as a pathwise integral (defined as a finite sum or a convergent series, for each ω almost surely) with respect to measure ν^p (all locally integrable jumps), and we give a L^2-sense (it cannot be pathwise!) to

$$\int_{\mathbb{R}^m_* \times]a.b]} f(z,s)\, (\nu - \nu^p)(\mathrm{d}z, \mathrm{d}s) = \sum_{i=1}^{\infty} \mathbb{1}_{a<\tau_i \leq b}\, [f(\zeta_i, \tau_i) - f_1(\tau_i)],$$

whenever the process

$$t \mapsto \sqrt{\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \leq t}\, [f(\zeta_i, \tau_i) - f_1(\tau_i)]^2}$$

is (locally) integrable. The compensator argument is used to define a measure ν^p, which agrees with ν on predictable processes and such that $\tilde{\nu} := \nu - \nu^p$ is a local martingale measure. Briefly, for each ω, we make use of a series with indices i such that either $\sum_i |\zeta_i| \wedge 1$ converges or such that the quadratic series $\sum_i |\zeta_i|^2 \wedge 1$ converges to define ν^p. All other indices are ignored.

Here, the martingale theory is used to define the stochastic integral with respect to $\tilde{\nu}$ for any predictable process (class of equivalence) $f(z,s)$ such that the monotone increasing process

$$t \mapsto \Big[\int_{\mathbb{R}^m_* \times]0.t]} |f(z,s)|^2 \, \nu^p(\mathrm{d}z, \mathrm{d}s) \Big]^{1/2}$$

is (locally) integrable. Moreover, we can require only that the following process

$$t \mapsto \sqrt{\sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \leq t} \, [f(\zeta_i, \tau_i) - f_1(\tau_i)]^2}$$

be (locally) integrable.

For a neat and deep study, the reader may consult Chung and Williams [45], while a comprehensive treatment can be found in Dellacherie and Meyer [58, Chapters V–VIII], Jacod and Shiryaev [117, Chapters 1 and 2]), Rogers and Williams [214, Volume 2]). Also, a more direct approach to stochastic integrals can be found in the book Protter [206], covering even discontinuous martingales.

4.2.5 Vector Valued Integrals

Firstly, recall that any local martingale M can be written in a unique form as the sum $M_0 + M^c + M^d$, where $M_0 = M(0)$ is a $\mathcal{F}$-measurable random variable, M^c is a continuous local martingale (and therefore locally square integrable) and M^d is a purely discontinuous local martingale, both $M^c(0) = M^d(0) = 0$. Also, any local martingale M with $M(0) = 0$ (in particular a purely discontinuous local martingale) can be written in a (non unique) form as the sum $M' + M''$, where both M' and M'' are local martingale, the jumps of M'' are bounded by a constant a (i.e., $|\delta M''| \leq a$ so that M'' is locally square integrable) and M' has locally integrable bounded variation paths. The predictable projection of a local martingale M is $(M(t-) : t > 0)$ so that a predictable local martingale is actually continuous. Finally, a continuous or predictable local martingale with locally bounded variation paths is necessarily a constant.

Secondly, recall the definitions of the predictable and the optional quadratic variation processes. Given real-valued local square integrable martingale M the predictable (increasing) quadratic variation process $t \mapsto \langle M \rangle(t)$ obtained via the Doob-Meyer decomposition Theorem 3.10 applied to $t \mapsto M^2(t)$ as a local sub-martingale of class (D). This is the only increasing predictable locally integrable process $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a martingale. However, the predictable quadratic variation process is generally used for continuous local martingales. For a real-valued (non necessarily continuous) local (non necessarily square integrable) martingale M, the optional (increasing) quadratic variation process $t \mapsto [M](t)$ is defined as $\langle M \rangle(t) + \sum_{s \leq t} |M(s) - M(s-)|^2$. This is the only increasing optional process (not necessarily locally integrable) $[M]$ such that $M^2 - [M]$ is a local martingale and $\delta[M] = (\delta M)^2$. The increasing optional process $\sqrt{[M]}$ is locally integrable, and if $[M]$ is locally integrable then it is a local sub-martingale of class (D) and again via the Doob-Meyer decomposition we

obtain a predictable increasing locally integrable $\langle M\rangle$ (called the compensator of $[M]$), which agrees with the predictable quadratic variation process previously defined for local square integrable martingales. Therefore, the predictable quadratic variation process $\langle M\rangle$ may not be defined for a discontinuous local martingale, but the optional quadratic variation $[M]$ is always defined. The concept of integer-valued random measures is useful to interpret $[M]$ as the increasing process associated with the integer-valued measure ν_M derived from M. Thus $\langle M\rangle$ is the increasing predictable process (not necessarily integrable) associated with the predictable compensator ν_M^p of ν_M. If M is quasi-left continuous then $\langle M\rangle$ is continuous, and therefore locally integrable. Next, for any two real-valued local martingale M and N the predictable and optional quadratic co-variation processes are defined by the formula $4\langle M,N\rangle = \langle M+N\rangle - \langle M-N\rangle$ and $4[M,N] = [M+N] - [M-N]$. Notice that

$$E\Big\{\int_{]a,b]} f(t)\mathrm{d}\langle M,N\rangle(t)\Big\} = E\Big\{\int_{]a,b]} f(t)\mathrm{d}[M,N](t)\Big\},$$

for every predictable process such that the above integrals are defined.

An important role is played by the *Kunita-Watanabe inequality*, namely for any two real-valued local martingales M and N and any two (extended) real-valued measurable processes α and β we have the inequality

$$\left\{\int_0^t |\alpha(s)|\,|\beta(s)|\,|\mathrm{d}[M,N](s)| \le \sqrt{\int_0^t |\alpha(s)|\mathrm{d}[M](s)}\times \right. \\ \left. \times\sqrt{\int_0^t |\beta(s)|\mathrm{d}[N](s)}, \right. \tag{4.72}$$

almost surely for every $t > 0$, where $|\mathrm{d}[M,N]|$ denotes the total variation of the signed measure $\mathrm{d}[M,N]$. Certainly, the same estimate is valid for the predictable quadratic co-variation process $\langle M,M\rangle$ instead of optional process $[M,N]$. The argument to prove estimate (4.72) is as follow. Since $[M+rN, M+rN] = [M] - 2r[M,N] + r^2[N]$ is an increasing process for every r, we deduce $(\mathrm{d}[M,N])^2 \le \mathrm{d}[M]\,\mathrm{d}[N]$. Next, Cauchy-Schwarz inequality yields (4.72) with $\mathrm{d}[M,N](s)$ instead of $|\mathrm{d}[M,N](s)|$. Finally, by means of the Radon-Nikodym derivative, i.e., replacing α by $\alpha = (\mathrm{d}[M,N]/|\mathrm{d}[M,N](s)|)\,\alpha$, we conclude. For instance, a full proof can be found in Durrett[67, Section 2.5, pp. 59–63] or Revuz and Yor [212, Proposition 1.15, Chapter, pp. 126–127].

Let $M = (M_1,\ldots,M_d)$ a d-dimensional continuous local martingale in a filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \ge 0)$, i.e., each component $(M_i(t) : t \ge 0)$, $i = 1,\ldots,d$, is a local continuous martingale in $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \ge 0)$. Recall that the predictable quadratic co-variation $\langle M\rangle = (\langle M_i, M_j\rangle : i,j = 1,\ldots,d)$ is a symmetric nonnegative matrix valued process. The stochastic integral with respect to M is defined for a d-dimensional progressively measurable process $f = (f_1,\ldots,f_d)$ if for some increasing sequence of stopping times $\{\tau_n : n \ge 1\}$

with $\tau_n \to \infty$ we have

$$\mathbb{E}\Big\{ \int_0^{\tau_n} \sum_{i,j=1}^{d} f_i(s) f_j(s) \mathrm{d}\langle M_i, M_j\rangle(s) \Big\} < \infty. \tag{4.73}$$

Based on (4.72), it is clear that, if each component f_i is locally square integrable with respect to $\langle M_i\rangle$, i.e.,

$$\mathbb{E}\Big\{ \int_0^{\tau_n} |f_i(s)|^2 \mathrm{d}\langle M_i\rangle(s) \Big\} < \infty,$$

then the above condition (4.73) is satisfied. However, the converse may be false, e.g., if $w = (w_1, w_2)$ is a two-dimensional standard Wiener process then set $M_1 := w_1$, $M_2 := kw_1 + (1-k)w_2$, where k is a $(0,1)$-valued predictable process. Choosing $f = (f_1, f_2) = (-\frac{k}{1-k}, \frac{1}{1-k})$, we have $\sum_{i,j} f_i f_j \mathrm{d}\langle M_i, M_j\rangle = \mathrm{d}\ell$, the Lebesgue measure, so we certainly have (4.73), but

$$\int_0^t |f_1(s)|^2 \mathrm{d}\langle M_1\rangle(s) = \int_0^t \Big| \frac{k(s)}{1-k(s)} \Big|^2 \mathrm{d}s < \infty \quad \text{a.s. } \forall t > 0,$$

may not be satisfied.

For a n-dimensional *continuous* local martingale $M = (M_1, \dots, M_n)$ and an adapted $\mathbb{R}^n$-valued (measurable) process $f = (f_1, \dots, f_n)$ we have the following estimate: for every $p > 0$ there exists a positive constant $C = C_p$ depending only on p, such that

$$\begin{cases} \mathbb{E}\Big\{ \sup\limits_{0\le t\le T} \Big[\sum\limits_{i=1}^{n} \Big| \int_0^t f_i(s) \mathrm{d}M_i(s) \Big|^2 \Big]^{p/2} \Big\} \le \\ \qquad \le C\, \mathbb{E}\Big\{ \Big[\sum\limits_{i,j=1}^{n} \int_0^T f_i(s)\, f_j(s) \mathrm{d}\langle M_i(s), M_j(s)\rangle \Big]^{p/2} \Big\}. \end{cases} \tag{4.74}$$

for any stopping time T. In particular, for a standard n-dimensional Wiener process $(w(t) : t \ge 0)$, we can write

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} \Big| \int_0^t f(s) \mathrm{d}w(s) \Big|^p \Big\} \le C\, \mathbb{E}\Big\{ \Big[\int_0^T |f(s)|^2 \mathrm{d}s \Big]^{p/2} \Big\}. \tag{4.75}$$

This estimate follows from Burkhölder-Davis-Gundy inequality (as in (3.31) of Chapter 3), e.g., see Karatzas and Shreve [124, Section 3.3.D, pp. 163–167]. Notice that we make take $C_1 = 3$ and $C_2 = 4$.

Regarding the stochastic integral with respect to a Poisson measure in the Polish space $\mathbb{R}_*^m$ (or more general in a Blackwell space), we should mention that the key elements are the compensated local martingale measure $\tilde{\nu}$ and the compensator ν^p, which is a predictable random measure. Both are constructed from an integer-valued random measure, which is naturally obtained from a optional locally integrable jump process or better a (purely discontinuous) local

(sub-)martingale. A posteriori, we may use a predictable real-valued process $\gamma(z,t)$ on $\mathbb{R}^m_* \times (0,\infty)$ such that

$$P\Big\{ \int_{]0,t]} \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma^2(z,s)\, \pi(\mathrm{d}z) < \infty \Big\} = 1, \quad \forall t > 0,$$

and use the stochastic integral to define a local martingale measure

$$\tilde\nu_\gamma(B\times]a,b]) := \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_B\, \gamma(z,s)\, \tilde\nu(\mathrm{d}z,\mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$ and $b > a \geq 0$, with a predictable quadratic variation (or *compensator*) given by

$$\nu^p_\gamma(B\times]a,b]) := \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_B\, \gamma^2(z,s)\, \nu^p(\mathrm{d}z,\mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$, $b > a \geq 0$, and for the case of the Poisson measure $\nu^p(\mathrm{d}z,\mathrm{d}s) = \pi(\mathrm{d}z)\,\mathrm{d}s$. Thus ν^p_γ has a density $\delta := \gamma^2$ with respect to ν^p.

The estimate (4.74) is also valid for a Poisson integral, with a small correction, namely, for any p in $(0,2]$ there exists a positive constant $C = C_p$ (actually $C_p := (4-p)/(2-p)$ if $0 < p < 2$ and $C_2 = 4$) such that for any adapted (measurable) process $f(\zeta,s)$ (actually, the predictable version is used) we have

$$\begin{cases} \mathbb{E}\Big\{ \sup\limits_{0\leq t\leq T} \Big| \int_{\mathbb{R}^m_*\times]0,t]} f(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \leq \\ \qquad\qquad \leq C\, \mathbb{E}\Big\{ \Big[\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big]^{p/2} \Big\}, \end{cases} \tag{4.76}$$

for every stopping time T. This follows immediately from estimate (3.30) of Chapter 3. The case $p > 2$ is a little more complicate and involves Itô formula as discussed in the next section.

For the sake of simplicity and to recall the fact that stochastic integral are defined in an L^2-sense, instead of using the natural notation $\bar{\mathcal{E}}_{M,\mathrm{loc}}$, $\bar{\mathcal{E}}_M$, $\bar{\mathcal{E}}_{\pi,\mathrm{loc}}$, $\bar{\mathcal{E}}_\pi$, $\bar{\mathcal{E}}_{\mathrm{loc}}$, $\bar{\mathcal{E}}$ of this Section 4.2 we adopt the following

Definition 4.21 (L^2-Integrand Space)**. (a)** Given a d-dimensional continuous square integrable martingale M with predictable quadratic variation process $\langle M\rangle$ in a filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \geq 0)$, we denote by $L^2(M)$ or $L^2(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \geq 0, M, \langle M\rangle)$ the equivalence class with respect to the completion of product measure $P \times \langle M\rangle$ of $\mathbb{R}^d$-valued square integrable predictable processes X, i.e. (4.73) with $\tau_n = \infty$. This is regarded as a closed subspace of the Hilbert space $L^2([0,\infty)\times\Omega, \bar{\mathcal{P}}, \langle M\rangle \times P)$, where $\bar{\mathcal{P}}$ is the $\langle M\rangle \times P$-completion of the predictable σ-algebra $\mathcal{P}$ as discussed at the beginning of this chapter.
(b) Given a $\mathbb{R}^m$-valued quasi-left continuous square integrable martingale M with integer-valued measure ν_M and compensated martingale random measure $\tilde\nu_M$ in the filtered space $(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \geq 0)$, we denote by $L^2(\tilde\nu_M)$ or $L^2(\Omega,\mathcal{F},P,\mathcal{F}(t) : t \geq 0, M, \tilde\nu_M)$ the equivalence class with respect to the

completion of product measure $\tilde{\nu}_M \times P$ of real-valued square integrable predictable processes X, i.e., as a closed subspace of the Hilbert space $L^2(\mathbb{R}^m_* \times [0,\infty) \times \Omega, \overline{\mathcal{B}(\mathbb{R}^m_*) \times \mathcal{P}}, \tilde{\nu}_M \times P)$, where $\mathcal{B}(\mathbb{R}^m_*)$ is the Borel σ-algebra in $\mathbb{R}^m_*$ and the *bar* means completion with respect to the product measure $\tilde{\nu}_M \times P$. If an integer-valued random measure ν is initially given with compensated martingale random measure $\tilde{\nu} = \nu - \nu^p$, where ν^p is the predictable compensator satisfying $\nu^p(\mathbb{R}^m_* \times \{t\}) = 0$ for every $t \geq 0$, then we use the notation $L^2(\tilde{\nu})$ or $L^2(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0, \tilde{\nu}_M)$. Moreover, the same applies if a predictable ν^p-locally integrable density δ is used, i.e., if $\tilde{\nu}$ and ν^p are replaced by $\tilde{\nu}_\delta := \sqrt{\delta}\,\tilde{\nu}$ and $\nu^p_\delta := \delta\,\tilde{\nu}$.

(c) Similarly, localized Hilbert spaces $L^2_{\rm loc}(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0, M, \langle M \rangle)$ or $L^2_{\rm loc}(M)$ and $L^2_{\rm loc}(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0, M, \tilde{\nu}_M)$ or $L^2_{\rm loc}(\tilde{\nu}_M)$ are defined. If M is only a local continuous martingale then X in $L^2_{\rm loc}(M)$ means that for some localizing sequence $\{\tau_n : n \geq 1\}$ the process $M_n : t \mapsto M(t \wedge \tau_n)$ is a square integrable martingale and $\mathbb{1}_{]0,\tau_n]}X$ belongs to $L^2(M_n)$, i.e, (4.73) holds for every $n \geq 1$. Similarly, if M is only a local quasi-left continuous square integrable martingale then X in $L^2_{\rm loc}(\tilde{\nu}_M)$ means that for some localizing sequence $\{\tau_n : n \geq 1\}$ the process $M_n : t \mapsto M(t \wedge \tau_n)$ is a square integrable martingale, with compensated martingale random measure denoted by $\tilde{\nu}_{M_n}$, and $\mathbb{1}_{]0,\tau_n]}X$ belongs to $L^2(\tilde{\nu}_{M_n})$, i.e., the M and X share the same localizing sequence of stopping times. □

Notice that we do not include the general case where M is a semi-martingale (in particular, local martingales which are neither quasi-left continuous nor local square integrable), since the passage to include these situation is essentially a pathwise argument covered by the measure theory. If the predictable quadratic variation process $\langle M \rangle$ gives a measure equivalent to the Lebesgue measure $\mathrm{d}\ell$ then the spaces $L^2(M)$ and $L^2_{\rm loc}(M)$ are equals to $\mathbb{P}^p(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$ and $\mathbb{P}^p_{\rm loc}(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, for $p = 2$, as defined at the beginning of this Section 4.2 in the one-dimensional case. If M is a (local) quasi-left continuous square integrable martingale then we can write (uniquely) $M = M^c + M^d$, where M^c is the continuous part and M^d the purely discontinuous part with $M^d(0) = 0$. Then, we may write $L^2_{\rm loc}(M^d) := L^2_{\rm loc}(\tilde{\nu}_{M^d})$, $L^2_{\rm loc}(M) := L^2_{\rm loc}(M^c) + L^2_{\rm loc}(M^d)$, and similarly without the localization. Furthermore, if predictable quadratic co-variation (matrix) process $\langle M \rangle$ or the predictable compensator ν^p is deterministic then the (local) space $L^2_{\rm loc}(M)$ or $L^2_{\rm loc}(\tilde{\nu})$ is characterized by the condition

$$P\Big\{ \int_{]0,t]} \sum_{i,j=1}^{d} f_i(s) f_j(s) \mathrm{d}\langle M_i, M_j \rangle(s) < \infty \Big\} = 1$$

or

$$P\Big\{ \int_{\mathbb{R}^m_* \times]0,t]} f^2(z,s) \nu^p(\mathrm{d}z, \mathrm{d}s) < \infty \Big\} = 1,$$

for every $t > 0$. This applies even if the local martingale M or the integer-valued random measure ν is not quasi-left continuous, in which case the predictable

quadratic co-variation process $\langle M_i, M_j\rangle(s)$ may be discontinuous or the predictable compensator measure ν^p may not vanish on $\mathbb{R}^m_* \times \{t\}$ for some $t > 0$, we must have $\nu^p(\mathbb{R}^m_* \times \{0\}) = 0$.

The Case of Semi-martingales

Another point to stress is the following fact. If M is a n-dimensional continuous local martingale and f is a $d \times n$ matrix-valued process in $L^2_{\rm loc}(M)$, i.e., each columns vector $f_{i\cdot} := (f_{ik} : k = 1, \dots, n)$ belongs to $L^2_{\rm loc}(M)$, then we can define d-dimensional continuous local martingale

$$(f \star M)_i(t) := \sum_{k=1}^{n} \int_0^t f_{ik}(s) \mathrm{d}M_k(s), \quad \forall t \geq 0,$$

and $i = 1, \dots, d$. The predictable quadratic co-variation process becomes

$$\langle (f \star M)_i, (f \star M)_j \rangle = \sum_{k,\ell=1}^{n} f_{ik} \langle M_k, M_\ell \rangle f_{j\ell}\,.$$

On the other hand, if $\tilde\nu$ is a local martingale measure with a predictable compensator ν^p in $\mathbb{R}^m_*$ and g is a d vector-valued process in $L^2_{\rm loc}(\tilde\nu)$, i.e., each component g_i belongs to $L^2_{\rm loc}(\tilde\nu)$, then we can define d-dimensional purely discontinuous local martingale

$$(g \star \tilde\nu)_i(t) := \int_{]0,t]} g_i(\zeta, s) \tilde\nu(\mathrm{d}\zeta, \mathrm{d}s) \quad \forall t \geq 0,$$

and $i = 1, \dots, d$. The local martingale measure $\tilde\nu_{(g\star\tilde\nu)}$ associated with $g \star \tilde\nu$ in $\mathcal{B}(\mathbb{R}^d_*)$ can be expressed as

$$\tilde\nu_{(g\star\tilde\nu)}(B\times]a, b]) = \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_{\{g(\zeta,s)\in B\}} \tilde\nu(\mathrm{d}\zeta, \mathrm{d}s)$$

with its predictable compensator $\nu^p_{(g\star\tilde\nu)}$

$$\tilde\nu^p_{(g\star\tilde\nu)}(B\times]a, b]) = \int_{\mathbb{R}^m_*\times]a,b]} \mathbb{1}_{\{g(\zeta,s)\in B\}} \nu^p(\mathrm{d}\zeta, \mathrm{d}s),$$

for every $b > a \geq 0$ and B in $\mathcal{B}(\mathbb{R}^d_*)$. In short we write $\tilde\nu_{(g\star\tilde\nu)} = g\tilde\nu$ and $\tilde\nu^p_{(g\star\tilde\nu)} = g\tilde\nu^p$. Note that the optional quadratic variation process is given by

$$[(g \star \tilde\nu)_i, (g \star \tilde\nu)_j](t) = \int_{\mathbb{R}^m_*\times]0,t]} g_i(\zeta, s) g_j(\zeta, s) \nu^p(\mathrm{d}\zeta, \mathrm{d}s),$$

for every $t \geq 0$.

Let $g(z, s)$ be a d-dimensional predictable process which is integrable in $\mathbb{R}^m_*$ with respect to the measure $\nu^p(\mathrm{d}z, \mathrm{d}s)$ almost surely, i.e.,

$$P\Big\{ \int_{\mathbb{R}^m_*\times]0,t]} |g(z, s)| \nu^p(\mathrm{d}z, \mathrm{d}s) < \infty \Big\} = 1, \quad \forall t > 0,$$

which is a classic pointwise integral in the Lebesgue-Stieltjes. Moreover, if $\{(\zeta_n, \tau_n) : n = 1, 2, \ldots\}$ are the atoms of ν (i.e., its associated point process) then

$$\begin{aligned}\int_{\mathbb{R}_*^m \times]0,t]} |g(z,s)| \nu^p(\mathrm{d}z, \mathrm{d}s) &= \int_{\mathbb{R}_*^m \times]0,t]} |g(z,s)| \nu(\mathrm{d}z, \mathrm{d}s) = \\ &= \sum_{0<\tau_n \leq t} |g(\zeta_n, \tau_n)|.\end{aligned}$$

Since

$$\sum_{0<\tau_n \leq t} |g(\zeta_n, \tau_n)|^2 \leq \max_{0<\tau_n \leq t} |g(\zeta_n, \tau_n)| \sum_{0<\tau_n \leq t} |g(\zeta_n, \tau_n)|,$$

the process $g(z,s)$ also belongs to $L^2_{\mathrm{loc}}(\tilde{\nu})$ and we have

$$\begin{aligned}\int_{\mathbb{R}_*^m \times]0,t]} g(z,s) \nu(\mathrm{d}z, \mathrm{d}s) = \int_{\mathbb{R}_*^m \times]0,t]} g(z,s) \tilde{\nu}(\mathrm{d}z, \mathrm{d}s) + \\ + \int_{\mathbb{R}_*^m \times]0,t]} g(z,s) \nu^p(\mathrm{d}z, \mathrm{d}s),\end{aligned}$$

for every $t > 0$.

When comparing both stochastic integrals, with respect to (1) a continuous local martingale (typically a Wiener process) and (2) a quasi-left continuous (cad-lag) purely jump local martingale (typically a Poisson compensated-jump or martingale measure) we have two notations, which are different only in form. If $w = (w_k(t) : t \geq 0,\ k \geq 1)$ is a (standard) Wiener process and $\sigma = (\sigma_k(s) : s \geq 0,\ k \geq 1)$ is a adapted process then

$$(\sigma \star w)_t := \sum_k \int_0^t \sigma_k(s) \mathrm{d}w_k(s)$$

makes sense as long as

$$\sum_k \int_0^t |\sigma_k(s)|^2 \mathrm{d}s < \infty, \quad \forall t \geq 0,$$

almost surely. On the other hand, if $\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)$ is a (standard) Poisson martingale measure with Lévy measure and $\gamma = (\gamma(\zeta, s) : s \geq 0,\ \zeta \in \mathbb{R}_*^m)$ is a adapted process then

$$(\gamma \star \tilde{\nu})_t := \int_{\mathbb{R}_*^m \times]0,t]} \gamma(\zeta, s) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)$$

makes sense as long as

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}_*^m} |\gamma(\zeta, s)|^2 \pi(\mathrm{d}\zeta) < \infty, \quad \forall t \geq 0,$$

almost surely. At this point it is clear the role of the parameters k and ζ in the integrands $\sigma_k(\cdot)$ and $\gamma(\zeta, \cdot)$, i.e., the sum in k and the integral in ζ with respect

to the Lévy measure $m(\cdot)$. Moreover, the integrands σ and γ can be considered as ℓ^2-valued processes, i.e.,

$$\sum_k |\sigma_k|^2 < \infty \quad \text{and} \quad \int_{\mathbb{R}_*^m} |\gamma(\zeta)|^2 \pi(\mathrm{d}\zeta) < \infty,$$

so that the parameters k and ζ play similar roles. The summation in k can be converted to an integral and the separable locally compact and locally convex space $\mathbb{R}_*^m$ can be replaced by any Polish (or Backwell) space.

In general, if the (local) martingale measure $\tilde{\nu}$ is known then the Lévy measure $\pi(\cdot)$ is found as its predictable quadratic variation, and therefore ν is constructed as the integer measure associated with the compensated-jump process

$$\tilde{p}(t) := \int_{\mathbb{R}_*^m \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall t \geq 0.$$

Hence, the integer measure ν, the (local) martingale measure $\tilde{\nu}$ and the $\mathbb{R}^m$-valued compensated-jump process $\tilde{p}$ can be regarded as different viewpoints of the same concept. Each one of them completely identifies the others.

To conclude this section we mention that any quasi-left continuous (cadlag) semi-martingale X can be expressed in a unique way as $X(t) = X(0) + A(t) + M(t) + z \star \tilde{\nu}_X$, where $X(0)$ is a $\mathcal{F}(0)$-measurable random variable, $A(0) = M(0) = 0$, A is a continuous process with locally integrable bounded variation paths, M is a continuous local martingale, and $z \star \tilde{\nu}_X$ is the stochastic integral of the process $(z, t, \omega) \mapsto z$ with respect to the local martingale measure $\tilde{\nu}_X$ associated with X.

4.2.6 Representation of Martingales

Given a one-dimensional standard Brownian motion $(w(t) : t \geq 0)$ we define as in Section 4.2.2 the stochastic integral for processes in the class $L^2_{\mathrm{loc}}(w)$ (or $\bar{\mathcal{E}}_{\mathrm{loc}}$), i.e., $\mathrm{d}\ell \times P(\mathrm{d}\omega)$-equivalence class of adapted processes with

$$P\Big\{ \int_0^t f^2(s)\mathrm{d}s < \infty \Big\} = 1, \quad \forall t \geq 0,$$

and $\mathrm{d}\ell$ the Lebesgue measure. The stochastic integral process $I(t, f)$,

$$t \mapsto \int_0^t f(s)\mathrm{d}w(s), \quad \forall t \geq 0$$

is a continuous local martingale with quadratic variation process

$$[I(\cdot, f)](t) = \langle I(\cdot, f) \rangle(t) = \int_0^t f^2(s)\mathrm{d}s,$$

which is an absolutely continuous function of t for almost every ω. Our first interest is in the converse of this statement as introduced by Doob [59].

It is clear that a $\mathbb{R}^d$-valued stochastic process $M = (M_1(t), \ldots, M_d(t))$ is called a d-dimensional continuous local martingale if each component $(M_i(t) :$

$t \geq 0$), $i = 1, \ldots, d$ is a (real-valued) continuous local martingale relative to (complete and right-continuous) filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$. Assume that $\langle M_i, M_j \rangle$ is (almost surely) absolutely continuous yielding a symmetric square matrix of constant rank, i.e.,

$$\begin{cases} \langle M_i, M_j \rangle(t) = \displaystyle\int_0^t z_{ij}(s)\mathrm{d}s, \\ z := (z_{ij} : i, j = 1, \ldots, d), \quad \mathrm{rank}(z) = r, \end{cases} \tag{4.77}$$

where $1 \leq r \leq d$ is a constant. The matrix z is symmetric, semi-positive and progressively measurable, and so it can be diagonalized by an orthogonal matrix $q = (q_{ij} : i, j = 1, \ldots, d)$, $q^{-1} = q^*$, $q^* z q = \lambda$, where $\lambda = (\lambda_{ij})$, $\lambda_{ii} > 0$ for $i = 1, \ldots, r$ and $\lambda_{ij} = 0$ otherwise.

The stochastic integral process

$$N_i(t) := \sum_{k=1}^{d} \int_0^t q_{ki}(s)\mathrm{d}M_k(s)$$

satisfies

$$\langle N_i, N_j \rangle(t) = \delta_{ij} \int_0^t \lambda_{ii}\mathrm{d}s,$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, so that $N_i = 0$ for $r < i \leq d$. Defining

$$\begin{cases} X_{ij} := q_{ij}\sqrt{\lambda_{jj}}, \quad i = 1, \ldots, d,\ j = 1, \ldots, r, \\ w_i(t) := \displaystyle\int_0^t \frac{1}{\sqrt{\lambda_{ii}}(s)}\mathrm{d}N_i(s), \quad i = 1, \ldots, r, \end{cases} \tag{4.78}$$

we deduce that $w = (w_1, \ldots, w_r)$ is a r-dimensional standard Wiener process such that

$$M_i(t) = \sum_{j=1}^{r} \int_0^t X_{ij}\mathrm{d}w_i(s), \quad \forall t \geq 0,\ i = 1, \ldots, d.$$

The condition on the constant rank r can be removed and the Wiener process is constructed in an *extension* of the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$.

Another interesting point is the time-change. Let $M = (M_1(t), \ldots, M_d(t))$ be a d-dimensional continuous local martingale and assume that

$$\begin{cases} \lim\limits_{t \to \infty} \langle M_i \rangle(t) = \infty, \quad \text{a.s.}, \\ \langle M_i, M_j \rangle = 0, \quad \forall i \neq j. \end{cases} \tag{4.79}$$

Certainly, this is not enough to ensure that M_i is independent of M_j for $i \neq j$. However, the time-change

$$\begin{cases} \tau_i(s) := \inf\big\{t \geq 0 : \langle M_i \rangle(t) > s\big\}, \\ B_i(s) := M_i(\tau_i(s)), \quad s \geq 0,\ i = 1, \ldots, d, \end{cases} \tag{4.80}$$

are independent, standard, one-dimensional Brownian motions.

Detail on the above statements and a proof of the following result can be found in Karatzas and Shreve [124, Section 3.4, pp. 170–190]. A representation result for continuous martingales is as follows:

Theorem 4.22. *Let $w = (w_1(t), \dots, w_d(t))$ be a d-dimensional standard Wiener process in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ where the filtration $(\mathcal{F}(t) : t \geq 0)$ is the increasing right-continuous and completed family of σ-algebras generated by w. Then any square-integrable (cad-lag) martingale $X = (X(t) : t \geq 0)$ in the filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ is a stochastic integral with respect to w and therefore continuous, i.e., there exist a unique element $f = (f_1, \dots, f_d)$ in Hilbert space $L^2(w)$ or $L^2(\Omega \times [0, \infty), \bar{\mathcal{P}}, P \times \mathrm{d}\ell)$, where $\bar{\mathcal{P}}$ is the completion of the predictable (or progressively) measurable processes with respect to the product measure $P \times \mathrm{d}\ell$, such that*

$$X(t) = \sum_{i=1}^{d} \int_0^t f_i(s) \mathrm{d}w_i(s),$$

for every $t \geq 0$. □

As proved in Jacod and Shiryaev [117, Section II.4c, pp. 172–179] the above representation can be extended to some local martingales, other than the Wiener process. In particular, if $M = (M_1, \dots, M_d)$ is a d-dimensional continuous local martingale and ν is a quasi-left continuous integer-valued random measure in $\mathbb{R}_*^m$, both defined in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ where the filtration $(\mathcal{F}(t) : t \geq 0)$ is the augmented σ-algebras generate by M and ν (i.e., completed, right-continuous and quasi-left continuous). Then any (local) square-integrable (cad-lag) martingale X with $X(0) = 0$ is quasi-left continuous and can be represented as the following stochastic integral

$$X(t) = \sum_{i=1}^{d} \int_{]0,t]} f_i(s) \mathrm{d}M_i(s) + \int_{\mathbb{R}_*^m \times]0,t]} g(z,s)\, \tilde{\nu}(\mathrm{d}z, \mathrm{d}s), \tag{4.81}$$

for every $t \geq 0$, where $f = (f_i, \dots, f_d)$ and g belong to $L^2(M)$ and $L^2(\tilde{\nu})$ (or $L^2_{\mathrm{loc}}(M)$ and $L^2_{\mathrm{loc}}(\tilde{\nu})$), respectively, see Definition 4.21.

It is clear that for any $\mathbb{R}^d$-valued (local) martingale M with $M(0) = 0$ we can define the optional quadratic co-variation

$$[M_i, M_j] := \langle M_i^c, M_j^c \rangle + \\ + \sum_{0 < s \leq \cdot} \big(M_i(s) - M_i(s-)\big)\big(M_j(s) - M_j(s-)\big),$$

where the predictable quadratic variation $\langle M_i^c, M_j^c \rangle$ (of the continuous part (local) martingale M^c of M) is the unique adapted continuous process with local bounded variation paths such that $M_i^c\, M_j^c - \langle M_i^c, M_j^c \rangle$ is a continuous martingale. Also, an (adapted) integer random measure $\nu = \nu_M$ (sometime

denoted by M_ν) is associated with the jumps of M as the numbers of jumps with values in a Borel subset of $\mathbb{R}^d_* := \mathbb{R}^d \setminus \{0\}$, i.e.,

$$\nu(B,]a, b]) := \sum_{a<t\le b} \mathbb{1}_{M(t)-M(t-)\in B}, \quad \forall B \in \mathcal{B}(\mathbb{R}^d_*),\ b > a \ge 0,$$

with its predictable compensator $\nu^p_M = \nu^p$ (sometime denoted by M^p_ν), defined as the unique predictable random measure such that $\mathbb{E}\{\nu(K\times]0,\tau]) - \nu^p(K\times]0,\tau])\} = 0$ for every compact subset K of $\mathbb{R}^d_*$ and any bounded stopping time τ or equivalently the process $t \mapsto \nu(K\times]0,t]) - \nu^p(K\times]0,t])$ is a martingale. Thus, defining $\tilde\nu := \nu - \nu^p$ as its associated (local) martingale measure we can write

$$M(t) = M^c(t) + \int_{\mathbb{R}^d_*\times]0,t]} z\tilde\nu(\mathrm{d}z, \mathrm{d}s) + \sum_{0<s\le t} \nu^p(\mathbb{R}^d_*, \{s\}),$$

for every $t \ge 0$. If the martingale M is locally square-integrable then the processes

$$(t,\omega) \mapsto \int_{\mathbb{R}^d_*\times]0,t]} |z| \wedge |z|^2 \nu^p(\mathrm{d}z, \mathrm{d}s),$$
$$(t,\omega) \mapsto \sum_{0<s\le t} \nu^p(\mathbb{R}^d_*, \{s\})$$

are locally integrable. Clearly, we have

$$\sum_{0<s\le t} \varphi\big(M(s) - M(s-)\big) = \int_{\mathbb{R}^d_*\times]0,t]} \varphi(z)\nu(\mathrm{d}z, \mathrm{d}s),$$

and the adapted stochastic process

$$(t,\omega) \mapsto \Big[\int_{\mathbb{R}^d_*\times]0,t]} \varphi(z)\tilde\nu(\mathrm{d}z, \mathrm{d}s)\Big]^2 - \int_{\mathbb{R}^d_*\times]0,t]} \varphi(z)\nu^p(\mathrm{d}z, \mathrm{d}s)$$

is a local martingale for any measurable function φ such that

$$\mathbb{E}\Big\{\int_{\mathbb{R}^d_*\times]0,t]} |\varphi(z)|^2 \nu^p(\mathrm{d}z, \mathrm{d}s)\Big\} < \infty, \quad \forall t > 0.$$

The (local) martingale M is quasi-left continuous if and only if $\nu^p(\mathbb{R}^d_*, \{s\}) = 0$ for every $s > 0$. Therefore, the optional quadratic variation can be written in term of the integer measure ν_M as

$$[M_i, M_j] = \langle M^c_i, M^c_j\rangle + \int_{\mathbb{R}^d_*\times]0,\cdot]} z_i\, z_j \nu_M(\mathrm{d}z, \mathrm{d}s),$$

and the predictable quadratic variation $\langle M^c_i, M^c_j\rangle$ and the compensator ν^p_M are uniquely determined by the local martingale M, so-called the *characteristic* of the martingale M.

The reader may find suitable representation theorems in the book Ikeda and Watanabe [110, Section II.7, pp. 84–96]. We will quote some of them. To this end, we first recall the meaning of *standard extension* of a filtered probability space.

Definition 4.23. A filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}, \tilde{\mathcal{F}}(t) : t \geq 0)$ is called an *extension* of another filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ if there exists a measurable mapping $\Upsilon : \tilde{\Omega} \to \Omega$ such that:
(1) $P(A) = \tilde{P}(\Upsilon^{-1}(A))$ for every A in $\mathcal{F}$,
(2) $\Upsilon^{-1}(\mathcal{F}(t)) \subset \tilde{\mathcal{F}}(t)$, for every $t \geq 0$,
(3) $\mathbb{E}\{X \mid \mathcal{F}(t)\} \circ \Upsilon = \tilde{\mathbb{E}}\{X \circ \Upsilon : \mid \tilde{\mathcal{F}}(t)\}$, $\tilde{P}$-almost surely in $\tilde{\Omega}$, for every $t \geq 0$ and any bounded random variable X in Ω, where $(X \circ \Upsilon)(\tilde{\omega}) = X(\Upsilon(\tilde{\omega}))$.
This extension is called *standard* if $\tilde{\Omega} = \Omega \times \Omega'$, $\tilde{\mathcal{F}} = \mathcal{F} \times \mathcal{F}'$, $\tilde{P} = P \times P'$, and $\Upsilon(\tilde{\omega}) = \omega$ for every $\tilde{\omega} = (\omega, \omega')$ in $\tilde{\Omega}$. All product σ–algebras are completed if necessary and any random variable X in Ω are identified with $X \circ \Upsilon$ as a random variable in $\tilde{\Omega}$. □

The following is Theorem 7.1' and a re-phasing of Theorem 7.4' in the above cited reference.

Theorem 4.24. *Let M be a d-dimensional continuous local martingale in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ such that there exist predictable processes $\sigma_k = (\sigma_{ik}(t) : i = 1, \ldots, d)$, $k = 1, \ldots, n$ satisfying*

$$\langle M_i, M_j \rangle = \sum_k \int_0^{\cdot} \sigma_{ik}(s)\sigma_{jk}(s)\mathrm{d}s,$$
$$P\Big\{ \sum_{ik} \int_0^t |\sigma_{ik}(s)|^2 \mathrm{d}s < \infty \Big\} = 1.$$

Then there exists a standard extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}, \tilde{\mathcal{F}}(t) : t \geq 0)$ of the filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ and a n-dimensional Wiener process $w = (w_1, \ldots, n)$ such that

$$\mathrm{d}M(t) = \sum_k \sigma_k(t)\mathrm{d}w_k(t)$$

or equivalently

$$M_i(t) = M_i(0) + \sum_k \int_0^t \sigma_{ik}(s)\mathrm{d}w_k(s),$$

for every $t \geq 0$. □

Recall that for any d-dimensional local martingale M we can associate an integer random measure

$$M_\nu(K,]0, t]) := \sum_{0 < s \leq t} \mathbb{1}_{M(s) - M(s-) \in K}$$

and there exists a unique random measure M_ν^p, which is called *jump compensator* and characterized by the following property:
(a) For every compact subset K of $\mathbb{R}^d_*$ the process

$$t \mapsto M_\nu(K,]0, t]) - M_\nu^p(K,]0, t])$$

is a local martingale

A quasi-left continuous local martingale M is called purely jumps if

$$M(t) = M(0) + \int_{\mathbb{R}^d_* \times]0,t]} z\tilde{M}_\nu(\mathrm{d}z, \mathrm{d}s),$$

for every $t > 0$, where $\tilde{M}_\nu = M_\nu - M^p_\nu$ is its martingale measure.

Theorem 4.25. *Let M be a quasi-left continuous purely jumps local martingale in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ such that there exist a Radon (i.e., finite on every compact) measure $\pi(\cdot)$ in $\mathbb{R}^m_*$ and a predictable process $\gamma = (\gamma_i(\zeta, t) : i = 1, \ldots, d)$, ζ in $\mathbb{R}^m_*$ satisfying*

$$M^p_\nu(B,]0, t]) = \int_0^t \pi\big(\{\zeta \in \mathbb{R}^m_* : \gamma(\zeta, s) \in B\}\big)\mathrm{d}s,$$
$$P\Big\{ \int_{\mathbb{R}^m_*} |\gamma(\zeta, t)|^2 \pi(\mathrm{d}\zeta) < \infty \Big\} = 1,$$

for every B in $\mathcal{B}(\mathbb{R}^m_*)$ and $t > 0$, where M^p_ν is the jump compensator of M. Then there exists a standard extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}, \tilde{\mathcal{F}}(t) : t \geq 0)$ of the filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ and a m-dimensional Poisson measure ν on $\mathbb{R}^m_*$ with Lévy measure $\pi(\cdot)$ such that

$$\mathrm{d}M(t) = \int_{\mathbb{R}^m_*} \gamma(\zeta, t)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t)$$

or equivalently

$$M_i(t) = M_i(0) + \int_{\mathbb{R}^m_* \times]0,t]} \gamma_i(\zeta, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)$$

for every $t \geq 0$, where $\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) := \nu(\mathrm{d}\zeta, \mathrm{d}s) - \pi(\mathrm{d}\zeta)\mathrm{d}s$ is the martingale measure associated with ν. □

The proof of the above results is rather technical and we refer to the previous book and its references. For instance, the reader may consult the papers Cinlar and Jacod [47] or Yin and SiTu [260], for a more detailed discussion.

4.3 Stochastic Differential

One of the most important tools used with stochastic integrals is the *change-of-variable rule* or better known as *Itô's formula*. This provides an integral-differential calculus for the sample paths.

To motivate our discussion, let us recall that at the end of Subsection 4.2.2 we established the identity

$$\int_{(0,t]} w(s)\mathrm{d}w(s) = \frac{w^2(t)}{2} - \frac{t}{2}, \quad \forall t \geq 0,$$

for a real-valued standard Wiener process $(w(t) : t \geq 0)$, where the presence of new term, $t/2$, is noticed, with respect to the classic calculus.

In general, Fubini's theorem proves that given two processes X and Y of locally bounded variation (cad-lag) we have the integration-by-part formula

$$\left\{ \begin{array}{l} X(b)Y(b) - X(a)Y(a) = \displaystyle\int_{(a,b]} X(t-)\mathrm{d}Y(t) + \\ \qquad + \displaystyle\int_{(a,b]} Y(t-)\mathrm{d}X(t) + \sum_{a<t\leq b} \delta X(t)\, \delta Y(t), \end{array} \right. \tag{4.82}$$

where $X(t-)$ and $Y(t-)$ are the left-limits at t and δ is the jump-operator, e.g., $\delta X(t) := X(t) - X(t-)$. Since the integrand $Y(t-)$ is left-continuous and the integrator $X(t)$ is right-continuous as above, the pathwise integral can be interpreted in the Riemann-Stieltjes sense or the Lebesgue-Stieltjes sense, indistinctly. Consider, for example, a Poisson process with parameter $c > 0$, i.e., $X = Y = (p(t) : t \geq 0)$, we have

$$\int_{(0,t]} p(s-)\mathrm{d}p(s) = \frac{p^2(t)}{2} - \frac{p(t)}{2}, \quad \forall t \geq 0,$$

because all jumps are equals to 1. However, strictly in the Lebesgue-Stieltjes sense we write

$$\int_{(0,t]} p(s)\mathrm{d}p(s) = \frac{p^2(t)}{2} + \frac{p(t)}{2}, \quad \forall t \geq 0.$$

Recall that the stochastic integral is initially defined as the L^2-limit of Riemann-Stieltjes sums, where the integrand is a predictable (essentially, left-continuous having right-limits) process and the integrator is a (local) square integrable martingale. The (local) bounded variation integral can be defined by either way with a unique value, as long as the integrand is the predictable member of its equivalence class of processes. Thus, as mentioned at the end of Subsection 4.2.4, the stochastic integral with respect to the compensated Poisson process (or martingale) $\bar{p}(t) := p(t) - ct$ satisfies

$$\int_{(0,t]} \bar{p}(s)\mathrm{d}\bar{p}(s) = \int_{(0,t]} \bar{p}(s-)\mathrm{d}\bar{p}(s), \quad \forall t \geq 0,$$

the expression in left-hand side is strictly understood only as a stochastic integral, because it makes non sense as a pathwise Riemann-Stieltjes integral and does not agree with one in the pathwise Lebesgue-Stieltjes sense. However, the expression in right-hand side can be interpreted either as a pathwise Riemann-Stieltjes integral or as a stochastic integral. Notice that the processes $(p(t) : t \geq 0)$ and $(p(t-) : t \geq 0)$ belong to the same equivalence class for the $\mathrm{d}t \times P(\mathrm{d}\omega)$ measure, under which the stochastic integral is defined.

We may calculate the stochastic integral as follows. For a given partition $\pi := (0 = t_0 < t_1 < \cdots < t_n = t)$ of $[0, t]$, with $\|\pi\| := \max_i(t_i - t_{i-1})$, consider

the Riemann-Stieltjes sums

$$S_\pi := \sum_{i=1}^{n} \bar{p}(t_{i-1})[\bar{p}(t_i) - \bar{p}(t_{i-1})] = \int_{]0,t]} \bar{p}_\pi(s)\mathrm{d}\bar{p}(s) = \\ = \int_{]0,t]} \bar{p}_\pi(s)\mathrm{d}p(s) - c\int_0^t \bar{p}_\pi(s)\mathrm{d}s,$$

for the predictable process $\bar{p}_\pi(s) = \bar{p}(t_{i-1})$ for any s in $]t_{i-1}, t_i]$. Since $\bar{p}_\pi(s) \to \bar{p}(s-)$ as $\|\pi\| \to 0$, we obtain

$$\int_{(0,t]} \bar{p}(s-)\mathrm{d}\bar{p}(s) = \int_{(0,t]} \bar{p}(s-)\mathrm{d}p(s) - c\int_0^t \bar{p}(s-)\mathrm{d}s,$$

which is a martingale null at time zero. For instance, because $\mathbb{E}\{p(t)\} = ct$ and $\mathbb{E}\{[p(t) - ct]^2\} = ct$ we have $\mathbb{E}\{p^2(t)\} = c^2t^2 + ct$, and therefore

$$\mathbb{E}\Big\{ \int_{(0,t]} p(s-)\mathrm{d}p(s) - c\int_0^t p(s-)\mathrm{d}s \Big\} = 0,$$

as expected.

Given a smooth real-valued function $\varphi = \varphi(t, x)$ defined on $[0, T] \times \mathbb{R}^d$ and a $\mathbb{R}^d$-valued semi-martingale $\{M(t) : t \geq 0\}$ we want to discuss the stochastic chain-rule for the real-valued process $\{\varphi(t, M(t)) : t \geq 0\}$. If φ is complex-valued then we can tread independently the real and the imaginary parts.

For a real-valued Wiener process $(w(t) : t \geq 0)$, we have deduced that

$$w^2(t) = 2\int_{(0,t]} w(s)\mathrm{d}w(s) + t, \quad \forall t \geq 0,$$

so that the standard chain-rule does not apply. This is also seen when Taylor formula is used, say taking mathematical expectation in

$$\varphi(w(t)) = \varphi(0) + \varphi'(0)w(t) + \varphi''(0)\frac{w^2(t)}{2} + \int_0^1 \varphi'''(sw(t))\frac{w^3(t)}{6}\mathrm{d}s,$$

we obtain

$$\mathbb{E}\varphi(w(t)) = \varphi(0) + \varphi''(0)\frac{t}{2} + \int_0^1 \mathbb{E}\{\varphi'''(sw(t))\frac{w^3(t)}{6}\}\mathrm{d}s,$$

where the error-term integral can be bounded by $2t^{3/2}\sup|\varphi|$. The second order derivative produces a term of order 1 in t.

Given a (cad-lag) locally integrable bounded variation process $A = (A(t) : t \geq 0)$ and a locally integrable process $X = (X(t) : t \geq 0)$ with respect to A, we can define the pathwise Lebesgue-Stieltjes integral

$$(X \star A)(t) := \int_{]0,t]} X(s)\mathrm{d}A(s), \quad \forall\, t \geq 0,$$

which produces a new (cad-lag) locally integrable bounded variation process $X \star A = ((X \star A)(t) : t \geq 0)$. The substitution formula establishes that for any

locally integrable process $Y = (Y(t) : t \geq 0)$ with respect to $X \star A$, the process $YX = (Y(t)X(t) : t \geq 0)$ is locally integrable process with respect to A and

$$\int_{]0,t]} Y(s)\mathrm{d}(X \star A)(s) = \int_{]0,t]} Y(s)X(s)\mathrm{d}A(s), \tag{4.83}$$

for every $t \geq 0$. Certainly, if the processes X and Y are left-continuous then the above integral can be interpreted in the (pathwise) Riemann-Stieltjes sense. Moreover, if both processes X and Y are predictable and A is adapted then the $\star$ symbol, representing the pathwise Lebesgue-Stieltjes, can be replaced by the $\diamond$ symbol, representing the stochastic integral relative to an adapted (cad-lag) process with locally integrable bounded variation.

Similarly, given a (cad-lag) local martingale $M = (M(t) : t \geq 0)$ and a locally integrable predictable process $X = (X(t) : t \geq 0)$ relative to M (i.e., there is a reducing sequence of stopping times $(\tau_n : n \geq 0)$ for both processes X and M, simultaneously), we can define the stochastic integral which produces a new (cad-lag) local martingale $X \diamond M = ((X \diamond M)(t) : t \geq 0)$. Let $Y = (Y(t) : t \geq 0)$ be a locally integrable predictable process relative to $X \diamond M$ (i.e., there is another reducing sequence of stopping times $(\bar{\tau}_n : n \geq 0)$ for both processes Y and $X \diamond M$). The stochastic substitution formula says that the predictable process $YX = (Y(t)X(t) : t \geq 0)$ is locally integrable with respect to M admitting the (minimum) reducing sequence $(\tau_n \wedge \bar{\tau}_n : n \geq 0)$ and

$$\int_{]0,t]} Y(s)\mathrm{d}(X \diamond M)(s) = \int_{]0,t]} Y(s)X(s)\mathrm{d}M(s), \tag{4.84}$$

for every $t \geq 0$.

The first step in the proof of the above stochastic substitution formula is to observe that by taking the minimum localizing sequence $(\tau_n \wedge \bar{\tau}_n : n \geq 0)$ it suffices to show the result for an L^2-martingales M. Secondly, it is clear that equality (4.84) holds for any elementary predictable processes Y and that because of the isometry

$$\int_{]0,t]} Y^2(s)\mathrm{d}[X \diamond M](s) = \int_{]0,t]} Y^2(s)X^2(s)\mathrm{d}[M](s), \quad \forall\, t \geq 0,$$

for every $t \geq 0$, where $[\cdot]$ denotes the (optimal) quadratic variation of a martingale (as in Section 4.2.4), the process YX is integrable with respect to M. Finally, by passing to the limit we deduce that (4.84) remains valid almost surely for every $t \geq 0$. Since both sides of the equal sign are cad-lag processes, we conclude. A detailed proof can be found in Chung and Williams [45, Theorem 2.12, Section 2.7, pp. 48–49].

Let M be a (real-valued) square integrable martingale with its associated optional and predictable integrable monotone increasing processes $[M]$ and $\langle M \rangle$. Recall that $M^2 - [M]$ and $M^2 - \langle M \rangle$ are uniformly integrable martingale, $[M](t) := \langle M_c \rangle(t) + \sum_{s \leq t} [M(s) - M(s-)]^2$, where M_c is the continuous part of M. Moreover, if $\langle M \rangle$ is continuous (i.e., the martingale is quasi-left continuous)

and ${}^{\mathrm{p}}\mathrm{var}_2(M,\pi)$ denotes the predictable quadratic variation operator defined by

$$ {}^{\mathrm{p}}\mathrm{var}_2(M,\pi_t) := \sum_{i=1}^{m} \mathbb{E}\{|M(t_i) - M(t_{i-1})|^2 \,|\, \mathcal{F}(t_{i-1})\}, \tag{4.85} $$

for $\pi_t = (0 = t_0 < t_1 < \cdots < t_m = t)$, then ${}^{\mathrm{p}}\mathrm{var}_2(M,\pi)$ converges in L^1 to $\langle M\rangle$ as the mesh (or norm) of the partition $\|\pi_t\| := \max_k(t_i - t_{i-1})$ goes to zero, see Theorem 3.11 in Chapter 3. Another key point is the study of the variation of M, as defined by the operator

$$ \mathrm{var}_\ell(M,\pi_t) := \sum_{i=1}^{m} |M(t_i) - M(t_{i-1})|^\ell, \tag{4.86} $$

as the mesh $\|\pi\|$ vanishes, the cases $\ell = 2$ (quadratic variation) and $\ell = 4$ are of particular interest. As we have seen, the quadratic variation plays an important role in the stochastic integral.

Proposition 4.26 (Quadratic Variation Convergence)**.** *If M is a (real-valued) continuous square integrable martingale then for every $\varepsilon, t > 0$ there exists $\delta > 0$ such that for any partition π_t of the interval $[0,t]$ with $\|\pi_t\| < \delta$ we have $P\{|\mathrm{var}_2(M,\pi_t) - \langle M\rangle(t)| > \varepsilon\} < \varepsilon$.*

Proof. We only give some details for the case when M is continuous and bounded in L^4. Indeed, the martingale property yields

$$ \begin{aligned} &\sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \,|\, \mathcal{F}(t_{i-1})\} = \\ &\qquad = \sum_{i=k+1}^{m} \mathbb{E}\{M^2(t_i) - M^2(t_{i-1}) \,|\, \mathcal{F}(t_{i-1})\} \le \mathbb{E}\{M^2(t_m)|\, \mathcal{F}(t_k)\}, \end{aligned} $$

so that

$$ \begin{aligned} &\sum_{k=1}^{m-1} \sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2\, [M(t_k) - M(t_{k-1})]^2\} = \\ &\quad = \sum_{k=1}^{m-1} \mathbb{E}\Big\{[M(t_k) - M(t_{k-1})]^2 \times \\ &\qquad\qquad \times \sum_{i=k+1}^{m} \mathbb{E}\{[M(t_i) - M(t_{i-1})]^2 \,|\, \mathcal{F}(t_{i-1})\}\Big\} \le \\ &\quad \le \sum_{k=1}^{m-1} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^2 \sum_{i=k+1}^{m} \mathbb{E}\{M^2(t_m)|\, \mathcal{F}(t_k)\}\} = \\ &\quad = \mathbb{E}\{M^2(t_m)\} \sum_{k=1}^{m-1} \mathbb{E}\{[M(t_k) - M(t_{k-1})]^2\}. \end{aligned} $$

Since

$$\sum_{k=1}^{m}\mathbb{E}\{[M(t_k)-M(t_{k-1})]^4\}\leq$$
$$\leq\mathbb{E}\Big\{\Big(\max_i[M(t_i)-M(t_{i-1})]^2\Big)\sum_{k=1}^{m}[M(t_k)-M(t_{k-1})]^2\Big\}\leq$$
$$\leq\big(\mathbb{E}\{\max_i[M(t_i)-M(t_{i-1})]^4\}\big)^{1/2}\times$$
$$\times\big(\mathbb{E}\big\{\big[\sum_{k=1}^{m}[M(t_k)-M(t_{k-1})]^2\big]^2\big\}\big)^{1/2},$$

we deduce

$$\mathbb{E}\{[\mathrm{var}_2(M,\pi_t)]^2\}=\sum_{k=1}^{m}\mathbb{E}\{[M(t_k)-M(t_{k-1})]^4\}+$$
$$+2\sum_{k=1}^{m-1}\sum_{i=k+1}^{m}\mathbb{E}\{[M(t_i)-M(t_{i-1})]^2[M(t_k)-M(t_{k-1})]^2\}\leq$$
$$\leq 2\mathbb{E}\{M^2(t)\}\,\mathbb{E}\{[\mathrm{var}_2(M,\pi_t)]\}+\mathbb{E}\{\max_i[M(t_i)-M(t_{i-1})]^4\},$$

after using Hölder inequality. This shows that

$$\sup_{0<s\leq t}\mathbb{E}\{|M(s)|^4\}<\infty\quad\Rightarrow\quad\mathbb{E}\{[\mathrm{var}_2(M,\pi_t)]^2\}<\infty,\tag{4.87}$$

and if M is continuous then $\mathbb{E}\{\mathrm{var}_4(M,\pi_t)\}\to 0$ as $\|\pi_t\|\to 0$.

Therefore, because $M^2-\langle M\rangle$ is a martingale we also have

$$\mathbb{E}\{[\mathrm{var}_2(M,\pi_t)-\langle M\rangle(t)]^2\}=$$
$$=\sum_{k=1}^{m}\mathbb{E}\{[(M(t_k)-M(t_{k-1}))^2-(\langle M\rangle(t_k)-\langle M\rangle(t_{k-1}))]^2\}\leq$$
$$\leq 2\sum_{k=1}^{m}\mathbb{E}\{[M(t_k)-M(t_{k-1})]^4-[\langle M\rangle(t_k)-\langle M\rangle(t_{k-1})]^2\}\leq$$
$$\leq 2\mathbb{E}\{\mathrm{var}_4(M,\pi_t)\}+2\mathbb{E}\{\langle M\rangle(t)\max_i[\langle M\rangle(t_i)-\langle M\rangle(t_{i-1})]\},$$

which proves that $\mathrm{var}_2(M,\pi)$ converges in L^2 to $\langle M\rangle$, whenever M is continuous and belongs to L^4. □

For instance, a complete proof of this result can be found in Chung and Williams [45, Theorem 4.1, Section 4.3, pp. 76–79] or Karatzas and Shreve [124, Theorem 5.8, Chapter 1, pp. 32–34].

4.3.1 Itô's processes

Let $(w(t) : t \geq 0)$ be a n-dimensional standard Wiener process in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, i.e., with $w(t) = (w_1(t), \dots, w_n(t))$ we have $w_k(t)$ and $w_k(t)w_\ell(t) - \mathbb{1}_{k=\ell}t$ are continuous martingales null at time zero (i.e., $w_i(0) = 0$) relative to the filtration $(\mathcal{F}_t : t \geq 0)$, for any $k, \ell = 1, \dots, n$. Thus $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$ is called a n-dimensional (standard) Wiener space.

A $\mathbb{R}^d$-valued stochastic process $(X(t) : t \geq 0)$ is called a d-dimensional Itô's process if there exist real-valued adapted processes $(a_i(t) : t \geq 0,\ i = 1, \dots, d)$ and $(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ such that for every $i = 1, \dots, d$ we have

$$\begin{cases} \mathbb{E}\Big\{ \displaystyle\int_0^{\tau_r} \Big[|a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 \Big] \mathrm{d}t \Big\} < \infty, \quad \forall\, r = 1, 2, \dots, \\ X_i(t) = X_i(0) + \displaystyle\int_0^t a_i(s)\mathrm{d}s + \sum_{k=1}^n \int_0^t b_{ik}(s)\mathrm{d}w_k(s), \quad \forall\, t \geq 0, \end{cases} \tag{4.88}$$

in some n-dimensional Wiener space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$, where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. In short we write $\mathrm{d}X(t) = a(t)\mathrm{d}t + b(t)\mathrm{d}w(t)$, for every $t \geq 0$, with a in L^1_{loc} and b in L^2_{loc}. Notice that for a Wiener process or in general for a continuous local martingale M, we write the stochastic integral

$$\int_{]0,t]} f(s)\mathrm{d}M(s) = \int_{(0,t]} f(s)\mathrm{d}M(s) = \int_0^t f(s)\mathrm{d}M(s),$$

indistinctly. Notice that any Itô process is a (special) semi-martingale, and a quasi-martingale whenever a and b belong to $L^1([0,T] \times \Omega)$ and $L^2([0,T] \times \Omega)$, for any $T > 0$, respectively.

Theorem 4.27 (Itô formula)**.** *Let $(X(t) : t \geq 0)$ be a d-dimensional Itô's process in a given Wiener space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \geq 0)$, i.e, (4.88), and $\varphi = \varphi(t, x)$ be a real-valued smooth function on $[0, \infty) \times \mathbb{R}^d$, i.e., C^1 in the first variable t on $[0, \infty)$ and C^2 in the second variable x on $\mathbb{R}^d$. Then $(\varphi(t, X(t)) : t \geq 0)$ is a (real-valued) Itô's process and*

$$\begin{cases} \varphi(t, X(t)) = \varphi(0, X(0)) + \displaystyle\int_0^t A(s, X)\varphi(s, X(s))\mathrm{d}s + \\ \qquad + \displaystyle\sum_{k=1}^n \int_0^t B_k(s, X)\varphi(s, X(s))\mathrm{d}w_k(s), \quad \forall\, t \geq 0, \end{cases} \tag{4.89}$$

where the linear differential operators $A(s, X)$ and $B(s, X) = (B_k(s, X) : k =$

$1, \dots, n)$ *are given by*

$$A(s,X)\varphi(t,x) := \partial_t\varphi(t,x) + \sum_{i=1}^{d} a_i(s)\,\partial_i\varphi(t,x) + \\ + \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sum_{k=1}^{n} b_{ik}(s)b_{jk}(s)\Big)\,\partial^2_{ij}\varphi(t,x),$$

$$B_k(s,X)\varphi(t,x) := \sum_{i=1}^{d} b_{ik}(s)\partial_i\varphi(t,x),$$

for any $s,t \geq 0$ *and* x *in* $\mathbb{R}^d$, *with* ∂_t, ∂_i *and* $\partial^2_{i,j}$ *denoting the partial derivatives with respect to the variable* t, x_i *and* x_j.

Proof. The first step is to localize, i.e., setting

$$T_r := \tau_r \wedge \inf\big\{t \geq 0 : |X(t)| \geq r\big\}$$

we have a non-decreasing sequence of stopping times satisfying $T_r \to \infty$ almost surely. Moreover, if $X_n(t) := X(t \wedge T_n)$ then X_n is a processes with values in the compact ball of radius r and therefore the processes $A(s) := A(s,X)\varphi(s,X_n(s))$ and $B_k(s) := B_k(s,X)\varphi(s,X_n(s))$ are in L^1 and L^2, respectively, i.e.,

$$\mathbb{E}\Big\{\int_0^{T_r}\Big[|A(t)| + \sum_{k=1}^{n}|B_k(t)|^2\Big]\mathrm{d}t\Big\} < \infty, \quad \forall\, r = 1,2,\dots,$$

so that the right-hand side of the so-called Itô formula or rule (4.89) is an real-valued Itô's process. This shows that without loss of generality, we may assume that the function φ has a compact support. Furthermore, details on the proof are only provided for the one-dimensional case, i.e., $d = 1$ and $n = 1$, with $X(t) = X(0) + A(t) + B(t)$ and

$$A(t) := \int_0^t a(s)\mathrm{d}s, \qquad B(t) := \int_0^t b(s)\mathrm{d}w(s), \tag{4.90}$$

$a(s)$ and $b(s)$ are predictable (actually, adapted is sufficient) processes such that

$$|B(t)| + \int_0^t [|a(s)| + |b(s)|^2]\mathrm{d}s \leq C,$$

for any $t \geq 0$ and some deterministic constant $C > 0$.

The second step is to apply Taylor formula for a smooth real-valued function $\varphi = \varphi(x)$ on $\mathbb{R}$, with a partition $\pi := (0 = t_0 < t_1 < \cdots < t_m = t)$ of $[0,t]$,

$$\begin{cases} \varphi(X(t)) - \varphi(X(0)) = \displaystyle\sum_{k=1}^{m}[\varphi(X(t_k)) - \varphi(X(t_{k-1})] = \\ \quad = \displaystyle\sum_{k=1}^{m}[X(t_k) - X(t_{k-1})]\varphi'_k + \frac{1}{2}\sum_{k=1}^{m}[X(t_k) - X(t_{k-1})]^2\varphi''_k, \end{cases} \tag{4.91}$$

where $X(t) = X(0) + A(t) + B(t)$ satisfying (4.90),

$$\varphi'_k := \varphi'(X(t_{k-1})),$$
$$\varphi''_k := \int_0^1 \varphi''((1-s)X(t_{k-1}) + sX(t_k))\mathrm{d}s,$$

and the mesh (or norm) $\|\pi\| := \max_i(t_i - t_{i-1})$ is destined to vanish.

Considering the predictable process $\varphi'_\pi(s) = \varphi'(X(t_{k-1}))$ for s belonging to $]t_{k-1}, t_k]$, we check that

$$\sum_{k=1}^m [X(t_k) - X(t_{k-1})]\varphi'_k = \int_{]0,t]} \varphi'_\pi(s)\mathrm{d}A(s) + \int_{]0,t]} \varphi'_\pi(s)\mathrm{d}B(s),$$

which converges in $L^1 + L^2$ (or pathwise for the first term and L^2 for the second term) to

$$\int_{]0,t]} \varphi'(X(s))\mathrm{d}A(s) + \int_{]0,t]} \varphi'(X(s))\mathrm{d}B(s)$$

where the first integral is in the Riemann-Stieltjes (or Lebesgue-Stieltjes) sense and the second term is a stochastic integral. By means of the substitution formula (4.83) and (4.84), the above limit can be rewritten as

$$\int_{]0,t]} \varphi'(X(s))a(s)\mathrm{d}s + \int_{]0,t]} \varphi'(X(s))b(s)\mathrm{d}w(s),$$

where the first integral is now in the Lebesgue sense, which agrees with the stochastic sense if a predictable version of the integrand is used.

To handle the quadratic variation in (4.91), we notice that

$$[X(t_k) - X(t_{k-1})]^2 = -2[A(t_k) - A(t_{k-1})]\,[B(t_k) - B(t_{k-1})] + \\ +[A(t_k) - A(t_{k-1})]^2 + [B(t_k) - B(t_{k-1})]^2,$$

and for any $k \geq 1$,

$$|\varphi''(X(t_{k-1})) - \varphi''_k| \leq \max_k \rho(\varphi'', |X(t_k) - X(t_{k-1})|),$$

where $\rho(\varphi'', r)$ is the modulus of continuity of φ'', i.e.,

$$\rho(\varphi'', r) := \sup_{|x-y|\leq r} |\varphi''(x) - \varphi''(y)|.$$

Therefore

$$\sum_{k=1}^m [X(t_k) - X(t_{k-1})]^2\varphi''_k = \\ = \sum_{k=1}^m \varphi''(X(t_{k-1}))[B(t_k) - B(t_{k-1})]^2 + \mathbf{o}(\|\pi\|)$$

where

$$|\mathbf{o}(\|\pi\|)| \le \max_k \big\{\rho(\varphi'', |X(t_k) - X(t_{k-1})|)\big\} \times \\ \times \Big\{ \sum_{k=1}^m [B(t_k) - B(t_{k-1})]^2 \Big\} + \\ + \max_k \big\{ [2|B(t_k) - B(t_{k-1})| + |A(t_k) - A(t_{k-1})|]\, |\varphi''_k| \big\} \times \\ \times \Big\{ \sum_{k=1}^m |A(t_k) - A(t_{k-1})| \Big\}.,$$

i.e., $\mathbf{o}(\|\pi\|)$ is bounded by a deterministic constant and $\mathbf{o}(\|\pi\|) \to 0$ as $\|\pi\| \to 0$, almost surely.

Since $\bar\varphi''_k := \varphi''(X(t_{k-1}))$ is $\mathcal{F}(t_k)$-measurable and

$$B^2(t) - \int_0^t |b(s)|^2 \mathrm{d}s,$$

is a martingale, we have

$$\mathbb{E}\Big\{ \Big\{ \sum_{k=1}^m [(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s]\, \bar\varphi''_k \Big\}^2 \Big\} = \\ = \mathbb{E}\Big\{ \sum_{k=1}^m \big[(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s\big]^2 |\bar\varphi''_k|^2 \Big\},$$

which is bounded by the expression

$$\big(\max_i \mathbb{E}\{|\bar\varphi''_k|^2\} \big)\, \mathbb{E}\Big\{ \sum_{k=1}^m \big[(B(t_k) - B(t_{k-1}))^2 - \int_{t_{k-1}}^{t_k} |b(s)|^2 \mathrm{d}s\big]^2 \Big\}.$$

In view of Proposition 4.26, we deduce that

$$\mathbb{E}\Big\{ \Big| \sum_{k=1}^m [B(t_k) - B(t_{k-1})]^2 \varphi''_k - \int_{]0,t]} |b(s)|^2 \bar\varphi''_\pi(s) \mathrm{d}s \Big|^2 \Big\} \to 0,$$

as $\|\pi\| \to 0$, where $\varphi''_\pi(s) = \varphi''(X(t_{k-1})) = \bar\varphi''_k$ for any s in $]t_{k-1}, t_k]$.

Thus, we have establish the one-dimensional Itô formula for a (real-valued) smooth function with compact support $\varphi(x)$, which conclude the proof. □

Notice the short vector notation for Itô formula when $\varphi = \varphi(x)$, namely,

$$\mathrm{d}\varphi(X(t)) = \nabla\varphi(X(t))\mathrm{d}X(t) + \frac{1}{2}\,\mathrm{Tr}[b(t)b^*(t)\nabla^2\varphi(x)]\mathrm{d}t \tag{4.92}$$

for every $t \ge 0$, where ∇ is the gradient operator and $\nabla^2\varphi$ is the matrix of second derivatives.

From the above proof, it is clear also that several generalizations of Itô formula are possible. Notice that it is not necessary to separate the t variable, since we may add one more dimension with $a(s) := 1$ and $b(s) := 0$ to pass from $\varphi(x)$ to $\varphi(t, x)$. By reviewing the previous steps and remarking the use of the

continuity and the quadratic variation of the martingale M, we can show the following rule.

Theorem 4.28. *Let $(X_i(t) : t \geq 0)$ be a continuous semi-martingale in a given filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, for each $i = 1, \dots, d$, and $\varphi = \varphi(x)$ be a real-valued C^2 function on $\mathbb{R}^d$. Then $(\varphi(X(t)) : t \geq 0)$, $X(t) := (X_1(t), \dots, X_d(t))$ is a continuous semi-martingale and*

$$\begin{cases} \varphi(X(t)) = \varphi(X(0)) + \displaystyle\sum_{i=1}^{d} \int_{]0,t]} \partial_i \varphi(X(s)) \mathrm{d}X_i(t) + \\ \qquad + \displaystyle\sum_{i,j=1}^{d} \int_{]0,t]} \partial^2_{ij} \varphi(X(s)) \mathrm{d}\langle X_i, X_j \rangle(s), \quad \forall\, t \geq 0, \end{cases} \tag{4.93}$$

where ∂_i and ∂^2_{ij} denote partial derivatives, and $\langle X_i, X_j \rangle(s)$ is the only predictable process with locally integrable bounded variation such that the expression $X_i X_j - \langle X_i, X_j \rangle$ is a martingale. □

We can also extend the integration-by-part formula (4.82) for two (cad-lag) real-valued semi-martingales $X = V_X + M_X$ and $Y = V_Y + M_Y$ where V_X, V_Y have locally bounded variation and M_X, M_Y are continuous local martingales as follows

$$\begin{cases} X(t)Y(t) - X(0)Y(0) = \langle M_X, M_Y \rangle(t) + \displaystyle\int_{(0,t]} X(s-) \mathrm{d}Y(s) + \\ \qquad + \displaystyle\int_{(0,t]} Y(s-) \mathrm{d}X(s) + \sum_{0<s\leq t} \delta V_X(s)\, \delta V_Y(s), \end{cases} \tag{4.94}$$

for every $t \geq 0$, where $X(t-)$ and $Y(t-)$ are the left limits at t, and δ is the jump-operator, e.g., $\delta X(t) := X(t) - X(t-)$. Notice that the correction term satisfies

$$\langle M_X, M_Y \rangle(t) + \sum_{0<s\leq t} \delta V_X(s)\, \delta V_Y(s) = [X, Y](t),$$

i.e., it is equal to the optional quadratic covariation process $[X, Y]$ associated with the semi-martingale X and Y.

As seen in (4.75) of the previous section, for a standard n-dimensional Wiener process $(w(t) : t \geq 0)$, for any adapted (measurable) process $f(s)$ and for any stopping time T, we can write

$$\mathbb{E}\Big\{ \sup_{0\leq t\leq T} \Big| \int_0^t f(s) \mathrm{d}w(s) \Big|^p \Big\} \leq C_p\, \mathbb{E}\Big\{ \Big[\int_0^T |f(s)|^2 \mathrm{d}s \Big]^{p/2} \Big\}. \tag{4.95}$$

for some constant positive C_p. Actually, for p in $(0, 2]$ the proof is very simple (see (3.30) of Chapter 3) and $C_p = (4 - p)/(2 - p)$ if $0 < p < 2$ and $C_2 = 4$. However, the proof for $p > 2$ involves Burkhölder-Davis-Gundy inequality. An alternative is to use Itô formula for the function $x \mapsto |x|^p$ and the process

$$X(t) := \int_0^t f(s) \mathrm{d}w(s), \quad \forall t \geq 0$$

to get

$$\mathbb{E}\{|X(t)|^p\} = \frac{p(p-1)}{2}\,\mathbb{E}\Big\{\int_0^t |X(s)|^{p-2}|f(s)|^2\mathrm{d}s\Big\}.$$

By means of the Doob's maximal inequality, for some constant $\tilde{C}_p$ depending only on p we have

$$\mathbb{E}\{\sup_{0\le t\le T}|X(t)|^p\} \le \tilde{C}_p\,\mathbb{E}\Big\{\Big(\sup_{0\le t\le T}|X(t)|^{p-2}\Big)\Big(\int_0^T |f(s)|^2\mathrm{d}s\Big)\Big\}$$

and in view of Hölder inequality with exponents $p/2$ and $p/(p-2)$, we deduce the desired estimate (4.95). Similarly, we can treat the multidimensional case.

4.3.2 Discontinuous Local Martingales

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t) : t \ge 0)$ be a n-dimensional (standard) Wiener space and $(p(B,]0,t]) : B \in \mathbb{R}_0^m,\ t \ge 0)$ be an independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{p(B,]0,t])\}/t$, which satisfies [3]

$$\int_{\mathbb{R}_*^m} \frac{|\zeta|^2}{1+|\zeta|}\pi(\mathrm{d}\zeta) < \infty,$$

and martingale measure $\tilde{p}(B,]0,t]) := p(B,]0,t]) - t\pi(B)$, as discussed in Sections 4.1.3 and 4.2.3. This is referred to as a (standard) Wiener-Poisson space. Clearly, a non-standard Wiener-Poisson space corresponds to a Poisson measure with (deterministic) intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$, which is not necessarily absolutely continuous (in the second variable $\mathrm{d}s$) with respect to the Lebesgue measure $\mathrm{d}s$, but $\Pi(\mathbb{R}_*^m, \{t\}) = 0$ for every $t \ge 0$. Also, an extended Wiener-Poisson space corresponds to an extended Poisson measure with (deterministic) intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$, which may have atoms of the form $\mathbb{R}_*^m \times \{t\}$. In any case, the deterministic intensity $\Pi(\mathrm{d}\zeta, \mathrm{d}s) := \mathbb{E}\{p(\mathrm{d}\zeta, \mathrm{d}s)\}$ is the (predictable) compensator of the optional random measure p.

So, a (standard) Wiener-Poisson space with Lévy measure $\pi(\cdot)$ is denoted by $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}_*^m,\ t \ge 0)$, and the (local) martingale measure $\tilde{p}$ is identified with the $\mathbb{R}^m$-valued compensated-jump (Poisson) process

$$\tilde{p}(t) := \int_{\mathbb{R}_*^m \times]0,t]} \zeta\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \ge 0,$$

which induces, on the canonical space $D := D([0,\infty[, \mathbb{R}^m)$ of cad-lag functions, a probability measure $P_{\tilde{\nu}}$, namely,

$$P_{\tilde{\nu}}(B) := P\Big\{\tilde{p}(\cdot) \in B\Big\}, \quad \forall B \in \mathcal{B}(D).$$

[3] the Polish space $\mathbb{R}_0^m := \mathbb{R}^m \setminus \{0\}$ may be replaced by a general Backwell space.

with the characteristic function (or Fourier transform) given by

$$\mathbb{E}\Big\{ \exp\Big[\mathtt{i} \int_{\mathbb{R}^m_* \times]0,t]} (z \cdot \zeta) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) \Big] \Big\} = \\ = \exp\Big[- t \int_{\mathbb{R}^m_*} \big(1 - \mathrm{e}^{\mathtt{i}\, z \cdot \zeta} + \mathtt{i}\, z \cdot \zeta \big) \pi(\mathrm{d}\zeta) \Big],$$

for every $t \geq 0$ and z in $\mathbb{R}^m$. Also note that the Wiener process w induces a probability measure P_w on the canonical space $C := C([0, \infty[, \mathbb{R}^n)$ of continuous functions, namely,

$$P_w(B) := P\big\{ w(\cdot) \in B \big\}, \quad \forall B \in \mathcal{B}(C).$$

and its the characteristic function (or Fourier transform) is given by

$$\mathbb{E}\big\{ \exp\big[\mathtt{i}\, \xi \cdot w(t) \big] \big\} = \exp\big(- t \frac{|\xi|^2}{2} \big),$$

for every $t \geq 0$ and ξ in $\mathbb{R}^n$. Therefore, a *canonical* (standard) Wiener-Poisson space with Lévy measure $\pi(\cdot)$ is a probability measure $P = P_w \times P_{\tilde{p}}$ on the Polish space $C([0, \infty[, \mathbb{R}^n) \times D([0, \infty[, \mathbb{R}^m)$. In this case, the projection map $(\omega_1, \omega_2) \mapsto \big(\omega_1(t), \omega_2(t)\big)$ on $\mathbb{R}^n \times \mathbb{R}^m$, for every $t \geq 0$, is denoted by $\big(X_w(t, \omega), X_{\tilde{p}}(t, \omega)\big)$, and under the probability P the canonical process $(X_w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener process and the canonical process $X_{\tilde{p}}(t)$ is a $\mathbb{R}^m$-valued compensated-jump Poisson process with Lévy measure $\pi(\cdot)$ on $\mathbb{R}^m_*$. The filtration $(\mathcal{F}_t : t \geq 0)$ is generated by the canonical process X_w and $X_{\tilde{p}}$ and completed with null sets with respect to the probability measure P. Note that since the Wiener process is continuous and the compensated-jump Poisson process is purely discontinuous, they are orthogonal (with zero-mean) so that they are independent, i.e., the product form of $P = P_w \times P_{\tilde{p}}$ is a consequences of the statistics imposed on the processes w and $\tilde{p}$.

Definition 4.29 (Itô process with jumps)**.** A $\mathbb{R}^d$-valued stochastic process $(X(t) : t \geq 0)$ is called a d-dimensional *Itô's process with jumps* if there exist real-valued adapted processes $(a_i(t) : t \geq 0,\ i = 1, \dots, d)$, $(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ and $(\gamma_i(\zeta, t) : t \geq 0,\ \zeta \in \mathbb{R}^m_*)$, such that for every $i = 1, \dots, d$ and any $r = 1, 2, \dots,$ we have

$$\begin{cases} \mathbb{E}\Big\{ \displaystyle\int_0^{\tau_r} \Big[|a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}^m_*} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t \Big\} < \infty, \\ X_i(t) = X_i(0) + \displaystyle\int_0^t a_i(s) \mathrm{d}s + \sum_{k=1}^n \int_0^t b_{ik}(s) \mathrm{d}w_k(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]0,t]} \gamma_i(\zeta, s) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0, \end{cases} \tag{4.96}$$

in some (standard) Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0),$$

with Lévy measure π, where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. In short we write

$$\mathrm{d}X(t) = a(t)\mathrm{d}t + b(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma(\zeta, t)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t),$$

for every $t \geq 0$, with a in L^1_{loc}, b in L^2_{loc} and γ in $L^2_{\mathrm{loc},\pi}$. The local martingale measure $\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) := p(\mathrm{d}\zeta, \mathrm{d}t) - E\{p(\mathrm{d}\zeta, \mathrm{d}t)\}$ is also referred to as the compensated jumps (martingale) measure. If the compensator has the form $\Pi(\mathrm{d}\zeta, \mathrm{d}s) = E\{p(\mathrm{d}\zeta, \mathrm{d}t)\}$ then the local integrability assumption on the coefficients γ_i should be changed accordingly, and γ_i should be progressively measurable. Moreover, if $\Pi(\mathbb{R}^m_* \times \{t\}) \neq 0$ for some t, then γ_i must be predictable. □

Notice that any Itô process with jumps is a quasi-left continuous (cad-lag) semi-martingale, and a quasi-martingale whenever a, b and γ belong to the spaces $L^1(]0,T[\times\Omega)$, $L^2(]0,T[\times\Omega)$ and $L^2_\pi(\mathbb{R}^m_* \times]0,T[\times\Omega)$, for any $T > 0$, respectively. Condition (4.96) is equivalent to

$$P\Big\{ \int_0^t \Big[|a(s)| + \mathrm{Tr}[b(s)b^*(s)] + \int_{\mathbb{R}^m_*} |\gamma(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}s < \infty \Big\} = 1, \quad (4.97)$$

for every $t \geq 0$, where $\mathrm{Tr}[\cdot]$ denotes the trace of a matrix and $|\cdot|$ is the Euclidean norm of a vector in $\mathbb{R}^m$. Again, for non-standard case, we modify all conditions accordingly to the the use of $\Pi(\mathrm{d}\zeta, \mathrm{d}s)$ in lieu of $\pi(\mathrm{d}\zeta)\mathrm{d}s$.

Theorem 4.30 (Itô formula with jumps)**.** *Let $(X(t) : t \geq 0)$ be a d-dimensional Itô's process with jumps in a Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{p}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$ with Lévy measure $\pi(\mathrm{d}\zeta)$, i.e., (4.96), and let $\varphi = \varphi(x)$ be a real-valued twice continuously differentiable function on $\mathbb{R}^d$, satisfying*

$$\left\{ \begin{aligned} &\mathbb{E}\Big\{ \int_0^{T_r} \mathrm{d}t \int_{\mathbb{R}^m_*} \big[|\varphi(X(t) + \gamma(\zeta, t)) - \varphi(X(t))|^2 + \\ &\qquad + \varphi(X(t) + \gamma(\zeta, t)) - \varphi(X(t)) - \\ &\qquad - \gamma(\zeta, t) \cdot \nabla\varphi(X(t)) \big] \pi(\mathrm{d}\zeta) \Big\} < \infty \end{aligned} \right. \quad (4.98)$$

for some increasing sequence $\{T_r : r \geq 1\}$ of stopping times such that $T_r \to \infty$ almost surely. Then $(\varphi(X(t)) : t \geq 0)$ is a (real-valued) Itô's process with jumps and

$$\left\{ \begin{aligned} \varphi(X(t)) &= \varphi(X(0)) + \int_0^t A(s, X)\varphi(X(s))\mathrm{d}s + \\ &\quad + \sum_{k=1}^n \int_0^t B_k(s, X)\varphi(X(s))\mathrm{d}w_k(s) + \\ &\quad + \int_{\mathbb{R}^m_* \times]0,t]} C(\zeta, s, X)\varphi(X(s))\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0, \end{aligned} \right. \quad (4.99)$$

where the linear integro-differential operators $A(s,X)$, $B(s,X) = (B_k(s,X) : k = 1,\dots,n)$ *and* $C(\zeta,s,X)$ *are given by*

$$\begin{aligned}
A(s,X)\varphi(x) &:= \sum_{i=1}^{d} a_i(s)\,\partial_i\varphi(x) + \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sum_{k=1}^{n} b_{ik}(s)b_{jk}(s)\Big)\,\partial^2_{ij}\varphi(x) + \\
&\quad + \int_{\mathbb{R}^m_*}[\varphi(x+\gamma(\zeta,s)) - \varphi(x) - \sum_{i=1}^{d}\gamma_i(\zeta,s)\,\partial_i\varphi(x)]\pi(\mathrm{d}\zeta), \\
B_k(s,X)\varphi(x) &:= \sum_{i=1}^{d} b_{ik}(s)\,\partial_i\varphi(x), \\
C(\zeta,s,X)\varphi(x) &:= \varphi(x+\gamma(\zeta,s)) - \varphi(x),
\end{aligned}$$

for any $s \geq 0$ *and* x *in* $\mathbb{R}^d$, *with* ∂_i, ∂_{ij}, *denoting the first and second partial derivatives with respect to the* i *and* j, *and* ∇ *being the gradient operator.*

Proof. First, we replace the coefficients $a(s)$, $b(s)$ and $\gamma(\zeta,s)$ by

$$a(s)\mathbb{1}_{s\leq\tau}, \qquad b(s)\mathbb{1}_{s\leq\tau}, \qquad \gamma(\zeta,s)\mathbb{1}_{s\leq\tau}\mathbb{1}_{\varepsilon<|\zeta|\leq 1/\varepsilon},$$

where $\tau = \tau_r$ and $\varepsilon > 0$. We could use elementary predictable processes a, b and γ to force concrete a calculation. Thus we can pass to the limit in $r \to \infty$ and $\varepsilon \to 0$ to revalidate (4.99), as long as the smooth function φ satisfies (4.98).

The continuity of the semi-martingale was needed in the proof of Theorem 4.27. Nevertheless, when $\gamma(\zeta,s) = 0$ for any $|\zeta| \leq \varepsilon$, the integer-valued measure of the Poisson measure used to integrate has bounded variation and the stochastic integral relative to the local martingale measure becomes a pathwise Lebesgue-Stieltjes integral. Then we can use the integration by parts formula (4.94) to get

$$\begin{cases} X(t)Y(t) - X(0)Y(0) = \displaystyle\int_{(0,t]} X(s-)\mathrm{d}Y(s) + \\ \qquad\qquad + \displaystyle\int_{(0,t]} Y(s-)\mathrm{d}X(s) + [X,Y](t), \quad \forall\, t \geq 0, \end{cases} \tag{4.100}$$

where $[X,Y]$ is the optional quadratic co-variation process. Actually, we may apply (4.94) for jumps with bounded variation and as ε vanishes, we deduce the validity of (4.100) for any two (real-valued) Itô's processes with jumps X and Y.

Notice that

$$\begin{aligned}
[X,Y](t) &:= \langle X^c, Y^c\rangle(t) + \sum_{0<s\leq t}\big(X(s) - X(s-)\big)\big(Y(s) - Y(s-)\big) \\
&= \langle X^c, Y^c\rangle(t) + \int_{\mathbb{R}^m_*\times]0,t]} \gamma^X(\zeta,s)\,\gamma^Y(\zeta,s)\,p(\mathrm{d}\zeta,\mathrm{d}s),
\end{aligned}$$

where $\langle\cdot,\cdot\rangle$ is the optional quadratic co-variation process, X^c and Y^c are the continuous parts of X and Y, e.g.,

$$X^c(t) := \int_0^t a^X(s)\mathrm{d}s + \int_0^t b^X(s)\mathrm{d}w(s),$$

and ν is the integer-valued measure, i.e., $\tilde{\nu}(\cdot,]0,t]) := \nu(\cdot,]0,t]) - t\,\pi(\cdot)$. We can rewrite (4.94) explicitly as

$$\begin{aligned} X(t)Y(t) - X(0)Y(0) &= \int_{(0,t]} X(s-)\mathrm{d}Y^c(s) + \\ &+ \int_{(0,t]} Y(s-)\mathrm{d}X^c(s) + \langle X^c, Y^c\rangle(t) + \\ &+ \int_{\mathbb{R}^m_*\times]0,t]} [X(t)\gamma^Y(\zeta,s) + Y(t)\gamma^X(\zeta,s)]\,\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) + \\ &+ \int_{\mathbb{R}^m_*\times]0,t]} \gamma^X(\zeta,s)\,\gamma^Y(\zeta,s)\,p(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall\, t \geq 0. \end{aligned}$$

In particular, if $X = Y$ we get

$$\begin{aligned} X^2(t) - X^2(0) &= 2\int_{(0,t]} X(s-)\mathrm{d}Y^c(s) + \langle X^c\rangle(t) + \\ &+ 2\int_{\mathbb{R}^m_*\times]0,t]} X(t)\gamma(\zeta,s)\,\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) + \int_{\mathbb{R}^m_*\times]0,t]} \gamma^2(\zeta,s)\,p(\mathrm{d}\zeta,\mathrm{d}s), \end{aligned}$$

for every $t \geq 0$, which exactly reproduces Itô formula (4.99) for $\varphi(x) = x^2$.

Iterating this argument, we check the validity of (4.99) for any multi-dimensional polynomial function $\varphi(x_1,\ldots,x_d)$, and by density, for any smooth function $\varphi(x)$.

Finally, for any smooth function satisfying (4.98) we may let $r \to \infty$ and $\varepsilon \to 0$ to conclude. □

Notice that we also have

$$\begin{cases} X(t)Y(t) - X(0)Y(0) = \displaystyle\int_{(0,t]} X(s-)\mathrm{d}Y(s) + \\ \qquad\qquad + \displaystyle\int_{(0,t]} Y(s-)\mathrm{d}X(s) + \langle X,Y\rangle(t), \quad \forall\, t \geq 0, \end{cases} \tag{4.101}$$

i.e., in the integration by parts the optional quadratic variation $[X,Y]$ may be replaced by the predictable quadratic variation $\langle X,Y\rangle$ associated with the whole *quasi-left continuous* square integrable semi-martingales X and Y. Also for a function $\varphi = \varphi(t,x)$, we do not need to require C^2 in the variable t. Also, when $\varphi = \varphi(x)$, we could use a short vector notation

$$\begin{cases} \mathrm{d}\varphi(X(t)) = \nabla\varphi(X(t))\mathrm{d}X^c(t) + [\varphi \diamond_\gamma \tilde{p}](\cdot,\mathrm{d}t)(t,X(t)) + \\ \qquad\qquad + \big[\frac{1}{2}\,\mathrm{Tr}[b(t)b^*(t)\nabla^2\varphi(x)] + [\varphi \bullet_\gamma \pi](t,X(t))\big]\mathrm{d}t, \end{cases} \tag{4.102}$$

for every $t \geq 0$, where

$$\begin{aligned} [\varphi \diamond_\gamma \tilde{p}(\cdot,\mathrm{d}t)](t,x) &:= \int_{\mathbb{R}^m_*} [\varphi(x+\gamma(\zeta,t)) - \varphi(x)]\,\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \\ [\varphi \bullet_\gamma \pi](t,x) &:= \int_{\mathbb{R}^m_*} [\varphi(x+\gamma(\zeta,t)) - \varphi(x) - \gamma(\zeta,t)\cdot\nabla\varphi(x)]\pi(\mathrm{d}\zeta), \end{aligned}$$

and ∇ and $\mathrm{Tr}[\cdot]$ are the gradient and trace operator, respectively. The above calculation remains valid for a Poisson measure not necessarily standard, i.e., the intensity or Lévy measure has the form $\Pi(\mathrm{d}\zeta, \mathrm{d}t) := \mathbb{E}\{p(\mathrm{d}\zeta, \mathrm{d}t)\}$ and $\Pi(\mathbb{R}^m_* \times \{t\}) = 0$ for every $t \geq 0$. For an extended Poisson measure, the process is no longer quasi-left continuous and the rule (4.99) needs a jump correction term, i.e., the expression $X(s)$ is replaced by $X(s-)$ inside the stochastic integrals. For instance, the reader may consult Bensoussan and Lions [17, Section 3.5, pp. 224–244] or Gikhman and Skorokhod [99, Chapter II.2, pp. 215–272] for more details on this approach.

Semi-martingale Viewpoint

In general, the integration by parts formula (4.100) is valid for any two semi-martingales X and Y, and we have the following general Itô formula for semi-martingales, e.g., Chung and Williams [45, Theorems 38.3 and 39.1, Chapter VI, pp. 392–394], Dellacherie and Meyer [58, Sections VIII.15–27, pp. 343–352], Jacod and Shiryaev [117, Theorem 4.57, Chapter 1, pp. 57–58].

Theorem 4.31. *Let $X = (X_1, \ldots, X_d)$ be a d-dimensional semi-martingale and φ be a complex-valued twice-continuously differentiable function on $\mathbb{R}^d$. Then $\varphi(X)$ is a semi-martingale and we have*

$$\begin{aligned}\varphi(X(t)) = \varphi(X(0)) &+ \sum_{i=1}^{d} \int_{]0,t]} \partial_i \varphi(X(s-)) \mathrm{d}X_i(s) + \\ &+ \frac{1}{2} \sum_{i,j=1}^{d} \int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X^c_i, X^c_j \rangle(s) + \\ &+ \sum_{0<s\leq t} \Big\{ \varphi(X(s)) - \varphi(X(s-)) - \sum_{i=1}^{d} \partial_i \varphi(X(s-)) \delta X(s) \Big\},\end{aligned}$$

where ∂_i and ∂^2_{ij} denotes partial derivatives, $\delta X(s) := [X_i(s) - X_i(s-)]$ and $X(s-)$ is the left limit at s and X^c_i is the continuous part. □

First remark that

$$\begin{aligned}\int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X_i, X_j \rangle(s) = &\sum_{0<s\leq t} \partial^2_{ij} \varphi(X(s-)) \big[\delta X(s)\big]^2 + \\ &+ \int_{]0,t]} \partial^2_{ij} \varphi(X(s-)) \mathrm{d}\langle X^c_i, X^c_j \rangle(s),\end{aligned}$$

where the integrals and series are absolutely convergent. Hence, the above formula can be rewritten using the predictable quadratic variation $\langle X_i, X_j \rangle$, i.e., the predictable processes obtained via the Doob-Meyer decomposition when X is locally square integrable or in general the predictable projection of the optional quadratic variation $[X_i, X_j]$.

Let X be a (special) quasi-left continuous semi-martingale written in the canonical form

$$X(t) = X(0) + X^c(t) + A(t) + \int_{\mathbb{R}^d_* \times]0,t]} z \tilde{\nu}(\mathrm{d}z, \mathrm{d}s), \forall t \geq 0,$$

where X^c is the continuous (local martingale) part, A is the predictable locally bounded variation (and continuous) part, and $\tilde{\nu}$ is the compensated (local martingale) random measure associated with the integer-valued measure $\nu = \nu_X$ of the process X with compensator ν^p. Then

$$\mathrm{d}X_i(s) = \mathrm{d}X_i^c(s) + \int_{\mathbb{R}_*^d} z_i \tilde{\nu}(\mathrm{d}z, \mathrm{d}s),$$

so that

$$\sum_{i=1}^d \int_{]0,t]} \partial_i \varphi(X(s-)) \mathrm{d}X_i(s) = \sum_{i=1}^d \int_{]0,t]} \partial_i \varphi(X(s-)) \mathrm{d}X_i^c(s) + \\ + \sum_{i=1}^d \int_{\mathbb{R}_*^d \times]0,t]} z_i \partial_i \varphi(X(s-)) \tilde{\nu}(\mathrm{d}z, \mathrm{d}s),$$

and the jump part can be written as

$$\sum_{0<s\le t} \big[\varphi(X(s)) - \varphi(X(s-)) - \sum_{i=1}^d \partial_i \varphi(X(s-)) \delta X(s)\big] = \\ = \int_{\mathbb{R}_*^d \times]0,t]} \big[\varphi(X(s-) + z) - \varphi(X(s-)) - \\ - \sum_{i=1}^d z_i \partial_i \varphi(X(s-))\big] \nu(\mathrm{d}z, \mathrm{d}s),$$

for every $t \ge 0$. Moreover, because $\nu^p(\mathbb{R}_*^m \times \{t\}) = 0$ for any $t \ge 0$, we can substitute $X(s-)$ for $X(s)$ in the above stochastic integral. Thus, combining the above jump parts we see that the expression (4.99) of Theorem 4.30 remains valid for any quasi-left continuous integer measure $\nu(\mathrm{d}z, \mathrm{d}s)$ with a local martingale measure $\tilde{\nu}(\mathrm{d}z, \mathrm{d}s)$ and compensator $\nu^p(\mathrm{d}z, \mathrm{d}s)$, which replaces the deterministic product measure $\pi(\mathrm{d}z) \times \mathrm{d}s$. The case of interest for us is when the predictable compensator measure $\nu^p(\mathrm{d}z, \mathrm{d}s)$ has a density with respect to the Lebesgue measure, i.e.,

$$\nu^p(B,]0,t]) = \int_0^t \mathtt{M}(B,s) \mathrm{d}s, \quad \forall B \in \mathcal{B}(\mathbb{R}_*^d),\ t \ge 0,$$

where the intensity kernel $\mathtt{M}$ is such that for every fixed B, the function $s \mapsto \mathtt{M}(B,s)$ defines a predictable process, while $B \mapsto \mathtt{M}(B,s)$ is a (random) measure for every fixed s. It is clear that Itô formula is suitable modified.

• *Remark* 4.32. In particular, Theorem 4.31 can be formulated as follows. Let $X = (X_1, \dots, X_d)$ be a semi-martingale, M be local martingale and g, a and $\mathtt{M}$ be local integrable predictable processes such that

$$X(t) - X(0) - M(t) = \int_0^t g(s) \mathrm{d}s, \quad \forall t \ge 0,$$
$$\langle M_i^c, M_j^c \rangle(t) = \int_0^t a_{ij}(s) \mathrm{d}s, \quad \forall t \ge 0,$$
$$\nu_M^p(B,]0,t]) = \int_0^t \mathtt{M}(B,s) \mathrm{d}s, \quad \forall B \in \mathcal{B}(\mathbb{R}_*^d),\ t \ge 0,$$

where M^c is the continuous part of M and ν_M^p is the compensator of the integer measure ν_M associated with M. Then

$$\varphi(X(t),t) = \varphi(X(0),0) + \int_0^t \big[(\partial_s + A_X)\varphi(X(s-),s)\big]\mathrm{d}s + \\ + \sum_{i=1}^d \int_0^t \partial_i\varphi(X(s-),s)\mathrm{d}M^c(s) + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\varphi(X(s-)+z,s) - \varphi(X(s-),s)\big]\tilde{\nu}_M(\mathrm{d}z,\mathrm{d}s),$$

where

$$(\partial_s + A_X)\varphi(\cdot,s) = \partial_s\varphi(\cdot,s) + \sum_{i=1}^d g_i(s)\partial_i\varphi(\cdot,s) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(s)\partial^2_{ij}\varphi(\cdot,s) + \\ + \int_{\mathbb{R}^d_*} \big[\varphi(\cdot+z,s) - \varphi(\cdot,s) - \sum_{i=1}^d z_i\partial_i\varphi(\cdot,s)\big]\mathtt{M}(\mathrm{d}z,s),$$

for every bounded function $\varphi(x,t)$ in $\mathbb{R}^d \times [0,\infty)$, which is twice continuously differentiable in x, once continuously differentiable in t with all derivatives bounded. In general, if the semi-martingale $X = V + M$, where V is a continuous process with local bounded variation and M a locally square-integrable martingale then $\phi(X(t)) = \phi(X(0)) + V_\phi(t) + M_\phi(t)$ is a semi-martingale with

$$V_\phi(t) = \int_0^t \nabla\phi(X(s-))\cdot\mathrm{d}V(s) + \frac{1}{2}\int_0^t \mathrm{Tr}\left(D^2\phi(X(s-))\mathrm{d}\langle M^c\rangle(s)\right] + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\phi(X(s-)+z) - \phi(X(s-)) - z\cdot\nabla\phi(X(s-))\big]\nu_M^p(\mathrm{d}z,\mathrm{d}s)$$

and

$$M_\phi(t) = \int_0^t \nabla\phi(X(s-))\cdot\mathrm{d}M^c(s) + \\ + \int_{\mathbb{R}^d_*\times]0,t]} \big[\phi(X(s-)+z) - \phi(X(s-))\big]\tilde{\nu}_M(\mathrm{d}z,\mathrm{d}s),$$

for any bounded twice continuously differentiable ϕ with all derivative bounded. This is usually referred to as the Itô formula for semi-martingales, which can be written as above, by means of the associated integer measure, or as in Theorem 4.31. □

• *Remark* 4.33. In general, if $\{x(t) : t \geq 0\}$ is a real-valued predictable process with local bounded variation (so $x(t+)$ and $x(t-)$ exist for every t) and $\{y(t) : t \geq 0\}$ is a (cad-lag) semi-martingale then we have

$$\mathrm{d}\big(x(t)y(t)\big) = x(t)\mathrm{d}y(t) + y(t-)\mathrm{d}x(t),$$
$$\mathrm{d}[x,y](t) = \big(x(t+) - x(t-)\big)\mathrm{d}y(t),$$
$$\mathrm{d}|y(t)|^2 = 2y(t-)\mathrm{d}y(t) + [y,y](t),$$

with the above notation. By the way, note that $\mathrm{d}x(t) = \mathrm{d}x(t+)$ and $x(t)\mathrm{d}y(t) = x(t-)\mathrm{d}y(t)$. □

Approximations and Comments

A double sequence $\{\tau_m(n) : n, m \geq 0\}$ of stopping times is called a *Riemann sequence* if $\tau_m(0,\omega) = 0$, $\tau_m(n,\omega) < \tau_m(n+1,\omega) < \infty$, for every $n = 0, 1, \ldots, N_m(\omega)$ and as $m \to 0$ we have

$$\sup_n \{\tau_m(n+1,\omega) \wedge t - \tau_m(n,\omega) \wedge t\} \to 0, \quad \forall\, t > 0,$$

for every ω, i.e., the mesh or norm of the partitions or subdivisions restricted to each interval $[0,t]$ goes to zero. A typical example is the dyadic partition $\tau_m(n) := n2^{-m}$, $m = 1, 2, \ldots$, and $n = 0, 1, \ldots, 2^m$, which is deterministic. We have the following general results:

Theorem 4.34 (Riemann sequence)**.** *Let X be a semi-martingale, Y be a cadlag adapted process and $\{\tau_m(n) : n, m \geq 0\}$ be a Riemann sequence. Then the sequence of Riemann-Stieltjes sums, $m \geq 0$,*

$$\sum_n Y(\tau_m(n)-)\big(X(\tau_m(n+1) \wedge t) - X(\tau_m(n) \wedge t)\big)$$

converges in probability, uniformly on each compact interval, to the stochastic integral

$$\int_{]0,t]} Y(s-)\mathrm{d}X(s).$$

Moreover, if Y is also a semi-martingale then the optional process

$$t \mapsto \sum_n \big(X(\tau_m(n+1) \wedge t) - X(\tau_m(n) \wedge t)\big) \times \\ \times \big(Y(\tau_m(n+1) \wedge t) - Y(\tau_m(n) \wedge t)\big)$$

converges in probability, uniformly on each compact interval, to the optional quadratic covariation process $[X, Y]$.

Proof. For instance to prove the first convergence, it suffices to see that the above Riemann-Stieltjes sums are equal to the stochastic integral

$$\int_{]0,t]} Y_m(s)\mathrm{d}X(s),$$

where $Y_m(s) := Y(\tau_m(n))$ for any s in the stochastic interval $\rrbracket \tau_m(n), \tau_m(n+1) \rrbracket$, is clearly a predictable left continuous process for each $m \geq 0$.

The proof of the second convergence is essentially based on the integration by part formula (4.100), which actually can be used to define the optional quadratic covariation process.

For instance, a full proof can be found in Jacod and Shiryaev [117, Proposition 4.44 and Theorem 4.47, Chapter 1, pp. 51–52]. □

The estimate (4.76) of the previous section for for Poisson integral, namely, for any p in $(0,2]$ there exists a positive constant $C = C_p$ (actually $C_p := (4-p)/(2-p)$ if $0 < p < 2$ and $C_2 = 4$) such that for any adapted (measurable) process $f(\zeta, s)$ (actually, the predictable version is used) we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \sup\limits_{0\le t\le T} \Big| \int\limits_{\mathbb{R}^m_*\times]0,t]} f(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \le \\ \qquad \le C\,\mathbb{E}\Big\{ \Big[\int\limits_0^T \mathrm{d}s \int\limits_{\mathbb{R}^m_*} |f(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \end{array} \right. \tag{4.103}$$

for every stopping time T. The case $p > 2$ is a little more complicate and involves Itô formula. Indeed, for the sake of simplicity let us consider the one-dimensional case, use Itô formula with the function $x \mapsto |x|^p$ and the process

$$X(t) := \int_{\mathbb{R}^m_*\times]0,t]} f(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall t \ge 0$$

to get

$$\begin{aligned} \mathbb{E}\{|X(t)|^p\} &= \mathbb{E}\Big\{ \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \Big[|X(s)+f(\zeta,s)|^p - |X(s)|^p - \\ &\quad -p\,|X(s)|^{p-2}X(s)f(\zeta,s)\Big]\pi(\mathrm{d}\zeta)\Big\} = p(p-1)\times \\ &\quad \times \mathbb{E}\Big\{ \int_0^t \mathrm{d}s \int_0^1 (1-\theta)\mathrm{d}\theta \int_{\mathbb{R}^m_*} |X(s)+\theta f(\zeta,s)|^{p-2}|f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big\}. \end{aligned}$$

The integrand is bounded as follows

$$|X(s)+\theta f(\zeta,s)|^{p-2}|f(\zeta,s)|^2 \le 2^{p-2}\big[|X(s)|^{p-2}|f(\zeta,s)|^2 + |f(\zeta,s)|^p\big],$$

and by means of the Doob's maximal inequality, we deduce

$$\begin{aligned} \mathbb{E}\{ \sup_{0\le t\le T} |X(t)|^p\} &\le \tilde{C}_p \Big[\mathbb{E}\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^p \pi(\mathrm{d}\zeta)\Big\} + \\ &\quad + \mathbb{E}\Big\{ \Big(\sup_{0\le t\le T} |X(t)|^{p-2}\Big) \Big(\int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big)\Big\}\Big], \end{aligned}$$

for some constant $\tilde{C}_p$ depending only on p. Hence, the simple inequality for any $a, b, \varepsilon \ge 0$,

$$ab \le \frac{p-2}{p}(\varepsilon a)^{p/(p-2)} + \frac{2}{p}\Big(\frac{a}{\varepsilon}\Big)^{p/2}$$

and the Hölder inequality yield the following variation of (4.103): for any $p > 2$ there exists a constant $C = C_p$ depending only on p such that

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \sup\limits_{0\le t\le T} \Big| \int\limits_{\mathbb{R}^m_*\times]0,t]} f(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \le \\ \qquad \le C\,\mathbb{E}\Big\{ \int\limits_0^T \mathrm{d}s \int\limits_{\mathbb{R}^m_*} |f(\zeta,s)|^p \pi(\mathrm{d}\zeta)\Big\} + \\ \qquad + \mathbb{E}\Big\{ \Big[\int\limits_0^T \mathrm{d}s \int\limits_{\mathbb{R}^m_*} |f(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \end{array} \right. \tag{4.104}$$

for any adapted (measurable) process $f(\zeta, s)$ and any stopping time T.

• *Remark* 4.35. These estimates for the moments of a stochastic integral can be partially generalized to some other type of integral, e.g., let M be a d-dimensional continuous square integrable martingale with predictable quadratic covariation process $\langle M_i, M_j \rangle = \mathrm{d}\ell$ if $i = j$ and $\langle M_i, M_j \rangle = 0$ if $i \neq j$, where ℓ is a continuous nondecreasing adapted process satisfying

$$\mathbb{E}\{\ell(t) - \ell(s) \,|\, \mathcal{F}(s)\} \leq h(t-s),$$

for every $t \geq s \geq 0$ and for some monotone function h form $[0, \infty)$ into itself. Using the integration by part formula

$$[\ell(t) - \ell(s)]^k = k \int_s^t [\ell(t) - \ell(r)]^{k-1} \, \mathrm{d}\ell(r)$$

and by induction on k, we can show that

$$\mathbb{E}\big\{[\ell(t) - \ell(s)]^k \,|\, \mathcal{F}(s)\big\} \leq k! \, [h(t-s)]^k,$$

for every $t \geq s \geq 0$ and any $k \geq 1$. Similarly, by means of Itô formula, the sub-martingale inequality and by induction, we can prove that for every positive integer k there exists a constant $C = C(k, d)$, depending only on k and the dimension d, such that

$$\mathbb{E}\big\{ \sup_{s \leq r \leq t} |M(r) - M(s)|^k \,|\, \mathcal{F}(s)\big\} \leq C(k, d) \, [h(t-s)]^{k/2},$$

for every $t \geq s \geq 0$ and any $k \geq 1$. □

It is clear that the above Itô calculus can be extended to non deterministic smooth functions, i.e., predictable processes $\varphi(t, x, \omega)$ which are continuously differentiable in t and twice-continuously differentiable in x. The rule given in this section is unchanged. As we may expect, if for each x the process $t \mapsto \varphi(t, x, \omega)$ is a local martingale (which has not bounded variation paths) then Itô calculus applies and another bracket $[\cdot, \cdot]$ with respect to this process should appear.

• *Remark* 4.36. In a given complete filtered space, an adapted increasing (locally integrable) cad-lag process A is called *natural* if for every bounded (not necessarily continuous) martingale M we have

$$\int_{[0,t]} M(s) \mathrm{d}A(s) = \int_{[0,t]} M(s-) \mathrm{d}A(s), \quad \forall \, t \geq 0.$$

This is equivalent to the concept of *predictable* process. On the other hand, a quasi left continuous (increasing or martingale) cad-lag process is also called *regular*. It turns out that an adapted increasing cad-lag process is continuous if and only if it is natural and regular. The reader is referred to the books Kallenberg [121] and Yeh [258] for a comprehensive treatment. □

• *Remark* 4.37. The operational Itô formula is better understood in its simplest product form, i.e., let X and Y be two d-dimensional Itô processes with jumps (see Definition 4.29), namely

$$\mathrm{d}X(t) = a^X(t)\mathrm{d}t + b^X(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma^X(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$
$$\mathrm{d}Y(t) = a^Y(t)\mathrm{d}t + b^Y(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma^Y(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$

then

$$\mathrm{d}\big(X_i(t)Y_j(t)\big) = X_i(t-)\mathrm{d}Y_j(t) + \big(\mathrm{d}X_i(t)\big)Y_j(t-) + \\ + \sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}t + \int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)p(\mathrm{d}\zeta,\mathrm{d}t),$$

for any $t \geq 0$. Note the independent role of the diffusion and jumps coefficients. Moreover, the last (jump) integral is not a *pure* stochastic integral, it is with respect to $p(\mathrm{d}\zeta,\mathrm{d}t)$ which can be written as $\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \pi(\mathrm{d}\zeta)\mathrm{d}t$. We can go further and make explicit each term, i.e.,

$$X_i(t)\mathrm{d}Y_j(t) = X_i(t-)\mathrm{d}Y_j(t) = X_i(t)a^Y_j(t)\mathrm{d}t + X_i(t)b^Y_j(t)\mathrm{d}w(t) + \\ + \int_{\mathbb{R}^m_*} X_i(t)\gamma^Y_j(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t),$$

where $X_i(t)$ goes inside the stochastic integral indistinctly as either $X_i(t)$ or it predictable projection $X_i(t-)$. □

Similarly to above Remark 4.37, a operational (generalized) Itô formula can be written for processes driven by local martingales. Let $M = M^c + M^d$ be a quasi-left continuous local square-integrable martingale in $\mathbb{R}^n$ written as the sum of a continuous local martingale $\{M^c_i : i = 1,\dots,n\}$ with predictable variation process $\{\langle M^c_i\rangle : i = 1,\dots,n\}$, satisfying $\langle M^c_i, M^c_j\rangle = 0$ if $i \neq j$, and a purely discontinuous local martingale $\{M^d_i : i = 1,\dots,n\}$ which yields an integer measure $\nu_{_M}$ with compensator $\nu^p_{_M}$ and (local) martingale measure $\tilde{\nu}_{_M} = \nu_{_M} - \nu^p_{_M}$. Note that

$$\int_{]0,t]} \alpha(s)\mathrm{d}M^d_i(s) = \int_{\mathbb{R}^d\times]0,t]} \alpha(s)\zeta_i\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}s).$$

Also let $\{V^c_i : i = 1,\dots,d\}$ be a local bounded variation continuous process, non-anticipating with respect to M. Now, if X and Y are two d-dimensional processes of the form

$$\mathrm{d}X(t) = a^X(t)\mathrm{d}V^c(t) + b^X(t)\mathrm{d}M^c(t) + \int_{\mathbb{R}^m_*} \gamma^X(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$
$$\mathrm{d}Y(t) = a^Y(t)\mathrm{d}V^c(t) + b^Y(t)\mathrm{d}M^c(t) + \int_{\mathbb{R}^m_*} \gamma^Y(\zeta,t)\tilde{\nu}_{_M}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \geq 0,$$

then

$$\mathrm{d}\big(X_i(t)Y_j(t)\big) = X_i(t-)\mathrm{d}Y_j(t) + \big(\mathrm{d}X_i(t)\big)Y_j(t-)+ \\ + \sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}\langle M^c_k\rangle(t) + \int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t),$$

for any $t \geq 0$. In particular, in term of the purely jumps (local) martingale M^d_k, i.e., $\gamma_i(\zeta,t) = \sum_k c_{ik}(t)\zeta_k$ for both processes, we have

$$\int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t) = \\ = \frac{1}{2}\sum_{k,\ell}\int_{]0,t]} \big(c^X_{ik}(s)c^Y_{j\ell}(s) + c^X_{i\ell}(s)c^Y_{jk}(s)\big)\mathrm{d}[M^d_k, M^d_\ell](s),$$

where $[M^d_k, M^d_\ell]$ is the optional quadratic (matrix) variation, i.e.,

$$[M^d_k, M^d_\ell](t) = \sum_{s\leq t}\big(M^d_k(s) - M^d_k(s-)\big)\big(M^d_\ell(s) - M^d_\ell(s-)\big),$$

Hence, if c^X_{ik} and $c^Y_{j\ell}$ are cad-lag then

$$\int_{\mathbb{R}^m_*} \gamma^X_i(\zeta,t)\gamma^Y_j(\zeta,t)\nu_{_M}(\mathrm{d}\zeta,\mathrm{d}t) = \frac{1}{2}\sum_{k,\ell}\sum_{0<s\leq t}\big(c^X_{ik}(s-)c^Y_{j\ell}(s-)+ \\ +c^X_{i\ell}(s-)c^Y_{jk}(s-)\big)\big(M^d_k(s) - M^d_k(s-)\big)\big(M^d_\ell(s) - M^d_\ell(s-)\big).$$

Moreover, if each coordinate is orthogonal to each other (i.e., $[M^d_i, M^d_j] = 0$, for $i \neq j$), equivalent to the condition that there are no simultaneous jumps of M^d_i and M^d_j, then only the terms $k = \ell$ and the $1/2$ is simplified. Clearly, there is only a countable number of jumps and

$$\mathbb{E}\Big\{\sum_{0<s\leq t\wedge\tau_n}\big[\big(c^X_{ik}(s-)\big)^2 + \big(c^Y_{jk}(s-)\big)^2\big]\big(M^d_k(s) - M^d_k(s-)\big)^2\Big\} < \infty,$$

for every $t > 0$, where $\{\tau_n\}$ is some sequence the stopping times increases to ∞ almost surely, i.e., the above series is absolutely convergence (localized) in the L^2-sense. If c^X_{ik} or c^Y_{jk} is not cad-lag, then a predictable version should be used in the series. Furthermore, if the initial continuous martingale M^c do not have orthogonal components then we may modify the drift and reduce to the above case, after using Gram-Schmidt orthogonalization procedure, or alternatively, we have a double (symmetric) sum,

$$\frac{1}{2}\sum_{k,\ell}[b^X_{ik}(t)b^Y_{j\ell}(t) + b^X_{i\ell}(t)b^Y_{jk}(t)]\mathrm{d}\langle M^c_k, M^c_\ell\rangle(t)$$

instead of the single sum in k. On the other hand, to include discontinuous process V or a non-necessarily quasi-left continuous local martingale, we need

to carefully consider possible deterministic jumps. Indeed, denoting by δ the jump operator, i.e., $\delta X(t) = \big(X(t) - X(t-)\big)$ for a cad-lag process X, the relation

$$\delta\big(X_i(t)Y_j(t)\big) = \big(\delta X_i(t)\big)Y_j(t-) + X_i(t-)\big(\delta Y_j(t)\big)\big) + \big(\delta X_i(t)\big)\big(\delta Y_j(t)\big)$$

shows the general expression

$$\begin{aligned}\mathrm{d}\big(X_i(t)Y_j(t)\big) &= X_i(t-)\mathrm{d}Y_j(t) + \big(\mathrm{d}X_i(t)\big)Y_j(t-) + \\ &+ \sum_k b^X_{ik}(t)b^Y_{jk}(t)\mathrm{d}\langle M^c_k\rangle(t) + \mathrm{d}\Big(\sum_{s\in]0,t]}\big(\delta X_i(s)\big)\big(\delta Y_j(s)\big)\Big),\end{aligned}$$

which makes sense as a stochastic integral after compensating the jumps. Since the jumps of $X_i(t)$ or $X_i(t)$ are due only to $V^d(t) = \sum_{0<s\le t}\delta V(s)$ and $M^d(t)$, we have

$$\begin{aligned}&\big(\delta X_i(t)\big)\big(\delta Y_j(t)\big) = \\ &\quad = \sum_{k,\ell}\big(a^X_{ik}(s)\delta V_k(s) + c^X_{ik}(s)\delta M^d_k(s)\big)\big(a^Y_{j\ell}(s)\delta V_\ell(s) + c^X_{j\ell}(s)\delta M^d_\ell(s)\big).\end{aligned}$$

Hence, without loss of generality, it seems better to take $V = V^c$ continuous and put all jumps into the integer measure ν, which may not be quasi-left continuous. This is the case of a special semi-martingale $S(t)$, $S(0) = 0$, written in its canonical form as $V + M^c + M^d$, where $V = V^c$ if S is quasi-left continuous. Essentially, this discontinuity (of V) imposes (implicitly) the condition that the drift must be continuous at each predictable jump (jumps non switchable to M^d, e.g., deterministic jumps), otherwise, the integrability of the drift with respect to a discontinuous V may be an issue, i.e., in the Stieltjes-Riemann sense may be not integrable and in the Stieltjes-Lebesgue sense may yield distinct values, depending on whether $a(s)$, $a(s+)$ or $a(s-)$ is used. This never arrive in the stochastic integral.

• *Remark* 4.38. Let X be a 1-dimensional Itô processes with jumps (see Definition 4.29), namely

$$\mathrm{d}X(t) = a(t)\mathrm{d}t + b(t)\mathrm{d}w(t) + \int_{\mathbb{R}^m_*}\gamma(\zeta,t)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t), \quad \forall t \ge 0,$$

with $X(0) = 0$, and such that almost surely we have $\gamma(\zeta,t) > -1$ or equivalently $\inf\big\{\delta X(t) : t > 0\big\} > -1$, where $\delta X(t) = X(t) - X(t-)$ is the jump at time t. Based on the inequalities $r - \ln(1+r) \ge 0$ if $r > -1$ and $r - \ln(1+r) \le r^2/2$ if $r \ge 0$, we deduce that the infinite product $\prod_{0\le s\le t}\big[1+\delta X(s)\big]\mathrm{e}^{-\delta X(s)}$ is almost surely finite and positive. Moreover, for every $t \ge 0$, either the exponential expression

$$\mathcal{E}_X(t) = \exp\Big\{X(t) - \frac{1}{2}\int_0^t\sum_{k=1}^n|b_k(s)|^2\mathrm{d}s\Big\}\prod_{0\le s\le t}\big[1+\delta X(s)\big]\mathrm{e}^{-\delta X(s)},$$

or the log-differential expression

$$\mathrm{d}\ln\big(\mathcal{E}_X(t)\big) = \big[a(t) - \frac{1}{2}|b(t)|^2\big]\mathrm{d}t + \int_{\mathbb{R}^m_*} \ln\big(1+\gamma(\zeta,t)\big)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \\ + \int_{\mathbb{R}^m_*} \big[\ln\big(1+\gamma(\zeta,t)\big) - \gamma(\zeta,t)\big]\pi(\mathrm{d}\zeta)$$

defines a 1-dimensional Itô processes with jumps satisfying

$$\mathrm{d}\mathcal{E}_X(t) = \mathcal{E}_X(t-)\,\mathrm{d}X(t),$$

which is called exponential martingale. Recall that $\tilde{p} = -\pi$ so that if γ has a finite π-integral (i.e., the jumps are of bounded variation) then

$$\int_{\mathbb{R}^m_*} \ln\big(1+\gamma(\zeta,t)\big)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + \int_{\mathbb{R}^m_*} \big[\ln\big(1+\gamma(\zeta,t)\big) - \gamma(\zeta,t)\big]\pi(\mathrm{d}\zeta) = \\ = \int_{\mathbb{R}^m_*} \ln\big(1+\gamma(\zeta,t)\big)p(\mathrm{d}\zeta,\mathrm{d}t) - \int_{\mathbb{R}^m_*} \gamma(\zeta,t)\pi(\mathrm{d}\zeta),$$

as formally expected. For instance, see Applebaum [6, Chapter 5, pp 246-291] or Jacod and Shiryaev [117, Section III.3, pp. 152–166]. □

4.3.3 Other Stochastic Integrals

First we recall some key facts about possibly discontinuous martingales and then we discuss Stratonovich (and other) stochastic integrals.

Refresh on Quasi-Martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a complete (relative to $\mathcal{F}$), right-continuous (not necessarily quasi-left continuous) filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$. Recall that an adapted cad-lag process X is called a quasi-martingale if $\mathbb{E}\{|X(t)|\} < \infty$ and ${}^{\mathrm{p}}\mathrm{Var}(X)(t) < \infty$, for every $t \geq 0$, where the *conditional variation* is defined by

$${}^{\mathrm{p}}\mathrm{var}_{\varpi}(X)(t,\cdot) = \sum_i \big|\mathbb{E}\{X(t_{i+1}\wedge t) - X(t_i\wedge t)\,|\,\mathcal{F}(t_i\wedge t)\}\big|,$$
$${}^{\mathrm{p}}\mathrm{Var}(X) = \sup\big\{{}^{\mathrm{p}}\mathrm{Var}_{\varpi}(X) : \varpi\big\}, \qquad {}^{\mathrm{p}}\mathrm{Var}_{\varpi}(X) = \mathbb{E}\{{}^{\mathrm{p}}\mathrm{var}_{\varpi}(X)\},$$

where the supremum is taken over all (deterministic) partitions $\varpi = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n < \cdots\}$ of $[0,\infty)$ with norm $|\varpi| = \sup\{t_i - t_{i-1} : i \geq 1\}$. An adapted processes X is a quasi-martingale if and only it can be decomposed as the difference $X = Y - Z$ of two positive (cad-lag) super-martingales Y and Z, or equivalently, it is a special semi-martingale, which yields the decomposition $X = M + A$ with M a local martingale and A a predictable local integrable finite variation process, i.e., $A = A_+ - A_-$, both predictable, local integrable and monotone increasing. In particular if X is an adapted local integrable monotone increasing (or finite variation) process then $X = M + A$, where M is a local martingale and A is a predictable local integrable monotone increasing

(or finite variation) process. The process A is called the predictable (jumps) compensator of X. Note that the essential different between quasi-martingales and semi-martingales is the integrability of the large jumps.

If X is a semimartingale then the optional quadratic variation of X is defined as

$$[X](t) = X^2(t) - X^2(0) - 2\int_{]0,t]} X(s-)\mathrm{d}X(s), \quad t \geq 0,$$

or equivalently, as

$$[X]_\varpi(t) = \sum_{i=1}^{n} \left|X(t_{i+1} \wedge t) - X(t_i \wedge t)\right|^2, \qquad [X](t) = \lim_{|\varpi|\to 0} [X]_\varpi(t).$$

However, the predictable quadratic variation $\langle X\rangle$ is the predictable (jumps) compensator of $[X]$, i.e., $\langle X\rangle$ is the unique predictable process with local integrable finite variation (increasing) vanishing at 0 such that $[X] - \langle X\rangle$ is a local martingale or equivalently $X^2 - \langle X\rangle$ is a local martingale. Because $[X]$ is an adapted increasing process we may define its continuous part

$$[X]^c(t) = [X](t) - \sum_{s\leq t} \delta[X](s),$$

where δ is the jump operator, $\delta Y(0) = 0$,

$$\delta Y(t) = Y(t+) - Y(t-), \quad t > 0,$$

defined for any process Y having no discontinuities of second kind.

For any quasi-martingale X we have:

(1) if X is continuous then $[X]$ and $\langle X\rangle$ are (the same) continuous processes,

(2) if X has local integrable finite variation then $[X]^c = 0$,

(3) if $[X] = 0$ then $X = X(0)$,

(4) if X is a local martingale satisfying $\langle X\rangle = 0$ then $X = X(0)$,

(5) X is quasi-left continuous if and only if $\langle X\rangle$ is continuous.

Moreover, any quasi-martingale X has a unique decomposition $X(0) + V_p(t) + M_c(t) + M_d(t)$, where $V_p(0) = M_c(0) = M_d(0) = 0$, V_p is a predictable process with local integrable finite variation, M_c is a continuous local martingale and M_d is a local martingale satisfying $[M_d]^c = 0$, also (1) $[X]^c = \langle M_c\rangle$, (2) V_p is continuous if $M_d = 0$, and (3) if X has also local integrable finite variation then $M_c = 0$.

Note that $[X](t) = \sum_{s\leq t}(X(t)-X(t-))^2$ for any process X of local bounded variation, and we have $\langle X\rangle = 0$ if X^2 is a local martingale. In particular if $X = N$ is a Poisson process then $X = V_p + M_d$, where $V_p(t) = \mathbb{E}\{X(t)\}$ is continuous, and $\sum_{s\leq t} M_d(s) = X(t)$, $[X] = [M_d] = \sum_{s\leq t}(X(t) - X(t-))^2$ and $\langle X\rangle = \langle M_d\rangle = \mathbb{E}\{X(t)\}$. In general, the sum of jumps $\sum_{s\leq t} X(s)$ of a local

martingale satisfying $[X]^c = 0$ may not be defined (i.e., the series of jumps may not be pathwise convergent) or it may converge not necessarily to X. The local martingale M_d contains predictable and unpredictable jumps, and $\langle M_d \rangle$ contains only the predictable jumps, but if $\langle M_d \rangle = 0$ then $M_d = 0$. Note that the square-bracket $[\cdot]$ is defined for any semi-martingale (and so for any quasi-martingale), while the angle-bracket $\langle \cdot \rangle$ is only define for local martingale.

Refresh on Stieltjes integrals

Let us consider the pathwise Riemann-Stieltjes integral for bounded variation integrator and integrand, which is defined as a limit on partitions of a compact interval $[a, b]$. Typically, the integral exists for a continuous integrand f and a bounded variation integrator g (or conversely), but if fails to exists if both f and g are discontinuous on the same side (either right or left). The integration by part formula is granted if one of the integral exists, namely,

$$f(b)g(b) - f(a)g(a) = \int_a^b f(t)\mathrm{d}g(t) + \int_a^b g(t)\mathrm{d}f(t).$$

However, we have

$$f(b)g(b) - f(a)g(a) \quad = \int_{]a,b]} f(t)\mathrm{d}g(t) + \int_{]a,b]} g(t-)\mathrm{d}f(t).$$

in the Lebesgue-Stieltjes sense, if both f and g are only right-continuous with finite variation. Indeed, if V is a cad-lag process with locally bounded variation and X is a cad-lag process then, for any $b > a \geq 0$ we have

$$\int_a^b X_-(t)\mathrm{d}V(t) = \int_a^b X_-(t)\mathrm{d}V^c(t) + \sum_{a \leq t < b} X_-(t)\delta V(t),$$
$$\int_a^b X(t)\mathrm{d}V_-(t) = \int_a^b X(t)\mathrm{d}V^c(t) + \sum_{a < t \leq b} X(t)\delta V(t),$$

where $X_-(t) = X(t-)$, $V_-(t) = V(t-)$, for every $t > 0$, and V^c is the continuous part of V, i.e.,

$$V(t) = V^c(t) + \sum_{0 < s \leq t} \delta V(s).$$

Note that $X = X_- + \delta X$ and that we may replace X with X_- as the integrand of $\mathrm{d}V^c$. Nevertheless, if $X = U$ is a cad-lag process with locally bounded variation then we can rewrite the integration by part formula as

$$\begin{cases} U(b)V(b) - U(a)V(a) = \int_{]a,b]} U_-(t)\mathrm{d}V(t) + \\ \qquad\qquad + \int_{]a,b]} V_-(t)\mathrm{d}U(t) + \sum_{a < t \leq b} \delta U(t)\delta V(t), \end{cases} \tag{4.105}$$

where all series are absolutely convergent and all integrals are considered pathwise, in either Riemann-Stieltjes (without including any possible jump at a, but including any possible jump at b) or Lebesgue-Stieltjes sense, i.e., if μ_V denotes the Lebesgue-Stieltjes measure generated by the cad-lag path function $t \mapsto V(t)$ then

$$\int_{]a,b]} X_-(t)\mathrm{d}V(t) = \lim_{\alpha\to a^+,\ \beta\to b^+} \int_\alpha^\beta X_-(t)\mathrm{d}V(t) = \int_{]a,b]} X_-(t)\mu_V(\mathrm{d}t),$$

actually, this is the definition of the integral in Riemann-Stieltjes sense over the semi-open interval $]a,b]$ for cag-lad (left continuous with right limits) integrands and cad-lag integrators.

A cad-lag process is integrable for the (signed) Lebesgue-Stieltjes measure μ_V (which can be expressed as the difference of two measures, the positive and negative variations) and for any $b > a \geq 0$, we have

$$\begin{aligned}\int_{]a,b]} X(t)\mu_V(\mathrm{d}t) &= \int_{]a,b]} X_-(t)\mu_V(t) + \int_{]a,b]} \delta X(t)\mu_V(\mathrm{d}t) = \\ &= \int_{]a,b]} X_-(t)\mathrm{d}V(t) + \sum_{a<t\leq b} \delta X(t)\delta V(t),\end{aligned}$$

Note that $\mu_V = \mu_V^c + \mu_V^d$, where μ_V^c is the continuous part of μ_V, i.e., when all atoms have been removed (or equivalently, the measure associated with the continuous part V^c of V). Moreover, $\mu_V^c = \mu_V^a + \mu_V^s$, where μ_V^a is absolutely continuous with respect to the Lebesgue measure and μ_V^s is singular (i.e., there exists a Borel measurable set S of Lebesgue measure zero such that for any measurable set N with Lebesgue measure zero we have $\mu_V^c(N \smallsetminus S) = 0$, and then we define $\mu_V^s(A) = \mu_V^c(A \cap S)$, for any measurable set A). Thus, any set of one point $\{t\}$ is μ_V^c-negligible and so is any countable set, i.e., $\delta X = 0$ μ_V^c-almost surely and the integral of X and X_- relative to μ_V^c coincide. It is cleat that, both X_- and X (and any bounded Borel measurable process) are integrable with respect to μ_V, but to recall that the integration by part formula (4.105) should be written with X_-, we use the Riemann-Stieltjes sense over the semi-open interval $]a,b]$. Certainly, the notation $\mathrm{d}V$ actually means $\mathrm{d}\mu_V$, when the integrands are not cag-lad processes.

On the other hand, if $F\colon \mathbb{R} \to \mathbb{R}$ is a locally Lipschitz function and V is a cad-lag process with locally bounded variation then $t \mapsto F\big(V(t)\big)$ is also a cad-lag process with locally bounded variation. Moreover, if f is also continuously differentiable then we have the change of variable formula

$$\left\{\begin{aligned} F\big(V(b)\big) - F\big(V(a)\big) &= \int_{]a,b]} F'\big(V_-(t)\big)\mathrm{d}V(t) + \\ &+ \sum_{a<t\leq b} \big\{F\big(V(t)\big) - F\big(V(t-)\big) - F'\big(V(t-)\big)\delta V(t)\big\}, \end{aligned}\right. \tag{4.106}$$

where F' denotes the derivative of F. Since F' is locally bounded and V has locally bounded variation, the above series can be written as

$$\sum_{a<t\leq b} \delta F\big(V(t)\big) + \sum_{a<t\leq b} F'\big(V_-(t)\big)\delta V(t)$$

and both series are absolutely convergent. Clearly, the change of variable (4.106) is usually written as

$$F\big(V(b)\big) - F\big(V(a)\big) = \int_{]a,b]} F'\big(V_-(t)\big)\mathrm{d}V^c(t) + \sum_{a<t\le b} \delta F\big(V(t)\big),$$

and we note that $\delta F\big(V(t)\big) > 0$ if and only if $\delta V(t) > 0$.

At this point, it is important to recognize that to capture the jumps of a $\mathbb{R}^m$-valued cad-lag process X we need to study its associate integer measure ν_X, which is defined as the extension of

$$\nu_X(K\times]a,b]) = \sum_{a<s\le b} \mathbb{1}_{\delta X(s)\in K}, \qquad \text{(a finite sum)},$$

for any compact set K in $\mathbb{R}^m_* = \mathbb{R}^m \setminus \{0\}$ and any $b \ge a \ge 0$. If X is cad-lag with bounded variation then X and ν_X are equivalent in the sense that from

$$X(t) = X(0) + \int_{\mathbb{R}^m_*} z\nu_X(]0,t], \mathrm{d}x)$$

we can reconstruct X from ν_X. However, if the process X is not necessarily of bounded variation then we need to make sense to the limits

$$X(t) = X(0) + \lim_{\varepsilon\to 0} \int_{\mathbb{R}^m_*} z\mathbb{1}_{\{|z|\ge\varepsilon\}}\nu_X(]0,t], \mathrm{d}z)$$

to be able to reconstruct X. Clearly, this limit makes sense as a stochastic integral if X is a local martingale.

Square-Brackets and Angle-Brackets

If X and Y are two semi-martingales then we define the square-bracket by

$$[X,Y] = XY - \int_{]0,\cdot]} X(s-)\mathrm{d}Y(s) - \int_{]0,\cdot]} Y(s-)\mathrm{d}X(s),$$

when $X(0) = Y(0) = 0$, or by polarization as

$$[X,Y] = \big([X+Y] - [X] - [Y]\big)/2 = \big([X+Y] - [X-Y]\big)/4.$$

Similarly with the angle-bracket $\langle X,Y\rangle$, which is defined only for local martingales. If X and Y are local square-integrable martingales then $[X,Y]$ is the unique cad-lag adapted process with integrable finite variation and vanishing at 0 such that (1) $XY - [X,Y]$ is a local martingale and (2) $\delta[X,Y] = \delta X\,\delta Y$, while $\langle X,Y\rangle$ is the unique cad-lag predictable process with integrable finite variation and vanishing at 0 such that $XY - \langle X,Y\rangle$ is a local martingale. For any quadratic pure jump semi-martingale X, i.e., satisfying $[X]^c = 0$, and any semi-martingale Y we have

$$[X,Y](t) = \sum_{s\le t} \delta X(s)\,\delta Y(s), \quad \forall t > 0.$$

A local martingale X is called purely discontinuous if $X(0) = 0$ and $\langle X, Y\rangle = 0$ for any continuous local martingale Y. Then (1) a local martingale X vanishing at 0 is purely discontinuous if and only if $[X]^c = 0$, (2) a local martingale with local finite variation and $X(0) = 0$ is purely discontinuous, (3) a continuous local martingale which is purely discontinuous is indeed null, and (4) a predictable local martingale is a continuous martingale.

Let X be a quasi-martingale and V be an adapted process with local integrable finite variation and $V(0) = 0$: we have (a)

$$[X, V] = \int_{]0,\cdot]} \delta X(s) \mathrm{d}V(s), \quad XV = \int_{]0,\cdot]} V_-(s) \mathrm{d}X(s) + \int_{]0,\cdot]} X(s) \mathrm{d}V(s),$$

and (b) if V is predictable then

$$[X, V] = \int_{]0,\cdot]} \delta V(s) \mathrm{d}X(s), \quad XV = \int_{]0,\cdot]} V(s) \mathrm{d}X(s) + \int_{]0,\cdot]} X_-(s) \mathrm{d}V(s).$$

Hence, we also have (c) if X is a local martingale and V is predictable then the optional covariance or square-bracket $[X, V]$ is a local martingale, and (d) we have $[X, V](t) = \sum_{s\le t} \delta X(t)\delta V(t)$ and so $[X, V] = 0$, if at least one (either X or V) is continuous. There are several useful estimates involving local martingales, e.g., Davis-Burkhölder-Gundy inequality for local martingales vanishing at the initial time, namely, for any $p \ge 1$ there exist constants $C_p > c_p > 0$ (recall that $C_1 = 3$ and $C_2 = 4$) such that for any stopping time T and any local martingale M with $M(0) = 0$, we have

$$c_p\, \mathbb{E}\{\big([M](T)\big)^{p/2}\} \le \mathbb{E}\{\sup_{t\le T} |M(t)|^p\} \le C_p\, \mathbb{E}\{\big([M](T)\big)^{p/2}\},$$

and Lenglart domination estimate, namely, for any cad-lag adapted process X dominated by an increasing cad-lag process A with $A(0) = 0$ (i.e., $\mathbb{E}\{|X(\tau)|\} \le \mathbb{E}\{A(\tau)\}$ for any bounded stopping time τ) we have

$$P\big\{\sup_{t\le T} |X(t)| \ge \varepsilon\big\} \le \frac{1}{\varepsilon}\Big[\eta + \mathbb{E}\big\{\sup_{t\le T} |A(t) - A(t-)|\big\}\Big] + P\big\{A(T) \ge \eta\big\},$$

for any positive constants ε, η and any stopping time T, and if A is predictable, we may drop the term with the jumps. However, for any p in $(0, 2]$ there exist a constant $C_p > 0$ (with $C_1 = 3$ and $C_2 = 4$) such that

$$\mathbb{E}\{\sup_{t\le T} |M(t)|^p\} \le C_p\, \mathbb{E}\{\big(\langle M\rangle(T)\big)^{p/2}\},$$

for any stopping time T and any local martingale M with $M(0) = 0$.

Let X and Y be two semi-martingales, and a and b be two adapted cag-lad (left continuous with right limits) then

$$\Big[\int_{]0,\cdot]} a(s) \mathrm{d}X(s), \int_{]0,\cdot]} b(s) \mathrm{d}Y(s)\Big] = \int_{]0,\cdot]} a(s)b(s) \mathrm{d}[X, Y](s),$$

and similarly with the angle-bracket $\langle\cdot,\cdot\rangle$, where the last integral is in either the Riemann-Stieltjes or Lebesgue-Stieltjes sense. Suppose that X_i, $i = 1,\dots,n$ and Y_j, $j = 1,\dots,m$ are semi-martingales, and that $\varphi(x)$ and $\psi(y)$ are smooth real-valued functions then Itô formula shows that $\varphi(X)$ and $\psi(Y)$ are also semi-martingales and

$$\begin{aligned}
\langle\varphi(X),\psi(Y)\rangle(t) &= \sum_{ij}\int_0^t \partial_i\varphi(X(s))\partial_j\psi(Y)(s)\mathrm{d}\langle X_i,Y_j\rangle(s),\\
[\varphi(X),\psi(Y)]^c(t) &= \sum_{ij}\int_0^t \partial_i\varphi(X(s))\partial_j\psi(Y)(s)\mathrm{d}[X_i,Y_j]^c(s),\\
[\varphi(X),\psi(Y)](t) - [\varphi(X),\psi(Y)]^c(t) &= \sum_{s\le t}\delta\varphi(X)(s)\delta\psi(Y)(s),
\end{aligned}$$

for any $t > 0$. Hence, let ν_{XY} denote the integer measure associated with the (joint) jumps of the $\mathbb{R}^{n+m}$-valued process (X,Y), namely,

$$\nu_{XY}(B,]a,b]) \text{ is the number of jumps } \big(\delta X(s),\delta Y(s)\big) \text{ in } B \text{ within the interval }]a,b], \text{ for any } B \text{ in } \mathcal{B}(\mathbb{R}^{n+m}_*) \text{ with } \bar{B}\cap\{0\} = \emptyset \text{ and } 0 \le a < b,$$

with a predictable jumps compensator $\nu^p_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t)$. Thus, the jumps part of the optional quadratic covariation, i.e., $[X,Y] - [X,Y]^c$, can be expressed as

$$\begin{aligned}
\sum_{s\le t}\delta\varphi(X)(s)\delta\psi(Y)(s) = \int_{\mathbb{R}^{n+m}\times]0,t]} &\big[\varphi(X(s-)+x) - \varphi(X(s-))\big]\times\\
&\times\big[\psi(Y(s-)+y) - \psi(Y(s-)\big]\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t).
\end{aligned}$$

The continuous part of a semi-martingale X is defined as the unique continuous semi-martingale X^c satisfying $[X - X^c, Z] = 0$, for any continuous semi-martingale Z. Then we have $[X^c,Y] = [X,Y]^c$. On the other hand, the processes X and Y are quasi-left continuous if only if there are no predictable jumps, i.e., $\nu^p_{XY}(\mathbb{R}^{n+m}_*\times\{t\}) = 0$, for any $t \ge 0$, or equivalently the predictable covariation $\langle X,Y\rangle$ is continuous. Note that if the jumps of X and Y have the form

$$X(t) = X^c(t) + \int_{Z\times]0,t]}\gamma^X(z,s)\tilde{\nu}(\mathrm{d}z,\mathrm{d}s), \quad \forall t \ge 0,$$

and similarly for Y, with the same martingale measure $\tilde{\nu}$ and continuous processes X^c and Y^c, then

$$\begin{aligned}
\int_{\mathbb{R}^{n+m}_*\times]0,t]} h(x,y,s)\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}s) &=\\
&= \int_{Z\times]0,t]} h(\gamma^X(z,s),\gamma^Y(z,s),s)\nu(\mathrm{d}z,\mathrm{d}s),
\end{aligned}$$

for any positive Borel measurable function h, and similarly for the predictable jump compensator measure.

In particular, let $M = M_c + M_d$ be a quasi-left continuous local square-integrable martingale in $\mathbb{R}^n$ written as the sum of a continuous local martingale

$\{M_{c,i} : i = 1, \dots, n\}$ with predictable variation process $\{\langle M_{c,i}\rangle : i = 1, \dots, n\}$, and a purely discontinuous local martingale $\{M_{d,i} : i = 1, \dots, n\}$ which yields an integer measure $\nu_{_M}$ with compensator $\nu^p_{_M}$ and martingale measure $\tilde\nu_{_M} = \nu_{_M} - \nu^p_{_M}$. Note that

$$\int_{]0,t]} \alpha(s)\mathrm{d}M_{d,i}(s) = \int_{\mathbb{R}^d\times]0,t]} \alpha(s)\zeta_i\tilde\nu_{_M}(\mathrm{d}\zeta, \mathrm{d}s), \quad i = 1, \dots, d,$$

and

$$\sum_{s\le t} h\big(s, \delta M_d(s)\big) = \int_{\mathbb{R}^d\times]0,t]} h(s,\zeta)\nu_{_M}(\mathrm{d}\zeta, \mathrm{d}s),$$

for any predictable integrable processes α and h. Thus, if X is a d-dimensional processes of the form

$$\mathrm{d}X(t) = a^{\scriptscriptstyle X}(t)\mathrm{d}V^c(t) + b^{\scriptscriptstyle X}(t)\mathrm{d}M_c(t) + \int_{\mathbb{R}^m_*} \gamma^{\scriptscriptstyle X}(\zeta,t)\tilde\nu_{_M}(\mathrm{d}\zeta, \mathrm{d}t), \quad \forall t \ge 0, ,$$

where V^c is an adapted continuous process with local integrable finite variation, and φ is real-valued smooth functions then Itó formula shows that the semi-martingales $\varphi(t) = \varphi(t, X(t))$ can be expressed in term of continuous part M_c, the compensated integer (or martingale) measure $\tilde\nu_{_M}$ and some continuous locally bounded variation processes V_φ, i.e.,

$$\mathrm{d}\varphi(t) = \mathrm{d}V_\varphi(t) + b^\varphi(t)\mathrm{d}M_c(t) + \int_{\mathbb{R}^m_*} \gamma^\varphi(\zeta,t)\tilde\nu_{_M}(\mathrm{d}\zeta, \mathrm{d}t), \quad \forall t \ge 0,$$

where

$$\begin{aligned}
\mathrm{d}V_\varphi(t) &= \partial_t\varphi(t,\cdot)\mathrm{d}t + \partial_x\varphi(t,\cdot)\mathrm{d}V^c(t) + \sum_{i,j,k} b^{\scriptscriptstyle X}_{ik}(t)b^{\scriptscriptstyle X}_{jk}(t)\partial_{ij}\varphi(t,\cdot)\mathrm{d}\langle M_{c,k}\rangle(t) + \\
&\quad + \int_{\mathbb{R}^m_*\times]0,t]} \big[\varphi(t, \cdot + \gamma^{\scriptscriptstyle X}(\zeta,t)) - \varphi(t,\cdot) - \gamma^{\scriptscriptstyle X}(\zeta,t)\partial_x\varphi(t,\cdot)\big]\nu^p_{_M}(\mathrm{d}\zeta, \mathrm{d}t), \\
b^\varphi_k(t) &= \sum_i \partial_i\varphi(t,\cdot)b^{\scriptscriptstyle X}_{ik}(t), \qquad \gamma^\varphi(\zeta,t) = \varphi(t, \cdot + \gamma^{\scriptscriptstyle X}(\zeta,t)) - \varphi(t,\cdot),
\end{aligned}$$

the dot $\cdot$ is replaced by $M(t-)$. Thus

$$\begin{aligned}
\langle\varphi_c, M_{c,k}\rangle &= \sum_{i,j} \int_0^{\cdot} \partial_i\varphi(t, M(t-))b^{\scriptscriptstyle X}_{ij}(t)\mathrm{d}\langle M_{c,j}, M_{c,k}\rangle(t), \\
\delta\varphi(t) &= \int_{\mathbb{R}^m_*} \big(\varphi(t, M(t-) + \gamma^{\scriptscriptstyle X}(\zeta,t)) - \varphi(t, M(t-))\big)\nu_{_M}(\mathrm{d}\zeta, \{t\}), \\
[\varphi, M_{d,k}] &= \int_{\mathbb{R}^m_*\times]0,\cdot]} \big(\varphi(t, M(t-) + \gamma^{\scriptscriptstyle X}(\zeta,t)) - \varphi(t, M(t-))\big)\zeta_k\nu_{_M}(\mathrm{d}\zeta, \mathrm{d}t),
\end{aligned}$$

which give full information on the covariance of φ and M. These calculations are neat for the particular case where M^c is a standard Wiener process and $\tilde\nu_{_M}$ is a compensated Poisson integer measure with predictable compensator $\nu^p_{_M}(\mathrm{d}\zeta, \mathrm{d}t) = \pi(\mathrm{d}\zeta)\mathrm{d}t$.

For instance, the reader is referred to Dellacherie and Meyer [58, Sections VI.37–42, pp. 105–112], He et al. [105, Chapter VIII, pp. 209–223], Jacod and Shiryaev [117, Chapter 1, pp. 1–63], Kallenberg [121, Chapter 26, pp. 515–536], Protter [206, Chapter III, pp. 43–86], or Sections 3.4–3.5 and 3.13 in Chapter 3, for more detail on the above statements.

Martingales Integrals

After refreshing the above martingales concepts, we may reframe the stochastic integral defined previously for a quasi-martingale X. Recall that a sequence of general processes $\{u_k\}$ is said to converge to u *uniformly on compacts in probability* abbreviated "ucp", if for any $\varepsilon > 0$ there exists $K > 0$ such that $P\{\sup_{0\le t\le 1/\varepsilon} |u_k(t) - u(t)| \ge \varepsilon\} \le \varepsilon$ for any $k \ge K$. Given a filtration $\mathbb{F} = \{\mathcal{F}(t) : t \ge 0\}$, denote by $\mathtt{S} = \mathtt{S}(\mathbb{F})$, $\mathtt{L} = \mathtt{L}(\mathbb{F})$ and $\mathtt{D} = \mathtt{D}(\mathbb{F})$ the vector space of simple predictable processes (i.e., $H(t) = h_{i-1}$ for t on the stochastic interval $\rrbracket\tau_{i-1}, \tau_i\rrbracket$, for $i = 1, 2, \ldots, n$, where $\tau_0 = 0$, $\tau_{i-1} \le \tau_i$ are finite stopping times and h_i is a $\mathbb{R}^d$-valued $\mathcal{F}(\tau_i)$-measurable random variable), adapted cag-lad (left continuous with right limits) processes, and cad-lag (right continuous with left limits). If we add the ucp-convergence and version of processes are considered equals, then we use $\mathtt{S}_{\text{ucp}}$, $\mathtt{L}_{\text{ucp}}$ and $\mathtt{D}_{\text{ucp}}$. Actually, any process X in $\mathtt{L}$ (or in $\mathtt{S}$) are technically defined on $(0, \infty)$, but we assume $X(0) = X(0+)$ and so X is defined on $[0, \infty)$; otherwise, we may decide to work on the whole space $(-\infty, +\infty)$ or to explicitly introduce a value at time $t = 0$. Moreover a better notation would be $\mathtt{S}_{\text{ucp}}(\mathbb{F})$, $\mathtt{L}_{\text{ucp}}(\mathbb{F})$ and $\mathtt{D}_{\text{ucp}}(\mathbb{F})$ to recall the dependency on the filtration $\mathbb{F}$, however this is implicitly assumed.

Remark that a modification (also called a version) of an element in $\mathtt{S}$, $\mathtt{L}$ or $\mathtt{D}$ does not necessarily belongs to $\mathtt{S}$, $\mathtt{L}$ or $\mathtt{D}$, it belongs to $\mathtt{S}_{\text{ucp}}$, $\mathtt{L}_{\text{ucp}}$ or $\mathtt{D}_{\text{ucp}}$. We may have an element u in $\mathtt{L}_{\text{ucp}}$ and an element v in $\mathtt{D}_{\text{ucp}}$ such that u is a version of v, i.e., not any version of a given element in $\mathtt{L}_{\text{ucp}}$ (or $\mathtt{D}_{\text{ucp}}$) can be considered an element in $\mathtt{L}$ (or $\mathtt{D}$). On the other hand, we are allowed to modified an element (in any of the three topological vector spaces $\mathtt{S}$, $\mathtt{L}$ or $\mathtt{D}$) on a evanescent set and still remain in the same space.

Note that $\mathtt{L}_{\text{ucp}}$ and $\mathtt{D}_{\text{ucp}}$ are complete metric spaces, and let us prove that $\mathtt{S}_{\text{ucp}}$ is dense in $\mathtt{L}_{\text{ucp}}$. Indeed, given a positive number η and a u in $\mathtt{L}_{\text{ucp}}$ we define an increasing sequence of stopping times $0 = T_0^\eta < T_1^\eta < T_2^\eta < \cdots$ by recurrence

$$T_{k+1}^\eta = \inf\{s > T_k^\eta : |u(s+) - u(T_k^\eta+)| > \eta\},$$

where $T_{k+1}^\eta = \infty$ if $|u(s+) - u(T_k^\eta+)| \le \eta$, for every $s \ge T_k^\eta$. Because $t \mapsto u(t+)$ is cad-lag, the sequence T_k^η is almost surely increasing to infinite, i.e., $P\{T_k^\eta \le r\} \to 0$ as $k \to \infty$, for every positive constants r. Clearly, $|u(s+) - u(T_k^\eta+)| \le \eta$, for any s such that $T_k^\eta \le s < T_{k+1}^\eta$, and by continuity, we have $|u(s) - u(T_k^\eta+)| \le \eta$, if $T_k^\eta < s \le T_{k+1}^\eta$. Hence, define $u_k^\eta(t) = u(n \wedge T_i^\eta+)$ if $k \wedge T_i^\eta < t \le k \wedge T_{i+1}^\eta$ with $i = 0, 1, \ldots, k$ to have

$$P\Big\{\sup_{0\le t\le r} |u_k^\eta(t) - u(t)| > \eta\Big\} \le P\big\{k \wedge T_k^\eta \le r\big\}.$$

Therefore, u_k^η belongs to $\mathtt{S}_{\rm ucp}$, and for suitable η and k we construct a sequence convergent to u.

For any H in $\mathtt{S}_{\rm ucp}$ and X in $\mathtt{D}_{\rm ucp}$ we define the simple integral

$$\Sigma(H,\mathrm{d}X)(t)=\int_{]0,t]}H(s)\mathrm{d}X(s)=\sum_{i=1}^{n}h_iX(\tau_i\wedge t)-X(\tau_{i-1}\wedge t),$$

if $H=\sum_{i=1}^n h_{i-1}\mathbb{1}_{]\!]\tau_{i-1},\tau_i]\!]}$. Now if X is a quasi-martingale then this linear operator $\Sigma(\cdot,\mathrm{d}X)$ is continuous from $\mathtt{S}_{\rm ucp}$ into $\mathtt{D}_{\rm ucp}$ and therefore it can be uniquely extended to $\mathtt{L}_{\rm ucp}$, i.e., for any H in $\mathtt{L}_{\rm ucp}$ there exists a sequence H_k in $\mathtt{S}_{\rm ucp}$ such that $H_k\to H$ in $\mathtt{L}_{\rm ucp}$ and $\Sigma(H_k,\mathrm{d}X)\to\Sigma(H,\mathrm{d}X)$ in $\mathtt{D}_{\rm ucp}$. Actually, the continuity property can be proved directly or by means of Lenglart dominate estimate, namely, for any positive constants ε,η, any stopping time T, and any H in $\mathtt{L}_{\rm ucp}$, we have (recall δ is the jump operator)

$$P\big\{\sup_{t\le T}\big|\Sigma(H,\mathrm{d}A)(t)\big|\ge\varepsilon\big\}\le P\big\{A(T)\ge\eta\big\}+\\ +\frac{1}{\varepsilon}\Big[\eta+\mathbb{E}\big\{\sup_{t\le T}|H(t)|\,|\delta A(t)|\big\}\Big],$$

if A is an adapted increasing integrable process, and

$$P\big\{\sup_{t\le T}\big|\Sigma(H,\mathrm{d}M)(t)\big|^2\ge\varepsilon\big\}\le\frac{\eta}{\varepsilon}+P\big\{\Sigma(|H|^2,\mathrm{d}\langle M\rangle)(T)\ge\eta\big\},$$

if M is a local martingale with predictable variance $\langle M\rangle$.

Moreover, if $\varpi=\{t_i\}$ is a partition of $[0,\infty)$, $t_0=0$, $t_{i-1}<t_i$, $t_i\to\infty$, and $|\varpi|=\sup_i(t_i-t_{i-1})$ (possible of stopping times) and H is an element of $\mathtt{L}_{\rm ucp}$, then we may define $H_{\varpi,n}(t)=H(t_{i-1})$ for t in $(t_{i-1},t_i]$, $i=1,\dots,n$. It is clear that $H_{\varpi,n}$ belongs to $\mathtt{S}_{\rm ucp}$, and $\sup_{0\le t\le T}|H_{\varpi,n}(t)|\le C(T)$, almost surely for a constant $C(T)$ independent of ϖ,n, and $H_{\varpi,n}(t)\to H(t)$, almost surely, for every $t>0$. Hence, after using Lenglart dominate estimate, we deduce that $\Sigma(H_{\varpi,n},\mathrm{d}M)\to\Sigma(H,\mathrm{d}M)$ in $\mathtt{L}_{\rm ucp}$, as $n\to\infty$ and $|\varpi|\to0$.

Note that if X is a process with locally bounded variation belonging to $\mathtt{D}_{\rm ucp}$ and H is any process in $\mathtt{L}_{\rm ucp}$ then $\Sigma(H,\mathrm{d}M)$ coincides with the (pathwise) Riemann-Stieltjes (or Lebesgue-Stieltjes) integral.

Clearly, this technique can be extended simple integral relative to martingale measures $\tilde\nu$, e.g.,

$$\Sigma(H,\mathrm{d}\tilde\nu)(t)=\int_{\mathbb{R}^m_*\times]0,t]}H(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)=\\ =\sum_{i=1}^{n}\sum_{j=1}^{k}H_{ij}\tilde\nu(B_j\times]\tau_{i-1}\wedge t,\tau_i\wedge t]),$$

where $H=\sum_{i=1}^n\sum_{j=1}^k H_{ij}\mathbb{1}_{]\!]\tau_{i-1},\tau_i]\!]}\mathbb{1}_{B_j}$, and B_j is a Borel set separated from the origin, i.e., the closure $\bar B_j\cap\{0\}=\emptyset$. As discussed early in this chapter, the

cad-lag processes $\Sigma(H, \mathrm{d}X)$ and $\Sigma(H, \mathrm{d}\tilde{\nu})$ are local martingales with

$$[\Sigma(H, \mathrm{d}X)](t) = \int_{]0,t]} |H(s)|^2 \mathrm{d}[X](s),$$
$$[\Sigma(H, \mathrm{d}\tilde{\nu})](t) = \int_{\mathbb{R}^m_* \times]0,t]} |H(\zeta, s)|^2 \nu(\mathrm{d}\zeta, \mathrm{d}s),$$

and

$$\langle \Sigma(H, \mathrm{d}X) \rangle(t) = \int_{]0,t]} |H(s)|^2 \mathrm{d}\langle X \rangle(s),$$
$$\langle \Sigma(H, \mathrm{d}\tilde{\nu}) \rangle(t) = \int_{\mathbb{R}^m_* \times]0,t]} |H(\zeta, s)|^2 \nu^p(\mathrm{d}\zeta, \mathrm{d}s),$$

where ν^p is the predictable compensator of martingale measure $\tilde{\nu}$, i.e., for any fixed Borel set B separated from the origin, the process $t \mapsto \nu^p(B,]0, t])$ is the compensator of the local martingale $t \mapsto \tilde{\nu}(B,]0, t])$, or $\tilde{\nu}$ is the martingale measure corresponding to an integer measure ν with predictable jumps compensator ν^p. Note that

$$\int_{\mathbb{R}^m_*} |H(\zeta, t)| \nu(\mathrm{d}\zeta, \{t\}) \quad \text{replaces} \quad |H(t)|\,|\delta A(t)| \quad \text{and}$$
$$\int_{\mathbb{R}^m_* \times]0,T]} |H(\zeta, t)|^2 \nu^p(\mathrm{d}\zeta, \mathrm{d}t) \quad \text{replaces} \quad \Sigma(|H|^2, \mathrm{d}\langle M \rangle)(T)$$

in Lenglart dominate estimate.

It should be clear that besides the probability measure P, the initial filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$ plays a fundamental role in the above construction. Perhaps, a full notation for spaces $\mathtt{S}_{\mathrm{ucp}}$, $\mathtt{L}_{\mathrm{ucp}}$ and $\mathtt{D}_{\mathrm{ucp}}$ should includes the filtration and the probability, e.g., $\mathtt{D}_{\mathrm{ucp}}(P, \mathbb{F})$. However, if another filtration $\mathbb{G} = \{\mathcal{G}(t) : t \geq 0\}$ is given and

$$H \in \mathtt{L}_{\mathrm{ucp}}(P, \mathbb{F}) \cap \mathtt{L}_{\mathrm{ucp}}(P, \mathbb{G}), \qquad X \in \mathtt{D}_{\mathrm{ucp}}(P, \mathbb{F}) \cap \mathtt{D}_{\mathrm{ucp}}(P, \mathbb{G})$$

then H can be approximate in $\mathtt{S}_{\mathrm{ucp}}$ with respect to both filtrations, which implies that the limit $\Sigma(H, \mathrm{d}X)$ is independent of the particular filtration used. Certainly, if the limit exists for a probability P then also it exits for any other probability Q which is absolutely continuous with respect to P.

Non-Martingales Integrals

Consider a partition ϖ of $[0, \infty)$ and for any two cad-lag processes X and Y define the symmetric square-bracket along ϖ,

$$[X, Y]_\varpi(t) = \sum_i \big(X(t_i \wedge t) - X(t_{i-1} \wedge t)\big)\big(Y(t_i \wedge t) - Y(t_{i-1} \wedge t)\big),$$

as well as the bilinear expressions (integrals along ϖ)

$$\Sigma^-_\varpi(X,\mathrm{d}Y)(t) = \sum_i X(t_{i-1}\wedge t)\big(Y(t_i\wedge t) - Y(t_{i-1}\wedge t)\big),$$
$$\Sigma^+_\varpi(X,\mathrm{d}Y)(t) = \sum_i X(t_i\wedge t)\big(Y(t_i\wedge t) - Y(t_{i-1}\wedge t)\big),$$
$$\Sigma^\circ_\varpi(X,\mathrm{d}Y)(t) = \sum_i \big(X(t_i\wedge t) + X(t_{i-1}\wedge t)\big)\big(Y(t_i\wedge t) - Y(t_{i-1}\wedge t)\big)/2,$$

which are finite sums of non-zero terms. Note the relations

$$\Sigma^\circ_\varpi(X,\mathrm{d}Y)(t) = \tfrac{1}{2}\big(\Sigma^-_\varpi(X,\mathrm{d}Y)(t) + \Sigma^+_\varpi(X,\mathrm{d}Y)(t)\big),$$
$$\Sigma^+_\varpi(X,\mathrm{d}Y)(t) - \Sigma^-_\varpi(X,\mathrm{d}Y)(t) = [X,Y]_\varpi(t) = [Y,X]_\varpi(t),$$

and

$$\Sigma^+_\varpi(X,\mathrm{d}Y)(t) + \Sigma^-_\varpi(Y,\mathrm{d}X)(t) = Y(t)X(t) - X(0)Y(0),$$
$$\Sigma^\mp_\varpi(X,\mathrm{d}Y)(t) + \Sigma^\mp_\varpi(Y,\mathrm{d}X)(t) \pm [X,Y]_\varpi(t) = Y(t)X(t) - X(0)Y(0),$$

where we use the telescopy sum

$$\sum_{i=1}^n a_i(b_i - b_{i-1}) + \sum_{i=1}^n b_{i-1}(a_i - a_{i-1}) = a_n b_n - a_0 b_0,$$

valid for any numbers a_i and b_i.

For any cad-lag process X, we can consider the cag-lad process X_- defined as the left-hand limits, i.e., $X_-(t) = X(t-)$. If δ is the jump operator then we have $\delta X = \delta X_-$, $X_- = X - \delta X$, and $X = X_- + \delta X$. Hence we have $\Sigma^-_\varpi(X_-,\mathrm{d}Y) \to \Sigma(X_-,\mathrm{d}Y)$ in $\mathtt{D}_{\mathrm{ucp}}$ as $|\varpi| \to 0$, for any X in $\mathtt{D}_{\mathrm{ucp}}$ and for any quasi-martingale Y.

If X and Y are quasi-left continuous then for any t there exist a null set N_t such that $\delta X(t,\omega) = 0$ and $\delta Y(t,\omega) = 0$, for any ω in $\Omega \setminus N_t$. Thus,

$$\begin{cases} [X,Y]_\varpi(t) + \Sigma^-_\varpi(X_-,\mathrm{d}Y)(t) + \Sigma^-_\varpi(Y_-,\mathrm{d}X)(t) = \\ \qquad\qquad = Y(t)X(t) - X(0)Y(0), \end{cases} \tag{4.107}$$

almost surely, for each t. In particular, this proves that $[X,Y]_\varpi \to [X,Y]$ in $\mathtt{D}_{\mathrm{ucp}}$ and that $\Sigma^\pm_\varpi(X_-,\mathrm{d}Y)$ and $\Sigma^\pm_\varpi(X,\mathrm{d}Y)$ have a common limit $\mathtt{D}_{\mathrm{ucp}}$, as $|\varpi| \to 0$, for any quasi-left continuous quasi-martingales X and Y.

Our interest is on processes where the jumps are only due to a local martingale, i.e., the finite variation part of X can be chosen continuous. Now, let π be a Lévy measure in $\mathbb{R}^d_*$, $\gamma(z)$ be a (deterministic) function in $L^2(\mathbb{R}^m_*,\pi)$ and X be a real-valued Itô process with jumps,

$$X(t) = \int_0^t a^X(s)\mathrm{d}v(s) + \int_0^t b^X(s)\mathrm{d}w(s) + \int_{\mathbb{R}^m_*\times]0,t]} c^X(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s), \tag{4.108}$$

for any $t \geq 0$, where v is a d-dimensional adapted *continuous* process with local integral finite variation, w is a d-dimensional standard Wiener process independent of the compensated Poisson point process $\tilde{p}$ with Levy measure π in $\mathbb{R}^m_*$, and the coefficients suitable predictable processes, i.e., a^x is locally integrable with respect to the variation process $|\mathrm{d}v|$, b^x is locally square integrable, and c^x is jointly locally square integrable relative to $\pi(\mathrm{d}\zeta) \times \mathrm{d}t$. Choose $Y = w_k$ or $Y = \tilde{p}_\gamma$, where

$$\tilde{p}_\gamma(b) - \tilde{p}_\gamma(a) = \tilde{p}(\gamma,]a, b]) = \int_{\mathbb{R}^m_* \times]a,b]} \gamma(z)\tilde{p}(\mathrm{d}z, \mathrm{d}t), \quad \forall b > a \geq 0. \tag{4.109}$$

Thus, the expressions $[X, w_k]_\varpi$, $[X, \tilde{p}_\gamma]_\varpi$, $\Sigma^\pm_\varpi(X, \mathrm{d}w_k)$, $\Sigma^\circ_\varpi(X, \mathrm{d}w_k)$, $\Sigma^\pm_\varpi(X, \mathrm{d}\tilde{p}_\gamma)$ and $\Sigma^\circ_\varpi(X, \mathrm{d}\tilde{p}_\gamma)$ are adapted quasi-left continuous and cad-lag processes, and we may replace X by X_- without any modifications. By means of Itô formula we can calculate the predictable and optional covariances

$$\begin{aligned}
&[X, v_k](t) = \langle X, v_k\rangle(t) = 0, \qquad [X, w_k](t) = \langle X, w_k\rangle(t) = \int_0^t b^x_k(t)\mathrm{d}t,\\
&\langle X, \tilde{p}_\gamma\rangle(t) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^d_*} c^x(z, s)\gamma(z)\pi(\mathrm{d}z),\\
&[X, \tilde{p}_\gamma](t) = \int_{\mathbb{R}^d_* \times]0,t]} c^x(z, s)\gamma(z)p(\mathrm{d}z, \mathrm{d}s),
\end{aligned}$$

and, for instance, Theorem 4.34 shows that

$$\begin{aligned}
&\lim_{|\varpi|\to 0} [X, w_k]_\varpi = [X, w_k], \qquad \lim_{|\varpi|\to 0} [X, \tilde{p}_\gamma]_\varpi = [X, \tilde{p}_\gamma],\\
&\lim_{|\varpi|\to 0} \Sigma^-_\varpi(X, \mathrm{d}w_k) = \int_0^\cdot X(t)\mathrm{d}w_k(t),\\
&\lim_{|\varpi|\to 0} \Sigma^-_\varpi(X_-, \mathrm{d}\tilde{p}_\gamma) = \int_{\mathbb{R}^d_* \times]0,\cdot]} X(t-)\gamma(z)\tilde{p}(\mathrm{d}z, \mathrm{d}t),
\end{aligned}$$

where the limits are uniformly on compacts in probability (i.e., in the ucp sense). Moreover, because the limits of the two last term of the left-hand side of the equality (4.107) converges to the stochastic integrals, we re-establish the convergence of the square-bracket to the optional covariation. Clearly, for the adapted continuous process having local bounded variation v we have

$$\begin{aligned}
&\lim_{|\varpi|\to 0} [X, v_k]_\varpi = 0,\\
&\lim_{|\varpi|\to 0} \Sigma^-_\varpi(X, \mathrm{d}v_k) = \int_0^\cdot X(t-)\mathrm{d}v_k(t),
\end{aligned}$$

where the integral is pathwise, in either Riemann-Stieltjes or Lebesgue-Stieltjes sense.

If $\{\gamma_j : j \geq 1\}$ is an orthonormal basis in $L^2(\mathbb{R}^m_*, \pi)$ then the jumps of X given by (4.108) can be expressed as

$$\int_{\mathbb{R}^m_* \times]0,t]} c^x(\zeta, s)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) = \sum_j \int_{]0,t]} c^x_j(s)\mathrm{d}\tilde{p}_j(s),$$

where

$$c_j^X(s) = \int_{\mathbb{R}_*^m} c^X(\zeta, s)\gamma_j(\zeta)\pi(\mathrm{d}\zeta), \qquad \tilde{p}_j(t) = \int_{\mathbb{R}_*^m \times]0,t]} \gamma_j(\zeta)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s),$$

$c_j^X(s)$ are predictable processes and $\tilde{p}_i(s)$ are purely discontinuous martingales, and

$$\mathrm{d}\langle \tilde{p}_i, \tilde{p}_j\rangle(t) = \Big(\int_{\mathbb{R}_*^m} \gamma_i(\zeta)\gamma_j(\zeta)\pi(\mathrm{d}\zeta)\Big)\mathrm{d}t,$$

i.e., $\langle \tilde{p}_i, \tilde{p}_j\rangle(t) = t$ if $i = j$ and $\langle \tilde{p}_i, \tilde{p}_j\rangle = 0$ otherwise. Thus, we may rewrite X as

$$X(t) = \int_0^t a^X(s)\mathrm{d}v(s) + \int_0^t b^X(s)\mathrm{d}w(s) + \sum_j \int_{]0,t]} c_j^X(s)\mathrm{d}\tilde{p}_j(s),$$

for any $t \geq 0$. Formally, we have $\tilde{p} = \sum_j \gamma_j \tilde{p}_j$, but

$$\mathbb{E}\big\{ |\sum_j \gamma_j \tilde{p}_j(t)|_\pi^2 \big\} = \mathbb{E}\big\{ \sum_j |\tilde{p}_j(t)|^2 \big\} = \mathbb{E}\big\{ t \sum_j |\gamma_j|_\pi^2 \big\} = \infty, \quad \forall t > 0,$$

i.e., the series cannot be considered as $L^2(\mathbb{R}_*^m, \pi)$-valued martingale. However, for as given convergent sequence of strictly positive numbers $\{\kappa_i\}$ we may consider the Hilbert subspace

$$H = H_{\kappa,\gamma,\pi} = \big\{ h \in L^2(\mathbb{R}_*^m, \pi) : \sum_i \kappa_i \Big| \int_{\mathbb{R}_*^m} h(\zeta)\gamma_i(\zeta)\pi(\mathrm{d}\zeta)\Big|^2 < \infty \big\}.$$

Hence, we may regard the series $\tilde{p}(s) = \sum_j \gamma_j \tilde{p}_j(s)$ and $c^X(s) = \sum_j \gamma_j c_j^X(s)$ as processes with values in H,

$$\|\tilde{p}(s)\|_H^2 = \sum_j \kappa_j |\tilde{p}_j(s)|^2 \leq \Big(\sum_j \kappa_j \Big) s,$$
$$\|c^X(s)\|_H^2 = \sum_j \kappa_j |c_j^X(s)|^2 \leq \Big(\sum_j \kappa_j \Big) \int_{\mathbb{R}_*^m} |c^X(\zeta, s)|^2 \pi(\mathrm{d}\zeta),$$

and $\tilde{p}$ is a local martingale, while $c^X(s)$ is predictable with values in the dual space H', via the functional Riesz representation, and the duality inclusion $H \subset L^2(\mathbb{R}_*^m, \pi) \subset H'$. Therefore, the stochastic integral with respect to the (local) martingale measure $\tilde{p}$ can be regarded as an stochastic integral with respect to a (local) martingale with values in the Hilbert space H and a predictable process with values in its dual space H'. Nevertheless, we may define

$$\langle\!\langle X, \tilde{p}\rangle\!\rangle := \sum_j \langle X, \tilde{p}_j\rangle = \sum_j \int_0^\cdot c_j^X(s)\mathrm{d}s =$$
$$= \sum_j \int_0^\cdot \mathrm{d}s \int_{\mathbb{R}_*^m} c^X(\zeta, s)\gamma_j(\zeta)\pi(\mathrm{d}\zeta),$$
$$[\![X, \tilde{p}]\!] := \sum_j [X, \tilde{p}_j] = \sum_j \sum_{s \leq \cdot} \delta X(s)\delta\tilde{p}_j(s) =$$
$$= \sum_j \int_{\mathbb{R}_*^m \times]0,\cdot]} c^X(\zeta, t)\gamma_j(\zeta)p(\mathrm{d}z, \mathrm{d}t),$$

if the coefficients are sufficiently smooth (in time) to make the above series convergent. Since the integrand is predictable,

$$\mathbb{E}\Big\{ \int_{\mathbb{R}_*^m \times]0,T]} c^X(\zeta,t)\gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}t)\Big\} = \\ = \mathbb{E}\Big\{ \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} c^X(\zeta,t)\gamma_j(\zeta)\pi(\mathrm{d}z)\Big\} = \mathbb{E}\Big\{ \int_0^T c_j^X(t)\mathrm{d}t\Big\},$$

for any $T > 0$.

Now, recall that a Poisson measure p is a sum of (random) Dirac measures, i.e., $p(K,]a,b]) = \sum_{a<s\le t} \mathbb{1}_{\delta p(s)\in K}$ where $\delta p(s)$ denotes the jumps at time s (i.e., the Poisson point process originating the Poisson measure p), and assume that γ_j is π-integrable so that

$$\int_{\mathbb{R}_*^m \times]0,T]} \gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}t) = \\ = \int_{\mathbb{R}_*^m \times]0,T]} \gamma_j(\zeta)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}t) + T\mathbb{E}\Big\{ \int_{\mathbb{R}_*^d} \gamma_j(\zeta)\pi(\mathrm{d}\zeta)\Big\}$$

can be defined. Therefore, the integer measure ν_j induced by the pathwise integral of γ_j over p, i.e.,

$$p(\gamma_j,t) = \int_{\mathbb{R}_*^d \times]0,t]} \gamma_j(\zeta)p(\mathrm{d}\zeta,\mathrm{d}s), \qquad \nu_j(K,]a,b]) = \sum_{a<s\le b} \mathbb{1}_{\{\delta p(\gamma_j,t)\in K\}}$$

are defined, and the jump satisfy $\delta\nu_j(t) = \delta p(\gamma_j,t) = \gamma_j(\delta p(t))$. Hence the integer measure ν_j is indeed a Poisson measure with Lévy measure $\pi_j(\mathrm{d}\zeta) = \gamma_j(\zeta)\pi(\mathrm{d}\zeta)$. Moreover, $t \mapsto \nu_j(\mathbb{R}_*^d,]0,t])$ is a composed (real valued) Poisson process with the finite measure π_j on $\mathbb{R}_*^d$ as parameter.

Thus, the stochastic integral with respect to either the initial Poisson measure $\{p(K,]a,b])\}$ or its associate Poisson point process $\{\delta p(t)\}$ can be written as an orthogonal series either $\{\nu_j\}$ or $\{\delta p(\gamma_j,t)\}$, i.e., with $\tilde{\nu}_j = \nu_j - \pi_j(\mathrm{d}\zeta)\mathrm{d}t$,

$$\int_{\mathbb{R}_*^m \times]0,t]} c^X(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s) = \sum_j \int_{\mathbb{R}_*^d \times]0,t]} c^X(\zeta,s)\tilde{\nu}_j(\mathrm{d}\zeta,\mathrm{d}s).$$

and

$$\int_{\mathbb{R}_*^d \times]0,t]} c^X(\zeta,s)\nu_j(\mathrm{d}\zeta,\mathrm{d}s) = \int_{\mathbb{R}_*^d \times]0,t]} c^X(\zeta,s)\gamma_j(\zeta)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s).$$

Note that we may write $\tilde{p}_j(t) = \tilde{p}(\gamma_j,]0,t])$ and in the proper meaning for the jumps operator δ we have $\delta\tilde{p}_j(t) = \delta p(\gamma_j,t)$.

Sometimes, it is convenient to use the following

Definition 4.39. Let X be Itô processes with jumps as above, satisfying (4.108). We define the *backward*, *forward* and *symmetric* (or Stratonovich)

stochastic integrals in term of the Itô stochastic integral as follows

$$\begin{aligned}
\int_{]0,T]} X(t-)\mathrm{d}^- w_k(t) &= \int_{]0,T]} X(t-)\mathrm{d}w_k(t),\\
\int_{]0,T]} X(t-)\mathrm{d}^+ w_k(t) &= \int_{]0,T]} X(t-)\mathrm{d}w_k(t) + \int_0^T b_k(t)\mathrm{d}t,\\
\int_{]0,T]} X(t-)\mathrm{d}^\circ w_k(t) &= \int_{]0,T]} X(t-)\mathrm{d}w_k(t) + \frac{1}{2}\int_0^T b_k(t)\mathrm{d}t
\end{aligned}$$

and in general, for any two given semimartingales M and N we define the *backward*, *forward* and *symmetric* (or Stratonovich) stochastic integrals in term of the Itô stochastic integral as follows

$$\begin{aligned}
\int_{]0,T]} M(t-)\mathrm{d}^- N(t) &= \int_{]0,T]} M(t-)\mathrm{d}N(t),\\
\int_{]0,T]} M(t-)\mathrm{d}^+ N(t) &= \int_{]0,T]} M(t-)\mathrm{d}N(t) + \big[M,N\big](T),\\
\int_{]0,T]} M(t-)\mathrm{d}^\circ N(t) &= \int_{]0,T]} M(t-)\mathrm{d}N(t) + \frac{1}{2}\big[M,N\big](T).
\end{aligned}$$

Clearly, this take place in a probability space $(\Omega, \mathcal{F}, P)$, with a completed (relative to $\mathcal{F}$), right-continuous and quasi-left continuous filtration $\mathbb{F}$. □

Remark that because the martingales are quasi-continuous and the local finite variation part is continuous, we are allow to use the square bracket $[\cdot,\cdot]$ instead of the angular bracket $\langle\cdot,\cdot\rangle$ as usually, without to much complication in the calculations, since jumps are deduced from the martingale measure $\tilde{\nu}$.

Thus, if M and N are two local martingales with values in $\mathbb{R}^d$ and $\mathbb{R}$ (not necessarily continuous, but quasi-continuous and relative to the same filtered space) and associated martingale measures $\tilde{\nu}_{M}(\mathrm{d}z,\mathrm{d}t) = c^{M}(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$ and $\tilde{\nu}_{N}(\mathrm{d}z,\mathrm{d}t) = c^{N}(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$, for some integer measure ν in $\mathbb{R}^m_*$, then

$$\int_{]0,T]} \varphi\big(M(t-)\big)\mathrm{d}^\circ N(t) = \int_{]0,T]} \varphi\big(M(t-)\big)\mathrm{d}N(t) + \frac{1}{2}\big[\varphi(M),N\big](T),$$

for any smooth function $\varphi(x)$, and

$$\begin{aligned}
\big[\varphi(M),N\big](T) = \sum_{i=1}^{d}\int_0^T \partial_i\varphi\big(M(t-)\big)\mathrm{d}\big[M_i,N\big]^c(t) +\\
+ \int_{\mathbb{R}^m_*\times]0,T]} \big[\varphi\big(M(t-)+c^{M}(z,t)\big) - \varphi\big(M(t-)\big)\big]c^{N}(z,t)\nu(\mathrm{d}z,\mathrm{d}t),
\end{aligned}$$

where $\partial_i\varphi$ denotes the derivative in x, and ν is the common integer measure. Clearly, the predictable covariance $\langle\varphi(M),N\rangle$ has an expression similar to the above with ν^p replacing the ν. In general, we may use the integer measure ν_{XY} in $\mathbb{R}^{d+1}_*$ associated with the (purely discontinuous part of the) $\mathbb{R}^{d+1}$-valued local martingale (M,N), where we replace the integer measure $\nu(\mathrm{d}z,\mathrm{d}t)$ with $\nu_{XY}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}t)$ and the coefficients $c^{M}(z,t)$ and $c^{N}(z,t)$ with the variables x and

y. In this case, the variable (x, y) belongs to $\mathbb{R}_*^{d+1}$ and the integral should be in $\mathbb{R}_*^{d+1}$. However, because ν_{XY} is an integer measure and the integrand function $[\varphi(\cdot + x) - \varphi(x)]y$ vanishes if $x = 0$ or $y = 0$, the integral is only on the region $\{(x, y) \in \mathbb{R}_*^{m+1} : x \neq 0,\ y \neq 0\} \times]0, T]$ as expected, i.e., when both martingales have jumps simultaneously.

It is clear that the *vector-form* is deduced from the above definition and the *operational* Itô rule becomes

$$X(b)Y(b) - X(a)Y(a) = \int_{]a,b]} X(t-)\mathrm{d}^\circ Y(t) + \int_{]a,b]} Y(t-)\mathrm{d}^\circ X(t),$$

i.e., as the deterministic case, with all the jumps incorporated into the integral. Note that the processes X and Y are cad-lag and quasi-left continuous, and that the bounded variation part v is a continuous process. In general,

$$\begin{aligned}\varphi(X(T)) - \varphi(X(0)) = &\int_{]0,T]} \partial_x \varphi(X(t-))\mathrm{d}X(t) + \int_{]0,T]} \partial_x \varphi(X(t-))\mathrm{d}[X]^c(t) + \\ &+ \sum_{0<t\leq T} \Big(\varphi(X(t)) - \varphi(X(t-)) - \nabla\varphi(X(t-)) \cdot \delta X(t)\Big),\end{aligned}$$

for any a smooth function φ. Clearly, we have

$$\begin{aligned}&\sum_{0<t\leq T} \Big(\varphi(X(t)) - \varphi(X(t-)) - \partial_x \varphi(X(t-))\delta X(t)\Big) = \\ &\quad = \int_{\mathbb{R}_*^d \times]0,T]} \big[\varphi(X(t-) + z) - \varphi(X(t-)) - z \cdot \nabla\varphi(X(t-))\big]\nu_X(\mathrm{d}z, \mathrm{d}t),\end{aligned}$$

where ν_X is the integer measure associated with X in $\mathbb{R}_*^d$. Clearly, $\nu_X = \tilde{\nu}_X + \nu_X^p$, where $\tilde{\nu}_X$ is the martingale measure (yielding the martingale stochastic integral) and ν_X^p is its predictable jump compensator.

The square-bracket $[\cdot, \cdot]$ and the integer measures can be defined for any cad-lag processes, non necessarily semimartingales. The previous relations between the backward, forward and symmetric integrals with the quadratic variation are essential for this analysis. The interested reader may consult for instance, Chao and Chou [38], Errami et al. [75], Fisk [82], Föllmer [86], Meyer [180], among others.

4.3.4 Non-Anticipative Processes

The concept of non-anticipative or non-anticipating is rather delicate, and usually it means adapted or strictly speaking, if a process is adapted then it should be non-anticipative. For instance, a random process x is called non-anticipative with respect to a Markov process y if the past of x is independent of the future of y given the present of y, this means that given a realization y of a Markov process in some probability space $(\Omega, \mathcal{F}, P)$ with values in a topological space Y then any process x with values in some topological space X is called non-anticipative with respect to y if for any bounded Borel measurable functions f,

g and h and times $s_1 < \cdots < s_n \le t \le t_1 < t_2 < \cdots < t_n$, we have

$$\begin{aligned}\mathbb{E}\big\{ f(x_{s_1}, \dots, x_{s_n})\, g(y_t)\, h(y_{t_1}, \dots, y_{t_n}) \big\} = \\ = \mathbb{E}\big\{ \mathbb{E}\{ f(x_{s_1}, \dots, x_{s_n}) \,|\, y_t \}\, g(y_t)\, \mathbb{E}\{ h(y_{t_1}, \dots, y_{t_n}) \,|\, y_t \} \big\},\end{aligned}$$

where n is arbitrary. Note that the three functions f, g and h may be taken only bounded continuous, as long as the Baire σ-algebra (the one generated by continuous functions on X and Y) and the Borel σ-algebra coincide, e.g., if (X, d) is a metric space then $F = \{x \in X : d(x, X) = \inf_{u \in F} d(x, u) = 0\}$ for any closed subset F of X, so $x \mapsto d(x, F)$ is continuous, and so both σ-algebras coincide. Since Jakubowski topology is weaker that a metrizable topology, the Baire and the Borel σ-algebras coincide in this case too. Usually, X and Y are some $\mathbb{R}^n$ and the processes x and y are at least stochastically right continuous. It is convenient to take a cad-lag version of x and y if possible.

On the other hand, if y is a random process with independent increments and $y_0 = 0$, then a non-anticipative process x is a process such that the past of x is independent of the increments of y given the present of y, i.e, for any bounded Borel measurable functions f, g and h and times $s_1 < \cdots < s_n \le t \le t_1 < t_2 < \cdots < t_n$, we have

$$\begin{aligned}\mathbb{E}\big\{ f(x_{s_1}, \dots, x_{s_n})\, g(y_t)\, h(y_{t_2} - y_{t_1}, \dots, y_{t_n} - y_{t_{n-1}}) \big\} = \\ = \mathbb{E}\big\{ \mathbb{E}\{ f(x_{s_1}, \dots, x_{s_n}) \,|\, y_t \}\, g(y_t)\, \mathbb{E}\{ h(y_{t_2} - y_{t_1}, \dots, y_{t_n} - y_{t_{n-1}}) \,|\, y_t \} \big\},\end{aligned}$$

where n is arbitrary. In any case, note that (contrary to the adapted case) if x_1 and x_2 are non-anticipative then the cartesian product (x_1, x_2) is not necessarily non-anticipative. Recall that y is a process of independent increments (i.e., $y(t_1), \dots, y(t_n)$ are independent of $y(s_2) - y(s_1)$, for any $t_1 < \cdots < t_n < s_1 < s_2$), if and only if $y = m + a$, where m is a semi-martingale (and a process of independent increments) and a is a deterministic cad-lag process (e.g., see Jacod and Shiryaev [117, Theorem II.5.1, p. 114]).

Perhaps a better concept is the following:

Definition 4.40. For a given a process y in a probability space $(\Omega, \mathcal{F}, P)$ we define the *non-anticipative* filtration $\mathbb{A} = \{\mathcal{A}(t) : t \ge 0\}$, where $\mathcal{A}(t)$ is the σ-algebra composed by all sets in $\mathcal{F}$ which are independent of $y_{t_1} - y_{t_0}, \dots y_{t_n} - y_{t_{n-1}}$, for any $t \le t_0 < t_1 < \cdots < t_n$, and $n \ge 1$. So a measurable process x is non-anticipative with respect to y if it is adapted to $\mathbb{A}$, i.e., if for any bounded Borel measurable functions f and g we have

$$\begin{aligned}\mathbb{E}\{ f(x_{s_1}, \dots, x_{s_n})\, g(y_{t_1} - y_{t_0}, \dots, y_{t_n} - y_{t_{n-1}}) \} = \\ = \mathbb{E}\{ f(x_{s_1}, \dots, x_{s_n}) \}\, \mathbb{E}\{ g(y_{t_1} - y_{t_0}, \dots, y_{t_n} - y_{t_{n-1}}) \},\end{aligned}$$

for any times $s_1 < \cdots < s_n \le t_0 < t_1 < \cdots < t_n$. □

Clearly, once the non-anticipative filtration $\mathbb{A}$ has been defined, the concept of a non-anticipative process reduces to being adapted to the non-anticipative filtration $\mathbb{A}$. However, a good part for this concept is the fact of being a finite-dimensional property, i.e., if x' and y' two processes in another probability space

$(\Omega', \mathcal{F}', P')$ with the same (joint) finite-dimensional distributions as x and y then x' is also non-anticipative with respect to y'.

Alternatively, if y is a random process with orthogonal (or uncorrelated) increments and $y_0 = 0$, then any random process x which is orthogonal (or uncorrelated) to the increments of y could be called weakly non-anticipative, i.e., if

$$\mathbb{E}\big\{x_s \cdot (y_{t_2} - y_{t_1})\big\} = \mathbb{E}\{x_s\} \cdot \mathbb{E}\{(y_{t_2} - y_{t_1})\},$$

for any $0 \le s \le t_1 < t_2$, where the $\cdot$ denotes the scalar product. Certainly, an orthogonal process x is weakly non-anticipative if x_t belongs to the closed linear span of the variables $y_{s_1}, \dots, y_{s_n}$, with $0 \le s_1 < \ldots < s_n \le t$. All this means that any information on x does not help to gain some extra information on the characteristics of y. However, the following concept seems better for martingales.

Recall that for a $\mathbb{R}^d$-valued stochastic process y, the martingale property reads as follows:

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(y(s_1), \dots, y(s_n)\big)\big\} = 0,$$

for any bounded continuous functions f and any times $s_1 < \cdots < s_n \le s \le t$. This is a property finite-dimensional (i.e., any other stochastic process y' satisfies the above martingale properties provided $\mathbb{E}\big\{f\big(y(s_1), \dots, y(s_n)\big)\big\} = \mathbb{E}'\big\{f\big(y'(s_1), \dots, y'(s_n)\big)\big\}$, for any bounded continuous functions f and any times $s_1 < \cdots < s_n$), which makes sense for processes satisfying $\mathbb{E}\{|y(t)|\} < \infty$ for every $t \ge 0$ (or for a time-localization, as in the case of local martingales). However, most of the useful results for martingale processes requires a separable martingale, and separability is not finite-dimensional property.

Thus, of particular interest for us is the case when y is a (local) martingale.

Definition 4.41. Let y be a $\mathbb{R}^d$-valued (separable) martingale (with zero mean) in some probability space $(\Omega, \mathcal{F}, P)$. A process x is called *weakly non-anticipative* with respect to y if for any bounded continuous functions f and g and any times $s_1 < \cdots < s_n \le s \le t$ and $s'_1 < \cdots < s'_n \le s$, we have

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(x(s_1), \dots, x(s_n)\big)\, g\big(y(s'_1), \dots, y(s'_n)\big)\big\} = 0.$$

If y is a martingale relative to a filtration $\mathbb{F} = (\mathcal{F}_t : t \ge 0)$ then we say that x is weakly non-anticipative with respect to y (and $\mathbb{F}$) if for any bounded continuous functions f and any times $s_1 < \cdots < s_n \le s \le t$, we have

$$\mathbb{E}\big\{\big(y(t) - y(s)\big)\, f\big(x(s_1), \dots, x(s_n)\big)\, z_s\big\} = 0.$$

where z_s is any bounded $\mathcal{F}_s$-measurable function. Clearly, this notion extends to local martingales or semi-martingales. This means that the stochastic process x does not change the martingale property of y. □

It is clear that weakly non-anticipative is a finite-dimensional distribution property when the filtration is not mentioned, i.e., if x' and y' two processes in another probability space $(\Omega', \mathcal{F}', P')$ with the same finite-dimensional distributions and y' being integrable, then y is a martingale and x is non-anticipative

with respect to y if and only if then x' is non-anticipative with respect to y'. Also, if $\mathcal{F}(x,t)$ denotes the σ-algebra generated by the random variables $x(s)$, $0 \le s \le t$, then x is non-anticipative with respect to y if $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ is orthogonal to the increments $y(b)-y(a)$, for any $b > a \ge t$, where $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ is the minimal σ-algebra containing both $\mathcal{F}(x,t)$ and $\mathcal{F}(y,t)$.

Recall that a general (local) martingale is a (local) integrable process y satisfying the martingale property, namely,

$$\mathbb{E}\{y(t) \,|\, \mathcal{F}(y,s)\} = y(s), \quad \forall t \ge s \ge 0,$$

or equivalently

$$\mathbb{E}\big\{\big(y(t)-y(s)\big)\, f\big(y(s_1),\dots,y(s_n)\big)\big\} = 0, \quad \forall 0 \le s_1 < \cdots < s_n \le s < t,$$

and any arbitrary bounded continuous function f. Note that when the prefix *general* (or separable) is used, we mean that no particular version (or that a separable version) has been chosen.

Thus, if x is an adapted process to a martingale y relative to the filtration $\mathbb{F}$ then $\mathcal{F}_t$ contains $\mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ and x results non-anticipative with respect to y and $\mathbb{F}$. Note that if x_1 and x_2 are two weakly non-anticipative processes then the cartesian product (x_1, x_2) is not necessarily weakly non-anticipative, clearly, this is not the case for adapted processes. Conversely, if x is weakly non-anticipative with respect to a general (local) martingale y we deduce that x is certainly adapted to $\mathcal{F}(t) = \mathcal{F}(x,t) \vee \mathcal{F}(y,t)$ and also that y satisfies the martingale property relative to $\mathcal{F}(t)$, instead of just $\mathcal{F}(y,t)$. Moreover, if y is cad-lag then the martingale property holds for $\mathcal{F}^+(t) = \cap_{\varepsilon>0}\mathcal{F}(t+\varepsilon)$.

Now, if we assume that y is a general martingale (non necessarily cad-lag) with $t \mapsto \mathbb{E}\{y(t)\}$ cad-lag (which is a finite-dimensional distribution property) then there is a cad-lag version of y, still denoted by y, where the above argument applies. Therefore, starting with a process x weakly non-anticipative with respect to y (satisfying the above conditions) we obtain a filtration $\{\mathcal{F}^+(t) : t \ge 0\}$ such that x is adapted and y is a (local) martingale. If the function $t \mapsto \mathbb{E}\{y(t)\}$ is continuous then the process y has also a cag-lad version (left continuous having right-hand limit) which is denoted by y_-, with $y_-(0) = y(0)$ and $y_-(t) = \lim_{\varepsilon\to 0} y(t-\varepsilon)$, $t > 0$. In this case, x is also weakly non-anticipative with respect to y_-, since any version of y can be used.

Recall that with the above notation, a process x is progressively measurable if $(t,\omega) \mapsto x(t,\omega)$, considered as defined on $[0,T]\times\Omega$ is measurable with respect to the product σ-algebra $\mathcal{B}([0,T]) \times \mathcal{F}(x,T)$ or $\mathcal{B}([0,T]) \times \mathcal{F}(T)$, if the family of increasing σ-algebra $\{\mathcal{F}(t) : t \ge 0\}$ is a priori given. Progressively measurability and predictability are not a finite-dimensional distribution property, but for a given filtration and assuming that x is adapted and stochastically left continuous, we can obtain a predictable version of x. Similarly, if x is adapted and stochastically right continuous then there exists a progressively measurable version.

Suppose that x and y are two weakly non-anticipative processes with respect to M, which is a cad-lag square-integrable martingale. Let M_c and ν_M be their

associated continuous part and integer measure, with predictable covariance $\ell_M = \langle M_c \rangle$, martingale measure $\tilde{\nu}_M$ and predictable jump compensator $\nu_{M,p} = \pi_M \mathrm{d}\varrho_M$, where π_M is a Levy measure and ϱ_M is a predictable continuous increasing process. If

$$P\Big\{ \int_0^t |x(s)|^2 \mathrm{d}\ell_M(s) < \infty \Big\} = 1$$

and

$$P\Big\{ \int_0^t \mathrm{d}\varrho_M(s) \int_{\mathbb{R}_*^m} |y(\zeta, s)|^2 \pi_M(\mathrm{d}\zeta) < \infty \Big\} = 1$$

then the stochastic integrals

$$\int_0^t x(s) \mathrm{d}M_c(s) \qquad \text{and} \qquad \int_{\mathbb{R}_*^m \times (0,t]} y(\zeta, s) \tilde{\nu}_M(\mathrm{d}\zeta, \mathrm{d}s)$$

can be defined. Now, assume that in some other probability space there are processes $(x', y', M', \ell'_M, \varrho'_M)$ having the same finite-dimensional distribution, where M' is cad-lag, ℓ'_M and ϱ'_M continuous (and increasing), and x and y are almost surely integrable with respect to $\mathrm{d}\ell'_M$ and $\mathrm{d}\pi_M \mathrm{d}\varrho'_M$, respectively. Thus, M' is a cad-lag martingale and $(x, y, \ell'_M, \varrho'_M)$ is weakly non-anticipative with respect to M', hence, for a suitable filtration $\mathbb{F}$ the process M' remains a martingale and x and y adapted processes, ℓ'_M and ϱ'_M are predictable processes. Then the associate continuous martingale M'_c and integer measure ν'_M have predictable covariance $\langle M_c \rangle = \ell'_M$ and predictable jump compensator $\nu'_{M',p} = \pi_M \mathrm{d}\varrho'_M$, where ℓ'_M and $\mathrm{d}\varrho'_M$ are continuous. Hence, the stochastic integrals

$$\int_0^t x'(s) \mathrm{d}M'_c(s) \qquad \text{and} \qquad \int_{\mathbb{R}_*^m \times (0,t]} y'(\zeta, s) \tilde{\nu}_{M'}(\mathrm{d}\zeta, \mathrm{d}s)$$

are defined and have the same finite-dimensional distributions. In this sense, the stochastic integral are preserved if the characteristics of the integrand and integrator are preserved.

4.3.5 Functional Representation

First we recall a basic result (due to Doob) about functional representation, e.g., see Kallenberg [121, Lemma 1.13, pp. 7-8]. Given a probability space, let b and m be two random variables with values in B and M, respectively, where $(B, \mathcal{B})$ is a Borel space (i.e., a measurable space isomorphic to a Borel subset of $[0, 1]$, e.g., a Polish space) and $(M, \mathcal{M})$ is a measurable space. Then b is m-measurable (i.e., measurable with respect to the σ-algebra generated by m) if and only if there exists a measurable function h from M into B such that $b = h(m)$.

In general, a Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}_*^m,\ t \geq 0)$, with Lévy measure $\pi(\cdot)$ is composed by a complete filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, the stochastic process $(w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener space and $(\nu(B,]0, t]) : B \in \mathbb{R}_*^m,\ t \geq 0)$ is an

independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{\nu(B,]0,t])\}/t$, which satisfies

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1+|\zeta|}\pi(\mathrm{d}\zeta) < \infty,$$

with martingale measure $\tilde{\nu}(B,]0,t]) := \nu(B,]0,t]) - t\pi(B)$. This martingale measure $\tilde{\nu}$ is identified with the $\mathbb{R}^m$-valued (Poisson) compensated-jump process

$$\tilde{p}(t) := \int_{\mathbb{R}^m_* \times]0,t]} \zeta\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0,$$

in the sense that given the Poisson integer measure ν we obtain the Poisson martingale measure $\tilde{\nu}$, which yields the Poisson compensated-jump process $\tilde{p}$, and conversely, starting from a Poisson compensated-jump process $\tilde{p}$ we may define a Poisson integer measure

$$\nu(B,]0,t]) = \sum_{0<s\leq t} \mathbb{1}_{\{\tilde{p}(s)-\tilde{p}(s-)\in B\}},$$

which yields the Poisson martingale measure $\tilde{\nu}$. Thus, only the p and $\tilde{p}$ is used instead of ν and $\tilde{\nu}$, i.e., the Poisson jump-compensated process $\tilde{p}$ and the Poisson martingale measure $\tilde{p}$ are used indistinctive, and differentiated from the context.

• *Remark* 4.42. Using $\tilde{p}$ instead of $\tilde{\nu}$ in the setting of the stochastic integral results in an integrand of the form

$$\gamma_i(\zeta,t) = \sum_j \tilde{\gamma}_i(t)\zeta_j,$$

i.e., particular cases, but sufficiently general for all considerations. □

It should be clear that a Wiener-Poisson space could be called a Gauss-Poisson space or a Lévy space since $\ell = w + \tilde{p}$ is a (centered) Lévy process, where w is its continuous or Gaussian part and $\tilde{p}$ is its purely jumps or Poisson part. We prefer to emphasize the fact that a Wiener process and a Poisson measure are the driven objects. Recalling that any continuous martingale is orthogonal to any purely discontinuous martingale (with respect to a common filtration), we deduce that the processes $\phi(w) - \phi(0)$ and $\psi(\tilde{p}) - \mathbb{E}\{\psi(\tilde{p})\}$ are orthogonal martingales for any smooth functions ϕ and ψ, i.e., w and $\tilde{p}$ (or ν) are independent. Then, as long as the filtration $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$ is given and w, $\tilde{p}$ (or ν) are martingales, the independence of the Wiener process and the Poisson measure is granted.

As mentioned early, the canonical Wiener-Poisson measure P is defined on canonical sample space

$$\mathbb{C}_n \times \mathbb{D}_m = C([0,\infty),\mathbb{R}^n) \times D([0,\infty),\mathbb{R}^m)$$

as having characteristic measure

$$\mathbb{E}\big\{\exp\big[\mathtt{i}\,\xi\cdot x(t)\big]\big\} = \exp\Big\{-t\Big[\frac{|\xi_1|^2}{2} + \int_{\mathbb{R}^m_*}\big(1 - \mathrm{e}^{\mathtt{i}\,\xi_2\cdot\zeta} + \mathtt{i}\,\xi_2\cdot\zeta\big)\pi(\mathrm{d}\zeta)\Big]\Big\},$$

for every $t \geq 0$ and $\xi = (\xi_1, \xi_2)$ in $\mathbb{R}^n \times \mathbb{R}^m$, where $x(t)$ is the projection (or evaluation) map from $\mathbb{C}_n \times \mathbb{D}_m$ into $\mathbb{R}^n \times \mathbb{R}^m$.

The canonical sample space $\mathbb{C}_n \times \mathbb{D}_m$ is a Polish space (with the locally uniform convergence in the first variable and the Skorokhod topology in the second variable) and a probability measure is then defined on the Borel σ-algebra $\mathcal{B}(\mathbb{C}_n \times \mathbb{D}_m)$ which coincides with the σ-algebra generated by the projections, i.e., $\mathcal{F}^0_\infty = \sigma(x_t : t \geq 0)$. Also, we have the (uncompleted) filtration $\mathbb{F}^0 = \{F^0_t : t \geq 0\}$ generated by the projection maps x, i.e., $\mathcal{F}^0_t = \sigma(x_s : 0 \leq s \leq t)$. This filtration induces a predictable σ-algebra $\mathcal{P}^0$ on $\mathbb{R}^+ \times \mathbb{C}_n \times \mathbb{D}_m$, i.e., $\mathcal{P}^0$ is the σ algebra generated by the sets of the form $\{0\} \times F_0$ or $(s,t] \times F_s$, for any F_s in $\mathcal{F}^0_s$, $t > s \geq 0$. Because we are working on the sample space of cad-lag processes, the predictable σ-algebra $\mathcal{P}^0$ is not the same as the *optional* σ-algebra $\mathcal{O}^0$ (also called well-measurable), generated by sets of the form $\{0\} \times F_0$ and $[s,t) \times F_s$ for any F_s in $\mathcal{F}^0_s$, any $t > s \geq 0$. Similarly, the σ-algebra $\mathcal{M}^0$ of *progressively measurable* sets is composed by all subsets A of $\Omega \times [0,\infty)$ such that $A \cap (\Omega \times [0,t])$ belongs to $\mathcal{F}^0(t) \times \mathcal{B}([0,t])$ for every $t \geq 0$. Clearly, on the sample space on $\mathbb{C}_k$ we have $\mathcal{P}^0 = \mathcal{O}^0 = \mathcal{M}^0$, while on $\mathbb{D}_k$ we have only $\mathcal{O}^0 = \mathcal{M}^0$ as expected. Sometimes, this predictable σ-algebra $\mathcal{P}^0$ is universally completed, i.e., one universally complete $\mathcal{F}^0_t$ to $\mathcal{F}^u_t$ and then $\mathcal{P}^u$ is constructed. We proceed similarly with $\mathcal{O}^0$ and $\mathcal{M}^0$ to get $\mathcal{O}^u$ and $\mathcal{M}^u$. The interested reader is referred to the book Bichteler [25], where various measurability questions are treated in great details.

• *Remark* 4.43. Let $(\Omega, \mathcal{F}, P)$ be a probability space with $\mathcal{F}$ not necessarily completed with respect to P. If y is a cad-lag process (i.e., a random variable with values in some $\mathbb{D}_k$) and $\mathcal{F}^0_t(y)$ denotes the σ-algebra generated by the random variables $\{y(s) : 0 \leq s \leq t\}$ then the filtration $\mathbb{F}^0(y) = \{\mathcal{F}^0_t(y) : t \geq 0\}$ is not necessarily neither right-continuous nor complete. However, if y is a Lévy process and we add all null sets then we obtain a complete (relative to $\mathcal{F}$) right-continuous filtration, i.e, if $\mathcal{N}$ denotes the σ-algebra of all the P-null sets in $\mathcal{F}$ then $\mathcal{F}_t(y) = \mathcal{F}^0_t(y) \vee \mathcal{N}$ satisfies $\mathcal{F}_t(y) = \cap_{s>t}\mathcal{F}_s(y)$, for any $t \geq 0$, see Proposition 3.22 in Chapter 3. In particular, if y is a Lévy process and z is a $\mathbb{R}^k$-valued stochastic process which is predictable, optional or progressively measurable relative to $\mathbb{F}(y) = \{\mathcal{F}_t(y) : t \geq 0\}$ then there exists a version of z which is predictable, optional or progressively measurable relative to $\mathbb{F}^0(y)$, and so $P\{z(t) = h(t, y|_{[0,t]})\} = 1$, for every $t \geq 0$, for some measurable function h from $\mathbb{R}^+ \times \mathbb{D}_k$ endowed with the σ-algebra $\mathcal{P}^0$, $\mathcal{O}^0$ or $\mathcal{M}^0$ into $\mathbb{R}^k$, where $y|_{[0,t]}$ means the random variable $\omega \mapsto y(\cdot \wedge t, \omega)$. □

Now we are ready to discuss the following

Definition 4.44. A *non-anticipating functional* is any Borel measurable function f from $\mathbb{C}_n \times \mathbb{D}_m$ into $\mathbb{C}_k \times \mathbb{D}_\ell$ such that the mapping $x \mapsto f(x)(t)$ with values in $\mathbb{R}^{k+\ell}$ is $\mathcal{F}^0_t$-measurable, for every $t \geq 0$. Similarly, a measurable function from $(\mathbb{R}^+ \times \mathbb{C}_n \times \mathbb{D}_m, \mathcal{P}^0)$ into $\mathbb{R}^{k+\ell}$ is called a *predictable functional*. Moreover, if the universally completed σ-algebra $\mathcal{F}^u_t$ or $\mathcal{P}^u$ is used instead of $\mathcal{F}^0_t$ or $\mathcal{P}^0$, then the prefix *universally* is added, e.g., an universally predictable functional. □

Because non-anticipating functionals take values in some $\mathbb{C}_k \times \mathbb{D}_\ell$, the notions of optional, progressively measurable and adapted functional coincide. Actually, another name for non-anticipating functionals could be progressively measurable or optional functionals. Furthermore, we may consider predictable functionals defined on $E \times \mathbb{R} \times \mathbb{C}_n \times \mathbb{D}_m$ or $\mathbb{R} \times \mathbb{C}_n \times \mathbb{D}_m \times E$, for any Polish space E, in particular $E = \mathbb{R}_*^m$ or $E = \mathbb{R}^d$. Clearly the identity map is a non-anticipating functional and the following function

$$(t, x) \mapsto x_-(t), \quad \text{where} \quad x_-(0) = 0, \quad x_-(t) = \lim_{s \to t^-} x(t), \quad t > 0,$$

is a predictable functional. Perhaps another typical example is the (stochastic) integral of a simple integrand, i.e., if $0 = t_0 < t_1 < t_2 < \cdots < t_n$ are given real numbers and g_i is a (real-valued) measurable function in $(\mathbb{C}_n \times \mathbb{D}_m, \mathcal{F}^0_{t_{i-1}})$, for every $i = 1, \ldots, n$, then

$$x \mapsto z, \qquad z(t) = \sum_{i=1}^{n} g_i(x)[x(t \wedge t_i) - x(t \wedge t_{i-1})], \; t \geq 0, \tag{4.110}$$

defines a non-anticipating functional, and $z(t) = z(t_n)$ if $t \geq t_n$. Moreover, if t_i are stopping times relative to the uncompleted filtration $\mathbb{F}^0$ then g_i should be (real-valued) $\mathcal{F}^0(t_{i-1})$-measurable functions. Furthermore, if f is a non-anticipating functional then the mapping $(t, x) \mapsto f_-(t, x)$ defined as $f_-(t, x) = f(x_-(t))$ is a predictable functional.

• *Remark* 4.45. Once a probability P is given in $\mathbb{C}_n \times \mathbb{D}_m$ we complete the predictable σ-algebra, i.e., we may complete first the filtration and then we generate the predictable σ-algebra. Thus, an integrand of stochastic integrals is a predictable process y, which is identified with its equivalence class, relative to the measure $\mathrm{d}t \times P(\mathrm{d}\omega)$, for the Wiener process, and to the measure $\pi(\mathrm{d}\zeta) \times \mathrm{d}t \times P(\mathrm{d}\omega)$, for the Poisson measure. In this case, any adapted (and measurable) process has a predictable process belonging to the same equivalence class, moreover, once a predictable (respect to the completed filtration) representative of the equivalence class has been chosen, there is a version which is predictable with respect to uncompleted filtration, i.e., a predictable functional. Hence, in the case of the canonical Wiener-Poisson integrals, any integrands may be assumed to be a predictable functionals. □

On the canonical Wiener-Poisson space, the filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ is the minimal completed filtration (and right-continuous) such that canonical process x is adapted. However, given a Wiener-Poisson space, the filtration is also assumed given and it may not be the one generated by the Wiener process w and the Poisson measure ν. Therefore, if in a given Wiener-Poisson space the filtration results to be the one generated by the Wiener process w and the Poisson measure ν, then we can consider the image measure and reduce to the canonical Wiener-Poisson space.

Suppose that on the canonical Wiener-Poisson space with Lévy measure π, we are given some real-valued adapted processes $(a_i(t) : t \geq 0, \; i = 1, \ldots, d)$,

$(b_{ik}(t) : t \geq 0,\ i = 1, \dots, d,\ k = 1, \dots, n)$ and $(\gamma_i(\zeta, t) : t \geq 0,\ \zeta \in \mathbb{R}_0^m)$, such that for every $i = 1, \dots, d$ and any $r = 1, 2, \dots$, we have

$$\int_0^T \Big[|a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}_*^m} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t < \infty, \tag{4.111}$$

P-almost surely for any $T > 0$. This means that a_i, b_{ik} and γ_j are real-valued predictable functionals $a_i(t, w, \tilde{p})$, $b_{ik}(t, w, \tilde{p})$ and $\gamma_i(\zeta, t, w, \tilde{p})$. Hence, an Itô process with jumps takes the form

$$\begin{cases} X_i(t) = \displaystyle\int_0^t a_i(s, w, \tilde{p}) \mathrm{d}s + \sum_{k=1}^n \int_0^t b_{ik}(s, w, \tilde{p}) \mathrm{d}w_k(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}_*^m \times]0,t]} \gamma_i(\zeta, s, w, \tilde{p}) \tilde{p}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq 0, \end{cases} \tag{4.112}$$

for any $i = 1, \dots, d$. We may use the notation $X(t) = X(t, \omega, w, \tilde{p})$, with ω in $\Omega = \mathbb{C}_n \times \mathbb{D}_m$, or just $X = X(w, \tilde{p})$ to emphasize the dependency on the Wiener process and the Poisson measure $\tilde{p}$.

Proposition 4.46. *Any Itô process with jumps of the form (4.112) is a non-anticipating functional on the canonical Wiener-Poisson space, namely, $X = F(w, \tilde{p})$, for some non-anticipating functional. Moreover, if $(\Omega', P', \mathbb{F}', w', p')$ is another Wiener-Poisson space then*

$$P'\big\{ X'(w', \tilde{p}') = F(w', \tilde{p}') \big\} = 1,$$

i.e., the stochastic integral is a non-anticipating functional on the Wiener-Poisson space.

Proof. This means that we should prove that any process of the form (4.112) is indistinguishable from a non-anticipating functional. As usual, by a localization argument, we may assume that the predictable functional coefficients satisfy

$$\int_0^T \mathbb{E}\Big\{ |a_i(t)| + \sum_{k=1}^n |b_{ik}(t)|^2 + \int_{\mathbb{R}_*^m} |\gamma_i(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big\} \mathrm{d}t < \infty.$$

Now, if the coefficients are piecewise constant (i.e., simple or elementary functions) then (as noted early) the stochastic integral is a non-anticipating functional.

In general, by a monotone class argument (or merely, by the proper definition of the stochastic integral) we may find a sequence of elementary predictable functionals a^k, b^k and γ^k such that

$$\mathbb{E}\Big\{ \int_0^T \Big[|a^k(t) - a(t)| + |b^k(t) - b(t)|^2 + \\ + \int_{\mathbb{R}_*^m} |\gamma^k(\zeta, t) - \gamma(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \Big] \mathrm{d}t \Big\} \to 0,$$

for any $T > 0$. Then, by passing to a subsequence if necessary, we have

$$\sup_{0\le t\le T} |X^k(t,w,\tilde{p}) - X(t,w,\tilde{p})| \to 0,$$

outside of a set N with $P(N) = 0$, for any $T > 0$, where $X^k(t,w,\tilde{p})$ denotes the stochastic integral with elementary integrands a^k, b^k and γ^k.

Hence, if F_k is a non-anticipating functional satisfying $X^k(w,\tilde{p}) = F_k(w,\tilde{p})$ then define

$$F(w,\tilde{p}) = \begin{cases} \lim_k F_k(w,\tilde{p}) & \text{in} \quad \Omega \setminus N, \\ 0 & \text{in} \quad N, \end{cases}$$

where the limit is uniformly on $[0,T]$, any $T > 0$. Actually, we can use the convergence in L^2-sup-norm to define the non-anticipating functional F. Thus $X = F(w,\tilde{p})$.

This procedure gives an approximation independent of the particular Wiener process and Poisson measure used, so that the same approximation yields the equality $X'(w',\tilde{p}') = F(w',\tilde{p}')$, P'-almost surely. □

Now, let η and ξ be two cad-lag non-anticipative processes relative to $(w,\tilde{p})$, see Definition 4.41, and assume that each component η_i of η is non-decreasing. The non-anticipative property imply that if $\mathbb{F}_{\eta,\xi} = \mathbb{F}(w,\tilde{p},\eta,\xi)$ is the minimum completed filtration such that $(w,\tilde{p},\eta,\xi)$ is adapted to, then $(w,\tilde{p})$ is a martingale, i.e., $(\Omega,P,\mathbb{F}_{\eta,\xi},w,\tilde{p})$ is a Wiener-Poisson space. Moreover, any $\mathbb{F}_{\eta,\xi}$-adapted process y can be represented by a predictable functional, i.e., $y(t) = y(t,w,\tilde{p},\eta,\xi)$, P-almost surely, for almost every t, where $(t,w,\tilde{p},\eta,\xi) \mapsto y$ is a measurable function from $\mathbb{R}\times\mathbb{C}_n\times\mathbb{D}_{m+r+d}$ into $\mathbb{R}^{k+\ell}$.

Proposition 4.47. *Let us assume that a_{ik}, b_{ik} and γ_i are real-valued predictable functional on $\mathbb{C}_n\times\mathbb{D}_{m+r+d}$ as above. Then the stochastic integral*

$$\begin{cases} X_i(t) = \xi_i(t) + \sum_{j=1}^{r}\int_0^t a_{ij}(s)\mathrm{d}\eta_k(s) + \sum_{k=1}^{n}\int_0^t b_{ik}(s)\mathrm{d}w_k(s) + \\ \qquad\qquad + \int_{\mathbb{R}^m_*\times]0,t]} \gamma_i(\zeta,s)\tilde{p}(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall\, t \ge 0, \end{cases} \tag{4.113}$$

defines a non-anticipating functional, i.e., $X = F(w,\tilde{p},\eta,\xi)$. Moreover, if the process η is also a non-anticipating functional $\eta(w,\tilde{p},\xi)$ then $X = G(w,\tilde{p},\xi)$. Furthermore, if $H(w,\tilde{p},\xi)$ denotes a non-anticipating functional corresponding to a deterministic process ξ, then for any Wiener-Poisson space $(\Omega',P',\mathbb{F}',w',p')$ with a cad-lag process ξ' independent of $(w',\tilde{p}')$ the stochastic integral process like (4.113) is clearly defined and denoted by $X'(t)$, and we have $X'(t) = H(w',\tilde{p}',\xi')$, P'-almost surely.

Proof. The arguments are essentially the same as in previous Proposition 4.46. Note that the functional $G(w',\tilde{p}',\xi')$ depends on the distribution P_ξ on $\mathbb{D}_d$.

Perhaps we should make some comments on the functional H. Indeed, if the coefficients are simple (or elementary) functions then the stochastic integral takes the form

$$X(t) = \xi(t) + \sum_{i=1}^{n} a^{i-1}[\eta(t \wedge t_i-) - \eta(t \wedge t_{i-1})] + \\ + \sum_{i=1}^{n} b^{i-1}[w(t \wedge t_i) - w(t \wedge t_{i-1})] + \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma^{i-1,j}\tilde{p}(K_j \times]t_{i-1}, t \wedge t_i]),$$

where a^i, b^i and γ^i are themselves predictable functionals depending on some parameter integer k. This defines a approximating functional $H_k(w, \tilde{p}, \xi)$, having the desired properties, which are preserved (P- or P'-) almost surely as k goes to infinite. □

Certainly, an important particular case is when the process $\xi(\cdot)$ is actually equal to a $\mathbb{R}^d$-valued random variable ξ, which is independent of the Wiener process and the Poisson measure p.

4.4 Convergence of Integral Processes

A crucial point is to find a convergent (in various topologies) subsequence from a given sequence of stochastic processes. In the following we collect various sufficient (and necessary in some cases) conditions to this end.

4.4.1 Standard Convergences

An important result related with stochastically (left or right) continuous processes can be found in Skorokhod [229, Section 1.6, pp. 9–14]

Theorem 4.48 (Skorokhod)**.** *Let $X_n = (X_n(t) : t \geq 0)$, $n = 1, 2, \ldots$ be a sequence of stochastically continuous processes with values in $\mathbb{R}^d$ in the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n, t, s satisfying $0 \leq t \leq 1/\varepsilon$, $0 \leq s \leq 1/\varepsilon$, $|t - s| < \delta$ we have*

$$P_n\{|X_n(t)| \geq 1/\delta\} + P_n\{|X_n(t) - X_n(s)| \geq \varepsilon\} \leq \varepsilon. \tag{4.114}$$

Then there exist a stochastically continuous process $\tilde{X} = (\tilde{X}(t) : t \geq 0)$ and a subsequence, indexed by N, of stochastic processes $\tilde{X}_n = (\tilde{X}_n(t) : t \geq 0)$, n in N, all with values in $\mathbb{R}^d$ and defined in another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, such that X_n and $\tilde{X}_n$ have the same finite-dimensional distributions for every n in N and

$$\lim_{n \in N} \sup_{0 \leq t \leq 1/\varepsilon} \tilde{P}\{|\tilde{X}_n(t) - \tilde{X}(t)| \geq \varepsilon\} = 0, \tag{4.115}$$

for every $\varepsilon > 0$. □

Certainly, the construction uses the canonical probability space $(\tilde\Omega, \tilde{\mathcal{F}}, \tilde P)$, where $\tilde\Omega = [0,1]$ with the Lebesgue measure $\tilde P$ on the Borel σ-algebra $\tilde{\mathcal{F}} = \mathcal{B}([0,1])$. It is clear that each process X_n or $\tilde X_n$ may be only left (or right) stochastically continuous and the result remain valid. Moreover, if the processes $\{X_n : n \in N\}$ are continuous or cad-lag then there are continuous or cad-lag version of the processes $\{\tilde X_n : n \in N\}$. Indeed, denote by $\tilde P_n^*$ the outer measure on the product space $(\mathbb{R}^d)^{[0,\infty)}$ associated with the process $\tilde X_n$, or equivalently to X_n. Since X_n is cad-lag, $\tilde P_n^*\{D([0,\infty),\mathbb{R}^d)\} = 1$, and therefore the set

$$\{\tilde\omega \,:\, \tilde X_n(\cdot,\tilde\omega) \notin D([0,\infty),\mathbb{R}^d)\}$$

has $\tilde P_n^*$-measure zero. However, the limit process $\tilde X$ may not be continuous nor cad-lag, since in (4.114) the sup is outside of the probability.

The Skorokhod Representation Theorem can be generalized to a metric space $(\mathbb{X},\rho)$ we have the following result, where on $[0,1]$ is considered with the usual Borel σ-algebra and the standard Lebesgue measure (sometimes referred to as the universal probability space).

Theorem 4.49. *Given a weak convergent sequence of probability measures on metric space* $\mathbb{X}$, $\mu_n \to \mu_0$, *assume that either* $\mathbb{X}_0$ *is* separable *or each* μ_n, $n = 0,1,\dots\}$ *is* tight. *Then there exist a sequence of random variables* $X_n\colon [0,1] \to \mathbb{X}_0$, $n = 0,1,\dots\}$, *such that* (1) *the image measures of* X_n *are the* μ_n *and* (2) $X_n(\theta) \to X(\theta)$, *for any* θ *in* $[0,1]$. □

Lévy processes are characterized by being stochastically continuous processes with a specific Lévy characteristic function (or Fourier transform), with drift vector b, covariance matrix a and Lévy measure (or jump intensity) π. Hence, if X_n are Lévy processes then so are the processes $\tilde X_n$, $\tilde X$, after choosing appropriate versions. In particular this applies to Wiener processes and Poisson measures.

Another point of view in this direction is to consider a $\mathbb{R}^d$-valued stochastic process as a probability measure in a canonical space such as $C([0,\infty),\mathbb{R}^d)$ or the space $D([0,\infty),\mathbb{R}^d)$, of continuous or cad-lag functions from $[0,\infty)$ into $\mathbb{R}^d$, which are Polish (i.e., separable, complete and metric) spaces. Thus, a continuous or cad-lag stochastic process is a random variable with values in either $C([0,\infty),\mathbb{R}^d)$ or $D([0,\infty),\mathbb{R}^d)$. The modulus of continuity and its equivalent for cad-lag process can be estimated as follows:

(1) if X is a separable process on $[0,T]$ such that there exist positive constants p, q, C such that

$$\mathbb{E}\big\{|X(t) - X(s)|^p\big\} \le C|t-s|^{1+q}, \quad \forall t,s \in [0,T],$$

then for every $0 < \alpha < q/p$ we have

$$\lim_{\varepsilon\to 0} \varepsilon^{-\alpha}\, \rho_{\scriptscriptstyle C}(\varepsilon, X, T) = 0,$$
$$\rho_{\scriptscriptstyle C}(\varepsilon, X, T) := \sup_{0\le t\le s\le t+\varepsilon\le T} \big\{|X(t) - X(s)|\big\},$$

almost surely.

(2) if X is a separable process on $[0,T]$ such that there exist positive constants p, q, C such that

$$\mathbb{E}\big\{\big[|X(t+\delta)-X(s)|\wedge|X(s)-X(t)|\big]^p\big\}\le C\delta^{1+q},\quad \forall\delta>0,$$

for every $0\le t\le s\le t+\delta\le T$ then for every $0<\alpha<q/p$ we have

$$\lim_{\varepsilon\to 0}\varepsilon^{-\alpha}\,\rho_{\scriptscriptstyle D}(\varepsilon,X,T)=0,$$
$$\rho_{\scriptscriptstyle D}(\varepsilon,X,T):=\sup_{0\le t\le s\le t+\varepsilon\le T}\big\{|X(t+\varepsilon)-X(s)|\wedge|X(s)-X(t)|\big\},$$

almost surely.

To check these statements, we consider the diadic numbers $\mathbb{D}=\{k2^{-n}:k=0,1,\dots,2^n,\ n\ge 0\}$ on the time interval $[0,T]=[0,1]$, for simplicity. For each path, define

$$Z_n:=\sum_{k=0}^{2^n-1}Z_n(k),\qquad Z_n(k):=|X((k+1)2^{-n})-X(k2^{-n})|,$$

to get

$$|X(t)-X(s)|\le Z_n,\quad \forall t,s\in\mathbb{D},\ |t-s|=2^{-n}.$$

Since $\mathbb{D}$ is a separant subset of X, this shows that

$$\rho_{\scriptscriptstyle C}(2^{-n},X,T)\le\sup_{m\ge n}Z_m\le\sum_{m\ge n}Z_m.$$

The assumption on the process X in (1) yields

$$P\big\{|X(t)-X(s)|\ge|t-s|^\alpha\big\}\le C|t-s|^{1+\beta},$$

for every t, s in $[0,1]$ and with $\beta:=p-q\alpha$. Therefore

$$P\big\{Z_m\ge 2^{-m\alpha}\big\}\le\sum_{k=0}^{2^m-1}P\big\{Z_m(k)\ge 2^{-m\alpha}\big\}\le 2^{-m\beta},$$

and

$$P\Big\{\sum_{m\ge n}Z_m\ge\sum_{m\ge n}2^{-m\alpha}\Big\}\le\sum_{m\ge n}2^{-m\beta}=\frac{2^{-n\beta}}{1-2^{-\beta}}.$$

Hence

$$P\big\{2^{n\alpha}\,\rho_{\scriptscriptstyle C}(2^{-n},X,T)\ge\frac{1}{1-2^{-\alpha}}\big\}\le\frac{2^{-n\beta}}{1-2^{-\beta}},$$

and by means of the Borel-Cantelli lemma, we deduce that

$$\limsup_{\varepsilon\to 0} \varepsilon^{-\alpha}\, \rho_{C}(\varepsilon, X, T) \le \frac{1}{1-2^{-\alpha}},$$

almost surely, i.e, statement (1) for any $0 < \alpha' < \alpha$. To show assertion (2), we may redefine

$$\begin{aligned} Z_n(k,\ell) &:= |X((k+1)2^{-n}) - X(k2^{-n})| \wedge \\ &\qquad \wedge |X(\ell 2^{-n}) - X((\ell-1)2^{-n})|, \\ Z_n &:= \sum_{0<\ell\le k<2^m} Z_n(k,\ell) \end{aligned}$$

to get

$$\rho_{D}(2^{-n}, X, T) \le \sum_{m\ge n} Z_m,$$

and then to conclude similarly as above.

Going back to previous Theorem 4.48, if the processes X_n are cad-lag and the condition (4.114) is replaced by the following assumption: for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$\begin{cases} P_n\{w(X_n,\delta,1/\delta) \ge \varepsilon\} + P_n\{\sup\limits_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} \le \varepsilon, \\ w(X_n, r, T) = \inf\limits_{t_i} \max\limits_{i} \sup\limits_{t_{i-1}\le s,t<t_i} |X_n(t) - X_n(s)| \end{cases} \tag{4.116}$$

where $0 = t_0 < t_1 < \cdots < t_{n-1} < T \le t_n$, $t_i - t_{i-1} \ge r$, $i = 1,\dots,n$, then the limit $\tilde{X}$ is a cad-lag process and the sequence of laws $\tilde{P}_n$ (of X_n or equivalently of $\tilde{X}_n$) on the canonical space $D([0,\infty),\mathbb{R}^d)$ converge weakly to the law of $\tilde{X}$. Similarly, if the processes X_n are continuous and the condition (4.114) is replaced by: for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} + P_n\{\sup_{T(\varepsilon,\delta)} |X_n(t) - X_n(s)| \ge \varepsilon\} \le \varepsilon, \tag{4.117}$$

where now $T(\varepsilon,\delta)$ is the subset of t, s satisfying $0 \le t \le 1/\varepsilon - \delta$, $0 \le s \le 1/\varepsilon$, $|t-s| \le \delta$, then the limit $\tilde{X}$ is a continuous process and the sequence of law $\tilde{P}_n$ on the canonical space $C([0,\infty),\mathbb{R}^d)$ converges weakly to the law of $\tilde{X}$.

Sometime the above criteria (of tightness) of a sequence X_n could be not usable or hard to meet, specially the condition relative the uniform sup-bound on the increments in either (4.116) or (4.117). The so-called Aldous' criterion for tightness is a suitable tool. This reads as follows

Theorem 4.50. *Let $X_n = (X_n(t) : t \ge 0)$, $n = 1,2,\dots$ be a sequence of adapted cad-lag processes with values in $\mathbb{R}^d$ in the filtered probability spaces $(\Omega_n, \mathcal{F}_n, P_n, \mathcal{F}_n(t) : t \ge 0)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such*

that for every n and stopping times τ, θ satisfying $\theta \le \tau \le 1/\varepsilon$, $\tau - \theta \le \delta$, we have

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \ge 1/\delta\} + P_n\{|X_n(\tau) - X_n(\theta)| \ge \varepsilon\} \le \varepsilon. \tag{4.118}$$

Denote by $\tilde{P}_n$ the probability law of the process X_n in the canonical space $D([0,\infty),\mathbb{R}^d)$ of cad-lag functions. Then there exist a probability measure $\tilde{P}$ in $D([0,\infty),\mathbb{R}^d)$ and a subsequence, indexed by N, of $\{\tilde{P}_n : n \ge 1\}$ such that

$$\lim_{n\in N} \tilde{P}_n(F) \le \tilde{P}(F), \quad \forall \text{ closed } F \in D([0,\infty),\mathbb{R}^d), \tag{4.119}$$

and we also have $\tilde{P}_n(h) \to \tilde{P}(h)$, for every bounded h which is $\tilde{P}$-almost surely continuous, i.e., $\tilde{P}_n$ converge weakly to $\tilde{P}$. Moreover, in some probability space $(\Omega,\mathcal{F},P)$ there are random variables $\tilde{X}_n$ and $\tilde{X}$ with values in $D([0,\infty),\mathbb{R}^d)$ and distributions $\tilde{P}_n$ and $\tilde{P}$, respectively, such that $\tilde{X}_n$ converges in probability $\tilde{X}$. Furthermore, if we assume that for every $\varepsilon > 0$ there exists an index n_ε such that

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t) - X_n(t-)| \ge \varepsilon\} \le \varepsilon \quad \forall n \ge n_\varepsilon \tag{4.120}$$

then the limiting probability measure $\tilde{P}$ satisfies $\tilde{P}\big(C([0,\infty),\mathbb{R}^d)\big) = 1$, i.e., $\tilde{P}$ defines a probability measure on the canonical space $C([0,\infty),\mathbb{R}^d)$ of continuous functions. □

It is clear that the statement regarding the $D([0,\infty),\mathbb{R}^d)$-valued random variables comes from Skorokhod theorem. Recall that, if $\rho_D(\cdot,\cdot)$ denotes the metric in the Polish space $D([0,\infty),\mathbb{R}^d)$, then $\tilde{X}_n$ converges in probability $\tilde{X}$ if and only if for every $\varepsilon > 0$ we have

$$\lim_{n\in N} \tilde{P}\{\rho_D(\tilde{X}_n,\tilde{X}) \ge \varepsilon\} = 0,$$

in particular

$$\lim_{n\in N} \tilde{P}\{\sup_{T(\varepsilon)} |\tilde{X}_n(t+\varepsilon) - \tilde{X}_n(s)| \wedge |\tilde{X}_n(s) - \tilde{X}_n(t)| \ge \varepsilon\} = 0,$$

where $T(\varepsilon)$ is the subset of t, s satisfying $0 \le s,t \le 1/\varepsilon$, $0 \le t \le s \le t+\varepsilon$.

Note that the filtration $\{\mathcal{F}_n(t) : t \ge 0\}$ is always right-continuous (in this case, not necessarily completed). It is customary to identify a cad-lag process X_n defined on the probability spaces $(\Omega_n,\mathcal{F}_n,P_n)$ with its probability law $\tilde{P}_n$ on $D([0,\infty),\mathbb{R}^d)$. Elements in the canonical space $D([0,\infty),\mathbb{R}^d)$ are denoted by ω and the canonical process $\omega \mapsto \omega(t)$, which is interpreted as the projection from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}^d$ or as the identity mapping from $D([0,\infty),\mathbb{R}^d)$ into itself is denoted by $x : (t,\omega) \mapsto \omega(t)$ or $x_t = x_t(\omega) = \omega(t)$ or $x(t) = x(t,\omega) = \omega(t)$ as long as no confusion may arrive. Recalling that $\omega_n \to \omega$ in the Skorokhod topology if and only if there exists a sequence λ_n of continuous and strictly increasing functions with $\lambda_n(0) = 0$ and $\lambda_n(\infty) = \infty$ such that

$$\sup_s |\lambda_n(s) - s| \to 0 \qquad \sup_{s\le T} |\omega_n(\lambda_n(s)) - \omega(s)| \to 0,$$

for every $T > 0$, we can prove that, besides the projection function x, the functions

$$\omega \mapsto \sup_{t\le T} |\omega(t)|, \qquad \omega \mapsto \sup_{t\le T} |\omega(t) - \omega(t-)|$$

are continuous from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}$ for any ω such that $\omega(T) = \omega(T-)$. Moreover,

$$\omega \mapsto \sum_{t\le T} h\big(\omega(t) - \omega(t-)\big),$$

with h a continuous function vanishing near zero, is also continuous. It is clear that if $\tilde{P}$ is quasi-left continuous then $\tilde{P}\{\omega(T) = \omega(T-)\} = 0$ for every $T > 0$, and the above functionals are $\tilde{P}$-almost surely continuous.

Also on $D([0,\infty),\mathbb{R}^d)$ there is another canonical (integer random measure) process $\omega \mapsto \nu(\mathrm{d}z, \mathrm{d}t, \omega)$, defined as

$$\nu(B,]a,b], \omega) := \sum_{a<t\le b} \mathbb{1}_{\{\omega(t)-\omega(t-)\in B\}}, \quad \forall B \in \mathcal{B}(\mathbb{R}^d_*),\ b > a \ge 0,$$

which is interpreted as the counting jumps measure. Once a probability P is given so that the canonical process x is a local martingale, then its continuous martingale part x^c, the predictable jump compensator measure ν^p and the local martingale measure $\tilde{\nu} := \nu - \nu^p$ are defined.

Let $h(t,x,v)$ be a real valued Borel measurable function which is bounded and locally uniform continuous in x. For every $T > 0$ consider the expression

$$\omega \mapsto \int_0^T h(t, x(t,\omega))\mathrm{d}t,$$

which is a continuous and bounded function from $D([0,\infty),\mathbb{R}^d)$ into $\mathbb{R}$. Then, with the notation of the previous Theorem 4.50 we have

$$\mathbb{E}_n\Big\{\int_0^T h(t, X_n(t))\mathrm{d}t\Big\} = \tilde{\mathbb{E}}_n\Big\{\int_0^T h(t, x(t))\mathrm{d}t\Big\} = \mathbb{E}\Big\{\int_0^T h(t, \tilde{X}_n(t))\mathrm{d}t\Big\},$$

Hence, either if $\tilde{P}_n$ is weakly convergent to $\tilde{P}$ or if $\tilde{X}_n$ converge in probability to $\tilde{X}$ we deduce that the above expression converges to

$$\tilde{\mathbb{E}}\Big\{\int_0^T h(t, x(t))\mathrm{d}t\Big\} = \mathbb{E}\Big\{\int_0^T h(t, \tilde{X}(t))\mathrm{d}t\Big\},$$

where $\mathbb{E}_n$ and $\mathbb{E}$ are the mathematical expectation in the probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and $(\Omega, \mathcal{F}, P)$, respectively, and $\tilde{\mathbb{E}}_n$ and $\tilde{\mathbb{E}}$ are the integral with respect to the probability laws $\tilde{P}_n$ and $\tilde{P}$, respectively. Moreover, the convergence holds true if we have a sequence $\{h_n(t,x)\}$ of measurable functions, which are equi-bounded in (t,x) and equi-locally uniform continuous in x, and pointwise convergent to some function $h(t,x)$.

There is a key class of discontinuous functions in $D([0,\infty),\mathbb{R}^d)$, namely, the so-called *counting* functions of the form

$$\sum_{t_i\le t}\mathbb{1}_{t_i\le s},\quad \forall t\ge 0,$$

for some strictly increasing sequence $0<t_i<t_{i+1}$, $t_i\to\infty$. Recall that a *point process* is a cad-lag process with counting functions as sample paths. The following result is sometime useful

Theorem 4.51. *Let $X_n=(X_n(t):t\ge 0)$, $n=1,2,\dots$ be a sequence of* increasing *cad-lag processes with values in $\mathbb{R}^d$ in the probability spaces $(\Omega_n,\mathcal{F}_n,P_n)$. Suppose that the distributions of $\big(X_n(t_1),\dots,X_n(t_m)\big)$ in $(\mathbb{R}^d)^m$ converges to $\big(X(t_1),\dots,X(t_m)\big)$ for every $t_1,\dots,t_m$ in some dense set of $[0,\infty)$, where $X=(X(t):t\ge 0)$ is an increasing cad-lag process with values in $\mathbb{R}^d$. If either X is continuous or all X_n and X are point processes then the law of X_n converges weakly to the probability law of X in the canonical space $D([0,\infty),\mathbb{R}^d)$.* □

The reader is referred to Proposition VI.3.26 and Theorems VI.3.37, VI.4.5 in the book Jacod and Shiryaev [117, Chapter VI, pp. 312–322].

Again, with the notation of Theorem 4.50, if the canonical process x is a local martingale relative to $\tilde P_n$ then its continuous part x_n^c and its local martingale measure $\tilde\nu_n$ are defined and the expressions

$$\omega\mapsto\int_0^T h(t,x(t))\mathrm{d}x_n^c(t)\quad\text{and}\quad\omega\mapsto\int_{]0,T]\times\mathbb{R}^d_*}\tilde h(t,x(t),z)\tilde\nu_n(\mathrm{d}z,\mathrm{d}t)$$

are $\tilde P_n$-almost surely continuous, as long as $\tilde h(t,x,z)$ is locally uniform continuous in x and uniformly integrable in z with respect to ν_n^p, the compensator of ν under $\tilde P_n$. However, to pass to the limit we will need the $\tilde P$-almost surely continuity. If X_n is a specific Lévy process then its characteristic function (or Fourier transform) is determined by the drift vector b, the covariance matrix a and the Lévy measure (or jump intensity) π (all independent of n or conveniently convergent as n tends to infinite). Hence, $x_n^c=X_n^c$ is its continuous local martingale part with predictable quadratic variation at and $\tilde\nu_n$ its local martingale measure with predictable jump compensator (Lévy measure) π. The limiting probability law $\tilde P$ has the same properties, so that the mathematical expectations

$$\tilde{\mathbb{E}}_n\Big\{\int_0^T h(t,x(t))\mathrm{d}x_n^c(t)\Big\}\quad\text{and}\quad\tilde{\mathbb{E}}_n\Big\{\int_{]0,T]\times\mathbb{R}^d_*}\tilde h(t,x(t),z)\tilde\nu_n(\mathrm{d}z,\mathrm{d}t)\Big\}$$

converge to

$$\tilde{\mathbb{E}}\Big\{\int_0^T h(t,x(t))\mathrm{d}x^c(t)\Big\}\quad\text{and}\quad\tilde{\mathbb{E}}\Big\{\int_{]0,T]\times\mathbb{R}^d_*}\tilde h(t,x(t),z)\tilde\nu(\mathrm{d}z,\mathrm{d}t)\Big\}.$$

Moreover, the processes $\tilde X_n$ and $\tilde X$ on the probability space $(\Omega,\mathcal{F},P)$ are Lévy processes with the same characteristic function. Since the distributions of $\tilde X_n$

and $\tilde X$ coincide with those of the canonical process x under $\tilde P_n$ and $\tilde P$, respectively, and the stochastic integrals are limits (in L^2) of finite sums, we deduce that the previous stochastic integrals can be considered in the probability space $(\Omega, \mathcal{F}, P)$. Thus

$$\mathbb{E}\Big\{ \int_0^T h(t, \tilde X_n(t)) \mathrm{d}\tilde X_n^c(t) \Big\} \quad \text{and} \quad \mathbb{E}\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde h(t, \tilde X_n(t), z) \tilde\nu_{\tilde X_n}(\mathrm{d}z, \mathrm{d}t) \Big\}$$

converge to

$$\mathbb{E}\Big\{ \int_0^T h(t, \tilde X(t)) \mathrm{d}\tilde X^c(t) \Big\} \quad \text{and} \quad \mathbb{E}\Big\{ \int_{]0,T]\times\mathbb{R}_*^d} \tilde h(t, \tilde X(t), z) \tilde\nu_{\tilde X}(\mathrm{d}z, \mathrm{d}t) \Big\},$$

under the previous conditions, i.e., $\tilde X_n$ is a Lévy process with $\tilde P_n$ (its probability law) that converges weakly to $\tilde P$, the probability law of $\tilde X$. More delicate arguments apply if X_n are local martingale with characteristics determined as continuous predictable functionals on the paths of X_n, see Jacod and Shiryaev [117, Chapter VII, pp. 348–387].

However, because the processes $\tilde X_n$ converge in probability to $\tilde X$, we can establish the above convergence independently. We rephrase the result as follows:

Theorem 4.52. *Let f_n, g_n and w_n, ν_n, $n = 1, 2, \dots$ be sequences of real-valued predictable processes in $[0,\infty)$ and $[0,\infty)\times\mathbb{R}_*^m$, d-dimensional Wiener processes and Poisson measures with Lévy measure π on $\mathbb{R}_*^m$, all defined in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \ge 0)$. Suppose that for some processes f, g, w and ν we have*

$$\int_0^T |f_n(t) - f(t)|^2 \mathrm{d}t \to 0 \quad \text{and} \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} |g_n(z,t) - g(z,t)|^2 \pi(\mathrm{d}z) \to 0$$

and

$$w_n(t) \to w(t), \qquad \nu_n(K\times]0,t]) \to \nu(K\times]0,t]),$$

in probability, for every t in $[0,T]$ and any compact subset K of $\mathbb{R}_*^m$, where it is implicitly assumed that

$$\int_0^T \big[|f_n(t)|^2 + |f(t)|^2\big] \mathrm{d}t < \infty,$$
$$\int_0^T \mathrm{d}t \int_{\mathbb{R}_*^m} \big[|g_n(z,t)|^2 + |g(z,t)|^2\big] \pi(\mathrm{d}z) < \infty,$$

almost surely. Then the stochastic integrals

$$\int_0^T f_n(t) \mathrm{d}w_n(t) \to \int_0^T f(t) \mathrm{d}w(t),$$
$$\int_{\mathbb{R}_*^m\times]0,T]} g_n(z,t) \tilde\nu_n(\mathrm{d}z, \mathrm{d}t) \to \int_{\mathbb{R}_*^m\times]0,T]} g(z,t) \tilde\nu(\mathrm{d}z, \mathrm{d}t)$$

in probability, where $\tilde\nu_n := \nu_n - \pi\mathrm{d}t$ and $\tilde\nu := \nu - \pi\mathrm{d}t$ are the Poisson (local) martingale measures associated with Poisson measures ν_n and ν.

Proof. We follows the arguments in Skorokhod [229, Section 2.3, pp. 29–34]. First, recall that elementary predictable processes have the form $h(t,\omega) = h_{i-1}(\omega)$ if $t_{i-1} < t \le t_i$ with some $i = 1,\dots,n$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers and h_{i-1} is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $h(t,\omega) = 0$ otherwise, or $h(z,t,\omega) = h_{i-1,j}(\omega)$ if $t_{i-1} < t \le t_i$ and z belongs to K_j with some $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 = t_0 < t_1 < \cdots < t_n$ are real numbers, K_j are disjoint compact subsets of $\mathbb{R}^m_*$ and $h_{i-1,j}$ is a $\mathcal{F}(t_{i-1})$ measurable bounded random variable for any i, and $h(t,\omega) = 0$ otherwise. Then, we find sequences of elementary predictable processes $f_{n,k}$, $g_{n,k}$, f_k and g_k, such that

$$\int_0^T |f_{n,k}(t) - f_n(t)|^2 \mathrm{d}t \to 0, \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}^m_*} |g_{n,k}(z,t) - g_n(z,t)|^2 \pi(\mathrm{d}z) \to 0$$
$$\int_0^T |f_k(t) - f(t)|^2 \mathrm{d}t \to 0 \quad \text{and} \quad \int_0^T \mathrm{d}t \int_{\mathbb{R}^m_*} |g_k(z,t) - g(z,t)|^2 \pi(\mathrm{d}z) \to 0$$

in probability as $k \to \infty$, for every n. It is clear that

$$\int_0^T f_{n,k}(t)\mathrm{d}w_n(t) \to \int_0^T f_k(t)\mathrm{d}w(t),$$
$$\int_{\mathbb{R}^m_* \times]0,T]} g_{n,k}(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t) \to \int_{\mathbb{R}^m_* \times]0,T]} g_k(z,t)\tilde{\nu}(\mathrm{d}z,\mathrm{d}t)$$

in probability for each k. Now, based on the inequalities

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_0^t h(s)\mathrm{d}w(s)\Big| \ge \varepsilon\Big\} \le \frac{\delta}{\varepsilon^2} + P\Big\{ \int_0^T |h(s)|^2 \mathrm{d}s \ge \delta\Big\},$$

and

$$P\Big\{ \sup_{0\le t\le T} \Big| \int_{\mathbb{R}^m_* \times (0,t]} h(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s)\Big| \ge \varepsilon\Big\} \le \frac{\delta}{\varepsilon^2} +$$
$$+ P\Big\{ \int_0^T \mathrm{d}s \int_{\mathbb{R}^m_*} |h(z,s)|^2 \pi(\mathrm{d}z) \ge \delta\Big\},$$

valid for every positive constant T, δ and ε, we deduce that

$$\int_0^T f_{n,k}(t)\mathrm{d}w_n(t) \to \int_0^T f_n(t)\mathrm{d}w_n(t),$$
$$\int_{\mathbb{R}^m_* \times]0,T]} g_{n,k}(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t) \to \int_{\mathbb{R}^m_* \times]0,T]} g_n(z,t)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t)$$

in probability as $k \to \infty$, uniformly in n, which complete the proof. □

Notice that in the context of the previous Theorem 4.52 the conditions $\nu_n \to \nu$ and $\tilde{\nu}_n \to \tilde{\nu}$ are equivalents. On the other hand, if $w_n(t)$ and $\nu_n(K\times]0,t])$ converge in probability uniformly for t in $[0,T]$ then the same is true for the stochastic integrals.

Let w_n and ν_n, $n = 1,2,\dots$ be sequences of d-dimensional (standard) Wiener processes and Poisson measures with Lévy measure π on $\mathbb{R}^m_*$, all defined in a

filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_n(t) : t \geq 0)$, where $\mathcal{F}_n(t)$ is the σ-algebra generated by $\{x_n(s) : s \leq t\}$, where x_n is a cad-lag process for each n. It is clear that w_n and $\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}t) := \nu(\mathrm{d}z, \mathrm{d}t) - \pi(\mathrm{d}z)\,\mathrm{d}t$ are martingale and martingale measures relative to $(\mathcal{F}_n(t) : t \geq 0)$ and $(\mathcal{F}_n(t+) : t \geq 0)$. If $w_n(t)$, $\nu_n(K,]0, t])$ and $x_n(t)$ converge in probability to $w(t)$, $\nu(K,]0, t])$ and $x(t)$, for every $t \geq 0$ and any compact subset K of $\mathbb{R}^m_*$, then w and ν are too, a d-dimensional (standard) Wiener process and a Poisson measure with Lévy measure π on $\mathbb{R}^m_*$, but with respect to the limiting filtration either $(\mathcal{F}(t) : t \geq 0)$ or $(\mathcal{F}(t+) : t \geq 0)$, where $\mathcal{F}(t)$ is the σ-algebra generated by $\{x(s) : s \leq t\}$. The above remark can be generalized with a random change of time, i.e., if $\ell(t)$ is a continuous nondecreasing (adapted) process and such that $\ell(0) = 0$ and $\ell(t)$ is a stopping time relative to each filtration $(\mathcal{F}_n(t) : t \geq 0)$ the we can change t into $\ell(t)$ for each process w_n, ν_n and x_n. This means that if w_n a square integrable martingale with ℓ as its predictable quadratic covariation (just one dimension to simplify notation) and ν_n has a jump compensator ν_n^p given by

$$\nu_n^p(B,]a, b]) = \pi(B)\,[\ell(b) - \ell(a)], \quad \forall B \in \mathcal{B}(\mathbb{R}^m_*),\ b > a \geq 0,$$

relative to $(\mathcal{F}_n(t) : t \geq 0)$, then the same is valid for the limiting process w, ν and x. Therefore, the previous Theorem 4.52 can be modified for this case, replacing $\mathrm{d}t$ with $\ell(t)$. However, if $\mathrm{d}\ell_n$ changes with n then the situation requires more details.

Recall the locally uniform and the Skorokhod's topologies given by the family of functions $\rho(\omega, \delta,]a, b])$ and $w(\omega, \delta,]a, b])$, which are defined for ω in the space of cad-lag functions $D([0, \infty), \mathbb{R}^d)$, by the expressions

$$\rho(\omega, \delta,]a, b]) := \sup\{|\omega(t) - \omega(s)| \ :\ a < s,\, t \leq b,\ |t - s| \leq \delta\},$$
$$w(\omega, \delta,]a, b]) := \inf_{\{t_i\}} \sup_i \sup\{|\omega(t) - \omega(s)| \ :\ t_{i-1} \leq s < t < t_i\},$$

where $\{t_i\}$ ranges over all partitions of the form $a = t_0 < t_1 < \cdots < t_{n-1} < b \leq t_n$, with $t_i - t_{i-1} \geq \delta$ and $n \geq 1$. Both family of functions $\rho(\omega, \delta,]a, b])$ and $w(\omega, \delta,]a, b])$, define the same topology on the space of continuous functions $C([0, \infty), \mathbb{R}^d)$. It is clear ρ and w is the oscillation (or variation) for continuous and cad-lag functions.

If $\ell(t)$ is a nondecreasing element in $C([0, \infty), \mathbb{R}^d)$ and $f(t)$ is another element in $D([0, \infty), \mathbb{R}^d)$, then the Riemann-Stieltjes integral

$$\int_0^T f(t)\mathrm{d}\ell(t), \quad \forall T \geq 0,$$

is defined as the limit of the Riemann sums

$$R(f, \ell, \pi, [0, T[) := \sum_{i=1}^n f(t_i^*)[\ell(t_i) - \ell(t_{i-1})],$$
$$\varpi = \{t_i,\ t_i^*\}, \quad 0 = t_0 < t_1 < \cdots < t_n = T, \quad t_{i-1} \leq t_i^* < t_i,$$

when the mesh of the partition $|\varpi| := \max_i\{t_i - t_{i-1}\}$ vanishes. Moreover, if we define

$$f^{\varpi}(t) := f(t_i^*) \quad \text{if} \quad t_{i-1} < t \le t_i,$$

then

$$\lim_{|\varpi|\to 0} \int_0^T |f(t) - f^{\varpi}(t)|\, \mathrm{d}\ell(t) = 0,$$

for every $T > 0$. Indeed, since $f(t) - f^{\varpi}(t) = f(t) - f(t_i^*)$ for some t in $]t_{i-1}, t_i]$ and t_i^* in $[t_{i-1}, t_i[$, we see that for every $\varepsilon > 0$, $i \ge 1$ and any t in $]t_{i-1}, t_i]$,

$$\begin{aligned} |f(t) - f^{\varpi}(t)| &\le w(f, |\varpi|,]0, T]) + \varepsilon + \\ &+ \max_{t_{i+1} < s \le t_i} \mathbb{1}_{\{|f(s) - f(s-)| > \varepsilon\}} |f(s) - f(s-)|, \end{aligned}$$

i.e., the variation (or oscillation) is bounded by its continuous variation, plus ε, plus the maximum jumps bigger than ε. Hence

$$\begin{aligned} \int_0^T |f(t) - f^{\varpi}(t)|\, \mathrm{d}\ell(t) &= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} |f(t) - f_{\varpi}(t)|\, \mathrm{d}\ell(t) \le \\ &\le \sum_{i=1}^n \Big(\sup_{t_{i-1} < t \le t_i} |f(t) - f_{\varpi}(t)| \Big) [\ell(t_i) - \ell(t_{i-1})] \end{aligned}$$

which yields

$$\begin{aligned} \int_0^T |f(t) - f^{\varpi}(t)|\, \mathrm{d}\ell(t) \le \big[w(f, |\varpi|,]0, T]) + \varepsilon \big] \big[\ell(T) - \ell(0) \big] + \\ + \rho(\ell, |\varpi|,]0, T]) \sum_{0 < s \le T} \mathbb{1}_{\{|f(s) - f(s-)| > \varepsilon\}} |f(s) - f(s-)|. \end{aligned}$$

From the definition of the cad-lag modulus of continuity w we have

$$\sum_{a < t \le b} \mathbb{1}_{\{|\omega(t) - \omega(t-)| \ge w(\omega, \delta,]a, b])\}} \le \frac{b - a}{\delta},$$

for every ω, $\delta > 0$, and $b > a \ge 0$. Therefore, for $\varepsilon = w(f, \delta,]0, T])$ we obtain

$$\left\{ \begin{aligned} \int_0^T |f(t) - f^{\varpi}(t)|\, \mathrm{d}\ell(t) &\le \frac{T}{\delta}\, \rho(\ell, |\varpi|,]0, T]) + \\ &+ \big[w(f, |\varpi|,]0, T]) + w(f, \delta,]0, T]) \big] \big[\ell(T) - \ell(0) \big]. \end{aligned} \right. \tag{4.121}$$

Actually, this estimate implies the following result.

Lemma 4.53. *Let $\{f_n\}$ be a family of cad-lag processes and $\{\ell_n\}$ be another family of continuous and nondecreasing processes, defined in a probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n*

$$P_n\{w(f_n, \delta,]0, 1/\varepsilon]) \ge \varepsilon\} + P_n\{ \sup_{0 \le t \le 1/\varepsilon} |f_n(t)| \ge 1/\delta\} \le \varepsilon$$

and

$$P_n\{|\ell_n(1/\varepsilon) - \ell_n(0)| > 1/\delta\} + P_n\{\rho(\ell_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon\} \leq \varepsilon.$$

Now, for any partition $\varpi = \{t_i,\ t_i^*\}$, $0 = t_0 < t_1 < \cdots < t_n = T$, $t_{i-1} \leq t_i^* < t_i$, *define* $f_n^\varpi(t) := f_n(t_i^*)$ *if* $t_{i-1} < t \leq t_i$ *as above. Then for every* $\varepsilon > 0$ *there is* $\delta > 0$ *such that for every* n

$$P_n\Big\{ \int_0^T |f_n(t) - f_n^\varpi(t)|\, \mathrm{d}\ell_n(t) \geq \varepsilon \Big\} \leq \varepsilon,$$

for every ϖ *with* $|\varpi| \leq \delta$.

Proof. Notice that the assumptions means that $\{f_n\}$ is tight (or pre-compact) in the space $D([0,\infty), \mathbb{R}^d)$ and $\{\ell_n\}$ is tight in $C([0,\infty), \mathbb{R}^d)$. The conclusion is the uniform convergence in probability of the integral processes, which is a direct consequence of the a priori estimate (4.121). □

If we are looking at processes $g_n(z,t)$ instead of just $f_n(t)$, with $t \geq 0$ and z in $\mathbb{R}_*^d$, we may consider g_n as having values in the function space $L^2_{\pi_n}(\mathbb{R}_*^m)$, i.e., we use the following definition of the cad-lag modulo

$$w^{\pi_n}(x, \delta,]a, b]) :=$$
$$:= \inf_{\{t_i\}} \sup_i \sup_{t_{i-1} \leq s < t < t_i} \Big\{ \Big(\int_{\mathbb{R}_*^m} |x(z,t) - x(z,s)|^2 \pi_n(\mathrm{d}z) \Big)^{1/2} \Big\},$$

where $\{t_i\}$ ranges over all partitions of the form $a = t_0 < t_1 < \cdots < t_{n-1} < b \leq t_n$, with $t_i - t_{i-1} \geq \delta$ and $n \geq 1$. Estimate (4.121) becomes

$$\begin{cases} \displaystyle\int_0^T \Big(\int_{R_*^d} |g(z,t) - g^\varpi(z,t)|^2\, \pi_n(\mathrm{d}z) \Big)^{1/2} \mathrm{d}\ell(t) \leq \\ \qquad \leq \dfrac{T}{\delta}\, \rho(\ell, |\varpi|,]0, T]) + \big[w^{\pi_n}(g, |\varpi|,]0, T]) + \\ \qquad\qquad + w^{\pi_n}(g, \delta,]0, T]) \big] \big[\ell(T) - \ell(0) \big], \end{cases} \tag{4.122}$$

and the previous Lemma 4.53 remain valid under the assumption that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for every n

$$P_n\{w^{\pi_n}(g_n, \delta,]0, 1/\varepsilon]) \geq \varepsilon\} + P_n\{ \sup_{0 \leq t \leq 1/\varepsilon} |g_n(\cdot, t)|_{\pi_n} \geq 1/\delta\} \leq \varepsilon,$$

where

$$|g_n(\cdot, t)|_{\pi_n} := \Big(\int_{\mathbb{R}_*^m} |g_n(z,t)|^2 \pi_n(\mathrm{d}z) \Big)^{1/2}.$$

The a priori estimate obtained is written as for every $\varepsilon > 0$ there is $\delta > 0$ such that for every n

$$P_n\Big\{ \int_0^T \mathrm{d}\ell_n(t) \int_{\mathbb{R}_*^m} |g_n(z,t) - g_n^\varpi(z,t)|^2\, \pi_n(\mathrm{d}z) \geq \varepsilon \Big\} \leq \varepsilon, \tag{4.123}$$

for every ϖ with $|\varpi| \le \delta$.

Now based on above Lemma 4.53 we are able to generalize Theorem 4.52 as follows

Theorem 4.54. *Let ℓ_n, w_n, ν_n, and x_n, $n = 1, 2, \dots$ be sequences of processes defined in a probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_n(t) : t \ge 0)$ be the filtration generated by x_n. Assume that (1) ℓ_n are continuous nondecreasing adapted processes, (2) w_n are a d-dimensional square integrable martingales with predictable quadratic covariation $\langle w_{n,i}, w_{n,j}\rangle = \ell_n$ if $i = j$ and $\langle w_{n,i}, w_{n,j}\rangle = 0$ if $i \ne j$. (3) ν_n are integer measures with jump compensator $\nu_n^p(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z)\,\mathrm{d}\ell_n(t)$, where π is a given Lévy measure in $\mathbb{R}_*^m$. Suppose that ℓ_n converges to ℓ, i.e., for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that*

$$P\{\sup_{0\le t\le 1/\varepsilon} |\ell_n(t) - \ell(t)| \ge \varepsilon\} \le \varepsilon, \quad \forall n \ge N(\varepsilon),$$

$\ell_n(0) = 0$ and that $x_n(t) \to x(t)$, $w_n(t) \to w(t)$ and $\nu_n(K\times]0,t]) \to \nu(K\times]0,t])$ in probability, for every $t \ge 0$ and any compact subset K of $\mathbb{R}_*^m$. Then (a) w is also a square integrable martingale with predictable quadratic covariation $\langle w_i, w_j\rangle = \ell$ if $i = j$ and $\langle w_i, w_j\rangle = 0$ if $i \ne j$, (b) ν is also an integer measure with jump compensator $\nu^p(\mathrm{d}z, \mathrm{d}t) = \pi(\mathrm{d}z)\,\mathrm{d}\ell(t)$, both relative to the limiting filtration $(\mathcal{F}(t) : t \ge 0)$ generated by x. Furthermore, if f_n and g_n are cad-lag adapted processes pointwise (on a dense set of time) convergent to f and g in probability and for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ satisfying

$$P\{w(f_n, \delta,]0, 1/\varepsilon]) + w^\pi(g_n, \delta,]0, 1/\varepsilon]) \ge \varepsilon\} \le \varepsilon, \quad \forall n \ge 1,$$

the limiting processes f and g are certainly cad-lag, and there exist sequences of partitions $\{\varpi_k = \varpi_k^f : k \ge 1\}$ and $\{\varpi_k = \varpi_k^g : k \ge 1\}$ with mesh $|\varpi_k^f| \to 0$ and $|\varpi_k^g| \to 0$ such that in probability we have

$$\int_0^T f_n^{\varpi_k}(t)\,\mathrm{d}\ell_n(t) \to \int_0^T f^{\varpi_k}(t)\,\mathrm{d}\ell(t),$$
$$\int_0^T \mathrm{d}\ell_n(t) \int_{\mathbb{R}_*^d} g_n^{\varpi_k}(z,t)\,\pi(\mathrm{d}z) \to \int_0^T \mathrm{d}\ell(t) \int_{\mathbb{R}_*^d} g^{\varpi_k}(z,t)\,\pi(\mathrm{d}z),$$

for every k and T, then the Riemann-Stieltjes integrals converge in probability, i.e.,

$$\lim_n P\Big\{\Big|\int_0^t f_n(s)\,\mathrm{d}\ell_n(s) - \int_0^t f(s)\,\mathrm{d}\ell(s)\Big| \ge \varepsilon\Big\} = 0,$$

and

$$\lim_n P\Big\{\Big|\int_0^t \mathrm{d}\ell_n(s)\int_{\mathbb{R}_*^d} g_n(z,s)\pi(\mathrm{d}z) - \int_0^t \mathrm{d}\ell(s)\int_{\mathbb{R}_*^d} g(z,s)\pi(\mathrm{d}z)\Big| \ge \varepsilon\Big\} = 0,$$

for every $t, \varepsilon > 0$. Also the stochastic integrals

$$M_n(t) := \int_{]0,t]} f_n(s)\,\mathrm{d}w_n(s),$$
$$J_n(t) := \int_{\mathbb{R}_*^m\times]0,t]} g_n(z,s)\,\tilde{\nu}_n(\mathrm{d}z, \mathrm{d}s),$$

converge also in probability to

$$M(t) := \int_{]0,t]} f(s)\,\mathrm{d}w(s),$$
$$J(t) := \int_{\mathbb{R}^m_* \times]0,t]} g(z,s)\,\tilde{\nu}(\mathrm{d}z,\mathrm{d}s),$$

for every $t > 0$, *where* $\tilde{\nu}_n := \nu_n - \pi \mathrm{d}\ell_n$ *and* $\tilde{\nu} := \nu - \pi \mathrm{d}\ell$ *are the (local) martingale measures associated with integer measures* ν_n *and* ν. □

It is also clear that under the conditions of the above Theorem and the assumption that f_n and g_n converge to f and g in probability as random variable with values in the Polish space $D([0,\infty),\mathbb{R}^d)$ and $D([0,\infty),L^2_\pi(\mathbb{R}^m_*))$, the stochastic integrals M_n and J_n converge to M and J in probability as random variable with values in the Polish space $D([0,\infty),\mathbb{R}^d)$ and $D([0,\infty),\mathbb{R})$. Moreover, if M_n are continuous then we can replace the cad-lag space $D([0,\infty),\mathbb{R}^d)$ with $C([0,\infty),\mathbb{R}^d)$. In any case, the Riemann-Stieltjes integral processes converge as random variables with values in the Polish space $C([0,\infty),\mathbb{R}),$, i.e., for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that

$$P\Big\{ \sup_{0\le t\le 1/\varepsilon} \Big| \int_0^t f_n(s)\,\mathrm{d}\ell_n(s) - \int_0^t f(s)\,\mathrm{d}\ell(s) \Big| \ge \varepsilon \Big\} \le \varepsilon,$$

for every $n \ge N(\varepsilon)$, see estimate (4.121).

It is possible to consider the Lévy measure π in Theorem 4.54 depending on n, i.e. π_n, provided some uniform integrability at the origin is imposed, e.g.,

$$\lim_{\varepsilon\to 0} \sup_n \int_{|z|\ge\varepsilon} |z|^2\,\pi_n(\mathrm{d}z) = 0,$$

or replacing the function $|z|^2$ with either $|z|^2 \wedge |z|$ or $|z|^2 \wedge 1$, depending on the integrability condition imposed on each π_n.

4.4.2 Other Convergence of Probabilities

Mainly, we discuss here Jakubowski convergence of probability measures. The canonical spaces $C([0,\infty),\mathbb{R}^d)$ and $\mathbb{D} = D([0,\infty),\mathbb{R}^d)$, of continuous and cad-lag functions, are Polish (complete separable metric) spaces, with the local uniformly convergence and the Skorokhod topology (usually referred to as the J_1-topology. Clearly, the addition and multiplication are continuous operation on $C([0,\infty),\mathbb{R}^d)$, but not on $\mathbb{D}$, i.e., $C([0,\infty),\mathbb{R}^d)$ is a topological vector space but $\mathbb{D}$ is not. Moreover, the topology in $D([0,\infty),\mathbb{R}^d)$ is strictly stronger that the product topology in $D([0,\infty),\mathbb{R}^{d_1}) \times D([0,\infty),\mathbb{R}^{d_2})$, $d = d_1 + d_2$.

Now, the spaces of probability measures on $C([0,\infty),\mathbb{R}^d)$ and $\mathbb{D}$, denoted respectively $\wp(C([0,\infty),\mathbb{R}^d))$ and $\wp(\mathbb{D})$, are Polish spaces, with the weak convergence topology, i.e. $\mu_n \to \nu$ if $\mu_n(f) \to \mu(f)$ for every bounded continuous function f from $C([0,\infty),\mathbb{R}^d)$ (or $\mathbb{D}$) into $\mathbb{R}$; moreover any probability measure is tight. The reader is referred to the book Jacod and Shiryaev [117, Chapter VI, pp. 288–347] for a comprehensive discussion.

The *operation stochastic integral* can be regarded as a functional on either $C([0,\infty),\mathbb{R}^d)$ or $\mathbb{D}$, i.e., given a probability measure in $\mathbb{D}$ with a certain number of properties (relative to some integrands and integrators), the law of the stochastic integral process defines another probability measure. Loosely speaking, if we have a sequence of integrands and integrators then we actually have a sequence of probability measures on $C([0,\infty),\mathbb{R}^d)$ or $\mathbb{D}$. Specifically, we are interested in the functional defined by an stochastic differential equation. When dealing with cad-lag processes of (local) bounded variation, the Skorokhod topology seems too strong for some cases of reflected stochastic differential equation, and a weak topology is necessary. One of the key difficulties is that we exit the framework of Polish spaces and we need to recall or review certain points of general topology.

Sequential Convergence

First, it is a necessary some basic terminology on sequential convergence. In a given topological space $(\mathbb{X},\tau)$ the closure of any subset of $\mathbb{X}$ could be defined as a map $A\to\bar{A}\colon 2^{\mathbb{X}}\to 2^{\mathbb{X}}$ with the following properties: (a) $\bar{\emptyset}=\emptyset$, (b) $A\subset\bar{A}$, (c)$\overline{A\cup B}=\bar{A}\cup\bar{B}$ and (d) $\bar{\bar{A}}=\bar{A}$. This previous four properties are called the Kuratowski axioms.

Suppose now to have defined on a set $\mathbb{X}$ (without a topology) a map on the subsets of $\mathbb{X}$, say $\kappa\colon 2^{\mathbb{X}}\to 2^{\mathbb{X}}$, such that: (1) $\kappa(\emptyset)=\emptyset$, (2) $A\subset\kappa(A)$ and (3) $\kappa(A\cup B)=\kappa(A)\cup\kappa(B)$. Then, we can endow $\mathbb{X}$ with a topology τ_κ by defining as "closed sets" those subsets F such that $F=\kappa(F)$. We can easily check that the properties (1), (2) and (3) imply that the family of the complements of "closed sets", just defined, is a topology. The closure operator with respect to this topology has the property $A\subset\kappa(A)\subset\bar{A}$. Thus, if for any subset A we have that $\kappa(A)=\bar{A}$, then we have also the property (4) $\kappa(\kappa(A))=\kappa(A)$. Hence, we can shows that if the map κ verifies (1), (2), (3) and (4) as above, then the above topology τ_κ is the unique topology such that $\bar{A}=\kappa(A)$, for any subset $A\subset\mathbb{X}$.

Convergent Sequences in a Given Topology

Now given a topological space $(\mathbb{X},\tau)$, the family of converging sequences $x_n\to x$ is determined. We can define the map $\kappa(A)=[A]_{\mathsf{seq}}$ as the set of all limits points of τ-converging sequences of points of A. It is easy to check that κ satisfies (1) $[\emptyset]_{\mathsf{seq}}=\emptyset$, (2) $A\subset[A]_{\mathsf{seq}}\subset\bar{A}^{\tau}$ and (3) $[A\cup B]_{\mathsf{seq}}=[A]_{\mathsf{seq}}\cup[B]_{\mathsf{seq}}$, but in general the point (4) is not true, i.e., we may have $[A]_{\mathsf{seq}}\subsetneq[[A]_{\mathsf{seq}}]_{\mathsf{seq}}$.

Thus we can introduce, as before, a topology $\tau_{\mathsf{seq}}(=\tau_\kappa)$, by defining the new closed sets as $F=[F]_{\mathsf{seq}}$ and we have that $\tau\subset\tau_{\mathsf{seq}}$. Obviously that the two topologies have the same converging sequences, moreover, there is also the weakest topology τ' with the same converging sequences of τ, and

$$A\subset[A]_{\mathsf{seq}}\subset\bar{A}^{\tau_{\mathsf{seq}}}\subset\bar{A}^{\tau}\subset\bar{A}^{\tau'}.$$

Hence, a topological space $(\mathbb{X}, \tau)$ is called *sequential space* if $\tau = \tau_{\mathsf{seq}}$, with $\kappa(A) = [A]_{\mathsf{seq}}$. Since it could happen that $[[A]_{\mathsf{seq}}]_{\mathsf{seq}} \neq [A]_{\mathsf{seq}}$, a topological space $(\mathbb{X}, \tau)$ is called *Fréchet-Urysohn space* if it a sequential space such that $[[A]_{\mathsf{seq}}]_{\mathsf{seq}} = [A]_{\mathsf{seq}}$ or equivalently $\bar{A}^{\tau_{\mathsf{seq}}} = \bar{A}^{\tau}$. Note that any metric space $\mathbb{X}$ is a Fréchet-Urysohn space.

Topology After Convergent Sequences

We can define the convergence of a sequence without introducing necessarily a topology, in other words we can define a convergence of a sequence not in terms of a given topology, as, for example, in the usual weak convergence of probability measures on topological spaces. Now, if we assume that a notion of convergence of sequences on a set (arbitrary) $\mathbb{X}$ is given, then to find a topology τ on the space $\mathbb{X}$ such that all the converging sequences converge also in this topology, we need to impose the the following properties to the family of converging sequences:

(i) The uniqueness of the limit holds.

(ii) For every $x \in \mathbb{X}$, the constant sequence $\{x, x, x, \ldots\}$ is convergent to x.

(iii) Given a convergent sequence $\{x_1, x_2, x_3, \ldots\}$ $(x_n \to x)$, then every subsequence is convergent to the same limit x.

These hypotheses imply that the sequential closure map $\kappa(A) = [A]_{\mathsf{seq}}$ (as the set of all limits points of converging sequences of points of A) verifies the properties (1), (2) and (3) above. Hence we can introduce the topology τ_{seq}, and all converging sequences are also convergent in this topology. But in general, there are more τ_{seq}-converging sequences than converging sequences (in the initial sense).

Since a sequence $\{x_n\}$ is τ_{seq}-converging to x_0 if and only if from any subsequence it is possible to extract a further subsequence convergent to the same x_0 (in the initial sense). This motivates the following further property, after the properties (i), (ii) and (iii),

(iv) a sequence $\{x_n\}$ is converge to x_0 if from any subsequence of $\{x_n\}$ it is possible to extract a further subsequence convergent to the same x_0.

Therefore, if (iv) holds then all converging sequences in the topology τ_{seq} are just the given converging sequences.

If in a set $\mathbb{X}$ we have defined (initially) the meaning of converging sequences satisfying (i), (ii) and (iii), then we say that we have space of type $\mathcal{L}$ or *sequential convergence of type* $\mathcal{L}$. Moreover, if also the property (iv) is satisfied then we called it a space of type $\mathcal{L}^*$ or *sequential convergence of type* $\mathcal{L}^*$.

Now, starting from a space $\mathbb{X}$ with sequential convergence of type $\mathcal{L}$, we can endow $\mathbb{X}$ of the corresponding τ_{seq} topology. Next, if we take *all* the convergent sequences in this τ_{seq} topology, which is called the $*$-convergence (relative to the initial convergence), then we have a sequential convergence of type $\mathcal{L}^*$. Clearly, if a sequence converges in the initial convergence then it also converges in the $*$-convergence, but not necessarily the converse. On the other hand, if we start

from a space $\mathbb{X}$ of type $\mathcal{L}^*$ and we endow $\mathbb{X}$ with the corresponding τ_{seq} topology, then this time *all* the convergent sequences in this τ_{seq} topology are exactly the same convergent sequences given initially, i.e., the initial convergence and the $*$-convergence are the same.

The simplest example is perhaps the space of real-valued Borel measurable functions $B([0,\infty),\mathbb{R})$ with the pointwise convergence (which yields a sequential topology), where all four properties are satisfied, i.e., a sequential convergence of type $\mathcal{L}^*$. However, it is clear that $B([0,\infty),\mathbb{R})$ is not a Fréchet-Urysohn space. Indeed, if continuous functions are called Baire function of class 0 or of at most class 0, then pointwise limit of Baire functions of at most class n are called Baire function of at most class $(n+1)$, and Baire functions of class $(n+1)$ those that are of at most class $(n+1)$ without being of at most class n. Clearly, all Baire functions are Borel measurable functions. Thus denoting by $[\cdot]_{\mathsf{seq}}$ the sequential closure, the Baire functions of at most class 1 is the closure $[C([0,\infty),\mathbb{R})]_{\mathsf{seq}}$ while the Baire functions of at most class 2 is the double closure $[[C([0,\infty),\mathbb{R})]_{\mathsf{seq}}]_{\mathsf{seq}}$. Their difference is the Baire functions of class 2, e.g., the Dirichlet function ($= 1$ for all rational and $= 0$ for all irrational) $\lim_{n\to\infty}\left(\lim_{k\to\infty}(\cos n!\,\pi x)^{2k}\right)$ is a Baire function of class 2. It is clear that similar remarks apply to the pointwise and bounded convergence. Actually, if T is an interval and X is a $\mathcal{L}^*$ space, so is $B(T,X)$ and $C(T,X)$. Clearly, the pointwise convergence makes $B(T,X)$ a Hausdorff topological spaces, which is neither a countable separated space nor a separable space.

Another interesting example is the space $L^0(\Omega,\mathcal{F},P)$ of the equivalence classes of real-valued random variables with the almost surely pointwise convergence. This space is of type $\mathcal{L}$, but is not of type $\mathcal{L}^*$. Moreover, the convergence (iv), i.e., the τ_{seq} convergence or $*$-convergence (due to the topology induced by the almost surely pointwise convergence) is actually the convergence in probability, i.e., in this case, $L^0(\Omega,\mathcal{F},P)$ with the $*$-convergence becomes a complete metric space.

Sequence of Probability Measures

The notions of *tightness* (or *boundedness in probability*), *regularity*, and of *weak convergence* (or *convergence in law*) of measures need that the underlying measure space be a topological space $\mathbb{X}$ with the corresponding Borel σ-algebra $\mathcal{B}(\mathbb{X})$.

Definition 4.55. Given a topological space $\mathbb{X}$ and its Borel σ-algebra $\mathcal{B}(\mathbb{X})$, a family of probability measures $P_i, i\in\mathcal{I}$ is *uniformly tight* if for any $\varepsilon>0$ there exists a compact K_ε such that, for every $i\in\mathcal{I}$, $P_i(K_\varepsilon)>1-\varepsilon$. □

Let us mention two key results on Borel (measures defined on the Borel σ-algebra) and Radon measures (measures finite on any compact set):

(a) Any probability measure P on a *metric space* $\mathbb{X}$ is *regular*, i.e., for any Borel set A and every $\varepsilon>0$ there exist a closed set F and a open set G such that $F\subset A\subset G$ and $P(G\smallsetminus F)<\varepsilon$.

(b) Any probability measure P on a *Polish space* (i.e., complete separable metric space) $\mathbb{X}$ is tight, i.e., for any $\varepsilon > 0$ there exists a compact K_ε such that $P(K_\varepsilon) > 1 - \varepsilon$.

In particular, any probability measure on a Polish space is regular and tight.

Weak convergence

Given a topological space $(\mathbb{X}, \tau_0)$ and its Borel σ-algebra $\mathcal{B}(\mathbb{X})$, we can consider the space of probability measures on $\mathbb{X}$, which is denoted by $\wp(\mathbb{X})$ and endowed with the weakest topology such that every linear functional of the form $\mu \to \mu(f)$ is continuous, when f is any bounded and continuous function on $\mathbb{X}$. This topological space is denoted by $\mathfrak{W}(\wp(\mathbb{X}))$ or simply $\mathfrak{W}$. Note that if $\mathbb{X}$ is a metric (or Polish) space then $\wp(\mathbb{X})$ results also a metric (or Polish) space with the Prohorov's distance.

Note that $\mu_n \xrightarrow{\mathfrak{W}} \mu$ implies that $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. But the converse may be false, i.e. we can have that $\mu_n(f) \to \mu(f)$ for every f bounded and continuous, but not converging in the $\mathfrak{W}$ topology.

Recall that usually we have the weak convergence defined by $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. This convergent yields a space of type $\mathcal{L}$, which is not necessarily $\mathfrak{W}$. Thus, it makes sense to introduce the sequential weak topology (the previous τ_{seq} topology) $\mathfrak{W}_{\mathsf{seq}}$, the weakest topology with respect to which we have $\mu_n(f) \to \mu(f)$ for every f bounded and continuous. Certainly, $\mathfrak{W} \subset \mathfrak{W}_{\mathsf{seq}}$. We have

Theorem 4.56. *The space of probability measures on a* Polish *space with the weak convergence is a space of type* $\mathcal{L}^*$ *with* $\mathfrak{W}_{\mathsf{seq}} = \mathfrak{W}$. □

On the other hand, starting with $(\mathbb{X}, \tau_0)$, let us suppose that there is another (weaker) topology τ_1 on $\mathbb{X}$, such that $\tau_1 \subset \tau_0$ and that the Borel σ-algebra generated by τ_1 is the same as that generated by τ_0. In such a case the space $\wp(\mathbb{X})$ is uniquely defined, with either τ_0 or τ_1. We have that $\mathfrak{W}(\tau_1) \subset \mathfrak{W}(\tau_0)$ and $\mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_0)$. If $(\mathbb{X}, \tau_0)$ is a Polish space then

$$\mathfrak{W}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}(\tau_0) = \mathfrak{W}_{\mathsf{seq}}(\tau_0).$$

In fact, $\mu_n \to \mu$ in $\mathfrak{W}_{\mathsf{seq}}(\tau_0)$ means that $\mu_n(f) \to \mu(f)$ for every bounded and τ_0-continuous function. Since a τ_1-continuous function is also τ_0-continuous, we deduce also that $\mu_n \to \mu$ in $\mathfrak{W}_{\mathsf{seq}}(\tau_1)$, i.e., $\mathfrak{W}_{\mathsf{seq}}(\tau_1) \subset \mathfrak{W}_{\mathsf{seq}}(\tau_0)$.

Prohorov Theorem

There are two implication, the direct and the converse:

(1) Given a metric space $\mathbb{X}$, a family of probability measures $\{P_i, i \in \mathcal{I}\}$ on $\mathbb{X}$, is uniformly tight if it is relatively compact with respect to weak convergence.

(2) Given a separable, complete metric space $\mathbb{X}$ (i.e. a Polish space), a family of probability measures $\{P_i, i \in \mathcal{I}\}$ on $\mathbb{X}$, is relatively compact with respect to weak convergence if it is uniformly tight.

For instance, see for example Dudley [62, Section 11.5, pp. 402–405].

Non Metric Case

Now, let $(\mathbb{X}, \tau)$ be a topological space countably separated, i.e., such that the following hypothesis holds:

> there exists a countable family $\{f_i : \mathbb{X} \to [-1, 1], i = 1, 2, \ldots\}$ of τ-continuous functions which separates the points of $\mathbb{X}$, that is for any two distinct points $x_1, x_2 \in \mathbb{X}$ there exists a function f_κ such that $f_\kappa(x_1) \neq f_\kappa(x_2)$.

Consider $[0, 1]$ is considered with the usual Borel σ-algebra and the standard Lebesgue measure (sometimes referred to as the universal probability space), see Theorem 4.49 to compare assumptions.

Theorem 4.57. *Let $\{\mu_n\}$ be a sequence of* tight *probability measures on a topological space $\mathbb{X}$, with also the previous hypothesis. Then there exist a subsequence $\{\mu_{n_k}\}$ and a sequence of random variables $X_k : [0, 1] \to \mathbb{X}$ and a further random variable $X : [0, 1] \to \mathbb{X}$, such that (1) the image measures of X_k are the μ_{n_k} and (2) $X_n(\theta) \to X(\theta)$, for any θ in $[0, 1]$.* □

Note that given any compact $K \subset \mathbb{X}$ the set

$$C_K = \{\theta \in [0, 1] : X_k(\theta) \to X(\theta)\} \cap \bigcap_{k=1}^{\infty} \{\theta \in [0, 1] : X_k(\theta) \in K\}$$

is Borel measurable in $[0, 1]$. Moreover, for any ε there exists a compact K_ε such that the Lebesgue measure of C_{K_ε} is greater or equal to $1 - \varepsilon$.

Star-convergence of Tight Probability on $\mathbb{X}$

Definition 4.58. Given a sequence of *tight* probability measures $\{\mu_n\}$ on $\mathbb{X}$, we say that $\mu_n \overset{*}{\Longrightarrow} \mu$ if from every subsequence $\{\mu_{n_k}\}$ there exist a further subsequence $\{\mu_{n_{k_i}}\}$ and a sequence of "random" variables $X_i : [0, 1] \to \mathbb{X}$, whose image measures are just the $\mu_{n_{k_i}}$'s and a further "random" variable $X : [0, 1] \to \mathbb{X}$, whose image measure is μ such that, for each $\theta \in [0, 1]$, $X_i(\theta) \to X(\theta)$ and for each $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset \mathbb{X}$ such that $\mathsf{Leb}\big(\bigcap_{i=1}^{\infty} \{\theta : X_i(\theta) \in K_\varepsilon\}\big) > 1 - \varepsilon$. □

This definition gives to the space of tight probability measures (denoted by $\wp^o(\mathbb{X}) \subset \wp(\mathbb{X})$) the structure of space of type $\mathcal{L}^*$. Hence we have the corresponding $\mathfrak{W}_{\mathsf{Jak}}$ sequential topology. Moreover, referring to $\wp^o(\mathbb{X})$ instead of the whole $\wp(\mathbb{X})$, this topology $\mathfrak{W}_{\mathsf{Jak}}$ is stronger than the sequential topology $\mathfrak{W}_{\mathsf{seq}}$, i.e. $\mathfrak{W}_{\mathsf{seq}} \subset \mathfrak{W}_{\mathsf{Jak}}$.

Theorem 4.59. *This $\mathfrak{W}_{\mathsf{Jak}}$ topology has the property that the family of relatively compact sets coincides with the family of relatively uniformly tight sets.* □

If $\mathbb{X}$ is a metric space then the weak topology and the $\mathfrak{W}_{\mathsf{Jak}}$ topology coincide.

4.4.3 Back to the Canonical Space

Consider the dual space of $C([0,T])$, $T>0$, which is the space of functions $v(\cdot)$ with bounded variation with the duality pairing

$$\langle \varphi, v\rangle_T = \int_{[0,T]} \varphi(t)\mathrm{d}v(t).$$

Since each function $v(\cdot)$ with bounded variation can be modified (without changing the dual pairing) so that $v(\cdot)$ is also cad-lag, we denote by $D_{\text{BV}}([0,T])$, the set of cad-lag functions with bounded variation, which is considered as a subspace of the canonical space $D([0,T])$. Because $D_{\text{BV}}([0,T])$ is the dual of the Banach space $C([0,T])$, we can use the weak* topology on $D_{\text{BV}}([0,T])$, where balls are weakly* compact. Thus, we are interested in a topology on the space $D([0,T])$ such that relatively to the subspace $D_{\text{BV}}([0,T])$ the convergence is similar to the weak* convergence and any set of equi-bounded variation functions is compact. The topology introduced by Jakubowski [118] has this property. Clearly, what is done for $D([0,T])$ can be extended to $D([0,\infty),\mathbb{R}^d)$.

Again, let $D_{\text{BV}}([0,\infty),\mathbb{R}^d)$ denote the space of functions x in $D([0,\infty),\mathbb{R}^d)$ that locally are of bounded variation, with the sup-norm

$$\|x\|_{T,\infty} = \sup\big\{|x(t)| \,:\, 0\le t\le T\big\}$$

and the variation-norm

$$\|x\|_{T,\text{BV}} = \sup\Big\{\sum_{i=0}^{n} |x(t_{i+1})-x(t_i)| \,:\, t_i<t_{i+1}\Big\}.$$

where the supremum is taken with respect to all partitions with $t_0=0$, $t_n=T$ and t_i belonging only to a dense subset of $(0,T)$. We consider the following convergence in $D([0,\infty),\mathbb{R}^d)$, as introduced by Jakubowski [118].

Definition 4.60. We say that the sequence $\{x_n\}$ in $D([0,\infty),\mathbb{R}^d)$ is convergent in the sense of Jakubowski, denoting by $x_n \xrightarrow{\text{Jak}} x$, if and only if, for any $\epsilon>0$ there exist a sequence $\{v_n^\epsilon\}$ and v^ϵ in $D_{\text{BV}}([0,\infty),\mathbb{R}^d)$ such that

$$\|v_n^\epsilon - x_n\|_{1/\varepsilon,\infty}\le\epsilon, \quad \forall n\ge 1, \qquad \|v^\epsilon - x\|_{1/\varepsilon,\infty}\le\epsilon$$

and

$$\int_0^{1/\varepsilon}\varphi(t)\,\mathrm{d}v_n^\epsilon \to \int_0^{1/\varepsilon}\varphi(t)\,\mathrm{d}v^\epsilon,$$

for any $\varphi\in C([0,1/\varepsilon])$. □

Star-convergence of Probability on $\mathbb{D}_S$

Now we look at $\mathbb{D}_S$, the canonical space $\mathbb{D}=D([0,\infty),\mathbb{R}^d)$ space of cad-lag functions with the S-topology defined below (i.e., the $*$-convergence, denoted by "$\xrightarrow{\text{Jak}^*}$", and derived from Definition 4.60).

It can be proved that the S-topology on $\mathbb{D}$ generates the same Borel sets that we have with the metric (J_1) topology, thus the probability measures are the same. Since the compact sets in the metric topology are also compact in the S-topology, we have that every probability measure is also tight in the S-topology. Hence, the sequential convergence is defined on the entire space of probability measures $\wp(\mathbb{D}_S) = \wp^o(\mathbb{D}_S)$, because all probability measures are tight in $\mathbb{D}_S$.This new topology $\mathfrak{W}_{\mathsf{Jak}}(\mathbb{D}_S)$ is stronger than the topology of weak convergence $\mathfrak{W}_{\mathsf{seq}}(\mathbb{D}_S)$ (where we consider the S-topology on $\mathbb{D}$), in other words if we have the $\mathfrak{W}_{\mathsf{Jak}}$ convergence then we have also the weak S-convergence. On the other hand the $\mathfrak{W}_{\mathsf{seq}}(\mathbb{D}_S)$-topology is weaker than the classical (metric) topology of weak convergence (that is with J_1 as topology on $\mathbb{D}$). However, in general, we cannot say anything (from only this information) on the classical weak convergence, hence every case needs a specific study: for example, the Łaukajtys-Słomiński paper [151] shows that we don't have the classical convergence, nevertheless they prove the weak S-convergence (proving namely the stronger $\mathfrak{W}_{\mathsf{Jak}}$ convergence).

In other words, rephrasing Definition 4.60), $x_n \xrightarrow{\mathsf{Jak}} x$ if and only if there exists a double sequence $\{v_{n,k}\}$ in $D_{\mathsf{BV}}([0,\infty),\mathbb{R}^d)$ such that (1) for every n, $v_{n,k} \to x_n$ locally uniform as $k \to \infty$, (2) for every k and any continuous function φ, $\langle \varphi, v_{n,k}\rangle \to \langle \varphi, v_k\rangle$ as $n \to \infty$, and (3) $v_k \to x$ locally uniform as $k \to \infty$.

Actually we can endow $D([0,\infty),\mathbb{R}^n)$ of the topology τ_{Jak} defined by the following family of open sets

> G is *open* if and only if for any sequence $\{x_n\}$, converging to a element $x \in G$ in the previous sense of Jakubowski, a *tail* of the sequence belongs to G, i.e. there exists an integer N such that $x_n \in G$ for any $n \geq \mathbb{N}$.

We now remark that the converging sequences in the sense of Jakubowski are *not the only sequences* that converge in the above topology! There are sequences that are convergent in the sense of τ_{Jak} topology without being convergent in the sense of Jakubowski. Thus, we will indicate this weaker convergence in the topology τ_{Jak} as $x_n \xrightarrow{\mathsf{Jak}^*} x$.

We can also endow $D([0,\infty),\mathbb{R}^n)$ of the usual Skorokhod topology (the so called J_1 topology), under which $D([0,\infty),\mathbb{R}^n)$ is a separable, complete metric space (Polish space).

Actually the topology τ_{Jak} is weaker of the usual Skorokhod topology, but the Borel sets with respect to τ_{Jak} are just the same Borel sets with respect to Skorokhod topology, both coinciding with the σ-algebra generated by the cylindrical sets.

Note the contrast, $D_{\mathsf{BV}}([0,\infty),\mathbb{R}^d)$ is dense in $\mathbb{D}_S$, but $C([0,\infty),\mathbb{R}^d)$ is closed in $D([0,\infty),\mathbb{R}^d)$, and $C([0,\infty),\mathbb{R}^d)$ is dense in $\mathbb{D}_S$. To check this, first recall that for any x in $D([0,\infty),\mathbb{R}^d)$ and any $\varepsilon > 0$ there exist $0 = t_0 < t_1 < \cdots < t_r = 1/\varepsilon$ such that for any $i = 1,\ldots,r$ and for any s,t in $[t_{i-1},t_i)$ we have $|x(t)-x(s)| < \varepsilon$. Indeed, by means of the right continuity property, we can define inductively $t_0 =$

$\inf\{t > 0 : |x(t) - x(0)| \geq \varepsilon/2\}$ and $t_i = \inf\{t > t_{i-1} : |x(t) - x(t_{i-1})| \geq \varepsilon/2\}$ for $i \geq 1$. This sequence $\{t_k\}$ is divergent, namely if $t_k \to \tilde{t}$ we would have also $\varepsilon/2 \leq |x(t_k) - x(t_{k-1}| \to |x(\tilde{t}-) - x(\tilde{t}-)| = 0$, in view of the existence of left-hand limits, which is a contradiction. Thus we can define

$$v^\varepsilon(t) = x(t_{i-1}) \quad \text{if} \quad t \in [t_{i-1}, t_i), \quad i = 1, \dots, r,$$

which is a piecewise constant function (so cad-lag with bounded variation) satisfying $\|v^\varepsilon - x\|_{1/\varepsilon,\infty} < \varepsilon$. Now defined

$$x_n(t) = n \int_t^{(t+1/n)} x(s)\mathrm{d}s \quad \text{and} \quad v_n^\varepsilon(t) = n \int_t^{(t+1/n)} v^\varepsilon(s)\mathrm{d}s,$$

which are absolutely continuous (and so continuous with bounded variation), $\|v_n^\varepsilon - x_n\|_{1/\varepsilon,\infty} < \varepsilon$, and as $n \to \infty$, converge pointwise (i.e., for each t fixed) to x and v^ε. This proves that the space of absolutely continuous functions is dense $D([0,\infty,\mathbb{R}^d)$ with the Jakubowski topology.

4.4.4 Uniform Tightness or UT Condition

The following statements described this topology:

(1) The space $D([0,\infty),\mathbb{R}^n)$ equipped with the sequential topology τ_{Jak} is a Hausdorff topological space which is not a metric space. Recall that with the Skorokhod topology, it is a complete separable metrizable space.

(2) There exists a countable family of τ_{Jak}-continuous functions which separate points in $D([0,\infty),\mathbb{R}^n)$.

(3) The addition is sequentially continuous with respect to convergence in the sense of Jakubowski. In particular, $x_n \xrightarrow{\mathsf{Jak}} x$ if and only if $x_n - x \xrightarrow{\mathsf{Jak}} 0$. Recall, this holds with the Skorokhod topology only if x is continuous.

(4) Compact subsets $K \subset D([0,\infty),\mathbb{R}^n)$ are metrizable spaces.

(5) A subset $K \subset D([0,\infty),\mathbb{R}^n)$ is relatively τ_{Jak}-compact if for any each $\varepsilon > 0$ there exists a constant C_ε such that for each $x \in K$ there exists $v_{x,\varepsilon}$ in $D_{\mathsf{BV}}([0,\infty),\mathbb{R}^n)$ such that

$$\|x - v_{x,\varepsilon}\|_{1/\varepsilon,\infty} \leq \varepsilon \quad \text{and} \quad \|v_{x,\varepsilon}\|_{1/\varepsilon,\mathsf{BV}} \leq C_\varepsilon \tag{4.124}$$

is satisfied.

(6) The evaluation or projection operators $x \mapsto x(t)$ from $D([0,\infty),\mathbb{R}^n)$ into $\mathbb{R}^d$ are nowhere continuous with the τ_{Jak} topology. However, the functionals

$$x \mapsto \frac{1}{\varepsilon}\int_t^{t+\varepsilon} x(s)\mathrm{d}s \quad \text{and} \quad x \mapsto \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} x(s)\mathrm{d}s$$

are continuous and converges to $x(t)$ and $x(t-)$ as $\varepsilon \to 0$. Thus, τ_{Jak}-Borel subsets $\mathcal{B}_{\mathsf{Jak}}$ coincide with the standard σ-algebra generated by evaluations (projections). This fact implies that any probability measure on $(D([0,\infty),\mathbb{R}^n),\mathcal{B}_{\mathsf{Jak}})$

is tight. Recall that with the Skorokhod topology, the evaluation operators are continuous at any continuity time of the limit point.

(7) τ_{Jak} is coarser (weaker) than the usual Skorokhod topology. Thus the cadlag space $D([0,\infty),\mathbb{R}^n)$ endowed with the τ_{Jak} Jakubowski topology is a Lusin space, i.e., it is a one-to-one continuous image of a Polish space. However, it is unknown if the space is completely regular (i.e., the topology is given by a family of pseudo-metric which is Hausdorff separated.)

(8) The subspace $D_{\text{BV}}([0,\infty),\mathbb{R}^n)$ is dense, indeed, if x is any point in in the space $D([0,\infty),\mathbb{R}^n)$ the singleton $\{x\}$ is compact, hence we can find a sequence of functions in $D_{\text{BV}}([0,\infty),\mathbb{R}^d)$ converging to x in the sense of Jakubowski, in view of (4.124). Actually, we proved above that the space of absolutely continuous is dense. We may consider $D_{\text{BV}}([0,\infty),\mathbb{R}^n)$ with the relative τ_{Jak} topology on $D([0,\infty),\mathbb{R}^n)$, but it is weaker than the weak* topology.

(9) Let $N_T^{a,b}(x)$ be the number of up-crossing (of x) of the interval $[a,b]$ in the time interval $[0,T]$, i.e., $N_T^{a,b}(x)\geq k$ if there exist $0\leq t_1<t_2<\ldots<t_{2k-1}<t_{2k}$ such that $x(t_{2i-1})<a<b<x(t_{2i})$ for any $i=1,2,\ldots k$. On the other hand, let $N_{T,\eta}(x)$ be the number of oscillations (of x) greater than η in the time interval $[0,T]$, i.e., $N_{T,\eta}(x)\geq k$ if there exist $0\leq t_1<t_2<\ldots<t_{2k-1}<t_{2k}$ such that $|x(t_{2i})-x(t_{2i-1})|>\eta$ for any $i=1,2,\ldots k$. Then, it can be proved that any uniformly equi-bounded subset K of $D([0,\infty),\mathbb{R}^n)$ (i.e., for any $T>0$ there exists a constant $C>0$ such that $\|x\|_{T,\infty}\leq C$ for every x in K) is compact (i.e., condition (4.124) holds) if and only if one of the following two conditions, for every $T>0$,

$$\sup_{x\in K} N_T^{a,b}(x)<+\infty \qquad \text{or} \qquad \sup_{x\in K} N_{T,\eta}(x)<+\infty,$$

for each $a<b$, or for each $\eta>0$, is satisfied.

The following result is useful to check the τ_{Jak} convergence. Let $\{x_n\}$ is relatively τ_{Jak}-compact. Then $x_n \xrightarrow{\text{Jak}^*} x$ if there exists a (countable) dense set Q in $[0,\infty)$ such that $x_n(q)\to x(q)$, as $n\to+\infty$, for every q in Q.

On the other hand, if $\varphi(t,x)$ is a locally bounded Charathéodory function (measurable in t and continuous in x) defined on $[0,\infty)\times\mathbb{R}^d$, and ℓ is a continuous bounded variation function, then the integral functionals

$$x\mapsto\int_{]0,T]}\varphi(t,x(t-))\mathrm{d}\ell(t) \quad\text{and}\quad x\mapsto\int_{]0,T]}\varphi(t,x(t))\mathrm{d}\ell(t)$$

are continuous the Jakubowski topology.

However, if we consider the sequence $k_n(t)=\mathbb{1}_{[1/2-1/n,1]}(t)$, the function $k(t)=1_{[1/2,1]}(t)$ and the sequence $x_n(t)=k(t)$, constant for any n, then $k_n\to k$ and $x_n\to k$ in the Skorokhod space $D([0,1],\mathbb{R})$ but

$$k(t)=\int_{]0,t]}k_n(s-)\mathrm{d}x_n(s)\not\to\int_{]0,t]}k(s-)\mathrm{d}k(s)=0.$$

Another way to get a convergence of integrals is to consider convergence in $D([0,1],\mathbb{R}^2)$. There is a general result from Jakubowski-Mémin-Pagès [119]: we have (note that in the following theorem the topology in $D([0,T],\mathbb{R}^n)$ is the Skorokhod topology J_1)

Now we take a look at

Definition 4.61 (UT Condition)**.** Given a sequence of stochastic processes (semi-martingales) X_n with respect to the stochastic basis $(\Omega^n,\mathcal{F}^n,\{\mathcal{F}_t^n\},P^n)\}$, the UT condition means that the family of all random variables of the form

$$\sum_{i=1}^{N} H_{t_{i-1}}^n (X_{t_i}^n - X_{t_{i-1}}^n)$$

is uniformly tight, where N is any integer, $0 = t_0 < t_1 < \cdots < t_N = T$ and $|H_{t_i}^n| \le 1$ with $H_{t_i}^n$ is $\mathcal{F}_{t_i}^n$-measurable for any i. □

Theorem 4.62. *Given a sequence of semi-martingales M_n with respect to the stochastic basis $(\Omega^n,\mathcal{F}^n,\{\mathcal{F}_t^n\},P^n)\}$ satisfying the UT condition and a sequence of stochastic processes K_n adapted to $\{\mathcal{F}_t^n\}$, with trajectories in $D([0,\infty),\mathbb{R}^d)$, let us suppose that*

$$(K_n, M_n) \to (K, M)$$

weakly in $D([0,\infty),\mathbb{R}^{2d})$. Then M is a semi-martingale with respect to the natural filtration generated by (K,M) and

$$\int_{]0,\cdot]} K_n(t-)\mathrm{d}M_n(t) \to \int_{]0,\cdot]} K(t-)\mathrm{d}M(t)$$

weakly in $D([0,\infty),\mathbb{R}^d)$, and

$$\Big(K_n, X_n, \int_{]0,\cdot]} K_n(t-)\mathrm{d}M_n(t)\Big) \to \Big(K, M, \int_{]0,\cdot]} K(t-)\mathrm{d}M(t)\Big)$$

weakly in $D([0,\infty),\mathbb{R}^{3d})$. □

Note that the convergence $(K_n,M_n)\to(K,M)$ in $D([0,\infty),\mathbb{R}^{2d})$ is strictly stronger than the convergence $(K_n,M_n)\to(K,M)$ in the product topology of $D([0,\infty),\mathbb{R}^d)\times D([0,\infty),\mathbb{R}^d)$. We have from Jakubowski [118]

Theorem 4.63. *The UT condition implies the tightness in the space of probability measures on $D([0,\infty),\mathbb{R}^n)$ with respect to $*$-convergence.* □

The above statements are valid sometimes also in some not Polish space. If Ω is topological space having a countable family of continuous functions separating points then Skorokhod representation and Prohorov's theorem hold, i.e., for any uniformly tight sequence $\{P_n : n \ge 1\}$ of probability measures on Ω there exist a subsequence of indexes $\{n_k\}$ and random variables $\{X_k : k\in\mathbb{N}\}$ and X on the universal (Lebesgue) probability space $([0,1],\mathcal{B}([0,1]),\ell)$ with values in Ω such that P_{n_k} is the distribution of X_k for any $k\in\mathbb{N}$ and $X_k(t)$ converges to $X(t)$

for every t in $[0,1]$. In particular this applies when Ω is the canonical space $D([0,\infty),\mathbb{R}^d)$ with the τ_{Jak} Jakubowski topology (which is then only a Lusin space).

Because this is based on Prohorov's theorem, the above result holds for the $D([0,\infty),\mathbb{R}^d)$ with τ_{Jak} Jakubowski topology (see Definition 4.60) if the criterium of compactness (a') and (b') are modified accordingly.

Theorem 4.64 (tight)**.** *Let $X_1, X_2, \dots$ be a sequence of random variables with values in $D([0,\infty),\mathbb{R}^d)$, and $P_1, P_2, \dots$ be its associated probability law on $D([0,\infty),\mathbb{R}^d) = \mathbb{D}_S$, endowed with the Jakubowski topology. Then the sequence $P_1, P_2, \dots$ is tight (hence relatively compact) in $\mathbb{D}_S$ if and only if the following two conditions hold:*

(a) *almost equi-bounded, i.e., for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have*

$$P_n\{\sup_{0\le t\le 1/\varepsilon} |X_n(t)| \le C\} \ge 1-\varepsilon,$$

(b) *equi-UT-condition, i.e., for any $\varepsilon > 0$ and for each $T, \eta > 0$, there exists $K > 0$ such that for any index n we have*

$$P_n\{N_{T,\eta}(X_n) \le K\} \ge 1-\varepsilon.$$

Moreover, if the sequence is tight, then it is weakly convergent if and only its finite-dimensional distributions converge. □

Certainly we can replace (b) with

(b') for any $\varepsilon > 0$ and for each $T > 0$, $b > a$, there exists $K > 0$ such that for any index n we have

$$P_n\{N_T^{a,b}(X_n) \le K\} \ge 1-\varepsilon,$$

where $N_T^{a,b}$ and $N_{T,\eta}(X_n)$ are as in (9) above.

For instance, if the processes (P_n, X_n) has local bounded variation, i.e., $X_n = X_n^+ - X_n^-$, with X_n^+ and X_n^- being increasing monotone, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\{|X_n^+(t)| + |X_n^-(t)| > C\} \le \varepsilon,$$

implies both (a) and (b) above, since $N_{T,\eta}$ is controlled by the variation $|X_n^+| + |X_n^-|$ process.

Similarly, if the processes (P_n, X_n) is a local continuous martingale with predictable variation process $\langle X_n\rangle$, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\big\{\big|\langle X_n\rangle(t)\big| > C\big\} \le \varepsilon,$$

implies both (a) and (b), since $N_T^{a,b}$ is controlled by the predictable variation process. Similarly, if the processes (P_n, X_n) is a local purely discontinuous (square-integrable) martingale with integer measure ν_n and predictable jumps compensator ν_n^p, then the condition: for any $\varepsilon > 0$ there exists $C > 0$ such that for any index n we have

$$\sup_{0\le t\le 1/\varepsilon} P_n\Big\{ \int_{\mathbb{R}^d_*\times]0,t]} \big(|z|^2 \wedge 1\big)\nu_n^p(\mathrm{d}z,\mathrm{d}s) > C\Big\} \le \varepsilon,$$

implies both (a) and (b), since $N_T^{a,b}$ is controlled by the predictable jumps compensator process. Note that

$$\int_{\mathbb{R}^d_*\times]0,t]} \big(|z|^2 \wedge 1\big)\nu_n^p(\mathrm{d}z,\mathrm{d}s) = \\ = \sum_{0<s\le t} \big[\mathbb{1}_{|X_n(s)-X_n(s-)|\ge 1} + |X_n(s)-X_n(s-)|^2\mathbb{1}_{|X_n(s)-X_n(s-)|<1}\big],$$

i.e., adding the number of jumps greater than 1 and the square of the small jumps. Actually, these martingale cases can be treated directly with the classic Skorokhod topology, since

$$P_n\{\sup_{a\le t,s\le b}|X_n(t)-X_n(s)|\ge\varepsilon\} \le \frac{\delta}{\varepsilon^2} + P_n\{\big|\langle X_n\rangle(b)-\langle X_n\rangle(a)\big|\ge\delta\},$$
$$P_n\{\sup_{a\le t,s\le b}|X_n(t)-X_n(s)|\ge\varepsilon\} \le \frac{\delta}{\varepsilon^2} + \\ + P_n\Big\{\int_{\mathbb{R}^d_*\times]a,b]}\big(|z|^2\wedge 1\big)\nu_n^p(\mathrm{d}z,\mathrm{d}s)\ge\delta\Big\},$$

for every $\varepsilon, \delta > 0$, in view of Lenglart dominate property, e.g., see Jacod and Shiryaev [117, Section I.3c, pp. 35–36]. Essentially, the local bounded variation processes are of main interest for the Jakubowski topology.

Another situation is the following, see Section 4.3.4. Let ℓ_n, υ_n, w_n and ν_n, $n = 1, 2, \ldots$ be sequences of processes defined in a probability space $(\Omega_n, \mathcal{F}_n, P_n)$. Assume that:

(1) ℓ_n and υ_n are cad-lag processes with values in $\mathbb{R}^d$ and non-anticipative relative to (w_n, ν_n), and ℓ_n are nondecreasing,

(2) w_n are a d-dimensional continuous square integrable martingales with predictable quadratic covariation $\langle w_{n,i}, w_{n,j}\rangle = \varsigma_{n,j}$ if $i = j$, and $\langle w_{n,i}, w_{n,j}\rangle = 0$ if $i \ne j$,

(3) ν_n are integer measures with jump compensator $\nu_n^p(\mathrm{d}z,\mathrm{d}t) = \pi_n(\mathrm{d}z)\,\mathrm{d}\varrho_n(t)$, where π_n is a given Lévy measure in $\mathbb{R}^m_*$, and q_n denotes the corresponding purely discontinuous square-integrable martingale, i.e.,

$$q_n(t) = \int_{\mathbb{R}^m_*\times]0,t]} |z|^2\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}s), \quad \forall t \ge 0,$$

with $\tilde{\nu}(\mathrm{d}z,\mathrm{d}t) = \nu(\mathrm{d}z,\mathrm{d}t) - \pi_n(\mathrm{d}z)\varrho_n(\mathrm{d}t)$. Suppose that ℓ_n, υ, ς_n and ϱ_n are equi-bounded in probability, i.e., for every $\varepsilon > 0$ there exists $C = C(\varepsilon)$ such that

$$\sup_{0\le t\le 1/\varepsilon} P_n\big\{\big[|\ell_n(t)| + |\upsilon_n(t)| + |\varsigma_n(t)| + |\varrho_n(t)|\big] \ge C\big\} \le \varepsilon, \quad \forall n,$$

and ℓ_n, ς_n and ϱ_n vanish at time $t = 0$.

(4) Also assume that υ_n satisfies the UT-condition, in term of the number of up-crossing $N_T^{a,b}$ or the number of oscillations $N_{T,\eta}$, e.g., for any $\varepsilon > 0$ and for each $T, \eta > 0$, there exists $K > 0$ such that for any index n we have

$$P_n\{N_{T,\eta}(\upsilon_n) \ge K\} \le \varepsilon;$$

also that ς_n and ϱ_n are equi-continuous in probability, i.e., for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P_n\big\{\sup_{0\le s,t\le 1/\varepsilon,\ |t-s|<\delta} \big[|\varsigma_n(t) - \varsigma_n(s)| + |\varrho_n(t) - \varrho_n(s)|\big] \ge \varepsilon\big\} \le \varepsilon, \quad \forall n;$$

and that $\{\pi_n\}$ is a uniformly integrable Lévy sequence, i.e., there is a constant $C >$ such that

$$\int_{\mathbb{R}_*^m} |z|^2 \pi_n(\mathrm{d}z) \le C, \quad \forall n,$$

and for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\int_{\{z:|z|<\delta\}\cup\{z:|z|>1/\delta\}} |z|^2 \pi_n(\mathrm{d}z) < \varepsilon, \quad \forall n.$$

Now, consider (a) the probability law Q_n defined by $(P_n, \ell_n, \upsilon_n, w_n, \varsigma_n, q_n, \varrho_n)$ in the canonical space $D([0,\infty), \mathbb{R}^{d_0})$, with $d_0 = 4d + m + 1$, (b) the canonical processes ℓ, υ, w, ς, q, ϱ, and (c) endowed with the Jakubowski topology in the first $2d$ variables (relative to ℓ and υ) and with the Skorokhod topology in the remaining variables. Actually, for the variables w, ς and ϱ, we could use the sample space $C([0,\infty), \mathbb{R}^{d_1})$, $d_1 = 2d + 1$, with the usual locally uniform convergence. Then we can extract a subsequence, still denoted by $\{Q_n, \pi_n\}$, which is weak convergent to Q, π.

Clearly, all limiting processes are cad-lag. Moreover w, ς and ϱ are also continuous. Then, relative to Q on $D([0,\infty), \mathbb{R}^{d_0})$, we have:

(a) ℓ and υ are non-anticipating processes relative to (w, ν), and ℓ is a nondecreasing,

(b) w is also a continuous square integrable martingale with predictable quadratic covariation $\langle w_i, w_j\rangle = \varsigma_i$ if $i = j$ and $\langle w_i, w_j\rangle = 0$ if $i \ne j$,

(c) the integer measure ν associated with q has $\nu^p(\mathrm{d}z,\mathrm{d}t) = \pi(\mathrm{d}z)\mathrm{d}\varrho(t)$ as it predictable jump compensator,

(d) on the universal probability space $([0,1],\mathcal{B}([0,1]),l)$, where l is the Lebesgue measure, there exist random variables x, x_n with values in the canonical space $D([0,\infty),\mathbb{R}^{d_0})$ such that first (i) x and x_n have the same finite distributions as $(\ell,\upsilon,w,\varsigma,q,\varrho)$ and $(\ell_n,\upsilon_n,w_n,\varsigma_n,q_n,\varrho_n)$, respectively, and secondly (ii) $x_n(\theta)\to x(\theta)$, for every θ in $[0,1]$.

At this point, we can take limit on any continuous functional defined on the space $\mathcal{D}([0,\infty),\mathbb{R}^{d_0})$, e.g.,

$$\int_{]0,T]} f(t,x_n)\mathrm{d}\ell_n(t), \quad \int_0^T f(t,x_n)\mathrm{d}w_n(t), \quad \int_{\mathbb{R}^m_*\times(0,T]} g(z,t,x_n)\tilde{\nu}_n(\mathrm{d}z,\mathrm{d}t),$$

preserved almost surely through finite-dimensional distributions. It is also clear that if the processes υ_n are equi-continuous in probability, we may use the Skorokhod topology in the variable υ instead of the weaker Jakubowski topology. Moreover, the cad-lag modulus of continuity can also be used. Note that $f(t,x)$ and $g(z,t,x)$ are regarded as deterministic (real or vector-valued) random fields with $t\geq 0$ and x in $\mathbb{D} = D([0,\infty),\mathbb{R}^{d_0})$, i.e., the mappings $(t,x)\mapsto f(t,x)$ and $(z,t,x)\mapsto g(z,t,x)$ are measurable with respect to the product Borel σ-algebras $\mathcal{B}([0,\infty)))\times\mathcal{B}(\mathbb{D})$ and $\mathcal{B}(\mathbb{R}^m_0)\times\mathcal{B}([0,\infty)))\times\mathcal{B}(\mathbb{D})$, plus some appropriate regularity conditions, e.g., continuity in x and causality, i.e., if $x(s)=y(s)$ for any $0\leq s<t$ then $f(t,x)=f(t,y)$. Typical $f(t,x)$ has the form $f(t,x(t))$ or $f(t,x(t-))$. It is perhaps important to recall that under the Skorokhod topology, the evaluation or projection functional $x\mapsto x(t)$ are continuous only at any point of continuity of x, however, under the Jakubowski topology, they are nowhere continuous. Thus, when the measures $\mathrm{d}\ell_n$ have atoms (i.e., the processes ℓ_n are discontinuous) some extra special care should be taken to ensure the passage to the limit in the Lebesgue-Stieltjes integral. Certainly all this applies to our case of interest, i.e., for a Wiener process or a Poisson measure.

4.5 Density Estimates

A very important point is to obtain a priori estimates of the distributions of non-degenerate stochastic integrals. Consider an Itô process in $\mathbb{R}^d$,

$$X(t) = x + \int_0^t b(s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(s)\mathrm{d}w_k(s), \quad \forall t\geq 0,$$

in some filtered probability space $(\Omega,F,P,\mathcal{F}(t):t\geq 0)$, and recall the Lebesgue spaces $L^p(D)$ or $L^p(D_T)$ of p-integrable functions $f(x)$, $g(t,x)$ defined on an open subset D of $\mathbb{R}^d$, or $D_T:=]0,T[\times D$ with the norm

$$\|f\|_{p,D} := \Big[\int_D |f(x)|^p\mathrm{d}x\Big]^{1/p},$$
$$\|g\|_{p,D_T} := \Big[\int_0^T \mathrm{d}t\int_D |g(t,x)|^p\mathrm{d}x\Big]^{1/p}$$

where $1 \le p < \infty$.

We quote Theorem 2.2.4 in the book Krylov [139, Sections 2.2 and 2.3, pages 51–67],

Theorem 4.65. *Let D be a bounded open subset of $\mathbb{R}^d$ and τ_D be the first exit time of the process $(X(t) : t \ge 0)$ from the open set D,*

$$\tau_D(\omega) := \inf\big\{t \ge 0 : X(t,\omega) \in D\big\}.$$

If there exists $\delta > 0$ such that

$$|b(t,\omega)| \le \frac{1}{\delta}, \qquad \sum_{ik} |\ell_i \sigma_{ik}(t,\omega)|^2 \ge \delta |\ell|^2,$$

for every ℓ in $\mathbb{R}^d$ and $0 \le t \le \tau_D(\omega)$, then there exists a constant $K = K(d, D, \delta)$ depending only on δ, the dimension d and the diameter of the region D, such that for any Borel measurable nonnegative functions $g(x)$ and $h(t,x)$ we have

$$\mathbb{E}\Big\{ \int_s^{s\vee\tau_D} g(X(t))\mathrm{d}t \,|\, \mathcal{F}(s)\Big\} \le K\Big[\int_D |g(x)|^d \mathrm{d}x\Big]^{1/d}$$

and

$$\mathbb{E}\Big\{ \int_s^{(s\vee\tau_D)\wedge T} h(t,X(t))\mathrm{d}t \,|\, \mathcal{F}(s)\Big\} \le K\Big[\int_s^T \mathrm{d}t \int_D |h(t,x)|^{d+1} \mathrm{d}x\Big]^{1/(d+1)},$$

for every $T > 0$. □

The proof of this Theorem (and some variations using the concept of *regular functions*) is based on the following result proved in Krylov [138] (see also the book Krylov [140]) and related to the weak maximum principle.

Let $\rho = \rho(t,x)$ be a nonnegative smooth function with support inside the compact set $\{(t,x) : 0 \le t \le 1,\ |x| \le 1\}$ of $(0,\infty) \times \mathbb{R}^d$ such that

$$\int_0^\infty \mathrm{d}t \int_{\mathbb{R}^d} \rho(t,x)\mathrm{d}x = 1.$$

For any locally integrable function h, denote by $h * \rho_\varepsilon$ the smooth mollification of h with respect to ρ, i.e.,

$$h * \rho_\varepsilon(t,x) := \varepsilon^{-(d+1)} \int_0^\infty \mathrm{d}s \int_{\mathbb{R}^d} h(t-s, x-y)\rho(s/\varepsilon, y/\varepsilon)\mathrm{d}y = \\ = \int_0^\infty \mathrm{d}s \int_{\mathbb{R}^d} h(t-\varepsilon s, x-\varepsilon y)\rho(s,y)\mathrm{d}y,$$

where

$$\rho_\varepsilon(t,x) := \varepsilon^{-(d+1)}\rho(t/\varepsilon, x/\varepsilon),$$
$$\int_0^\infty \mathrm{d}s \int_{\mathbb{R}^d} \rho_\varepsilon(s,y)\mathrm{d}y = \int_0^\infty \mathrm{d}s \int_{\mathbb{R}^d} \rho(s,y)\mathrm{d}y = 1,$$

for every (t,x) in $(0,\infty) \times \mathbb{R}^d$.

Theorem 4.66. *For each positive constants λ, ε, and any nonnegative continuous function $h = h(t,x)$ on $[0,\infty) \times \mathbb{R}^d$, where $h_\varepsilon := h * \rho_\varepsilon$ denote its mollification, there exists a smooth function u satisfying the following properties:*

$$\text{(a)} \quad |\nabla u(t,x)| \le \sqrt{\lambda}\, u(t,x), \quad \forall x \in \mathbb{R}^d,\ t \ge 0,$$

where ∇ is the gradient in the variable x,

$$\text{(b)} \quad \sum_{ij=1}^{d} a_{ij}\partial_{ij} u(t,x) - \lambda \operatorname{Tr}(a) u(t,x) \le 0,$$

$$\text{(c)} \quad \partial_t u(t,x) + \sum_{ij=1}^{d} a_{ij}\partial_{ij} u(t,x) - \lambda(\operatorname{Tr}(a)+1)u(t,x) \le \\ \le - \sqrt[d+1]{\det(a)}\, h_\varepsilon(t,x),$$

for any $t \ge 0$, x in $\mathbb{R}^d$ and for every symmetric nonnegative definite $d \times d$ matrix $a = (a_{ij})$, where $\operatorname{Tr}(\cdot)$ and $\det(\cdot)$ denote the trace and the determinant of a matrix, and

$$\text{(d)} \quad u(t,x) \le C(p,d,\lambda)\Big[\int_t^\infty \mathrm{e}^{-\lambda(d+1)(s-t)}\mathrm{d}s \int_{\mathbb{R}^d} |h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},$$

where the constant $C(p,d,\lambda)$ is equal to

$$(p+1)^{d/(p+1)}[d!\,\omega_d]^{-1/(p+1)}\lambda^{(d-2p)/(2p+2)}(d+1)^{-p/(p+1)},$$

for every $p \ge d$, x in $\mathbb{R}^d$ and $t \ge 0$, where ω_d is the volume of the unit sphere in $\mathbb{R}^d$, i.e., $\omega_d = \pi^{d/2}/\Gamma(d/2+1)$, with Γ being the gamma function. Furthermore, if f is independent of the variable t so is u. □

In the previous theorem, u is the mollification of a bounded function (obtained after some geometric arguments) with respect to the same smooth function ρ_ε. Also note that $\omega_d =$, is the volume of the unit sphere in $\mathbb{R}^d$. Notice that the property (b) yields

$$\text{(e)} \quad \sum_{ij=1}^{d} z_i z_j \partial_{ij} u(t,x) \le \lambda |z|^2 u(t,x), \quad \forall z = (z_1,\dots,z_d),$$

for every x in $\mathbb{R}^d$ and $t \ge 0$. Actually, if we express the matrix $a = \sigma\sigma^*$ then we see that (b) and (e) are equivalent.

4.5.1 In the Whole Space

Consider an Itô process with jumps in $\mathbb{R}^d$,

$$\xi(t) = x + \int_0^t b(s)\mathrm{d}s + \sum_k \int_0^t \sigma_k(s)\mathrm{d}w_k(s) + \int_{\mathbb{R}^d_*\times]0,t]} z\tilde{\nu}_\xi(\mathrm{d}z,\mathrm{d}s), \quad (4.125)$$

for every $t \geq 0$, in some filtered probability space $(\Omega, F, P, \mathcal{F}(t) : t \geq 0)$, where $\tilde{\nu}_\xi$ is the local martingale measure associated with the integer valued measure ν_ξ (with predictable compensator ν^p_ξ), corresponding to the quasi-left continuous process ξ. The compensator ν^p_ξ is assumed absolutely continuous with respect to the Lebesgue measure, i.e.,

$$\nu^p_\xi(B, (0,t]) := \int_0^t \mathtt{M}_\xi(B, s)\mathrm{d}s, \quad \forall B \in \mathcal{B}(\mathbb{R}^d_*),$$

where the intensity kernel $\mathtt{M}$ satisfies: (1) $s \mapsto \mathtt{M}_\xi(B, s)$ is predictable for every B and (2) $B \mapsto \mathtt{M}_\xi(B, s)$ is a measure in $\mathbb{R}^d_*$ for every s. The coefficients $b(s)$ and $\sigma(s)$ are predictable processes. Also, for a given predictable nonnegative process $\dot{\tau}(s)$ define the following (integral) process

$$\tau(s) := \int_0^s \dot{\tau}(r)\mathrm{d}r,$$

and also the processes

$$a(s) := \big(\frac{1}{2}\sum_k \sigma_{ik}(s)\sigma_{jk}(s)\big), \qquad \rho(s) := \int_0^s \mathrm{Tr}(a(r))\mathrm{d}r,$$

where $\rho(\cdot)$ has nonnegative real values and $a(\cdot)$ has symmetric nonnegative $d \times d$ matrix values. Note that

$$\mathrm{Tr}(a(s)) = \frac{1}{2}\sum_{ik} |\sigma_{ik}(s)|^2, \qquad \frac{\mathrm{d}\rho}{\mathrm{d}s}(s) = \dot{\rho}(s) = \mathrm{Tr}(a(s))$$

and usually $\dot{\tau} = 1$. Now we have

Theorem 4.67. *Let $\xi(\cdot) = \xi_x(\cdot)$ be an Itô process with jumps as in (4.125) in $\mathbb{R}^d$ and $\dot{\tau}$ be a predictable nonnegative processes. Define the processes $\tau(\cdot)$, $\rho(\cdot)$ and $a(\cdot)$ as above and suppose that for some positive constants $K = K_\xi$, and $\lambda > 0$, and any s, ω we have*

$$2\sqrt{\lambda}\,|b(s,\omega)| + \lambda \int_{\mathbb{R}^d_*} |z|^2 \mathtt{M}_\xi(\mathrm{d}z, s, \omega) \leq 2\,K\,\lambda\ \mathrm{Tr}\,\big(a(s,\omega)\big).$$

Then for every $p \geq d$ there exists a constant C depending only on p, d, λ and K such that the a priori parabolic estimate

$$\mathbb{E}\Big\{ \int_0^T \mathrm{e}^{-\lambda[\rho(s)+\tau(s)]\ \sqrt[d+1]{\det(a(s))\,\dot{\tau}(s)}}\,|h(\tau(s), \xi_x(s))|\mathrm{d}s \Big\} \leq$$
$$\leq C\Big[\int_0^\infty \mathrm{e}^{-\lambda(d+1)s}\mathrm{d}s \int_{\mathbb{R}^d} |h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},$$

and elliptic estimate

$$\mathbb{E}\Big\{ \int_0^T \mathrm{e}^{-\lambda\rho(s)}\ \sqrt[d]{\det(a(s))}\,\big|g(\xi_x(s))\big|\mathrm{d}s \Big\} \leq C\Big[\int_{\mathbb{R}^d} |g(y)|^p \mathrm{d}y\Big]^{1/p},$$

hold, for any x in $\mathbb{R}^d$, any stopping time T and any Borel functions $g(y)$ and $h(s,y)$.

Proof. For the sake of simplicity, we set $\varrho(s) := \rho(s) + \tau(s)$, we assume T deterministic, and we choose $h(s,y) = 0$ for $s > T$, and u as in the previous Theorem 4.66. All other cases are deduced from this particular choise. For a fixed x, apply Itô formula to the process $(\varrho(s), \tau(s), \xi(s))$ and the function $\exp(-\lambda\varrho)\, u(\tau,\xi)$, between 0 and T to get

$$\mathbb{E}\{\mathrm{e}^{-\lambda\varrho(T)}\, u(\tau(T),\xi(T))\} - u(0,x) = \mathbb{E}\{\int_0^T \mathrm{e}^{-\lambda\varrho(s)}\, L(s,\tau(s),\xi(s),u)\mathrm{d}s\},$$

where

$$\begin{aligned} L(s,\tau(s),\xi(s),u) &:= I(s,\tau(s),\xi(s),u) + \sum_i b_i(s)\partial_i u(\tau(s),\xi(s)) + \\ &+ \sum_{ij} a_{ij}(s)\partial_{ij} u(\tau(s),\xi(s)) + \\ &+ \dot\tau(s)\partial_t u(\tau(s),\xi(s)) - \lambda[\mathrm{Tr}(a(s)) + \dot\tau(s)]u(s,\xi(s)), \\ I(s,\tau(s),\xi(s),u) &:= \int_{\mathbb{R}^d_*} \big[u(\tau(s),\xi(s)+z) - u(\tau(s),\xi(s)) - \\ &- z\cdot\nabla u(\tau(s),\xi(s))\big]\mathtt{M}_\xi(\mathrm{d}z,s). \end{aligned}$$

In view of the properties (a) and (b) (or (e) later) in Theorem 4.66 we have

$$\begin{aligned} I(s,\tau(s),\xi(s),u) &= \int_0^1 (1-r)\mathrm{d}r \int_{\mathbb{R}^d_*} z_i z_j \partial_{ij} u(\tau(s),\xi(s)+rz)\mathtt{M}_\xi(\mathrm{d}z,s) \le \\ &\le \frac{\lambda}{2} \sup_{x\in\mathbb{R}^d} u(\tau(s),x) \int_{\mathbb{R}^d_*} |z|^2 \mathtt{M}_\xi(\mathrm{d}z,s), \end{aligned}$$

and

$$\Big|\sum_i b_i(s)\partial_i u(\tau(s),\xi(s))\Big| \le \sqrt{\lambda}\,|b(s)| \sup_{x\in\mathbb{R}^d} u(\tau(s),x),$$

while the estimate (d) yields

$$\begin{aligned} &\int_0^T \mathrm{e}^{-\lambda\varrho(s)} \sup_{x\in\mathbb{R}^d} u(\tau(s),x)\ \mathrm{Tr}(a(s))\mathrm{d}s \le \\ &\qquad \le \int_0^T \mathrm{e}^{-\lambda\varrho(s)}\ \mathrm{Tr}(a(s))\, q(\tau(s),h)\, \mathrm{e}^{\tau(s)\lambda(d+1)/(p+1)}\mathrm{d}s, \\ &q(s,h) := C(p,d,\lambda)\Big[\int_s^T \mathrm{e}^{-\lambda(d+1)r}\mathrm{d}r \int_{\mathbb{R}^d} |h(r,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)}, \end{aligned}$$

where $C(p,d,\lambda)$ is an in Theorem 4.66. Since $p \ge d$, we integrate by parts to

get

$$\begin{aligned}\int_0^T \mathrm{e}^{-\lambda\varrho(s)} \sup_{x\in\mathbb{R}^d} u(\tau(s),x)\ \mathrm{Tr}(a(s))\mathrm{d}s \leq \\ \leq \int_0^T \exp\Big(-\lambda\int_0^s \mathrm{Tr}(a(t))\mathrm{d}t\Big)\ \mathrm{Tr}(a(s))\, q(\tau(s),h)\mathrm{d}s = \\ = -\frac{1}{\lambda}\int_0^T q(\tau(s),h)\,\mathrm{d}\exp\Big(-\lambda\int_0^s \mathrm{Tr}(a(t))\mathrm{d}t\Big) \leq \\ \leq \frac{1}{\lambda}\Big[q(0,h)+\int_0^T \exp\Big(-\lambda\int_0^s \mathrm{Tr}(a(t))\mathrm{d}t\Big)\mathrm{d}q(\tau(s),h)\Big] \leq \\ \leq \frac{1}{\lambda}\, q(0,h).\end{aligned}$$

Hence, the assumption on the coefficients b and M implies

$$\begin{aligned}\int_0^T \mathrm{e}^{-\lambda\varrho(s)}\big[I(s,\tau(s),\xi(s),u)+\big|b(s)\cdot\nabla u(\tau(s),\xi(s))\big|\big]\mathrm{d}s \leq \\ \leq K\, C(p,d,\lambda)\Big[\int_0^T \mathrm{e}^{-\lambda(d+1)s}\mathrm{d}s\int_{\mathbb{R}^d}|h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)}.\end{aligned}$$

By means of the property (c) of the function u in Theorem 4.66, we also have

$$\begin{aligned}\sum_{ij} a_{ij}(s)\partial_{ij}u(\tau(s),\xi(s))+\dot\tau(s)\,\partial_t u(\tau(s),\xi(s))- \\ -\lambda[\mathrm{Tr}(a(s))+\dot\tau(s)]\,u(\tau(s),\xi(s)) \leq \\ \leq -\sqrt[d+1]{\det(a(s))\,\dot\tau(s)}\,h_\varepsilon(\tau(s),\xi(s)),\end{aligned}$$

so that Itô formula yields

$$\begin{aligned}\mathbb{E}\Big\{\int_0^T \mathrm{e}^{-\lambda\varrho(s)}\ \sqrt[d+1]{\det(a(s))\,\dot\tau(s)}\,h_\varepsilon(\tau(s),\xi(s))\mathrm{d}s\Big\} \leq \\ \leq (1+K)\,C(p,d,\lambda)\Big[\int_0^T \mathrm{e}^{-\lambda(d+1)s}\mathrm{d}s\int_{\mathbb{R}^d}|h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},\end{aligned}$$

after using again the estimate (d) on $u(0,x)$. Then, as ε vanishes we obtain the parabolic estimate with $C := (1+K)\,C(p,d,\lambda)$,

$$C(p,d,\lambda) := (p+1)^{d/(p+1)}[d!\,\omega_d]^{-1/(p+1)}\lambda^{(d-2p)/(2p+2)}(d+1)^{-p/(p+1)},$$

where ω_d is the volume of the unit sphere in $\mathbb{R}^d$. Moreover, if we set

$$\rho(t,s) := \int_t^s \mathrm{Tr}\big(a(r)\big)\mathrm{d}r, \qquad \tau(t,s) := \int_t^s \dot\tau(r)\mathrm{d}r,$$

for every $s \geq t$, then we also have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \displaystyle\int_t^T \mathrm{e}^{-\lambda[\rho(t,s)+\tau(t,s)]} \ \sqrt[d+1]{\det(a(s))\,\dot\tau(s)} \ \times \\ \qquad \times h(\tau(t,s),\xi(s))\mathrm{d}s\Big\} \leq (1+K)\,C(p,d,\lambda) \times \\ \qquad \times \Big[\displaystyle\int_t^\infty \mathrm{e}^{-\lambda(d+1)(s-t)}\mathrm{d}s \int_{\mathbb{R}^d} |h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)}. \end{array} \right. \tag{4.126}$$

for every stopping time $T \geq t \geq 0$.

To prove the a priori elliptic estimate we take a particular process $\dot\tau(s)$ as follows. Since $\sqrt[d]{\det(a(s))} \leq \mathrm{Tr}(a(s))$ we have

$$\sup_{(t,\omega)} \mathbb{E}\Big\{ \int_t^T \mathrm{e}^{-\lambda\rho(t,s)} \ \sqrt[d]{\det(a(s))}\,\big|g(\xi(s))\big|\mathrm{d}s \,|\, \mathcal{F}(t)\Big\} \leq \frac{\sup_x |g(x)|}{\lambda}.$$

so that, as long as g is bounded, we may set

$$\begin{array}{l} c := \displaystyle\sup_{(t,\omega)} \mathbb{E}\Big\{ \int_t^T \mathrm{e}^{-\lambda\rho(t,s)} \ \sqrt[d]{\det(a(s))}\,\big|g(\xi(s))\big|\mathrm{d}s \,|\, \mathcal{F}(t)\Big\}, \\ \dot\tau(s) := h(s)/c, \quad \text{and} \quad h(s) := \sqrt[d]{\det(a(s))}\,\big|g(\xi(s))\big|, \end{array}$$

to show that an integration by parts yields

$$\begin{array}{l} \displaystyle\int_t^T h(s)\,\mathrm{e}^{-\lambda\rho(t,s)}\mathrm{d}s = \int_t^T h(s)\,\mathrm{e}^{-\lambda[\rho(t,s)+2\tau(t,s)]}\mathrm{d}s + \\ \qquad\qquad +2\lambda \displaystyle\int_t^T \Big(\int_t^T \mathrm{e}^{-\lambda\rho(t,s)}\,h(s)\mathrm{d}s\Big)\mathrm{e}^{-2\lambda\tau(t,r)}\,\dot\tau(r)\mathrm{d}r. \end{array}$$

After taking the conditional mathematical expectation, the last term is equal to

$$\begin{array}{l} \mathbb{E}\Big\{ \displaystyle\int_t^T \Big(\int_t^T \mathrm{e}^{-\lambda\rho(t,s)}\,h(s)\mathrm{d}s \,|\, \mathcal{F}(t)\Big)\mathrm{e}^{-\lambda[\rho(t,r)+2\tau(t,r)]}\,\dot\tau(r)\mathrm{d}r\Big\} \leq \\ \qquad \leq \mathbb{E}\Big\{ \displaystyle\int_t^T \mathrm{e}^{-\lambda[\rho(t,r)+2\tau(t,r)]}\,h(r)\mathrm{d}r\Big\}, \end{array}$$

where we have used the definition of the constant c and the process $\dot\tau(s)$ to obtain the inequality. Hence

$$\begin{array}{l} \mathbb{E}\Big\{ \displaystyle\int_t^T \mathrm{e}^{-\lambda\rho(t,s)} \ \sqrt[d]{\det(a(s))}\,\big|g(\xi(s))\big|\mathrm{d}s \,|\, \mathcal{F}(t)\Big\} = \\ \qquad = \mathbb{E}\Big\{ \displaystyle\int_t^T h(s)\,\mathrm{e}^{-\lambda\rho(t,s)}\mathrm{d}s \,|\, \mathcal{F}(t)\Big\} \leq \\ \qquad \leq (1+2\lambda)\,\mathbb{E}\Big\{ \displaystyle\int_t^T h(s)\,\mathrm{e}^{-\lambda[\rho(t,s)+2\tau(t,s)]}\mathrm{d}s \,|\, \mathcal{F}(t)\Big\}, \end{array}$$

and this last term is equal to

$$\sqrt[d+1]{c}\,\mathbb{E}\Big\{ \int_t^T \mathrm{e}^{-\lambda[\rho(t,s)+\tau(t,s)]} \ \sqrt[d+1]{\det(a(s))\,\dot\tau(s)}\,f(\tau(t,s),\xi(s))\mathrm{d}s \,|\, \mathcal{F}(t)\Big\},$$

with

$$f(t,x) := (1+2\lambda)\,\mathrm{e}^{-\lambda t}\,\big|g(x)\big|^{d/(d+1)}.$$

Now, by means of the parabolic estimate we obtain

$$\mathbb{E}\Big\{\int_t^T \mathrm{e}^{-\lambda\rho(t,s)}\,\sqrt[d]{\det(a(s))}\,\big|g(\xi(s))\big|\mathrm{d}s\,|\,\mathcal{F}(t)\Big\} \leq \\ \leq (1+K)C'\,c^{1/(d+1)}\Big[\int_{\mathbb{R}^d}|g(x)|^{(p+1)d/(d+1)}\mathrm{d}x\Big]^{1/(p+1)},$$

for every $p \geq d$ and some constant C' depending only on p, d and λ). After taking the supremum in (t,ω), we deduce the elliptic estimate with

$$C = \Big[(1+K)C(p,d,\lambda)\frac{1+2\lambda}{\lambda(d+1-p)}\Big]^{1+1/d},$$

and $C(p,d,\lambda)$ as above. This completes the proof. □

The above prove is a modification of Anulova and Pragarauskas [5]. Note that if the non-local integral I is written as (to simplify we are taking $\dot\tau = 1$)

$$I(s,\xi(s),u) = \int_0^1 (1-r)\mathrm{d}r\int_{|z|<\varepsilon} z_i z_j \partial_{ij}u(s,\xi(s)+rz)\mathtt{M}_\xi(\mathrm{d}z,s) + \\ + \int_{|z|\geq\varepsilon}\big[u(s,\xi(s)+z)-u(s,\xi(s))-z\cdot\nabla u(s,\xi(s))\big]\mathtt{M}_\xi(\mathrm{d}z,s) \leq \\ \leq \Big[\frac{\lambda}{2}\int_{|z|<\varepsilon}|z|^2\mathtt{M}_\xi(\mathrm{d}z,s) + \int_{|z|\geq\varepsilon}\mathtt{M}_\xi(\mathrm{d}z,s)\Big]\sup_{x\in\mathbb{R}^d}u(s,x),$$

which allows us to replace the condition on the coefficients b and $\mathtt{M}$ with the assumption

$$\int_{|z|\geq\varepsilon}\mathtt{M}_\xi(\mathrm{d}z,s,\omega) + 2\sqrt{\lambda}\,|b(s,\omega)| + \lambda\int_{|z|<\varepsilon}|z|^2\mathtt{M}_\xi(\mathrm{d}z,s,\omega) \leq \\ \leq 2\,K\,\lambda\,\mathrm{Tr}\,\big(a(s)\big),$$

for some positive constants ε, λ and K, to get the same a priori estimate of Theorem 4.67.

An important consequence is the following estimates

Corollary 4.68. *Let $\xi = \xi_x$ be an Itô process with jumps in $\mathbb{R}^d$ as in* Theorem 4.67 *with bounded coefficients and non-degenerate diffusion, i.e,*

$$\begin{cases} |b(s,\omega)|^2 + \sum_k |\sigma_k(s,\omega)|^2 + \int_{\mathbb{R}^d_*}|z|^2\mathtt{M}_\xi(\mathrm{d}z,s,\omega) \leq \dfrac{1}{\delta}, \\ \sum_{ik}|\zeta_i\,\sigma_{ik}(s,\omega)|^2 \geq \delta|\zeta|^2, \end{cases}$$

for some $\delta > 0$ and for every ζ in $\mathbb{R}^d$, and s,ω. Then for every $p \geq d$, x in $\mathbb{R}^d$, $T \geq t \geq 0$, and any Borel function $h(s,y)$

$$\mathbb{E}\Big\{\int_t^T |h(s,\xi_x(s))|\mathrm{d}s\,\Big|\,\mathcal{F}(t)\Big\} \leq K_T\Big[\int_{(t,T)\times\mathbb{R}^d}|h(s,y)|^{p+1}\,\mathrm{d}s\,\mathrm{d}y\Big]^{1/(p+1)},$$

for some constant $K_T = K_T(d,p,\delta)$ depending only on δ, p, the dimension d and T. Moreover, for any stopping time $\theta \geq t$ and any Borel measurable function $g(y)$ we have

$$\mathbb{E}\Big\{ \int_t^\theta \exp\Big(-\lambda \int_t^s \big[\mathrm{Tr}(a(r)) + 1 \big] \mathrm{d}r \Big) |h(s,\xi_x(s))| \mathrm{d}s \;\Big|\; \mathcal{F}(t) \Big\} \leq \\ \leq K_\lambda \Big[\int_{(t,\infty)\times\mathbb{R}^d} \mathrm{e}^{-\lambda(d+1)(s-t)} \, |h(s,y)|^{p+1} \, \mathrm{d}s \, \mathrm{d}y \Big]^{1/(p+1)},$$

and

$$\mathbb{E}\Big\{ \int_t^\theta \mathrm{e}^{-\lambda(s-t)} \, |g(\xi_x(s))| \mathrm{d}s \;\Big|\; \mathcal{F}(t) \Big\} \leq K_\lambda \Big[\int_{\mathbb{R}^d} |g(y)|^p \, \mathrm{d}y \Big]^{1/p},$$

where $K_\lambda = K_\lambda(d,p,\delta)$ depending only on δ, p, the dimension d and λ. □

Note that if the stopping time $\theta = \theta_D$ is taken to be the first exit time of the process $\xi(t)$ from a bounded domain D then we can take $g(y) = 0$ outside of D to get

$$\mathbb{E}\Big\{ \int_0^{\theta_D} \mathrm{e}^{-\lambda s} \, |g(\xi(s))| \mathrm{d}s \Big\} \leq K_\lambda \Big[\int_D |g(y)|^p \, \mathrm{d}y \Big]^{1/p}.$$

To be able to take $\lambda = 0$ we need to use *barrier* functions of the type $\beta - \cosh(\alpha|x|)$ for some $\alpha > 0$, $\beta > \cosh(\alpha R)$ and $D \subset \{x \in \mathbb{R}^d \,:\, |x| \leq R\}$. The arguments in Krylov [139, Sections 2.2, pages 51–61] can be extended to show the validity of the estimates in the above Corollary for $\theta = \theta_D$ and $\lambda = 0$, as long as D is bounded and we suppose

$$\lim_{\varepsilon\to 0} \int_{|z|<\varepsilon} |z|^2 \mathtt{M}_\xi(\mathrm{d}z, s, \omega) = 0,$$

uniformly in (s,ω), besides the assumptions of Corollary 4.68 on the drift coefficients b, the diffusion σ and Lévy kernel $\mathtt{M}$, i.e., Theorem 4.65 for the Itô process with jumps $\xi(\cdot)$.

The relevance of the above a priori estimates is clear, because of the density of smooth functions into the Lebesgue spaces L^p, each time we have to deal with a limit involving an expectation with respect to an Itô process with jumps, e.g., an expression of the type

$$\mathbb{E}\Big\{ \int_0^T |h(s,\xi(s))| \mathrm{d}s \Big\},$$

we replace the Borel measurable function $h(s,y)$ with a smooth function $h_\varepsilon(s,y)$ (e.g., a mollification of h), we pass to the limit with $h_\varepsilon(s,y)$, and as long as $h_\varepsilon \to h$ in L^p for some $p \geq d$, we obtain the limit with h. In particular, Itô formula remains valid for C^1 functions with second derivatives (locally) in $L^{d+1}(]0,\infty[\times\mathbb{R}^d)$.

Theorem 4.69. *Let $(\xi_x^n(t) : t \geq 0,\, n \geq 1)$ be a sequence of Itô processes with jumps of the form (4.125), i.e.,*

$$\xi_x^n(t) = x + \int_0^t b^n(s)\mathrm{d}s + \sum_k \int_0^t \sigma_k^n(s)\mathrm{d}w_k(s) + \int_{\mathbb{R}_*^d \times]0,t]} z\tilde{\nu}_\xi^n(\mathrm{d}z, \mathrm{d}s),$$

with intensity kernel $\mathtt{M}_\xi^n(B, s)$. Assume that the probability law P_x^n on the canonical space $D([0,\infty), \mathbb{R}^d)$, induced by the processes $\xi_x^n(\cdot)$, converges weakly to a probability measure P_x, and that the first exit time τ_r^n from the closed ball $\{y \in \mathbb{R}^d \;:\; |y| \leq r\}$ of the process $\xi_x^n(\cdot)$ satisfies the condition: there exist constants C_r and c_r such that except in a set of probability zero and for any r sufficiently large, we have

$$\begin{cases} |b^n(s)|^2 + \sum_k |\sigma_k^n(s)|^2 + \int_{\mathbb{R}_*^d} |z|^2 \mathtt{M}_\xi^n(\mathrm{d}z, s) \leq C_r, \\ \det\Big(\sum_k \sigma_k^n(s)(\sigma_k^n(s))^*\Big) \geq c_r > 0, \end{cases}$$

for every $n \geq 1$ and for every $0 \leq s \leq \tau_r^n$. Suppose there is a equi-bounded sequence $\{\varphi_n : n \geq 1\}$ of real-valued measurable functions in $\mathbb{R}^d$ converging to some function φ in $L^d_{\mathrm{loc}}(\mathbb{R}^d)$, i.e.,

$$\lim_n \int_{\{y\,:\,|y|\leq r\}} |\varphi_n(y) - \varphi(y)|^d \,\mathrm{d}y = 0,$$

for every $r > 0$. Then

$$\lim_n \mathbb{E}\Big\{ \int_t^T \varphi_n\big(\xi_x^n(s)\big)\mathrm{d}s \;\Big|\; \mathcal{F}(t)\Big\} = \mathbb{E}_x\Big\{ \int_0^T \varphi\big(\omega(s)\big)\mathrm{d}s \;\Big|\; \mathcal{F}(t)\Big\},$$

for every x in $\mathbb{R}^d$ and $T > t \geq 0$, where $\mathbb{E}_x$ denotes the mathematical expectation with respect to the probability measure P_x. Similarly, if $\{\psi_n(s,y) : n \geq 1\}$ is an equi-bounded sequence of measurable functions which converges to $\psi(s,y)$ in $L^{d+1}_{\mathrm{loc}}(]0,\infty[\times\mathbb{R}^d)$, then

$$\lim_n \mathbb{E}\Big\{ \int_t^T \psi_n\big(s, \xi_x^n(s)\big)\mathrm{d}s \;\Big|\; \mathcal{F}(t)\Big\} = \mathbb{E}_x\Big\{ \int_t^T \psi\big(s, \omega(s)\big)\mathrm{d}s \;\Big|\; \mathcal{F}(t)\Big\},$$

for every x in $\mathbb{R}^d$ and $T > t \geq 0$.

Proof. Only the case with $\varphi(y)$ is considered, since the same arguments are applicable for the case $\psi(s,y)$. Moreover, we may take $t = 0$ without loss of generality.

If $\mathbb{E}_x^n$ denotes the mathematical expectation with respect to the probability measure P_x^n, then we have

$$\mathbb{E}\Big\{ \int_0^T \varphi_n\big(\xi_x^n(s)\big)\mathrm{d}s\Big\} = \mathbb{E}_x^n\Big\{ \int_0^T \varphi_n\big(\omega(s)\big)\mathrm{d}s\Big\}.$$

Since P_x^n converges weakly to P_x, we deduce

$$\lim_{r\to\infty} P\{\inf_n \tau_r^n \geq T\} = 0,$$

for every $T \geq 0$. Hence, because there exists a constant $C > 0$ such that $|\varphi_n(y)| \leq C$, for every y in $\mathbb{R}^d$ and $n \geq 1$, it suffices to show

$$\lim_n \mathbb{E}\Big\{ \int_0^{\tau_r^n \wedge T} \varphi_n\big(\xi_x^n(s)\big)\mathrm{d}s \Big\} = \mathbb{E}_x\Big\{ \int_0^{\tau_r \wedge T} \varphi\big(\omega(s)\big)\mathrm{d}s \Big\},$$

for every $r > 0$, to establish the result. Note that τ_r is the first exit time from the closed ball $\{y \in \mathbb{R}^d \,:\, |y| \leq r\}$ of the canonical process $x(t) := \omega(t)$.

In view of the previous Corollary 4.68 and because we may take $\varphi_n(y) = 0$ if $|y| > r$, for each $T > 0$ and $r > 0$ there exists a constant $K = K(T, r, d)$ such that

$$\begin{cases} \mathbb{E}_x^n\Big\{ \displaystyle\int_0^{\tau_r \wedge T} \big|\varphi_n\big(\omega(s)\big) - \tilde\varphi_m\big(\omega(s)\big)\big|\mathrm{d}s \Big\} \leq \\ \qquad\qquad \leq K\Big[\displaystyle\int_{\{y\,:\,|y|\leq r\}} |\varphi_n(y) - \tilde\varphi_m(y)|^d \,\mathrm{d}y \Big]^{1/d}, \end{cases} \tag{4.127}$$

for every $m \geq 1$ and any sequence of functions $\tilde\varphi_m$. In particular, we may choose continuous and equi-bounded functions $\tilde\varphi_m$ such that

$$\lim_m \varepsilon(m, r) = 0,$$
$$\varepsilon(m, r) := \sup_{n \geq m} \Big[\int_{\{y\,:\,|y|\leq r\}} |\varphi_n(y) - \tilde\varphi_m(y)|^d \,\mathrm{d}y \Big]^{1/d},$$

for every $r > 0$. Since $\tilde\varphi_n$ also converges to φ in $L^d_{\mathrm{loc}}(\mathbb{R}^d)$, we have

$$\mathbb{E}_x^n\Big\{ \int_0^{\tau_r \wedge T} \big|\tilde\varphi_m\big(\omega(s)\big) - \varphi\big(\omega(s)\big)\big|\,\mathrm{d}s \Big\} \leq K\varepsilon(m, r),$$

for every $n, m \geq 1$, $T > 0$ and $r > 0$.

On the other hand, for every continuous and bounded function g we have

$$\lim_n \mathbb{E}_x^n\Big\{ \int_0^{\tau_r \wedge T} g\big(\omega(s)\big)\mathrm{d}s \Big\} = \mathbb{E}_x\Big\{ \int_0^{\tau_r \wedge T} g\big(\omega(s)\big)\mathrm{d}s \Big\},$$

which proves that estimate (4.127) holds for the limiting $\mathbb{E}_x\{\cdot\}$ in lieu of $\mathbb{E}_x^n\{\cdot\}$ and g instead of $\varphi_n - \tilde\varphi_m$. A posteriori, this extends to any measurable function g in $L^d(\mathbb{R}^d)$ and we have

$$\mathbb{E}_x\Big\{ \int_0^{\tau_r \wedge T} \big|\varphi\big(\omega(s)\big) - \tilde\varphi_m\big(\omega(s)\big)\big|\mathrm{d}s \Big\} \leq K\varepsilon(m, r),$$

for every $m \geq 1$, $T > 0$, $r > 0$, and the same constant $K = K(T, r, d)$ as in (4.127).

Since $\tilde\varphi_m$ is continuous and bounded, we have

$$\lim_n \mathbb{E}_x^n\Big\{ \int_0^{\tau_r \wedge T} \tilde\varphi_m\big(\omega(s)\big)\mathrm{d}s \Big\} = \mathbb{E}_x\Big\{ \int_0^{\tau_r \wedge T} \tilde\varphi_m\big(\omega(s)\big)\mathrm{d}s \Big\},$$

for every $m \geq 1$. Collecting all, we get

$$\lim_n \Big|\mathbb{E}\Big\{\int_0^T \varphi_n\big(\xi_x^n(s)\big)\mathrm{d}s\Big\} - \mathbb{E}_x\Big\{\int_0^T \varphi\big(\omega(s)\big)\mathrm{d}s\Big\}\Big| \leq$$
$$\leq 2K(T,r,d)\,\varepsilon(m,r) + 2\Big(\sup_{y\in\mathbb{R}^d}|\varphi(y)|\Big)\,P\{\inf_n \tau_r^n \geq T\},$$

and, as $m \to \infty$ and then $r \to \infty$, we conclude. □

It is clear that we can replace the equi-bounded condition on the sequence of functions with a uniform bounded assumption on the coefficients b^n, σ_k^n and $\mathtt{M}_\xi^n$. In this case, the stopping time functional τ_r is not needed.

Note that all conditions are actually given in term of the diffusion matrix

$$a := \frac{1}{2}\sum_k \sigma_k\,\sigma_k^*,$$

and not directly on the diffusion coefficients σ_k.

4.5.2 With Bounded Variation Processes

Now, consider an stochastic integral process (with jumps) in $\mathbb{R}^d$ of the form

$$\begin{cases} \xi(t) = x + \displaystyle\sum_k \int_0^t \beta_k(s)\mathrm{d}v_k(s) + \sum_k \int_0^t \sigma_k(s)\mathrm{d}w_k(s) + \\ \qquad\qquad + \displaystyle\sum_k \int_0^t \varsigma_k(s)\mathrm{d}w_k^\vartheta(s) + \int_{\mathbb{R}_*^d\times]0,t]} z\tilde{\nu}_\xi(\mathrm{d}z,\mathrm{d}s), \end{cases} \tag{4.128}$$

for every $t \geq 0$, in some filtered probability space $(\Omega, F, P, \mathcal{F}(t) : t \geq 0)$, where (v_k) is a continuous nondecreasing adapted process with $v_k(0) = 0$, (w_k) is a (standard) Wiener process, (w_k^ϑ) is a continuous square integrable martingale with predictable quadratic covariation $\langle w_h^\vartheta, w_k^\vartheta\rangle = \vartheta_k$ if $h = k$, $\langle w_h^\vartheta, w_k^\vartheta\rangle = 0$ if $h \neq k$, and $\langle w_h, w_k^\vartheta\rangle = 0$ for every h, k (so that ϑ_k is a continuous nondecreasing adapted process with $\vartheta_k(0) = 0$), and $\tilde{\nu}_\xi$ is the local martingale measure associated with the integer valued measure ν_ξ (with predictable compensator ν_ξ^p), corresponding to the quasi-left continuous process ξ. The compensator ν_ξ^p is assumed to have the form

$$\nu_\xi^p(B,(0,t]) := \int_0^t \mathtt{M}_\xi(B,s)\mathrm{d}\kappa(s),$$

where the intensity kernel $\mathtt{M}$ satisfies: (1) $s \mapsto \mathtt{M}_\xi(B,s)$ is predictable for every B, (2) $B \mapsto \mathtt{M}_\xi(B,s)$ is a measure in $\mathbb{R}_*^d$ for every s, and (3) $s \mapsto \kappa(s)$ is a continuous nondecreasing adapted process with $\kappa(0) = 0$. The coefficients $\beta_k(s)$, $\sigma_k(s)$ and $\varsigma_k(s)$ are predictable processes. Also, for a given predictable nonnegative process $\dot{\tau}(s)$ define the following (integral) process

$$\tau(s) := \int_0^s \dot{\tau}(r)\mathrm{d}r,$$

and also the processes

$$a(s) := \big(\frac{1}{2}\sum_k \sigma_{ik}(s)\sigma_{jk}(s)\big), \qquad \rho(s) := \int_0^s \mathrm{Tr}(a(r))\mathrm{d}r,$$

where $\rho(\cdot)$ has nonnegative real values and $a(\cdot)$ has symmetric nonnegative $d\times d$ matrix values. Note that

$$\mathrm{Tr}(a(s)) = \frac{1}{2}\sum_{ik}|\sigma_{ik}(s)|^2, \qquad \frac{\mathrm{d}\rho}{\mathrm{d}s}(s) = \dot\rho(s) = \mathrm{Tr}(a(s))$$

and usually $\dot\tau = 1$. Now we have

Theorem 4.70. *Let $\xi(\cdot) = \xi_x(\cdot)$ be a stochastic integral process with jumps as in (4.128) in $\mathbb{R}^d$ and $\dot\tau$ be a predictable nonnegative processes. Define the processes $\tau(\cdot)$, $\rho(\cdot)$ and $a(\cdot)$ as above and suppose that for some positive constants $K = K_\xi$, and $\lambda > 0$, and any s, ω we have*

$$2\sqrt{\lambda}\sum_k|\beta_k(s,\omega)| + \lambda\int_{\mathbb{R}^d_*}|z|^2\mathtt{M}_\xi(\mathrm{d}z,s,\omega) + \lambda\sum_{ik}|\varsigma_{ik}(s,\omega|^2 \le$$
$$\le K\,\lambda\sum_{ik}|\sigma_{ik}(s,\omega)|^2,$$

Then for every $p \ge d$ there exists a constant C depending only on p, d, λ and K such that, with

$$C_{\kappa,\vartheta,v} := C\,\mathbb{E}\Big\{\lambda\int_0^T \mathrm{e}^{-\lambda\rho(s)}\ \mathrm{Tr}\big(a(s)\big)\mathrm{d}\Big[\kappa(s) + \sum_k \vartheta_k(s) + \sum_k v_k(s)\Big]\Big\},$$

the a priori parabolic estimate

$$\mathbb{E}\Big\{\int_0^T \mathrm{e}^{-\lambda[\rho(s)+\tau(s)]\ \frac{d+1}{}\sqrt{\det(a(s))\,\dot\tau(s)}}\,|h(\tau(s),\xi_x(s))|\mathrm{d}s\Big\} \le$$
$$\le C_{\kappa,\vartheta,v}\Big[\int_0^\infty \mathrm{e}^{-\lambda(d+1)s}\mathrm{d}s\int_{\mathbb{R}^d}|h(s,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},$$

and elliptic estimate

$$\mathbb{E}\Big\{\int_0^T \mathrm{e}^{-\lambda\rho(s)}\ \sqrt[d]{\det(a(s))}\,\big|g(\xi_x(s))\big|\mathrm{d}s\Big\} \le C_{\kappa,\vartheta,v}\Big[\int_{\mathbb{R}^d}|g(y)|^p\mathrm{d}y\Big]^{1/p},$$

hold, for any x in $\mathbb{R}^d$, any stopping time T and any Borel functions $g(y)$ and $h(s,y)$.

Proof. First notice that we need to consider only the case when the constant $C_{\kappa,\vartheta,v}$ is finite. Reviewing the proof in Theorem 4.67, we see that the first point is to realize that by means of Itô formula new terms appear, where $\mathrm{d}s$ is replaced by $\mathrm{d}\kappa(s)$, $\mathrm{d}\vartheta_k(s)$ and $\mathrm{d}v_k(s)$. Thus, the key point to discuss is how to estimate an expression of the form

$$R_u := \int_0^T \mathrm{e}^{-\lambda\varrho(s)}\big[\sup_{x\in\mathbb{R}^d} u(\tau(s),x)\big]\,\mathrm{Tr}(a(s))\mathrm{d}\alpha(s),$$

where $\alpha(s)$ is a continuous nondecreasing adapted process. To this purpose, we have

$$R_u \le \int_0^T \mathrm{e}^{-\lambda\varrho(s)}\ \mathrm{Tr}(a(s))\, q(\tau(s),h)\, \mathrm{e}^{\tau(s)\lambda(d+1)/(p+1)}\mathrm{d}\vartheta(s),$$
$$q(s,h) := C(p,d,\lambda)\Big[\int_s^T \mathrm{e}^{-\lambda(d+1)r}\mathrm{d}r \int_{\mathbb{R}^d} |h(r,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},$$

where $C(p,d,\lambda)$ is an in Theorem 4.66. Since $p \ge d$, define

$$p(t) := \int_t^T \exp\Big(-\lambda\int_0^s \mathrm{Tr}(a(t))\mathrm{d}t\Big)\,\mathrm{Tr}(a(s))\mathrm{d}\alpha(s),$$

to integrate by parts and to get

$$R_u \le \int_0^T \exp\Big(-\lambda\int_0^s \mathrm{Tr}(a(t))\mathrm{d}t\Big)\ \mathrm{Tr}(a(s))\, q(\tau(s),h)\mathrm{d}s =$$
$$= -\int_0^T q(\tau(s),h)\,\mathrm{d}p(s) \le p(0)\, q(0,h).$$

Thus

$$\mathbb{E}\Big\{\int_0^T \mathrm{e}^{-\lambda\varrho(s)}\big[\sup_{x\in\mathbb{R}^d} u(\tau(s),x)\big]\,\mathrm{Tr}\,\big(a(s)\big)\mathrm{d}\alpha(s)\Big\} \le$$
$$\le C_\alpha(p,d,\lambda)\Big[\int_s^T \mathrm{e}^{-\lambda(d+1)r}\mathrm{d}r\int_{\mathbb{R}^d}|h(r,y)|^{p+1}\mathrm{d}y\Big]^{1/(p+1)},$$

where

$$C_\alpha(p,d,\lambda) := C(p,d,\lambda)\,\mathbb{E}\Big\{\int_0^T \mathrm{e}^{-\lambda\rho(s)}\ \mathrm{Tr}\,\big(a(s)\big)\mathrm{d}\alpha(s)\Big\}.$$

Notice that if $\alpha(s) = s$ then $\lambda\, C_\alpha(p,d,\lambda) \le C(p,d,\lambda)$. This estimate and the property (b) of Theorem 4.66, namely

$$\sum_{ijk}\varsigma_{ik}(s)\,\varsigma_{jk}(s)\,\partial_{ij}u(t,x) \le \lambda\sum_{ik}|\varsigma_{ik}(s)|^2\,u(t,x),$$

allows us to bounds the extra terms due to the new form (with respect to Theorem 4.67) of the stochastic integral process. Hence, the parabolic estimate is established, which yields the elliptic estimate. □

The interest of the above parabolic and elliptic estimates is limited without more conditions on the processes κ, ϑ and v. Typically, the process ξ belongs to some domain $\mathcal{O}$ with a boundary $\partial\mathcal{O}$ where an instantaneous reflection is made, i.e., the nondecreasing processes satisfy

$$\mathtt{M}_\xi(\mathrm{d}z,s) = \mathbb{1}_{\bar{\mathcal{O}}}(\xi(s-)+z)\,\mathtt{M}_\xi(\mathrm{d}z,s),$$
$$\mathrm{d}\ell(s) = \mathbb{1}_{\partial\mathcal{O}}(\xi(s))\mathrm{d}\ell(s),$$
$$\sum_k \beta_k(s)\mathrm{d}v_k = \mathbb{1}_{\mathcal{O}}(\xi(s))\,b(s)\,\mathrm{d}s + \mathbb{1}_{\partial\mathcal{O}}(\xi(s))\,c(s)\,\mathrm{d}\ell(s),$$

for every $s \geq 0$, where the process ℓ is the *local time* on the boundary, i.e., $\kappa(s) = t$ and $\vartheta_k(s) = 0$. A natural assumption is to suppose that there is a smooth (positive on $\mathcal{O}$) function $\varphi = \varphi_{\mathcal{O}}(x)$ with bounded second derivatives such that

$$\mathbb{1}_{\mathcal{O}}(\xi(s-))\big[A_\xi(s)\varphi(\xi(s-))\big] \leq \lambda\,\varphi(\xi(s-))\sum_{ik}|\sigma_{ik}(s)|^2,$$
$$\mathbb{1}_{\partial\mathcal{O}}(\xi(s-))\big[B_\xi(s)\varphi(\xi(s-))\big] \geq c_0 > 0,$$

for every $s > 0$, where

$$A_\xi(s)\varphi(x) := \sum_i b_i(s)\partial_i\varphi(x) + \frac{1}{2}\sum_{ijk}\sigma_{ik}(s)\sigma_{jk}(s)\partial_{ij}\varphi(x) + \\ + \int_{\mathbb{R}^d_*}\big[\varphi(x+z) - \varphi(x) - z\sum_i \partial_i\varphi(x)\big]\mathtt{M}_\xi(\mathrm{d}z, s),$$
$$B_\xi(s)\varphi(x) := \sum_i c_i(s)\partial_i\varphi(x),$$

are defined for any $s \geq 0$ and x in $\mathbb{R}^d$. Thus, by applying Itô formula to the function $(s,x) \mapsto \exp(-\lambda\rho(s))\,\varphi(x)$ we get

$$\mathbb{E}\Big\{\lambda\int_t^T \mathrm{e}^{-\lambda\rho(s)}\ \mathrm{Tr}\,\big(a(s)\big)\mathrm{d}\ell(s)\Big\} \leq \frac{1}{c_0}\,\mathbb{E}\big\{\varphi(\xi_t)\big\}, \tag{4.129}$$

for every $T > t \geq 0$, which complement the estimates in Theorem 4.70. The technique of Remark 4.35 helps us to establish estimates on higher moments of the local time ℓ. Usually, $c(s)$ points in the normal direction at $\xi(s-)$ and the function φ is a smooth extension of the function $x \mapsto d(x)$, where $d(x)$ is the distance to the boundary $\partial\mathcal{O}$.

The above Theorem 4.70 includes an stochastic integral process (with jumps) in $\mathbb{R}^d$ of the form

$$\begin{cases} \xi(t) = x + \displaystyle\int_0^t b(s)\mathrm{d}s + \sum_k\int_0^t \sigma_k(s)\mathrm{d}w_k(s) + \int_0^t c(s)\mathrm{d}\ell(s) + \\ \qquad\qquad + \displaystyle\sum_k\int_0^t \varsigma_k(s)\mathrm{d}w^\ell_k(s) + \int_{\mathbb{R}^d_*\times]0,t]} z\tilde{\nu}_\xi(\mathrm{d}z,\mathrm{d}s), \end{cases} \tag{4.130}$$

for every $t \geq 0$, in some filtered probability space $(\Omega, F, P, \mathcal{F}(t) : t \geq 0)$, where ℓ is a continuous nondecreasing adapted process with $\ell(0) = 0$, (w_k) is a (standard) Wiener process, (w^ℓ_k) is a continuous square integrable martingale with predictable quadratic covariation $\langle w^\ell_h, w^\ell_k\rangle = \ell$ if $h = k$, $\langle w^\ell_h, w^\ell_k\rangle = 0$ if $h \neq k$ and $\langle w_h, w^\ell_k\rangle = 0$ for every h, k, and $\tilde{\nu}_\xi$ is the local martingale measure associated with the integer valued measure ν_ξ (with predictable compensator ν^p_ξ), corresponding to the quasi-left continuous process ξ. The compensator ν^p_ξ is assumed to have the form

$$\nu^p_\xi(B,(0,t]) := \int_0^t \mathtt{M}_\xi(B,s)\mathrm{d}s + \int_0^t \mathtt{N}_\xi(B,s)\mathrm{d}\ell(s),$$

for every $B \in \mathcal{B}(\mathbb{R}^d_*)$, where the intensity kernels $\mathtt{M}$ and $\mathtt{N}$ satisfies: (1) $s \mapsto \mathtt{M}_\xi(B,s)$ and $s \mapsto \mathtt{N}_\xi(B,s)$ are predictable for every B and (2) $B \mapsto \mathtt{M}_\xi(B,s)$ and $B \mapsto \mathtt{N}_\xi(B,s)$ are measures in $\mathbb{R}^d_*$ for every s. The coefficients $b(s)$, $c(s)$, $\sigma(s)$ and $\varsigma(s)$ are predictable processes. In this case we replace $C_{\kappa,\ell,v}$ with

$$C_\ell := C + C\,\mathbb{E}\Big\{\lambda \int_0^T \mathrm{e}^{-\lambda\rho(s)}\ \mathrm{Tr}\,\big(a(s)\big)\mathrm{d}\ell(s)\Big\},$$

in the a priori parabolic and elliptic estimates. When $\mathcal{O} = \mathbb{R}^d_+ := \{x \in \mathbb{R}^d : x_d > 0\}$ and under the some suitable assumptions, including

$$b_d(s) \geq c_0 > 0, \quad \forall s \geq 0,$$

we can use $\varphi(y) := y$ or $\varphi(y) := y/(1+y^2)$ and $y = x_d$ to get an estimate on the local time ℓ, which is more general than the instantaneous reflection at the boundary, i.e., diffusions, jumps and sojourn are also allowed on the boundary $\partial\mathbb{R}^d_*$. This type of bounds are useful when dealing with boundary conditions, e.g., see the papers Anulova [3, 4].

It is clear that after getting an estimate on $C_{\kappa,\vartheta,v}$ we deduce a bound as in Corollary 4.68 for the above stochastic integral process (4.128).

Also, it is clear that a convergence result similar to Theorem 4.69 holds for a stochastic integral process of the form (4.128) or (4.130) as in Theorem 4.70. The interest reader may take a look at the arguments in the paper Anulova [3] for the passage to the limit in stochastic integrals.

Chapter 5

Stochastic Differential Equations

In the physical and engineering sciences stochastic differential equations arise in a quite naturally way in the description of systems with disturbances, e.g., see Schuss [223], Wong [255]. On the other hand, the theory of stochastic differential equations was originally developed as a tool for explicit construction of the trajectories of diffusion processes with given drift and diffusion coefficients. This approach was extended to diffusion processes with jumps, where jumps coefficients are also given.

Sometimes the equation takes the form of an ordinary differential equation with random coefficients, i.e.,

$$\dot{x}(t) = g(t, x(t), \eta(t)), \quad x(t_0) = x_0,$$

where the vector field g is a deterministic function, but $\eta = \eta(t, \omega)$ is a random process (disturbances) and x_0 is the initial condition (possible random). Depending on the assumptions on the data g, η and x_0, solutions are found as stochastic processes $x(t)$ which paths are at least locally absolutely continuous. Most of the results here are obtained with classic methods from the theory of ordinary differential equations, e.g., see Ladde and Lakshmikantham [146]. The situation is quite different if the random disturbances are modelled as *white-noise*, e.g., see Wong and Hajek [256]. The simplest case is when g is linear in η, so that the equation becomes

$$\dot{x}(t) = g(t, x(t)) + \sigma(t, x(t))\dot{w}(t), \quad x(t_0) = x_0,$$

where $\dot{w}(t)$ is the white-noise, which is conceived as a stationary Gaussian stochastic process with mean value zero and a constant spectral density on the entire space. Such a process does not exist in the conventional sense, since the Dirac delta *function* would be its covariance function, independent values at all points with an infinite variance. However, the white-noise is a very useful mathematical idealization for describing random influences that fluctuate rapidly and

hence are virtually uncorrelated for different instants of time. The white-noise can be realized as a generalized stochastic process, but to keep working with stochastic processes, the white-noise is thought as the formal time-derivative of a standard Wiener process $(w(t) : t \geq 0)$, namely $(\dot{w}(t) : t \geq 0)$, e.g., see Itô and McKean [113]. For a stochastic dynamical system without after-effects or without memory, the law of the motion of the state can be described by a more general equation, namely

$$\mathrm{d}x(t) = G(t, x(t), \mathrm{d}t),$$

and if the fluctuations or disturbances η influence the system additively then

$$G(t, x(t), \mathrm{d}t) := g(t, x(t))\mathrm{d}t + \sigma(t, x(t))\eta(\mathrm{d}t),$$

where $(\eta(t) : t \geq 0)$ is a process with independent increments. Typically, η is expressed as the sum of a standard Wiener process and a standard Poisson measure, which produces a stochastic process $(x(t) : t \geq 0)$, so-called, diffusion with jumps. Moreover, in stochastic control theory we have an equation of the form

$$\mathrm{d}x(t) = G(t, x(t), v(t), \mathrm{d}t),$$

where $v(t)$ is the control process and

$$G(t, x(t), v(t), \mathrm{d}t) := g(t, x(t), v(t))\mathrm{d}t + \sigma(t, x(t), v(t))\eta(\mathrm{d}t).$$

The paths of the noise $t \mapsto \eta(t)$ are usually continuous or cad-lag, but always we expect the continuity in probability (or stochastic continuity) of both, the disturbance process and the state process $(x(t) : t \geq 0)$, however, no path regularity is expected for the control process $(v(t) : t \geq 0)$. On the other hand, to impose causality, the state and the control processes must be adapted to the disturbances.

Usually, the state variable x belongs to an Euclidean space, say $\mathbb{R}^d$, or to a region in $\mathcal{O}$ of $\mathbb{R}^d$. If $\mathcal{O}$ is a manifold with boundary, then some conditions on the boundary $\partial\mathcal{O}$ are necessary. The most studied boundary actions are either to stop or to reflect the process, which yields either Dirichlet or Neumann (or oblique) boundary conditions. Sometimes, the state variable x belongs to some (infinite dimensional) Hilbert or Banach space, and so, the above stochastic ordinary differential equation becomes a stochastic partial differential equation.

For a self contained course on *stochastic differential equations* (and more) we refer to the books Arnold [7], Chung and Williams [45], Da Prato and Zabczyk [51], Freidlin[87], Friedman [90], Gihman and Skorohod [99], Ikeda and Watanabe [110], Kunita [143], Mao [165], Oksendal [190], Protter [206], among others. On the other hand, some books in stochastic control contain a short introduction to stochastic differential equations, e.g., Fleming and Rishel [83], Morimoto [184], Yong and Zhou [261], among many others.

Stochastic ordinary differential equation are associated with three coefficients (drift, diffusion and jump terms) or directly with the integro-differential operator (first-order and second-order terms, and the jump kernel). There are three distinguished setups of essentially the same problem:

(1) the strong formulation, where the probability space with a Wiener and a Poisson measure, and the coefficients are given, and we look for a process satisfying the equation;

(2) the weak formulation, where only the coefficients are given, and we look for the probability space with a Wiener and a Poisson measure, and a process satisfying the equation;

(3) the martingale problem, where the integro-differential operator is given, and we look for a probability measure on the canonical space such that a process (depending on the operator applied to smooth functions) is a martingale for any choices of smooth functions.

Clearly, we can always go from (1) to (2) to (3), and under certain conditions, we may move backward, from (3) to (2) to (1). As discussed later, the *weak* character of the solution is directly related to functional representation of the solution. The martingale setting (3) can be viewed as looking for a semi-martingale with prescribed characteristics (drift, diffusion and integer measure). Extracting the coefficients (with the appropriated degree of regularity in the space and time variables) of a given second-order integro-differential operator is not always simple, the diffusion term involves a square-root of a matrix and jumps term is harder, a representation of the jump kernel (o Lévy kernel or measure) is necessary. These points should become clear with the specific details of what follows.

It is clear that we have in mind to model the disturbance with an underlaying Lévy process, but we need to distinguish the continuous part from the jump part. The jump part is certainly described by the Lévy measure (or jumps kernel). Two quick difficulties appear, one situation is when the Lévy measure is locally integrable (it may be no integrable at infinite). In this case, we have to deal with process with possible infinite first moment, but the construction is simple (like a Poisson process). We are more interested in the second case, when the Lévy measure at least integrates the function $1 \wedge |x|^2$.

The continuous part is well modeled by a n-dimensional Wiener process, which can be seen a n independent one-dimensional Wiener processes. The jump part is modeled as a m-dimensional Poisson measure, which corresponds to the jump of a m-dimensional Lévy process, and in general cannot be thought as coming from m (independent) Lévy processes (this would represent the case of no simultaneous jumps). Moreover, if we are willing to use martingales in a general Hilbert space, we can including even the case of (local) martingale measures as a particular case. Therefore, instead of refer to as noise with Lévy processes, we use the terminology *Wiener-Poisson space with Lévy measure*.

Modeling dynamical systems by a SDE includes only one time-scale, in our case, the continuous and the jump parts are measured with the same time-scale. However, more practical models include continuous (e.g., mechanical) and discrete (e.g., digital) variables (e.g., measurements), which are usually referred to as hybrid systems (e.g., see Bensoussan and Menaldi [18], Menaldi [172] and references). A typical model is an automaton, where the system (running

in a continuous time-scale) makes jumps or adjust it evolution according to a prescribed rule, in short, switching between differential equations. In more details, this is a switchings system between diffusions accordingly to a Markov chain. Essentially this can be viewed as a diffusion with jumps, with a bounded Lévy measure, assuming some details in the description are ignored.

The reader will find that there are a lot of details missing in this summary on stochastic integrals and martingales. Many excellent books (classic and recent) cover all the subjects mentioned above and much more, e.g., the reader may check Applebaum [6], Chung and Williams [45], Bichteler [25], Kallenberg[121], Medvegyev [168], Peszat and Zabczyk [197], Protter [206], among others, for various presentations on stochastic analysis.

5.1 Existence and Uniqueness of Solutions

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t), \zeta \in \mathbb{R}^m_*, t \geq 0)$ be a (standard) $n \times m$ Wiener-Poisson space with Lévy measure $\pi(\cdot)$, i.e., in a (complete) filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, t \geq 0)$, the stochastic process $(w(t), t \geq 0)$ is a n-dimensional (standard) Wiener space and $(\nu(B,]0, t]), B \in \mathbb{R}^m_*, t \geq 0)$ is an independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{\nu(B,]0, t])\}/t$, which satisfies

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1 + |\zeta|} \pi(\mathrm{d}\zeta) < \infty,$$

with martingale measure $\tilde{\nu}(B,]0, t]) := \nu(B,]0, t]) - t\pi(B)$, as discussed in the previous chapter. This martingale measure $\tilde{\nu}$ is identified with the $\mathbb{R}^m$-valued compensated-jump (Poisson) process

$$\tilde{p}(t) := \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0,$$

which induces a probability measure $P_{\tilde{\nu}}$ (also denoted by P_π) on the canonical space $D := D([0, \infty[, \mathbb{R}^m)$ of cad-lag functions, namely,

$$P_{\tilde{\nu}}(A) := P\Big\{\tilde{p}(\cdot) \in A\Big\}, \quad \forall A \in \mathcal{B}(D). \tag{5.1}$$

with its characteristic function (or Fourier transform) given by

$$\mathbb{E}\Big\{ \exp\Big[\mathtt{i} \int_{\mathbb{R}^m_* \times]0,t]} (z \cdot \zeta) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)\Big]\Big\} = \\ = \exp\Big[- t \int_{\mathbb{R}^m_*} \big(1 - \mathrm{e}^{\mathtt{i}\, z \cdot \zeta} + \mathtt{i}\, z \cdot \zeta\big) \pi(\mathrm{d}\zeta)\Big],$$

for every $t \geq 0$ and z in $\mathbb{R}^m$. Also note that the Wiener process w induces a probability measure P_w on the canonical space $C := C([0, \infty[, \mathbb{R}^m)$ of continuous functions, namely,

$$P_w(A) := P\big\{w(\cdot) \in A\big\}, \quad \forall A \in \mathcal{B}(C). \tag{5.2}$$

and its characteristic function (or Fourier transform) is given by

$$\mathbb{E}\{\exp[\mathtt{i}\,\xi\cdot w(t)]\}=\exp\big(-t\frac{|\xi|^2}{2}\big),$$

for every $t\geq 0$ and ξ in $\mathbb{R}^n$.

In other words, $\ell=w+\tilde{p}$ is a (centered) Lévy process, where w is its continuous or Gaussian part and $\tilde{p}$ is its purely jumps or Poisson part. Recalling that any continuous martingale is orthogonal to any purely discontinuous martingale (with respect to a common filtration), we deduce that the processes $\phi(w)-\phi(0)$ and $\psi(\tilde{p})-\mathbb{E}\{\psi(\tilde{p})\}$ are orthogonal martingales for any smooth functions ϕ and ψ, i.e., w and $\tilde{p}$ (or ν) are independent. Hence, as long as the filtration $\mathbb{F}=(\mathcal{F}_t:t\geq 0)$ is given and w, $\tilde{p}$ (or $\tilde{\nu}$) are $\mathbb{F}$-martingales, the independence of the Wiener process and the Poisson measure is granted. This Wiener-Poisson space is fixed throughout this section (unless otherwise mentioned).

Alternatively, we may begin with a Lévy process ℓ with characteristic $(0,I,\pi)$, i.e., with the continuous part being a standard Wiener process and with Levy measure π. Then, from the jumps of ℓ we obtain its associated integer measure ν and martingale measure $\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t)=\nu(\mathrm{d}\zeta,\mathrm{d}t)-t\pi(\mathrm{d}\zeta)$, so that we are back to in our initial setting. Here, we could use Lévy measures that integrate only $|\zeta|^2\wedge 1$ instead of $|\zeta|^2\wedge|\zeta|$, but our real interest is on Lévy processes having moment of all order, i.e., Lévy measures capable of integrating $1+|\zeta|^p$ for every $p\geq 0$. Other cases can be treated with a direct (or interlacing) construction of the jumps similarly to the case of piecewise deterministic processes, for instance, see also Applebaum [6, Chapter 6, pp. 292-357].

It should be clear that the setting of the stochastic differential equation is materialized in a Wiener-Poisson space as described early, which is constructed from its characteristics, namely, the identity matrix used as co-variance matrix of the Wiener process and the Lévy measure determining the Poisson measure. The coefficients (drift g, diffusion σ and jumps γ) are the key data, but not the only data. The underlying noise is represented by the Wiener process w and the centered Poisson measure $\tilde{\nu}$, and contrary to deterministic models, the characteristics elements of the noise also play. When jumps are not involved, the co-variance matrix (taken to be the identity) determine the Wiener process, but when dealing with jumps, there are much larger choices since the Lévy measure π carries all the intrinsic properties of the jumps. At this time, it may useful to recall that a linear combination of independent Wiener (or Lévy) processes produces another Wiener (or Lévy) process, but only a sum of independent Poisson measures yields another Poisson measure.

Note the role of the filtration $\mathbb{F}=(\mathcal{F}_t:t\geq 0)$ in the Wiener-Poisson space. This is to say that given the dimension n and the Lévy measure π on $\mathbb{R}^m_*$, the canonical construction mentioned early on the canonical space $\Omega=C([0,\infty[,\mathbb{R}^n)\times D([0,\infty[,\mathbb{R}^m)$ of the probability measure $P_w\times P_\pi$ (on the Borel σ-algebra $\mathcal{F}$ of Ω, considered as a Polish space) provides a basic Wiener-Poisson space, where $\mathbb{F}$ is the canonical filtration, i.e. first $\mathcal{F}^0_t$ is generated by the projection maps $\omega\mapsto\omega(s)$ for $0\leq s\leq t$ and then $\mathcal{F}_t=\mathcal{F}^0\cup\bigcap_{s>t}\mathcal{F}^0_s$, where $\mathcal{F}^0$ is the σ-algebra generated by all Borel sets with $P_w\times P_\pi$-zero measure.

However, this leaves the possibility of enlarging the filtration $\mathbb{F}$ and the σ-algebra $\mathcal{F}$, for instance, completing $\mathcal{F}$ with all $P_w \times P_\pi$-negligible sets or using the universally completed σ-algebra $\mathcal{F}^u$ (i.e., adding all sets which are negligible for any probability measure on Ω), and modifying the filtration $\mathbb{F}$ in a way that the martingale properties of the Wiener process and the Poisson measure are preserved. As seen later, these changes are necessary when discussing the so-called either weak formulation or martingale formulation. As usual, recall that the notations either x_t and $x(t)$ for stochastic processes, or $\mathcal{F}_t$ and $\mathcal{F}(t)$ for σ-algebras, are considered interchangeable everywhere in the text.

5.1.1 Lipschitz Coefficients

Suppose we are given the data, i.e., an initial condition $x^0 = (x_i^0,\ i = 1, \dots, d)$, an adapted (predictable) process $(v(t) : t \geq 0)$, and functions $g(t, x, v) := (g_i(t, x, v),\ i = 1, \dots, d)$, $\sigma(t, x, v) := (\sigma_{ik}(t, x, v),\ i = 1, \dots, d,\ k = 1, \dots, n)$ and $\gamma(\zeta, t, x, v) := (\gamma_i(\zeta, t, x, v),\ i = 1, \dots, d)$. The stochastic ordinary differential equation takes the form

$$\left\{ \begin{array}{l} x(t) = x^0 + \displaystyle\int_{t_0}^{t} g(s, x(s), v(s))\mathrm{d}s + \int_{t_0}^{t} \sigma(s, x(s), v(s))\mathrm{d}w(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m \times]t_0, t]} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0, \end{array} \right. \tag{5.3}$$

which is written by components, $i = 1, \dots, d$, as

$$\left\{ \begin{array}{l} x_i(t) = x_i^0 + \displaystyle\int_{t_0}^{t} g_i(s, x(s), v(s))\mathrm{d}s + \\ \qquad + \displaystyle\sum_{k=1}^{n} \int_{t_0}^{t} \sigma_{ik}(s, x(s), v(s))\mathrm{d}w_k(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m \times]t_0, t]} \gamma_i(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0, \end{array} \right. \tag{5.4}$$

or in differential form as

$$\left\{ \begin{array}{l} \mathrm{d}x(t) = g(t, x(t), v(t))\mathrm{d}t + \sigma(t, x(t), v(t))\mathrm{d}w(t) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m} \gamma(\zeta, t, x(t), v(t))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t), \end{array} \right. \tag{5.5}$$

plus the initial condition $x(t_0) = x^0$. Note the use of the stochastic integral with respect to a martingale.

Perhaps of key importance is to recall the martingale-type (or sup-type) inequalities for the stochastic integrals, namely, the Wiener integral,

$$\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1} \Big| \int_{t_0}^{t} \sigma(s)\mathrm{d}w(s) \Big|^p \Big\} \leq C_p \mathbb{E}\Big\{ \Big[\int_{t_0}^{t_1} |\sigma(s)|^2 \mathrm{d}s \Big]^{p/2} \Big\},$$

for any $p \geq 1$, and the Poisson integral

$$\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1} \Big| \int_{\mathbb{R}^m_* \times]t_0,t]} \gamma(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p \Big\} \leq$$
$$\leq C_p \mathbb{E}\Big\{ \Big[\int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^2 \pi(\mathrm{d}\zeta)\Big]^{p/2} \Big\},$$

if $1 \leq p \leq 2$, while a term of the form

$$C_p \mathbb{E}\Big\{ \Big[\int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^p \pi(\mathrm{d}\zeta)\Big]^{p} \Big\}, \quad \text{only if} \quad p \geq 2,$$

should be also added. In any case, the constant C_p depends only on p and in particular, $C_1 = 3$ and $C_2 = 4$.

Thus the expression (5.3) makes sense as long as following conditions hold:

(a) the process $(x(t) : t \geq 0)$ is measurable, adapted, and locally bounded, i.e.,

$$\sup_{t_0 \leq s \leq t} |x(s)| < \infty,$$

with probability one for any $t \geq t_0$.

(b) the (Lebesgue and stochastic) integrals

$$\int_{t_0}^{t} |g(s)|\mathrm{d}s, \quad \int_{t_0}^{t} |\sigma(s)|^2 \mathrm{d}s, \quad \int_{t_0}^{t} \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma(\zeta,s)|^2 \pi(\mathrm{d}\zeta),$$

are finite with probability one for any $t \geq t_0$, where the integrand processes ($\mathbb{R}^d$-, $\mathbb{R}^d \times \mathbb{R}^n$- and $\mathbb{R}^d$-valued) are defined by $g(t) := g(t, x(t), v(t))$, $\sigma(t) := \sigma(t, x(t), v(t))$ and $\gamma(\zeta, t) := \gamma(\zeta, t, x(t), v(t))$.

Hence, when we say that equation (5.3) is satisfied we means that (a) and (b) above are true, and the equality (5.3) holds for every $t \geq t_0$ with probability one. Since the initial condition x^0 is a $\mathcal{F}(t_0)$-measurable random variable and $v(t)$ is adapted, the processes $g(t)$, $\sigma(t)$ and $\gamma(t,\zeta)$ are necessarily measurable and adapted (if the solution $x(t)$ is so). A posteriori (once a locally bounded solution is found), by taking the right-hand-term of equality (5.3), another solution can be defined, which is a cad-lag (and stochastically continuous) version of the previous one. Therefore, uniqueness means *uniqueness up to a version* and and *solution* means *a cad-lag version* (with is also stochastically continuous, if needed). If the predictable version of the integrand in the Poisson stochastic integral is used, then $\gamma(\zeta, t, x(t), v(t))$ should be replaced by $\gamma(\zeta, t, x(t-), v(t))$.

• *Remark* 5.1. From the modeling or control viewpoint it is perhaps important to emphasis the various type between *data*, namely, (a) the deterministic coefficients $g(t,x,v)$ (drift), $\sigma(t,x,v)$ (diffusion), $\gamma(t,x,v)$ (jumps), and $\pi(\delta\zeta)$ (jump-measure); (b) the random elements t_0, $x(t_0)$ (initial condition) and $v(\cdot)$ (control process); (c) the Wiener-Poisson space composed by a filtered probability space $(\Omega, \mathbb{F}, P)$ with a standard Wiener process w and a Poisson measure

ν with compensated (or martingale) measure $\tilde{\nu}$, all relative to the filtration $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$. Therefore, the stochastic setting includes all the pieces mentioned in (c), which are constructed a priori, based on the jump-measure (Lévy measure) π. The random elements in (b) could be taken deterministic initially, e.g., our current model uses a deterministic initial time t_0, a possible random initial state $x(t_0)$ (which is mainly regarded as deterministic to all effects), and a adapted control process $v(\cdot)$ (which is only a predictable process) with not a priori regularity on its paths. If $v(\cdot)$ is assumed deterministic then there is not need to explicitly mention this variable v. Our model is based on the deterministic coefficients in (a), the drift g is the mean-evolution and (σ, γ) describes the noise-evolution, the continuous (or Gaussian) and the jumps (or Poisson) components. However, if the variable v is not explicitly mentioned then (to be able to accommodate the control process) the coefficients should include the variable ω, i.e.,$g(t, x, \omega) = g(t, x, v(t, \omega))$, $\sigma(t, x, \omega) = \sigma(t, x, v(t, \omega))$, $\gamma(t, x, \omega) = \gamma(t, x, v(t, \omega))$. Thus, the intrinsic characteristic of data (a) are lost, in relation to the data (b) and (c). □

Definition 5.2 (solution)**.** An adapted cad-lag process $(x(t) : t \geq t_0)$ is a solution of the d-dimensional stochastic ordinary differential equation if (5.3) is satisfied for every $t \geq t_0$. Actually, a solution means an equivalence class of processes represented by a cad-lag element. Thus, we say that the *uniqueness* holds whenever two solutions are each one version of each other, i.e., if x and y are two solutions then $P\{x(t) = y(t)\} = 1$ for every $t \geq t_0$. □

Note that if an adapted process $(x(t) : t \geq t_0)$ satisfying (5.3) is found, then a cad-lag version exists and any cad-lag version is a solution. Thus, uniqueness of solution means a unique equivalence class of solutions. It is clear that we may consider solution defined on a bounded interval $[t_0, t_1]$ or $[t_0, t_1)$ instead of $[t_0, \infty)$.

Regarding this model, it may be important to remark that deterministic jumps are not allowed, namely, the solution is also quasi-continuous from the left, i.e., for any sequence $\{\tau_k : k \geq 1\}$ stopping times converging to τ (which is then a predictable time) and satisfying $\tau_k < \tau < \infty$ almost surely, we have $x(\tau_k) \to x(\tau)$ in probability. Moreover, if $\tilde{p}$ is the purely jumps part of the underlying Lévy process (i.e., $\tilde{p}$ is the compensated Poisson process associated with the Poisson measure ν with martingale measure $\tilde{\nu}$ and Lévy measure π) as above, then the jumps $\delta x(t) := x(t) - x(t-) = \gamma\big(\delta\tilde{p}(t), t, x(t-), v(t)\big)\mathbb{1}_{\{|\tilde{p}(t)|>0\}}$, i.e., jumps are only generated by the jumps of the underlying Lévy process and they are modified by the coefficient γ. Moreover, if π integrates $|\gamma(\zeta, t, x(t), v(t))|$ almost surely on $\mathbb{R}^m_* \times]0, T]$, for every $T > 0$, then we have

$$\sum_{0<s\leq t} \big(x(s) - x(s-)\big) = \int_{\mathbb{R}^m_* \times]t_0,t]} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma(\zeta, t, x(t), v(t))\pi(\mathrm{d}\zeta).$$

Deterministic jumps could be added, by means of other *deterministic* tools.

Certainly, the mean vector process $\mu^x(t) = \mathbb{E}\{x(t)\}$ is given by

$$\mu^x(t) = x^0 + \int_{t_0}^{t} \mathbb{E}\{g(s, x(s), v(s))\}\mathrm{d}s, \quad \forall t \geq t_0,$$

and from Itô formula we deduce that the covariance matrix process $Q_x(t)$, with entries $q_{ij}^x(t) = \mathbb{E}\{[x_i(t) - \mu_i^x(t)][x_j(t) - \mu_j^x(t)]\}$, satisfies

$$\begin{aligned} q_{ij}^x(t) = \frac{1}{2}\sum_{k=1}^{n} \int_{t_0}^{t} \mathbb{E}\{\sigma_{ik}(s, x(s), v(s))\sigma_{jk}(s, x(s), v(s))\}\mathrm{d}s + \\ + \int_{t_0}^{t} \mathrm{d}s \int_{\mathbb{R}_*^m} \mathbb{E}\{\gamma_i(\zeta, s, x(s), v(s))\gamma_j(\zeta, s, x(s), v(s))\}\pi(\mathrm{d}\zeta), \end{aligned}$$

for every $t \geq t_0$. Actually, Itô formula can be used to obtain the expression of the characteristic function $y \mapsto \mathbb{E}\{\mathrm{e}^{\mathrm{i}\, y\cdot x(t)}\}$. It is clear that, unless the coefficients g, σ, γ have a particular form (e.g., when g is linear in x and, σ and γ are constant in x), we cannot solve for $\mu^x(t)$ or $Q_x(t)$ from a close ordinary differential equation and avoid dealing with the stochastic differential expression.

There are various sets of assumptions used to study the above d-dimensional stochastic ordinary differential equation. The coefficients $g(t, x, v)$, $\sigma(t, x, v)$ and $\gamma(\zeta, t, x, v)$ are always supposed Borel measurable, and because we are interested in global solutions defined on a prescribed bounded interval, say $[t_0, t_1]$, we impose a linear growth condition, namely, there exists a constant $C > 0$ such that

$$\left\{ \begin{aligned} |g(t, x, v)|^2 + |\sigma(t, x, v)|^2 + \int_{\mathbb{R}_*^m} |\gamma(\zeta, t, x, v)|^2 \pi(\mathrm{d}\zeta) \leq \\ \leq C(1 + |x|^2), \end{aligned} \right. \tag{5.6}$$

for every (t, x, v) in $[t_0, \infty) \times \mathbb{R}^d \times \mathbb{R}^q$. Thus, the initial condition x^0 must be an $\mathcal{F}(t_0)$-measurable random variable (most of the time, a deterministic value), and except for adaptability or predictability, no other conditions are assumed on the (stochastic) parameter or control process $(v(t) : t \geq 0)$.

Main Results

A clean existence and uniqueness theory is developed adding a uniform locally Lipschitz condition in the variable x, namely, for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\left\{ \begin{aligned} |g(t, x, v) - g(t, x', v)|^2 + |\sigma(t, x, v) - \sigma(t, x', v)|^2 + \\ + \int_{\mathbb{R}_*^m} |\gamma(\zeta, t, x, v) - \gamma(\zeta, t, x', v)|^2 \pi(\mathrm{d}\zeta) \leq M|x - x'|^2, \end{aligned} \right. \tag{5.7}$$

for every (t, x, v), (t, x', v) in $[t_0, \infty) \times \mathbb{R}^d \times \mathbb{R}^q$ with $t \leq t_0 + r$, $|x| \leq r$ and $|x'| \leq r$.

Theorem 5.3 (existence and uniqueness)**.** *Let t_0 be an initial time, x^0 be a $\mathcal{F}(t_0)$-measurable random variable with values in $\mathbb{R}^d$ and $(v(t) : t \geq t_0)$ be a $\mathbb{R}^q$-valued predictable*[1] *process. Assume that the coefficients*[2] *satisfy (5.6) and (5.7). Then, there exists one and only one solution $(t, \omega) \mapsto x(t, \omega)$ of the d-dimensional stochastic ordinary differential equation (5.3) on the time interval $[t_0, \infty)$, which is defined as a d-dimensional process (equivalence class of processes) on the standard $n \times m$ Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$.*

Proof. The arguments of the proof are essentially the same as in the deterministic case (i.e., $\sigma = 0$ and $\gamma = 0$). The solution is found as a unique fixed point of some contraction map on a Banach space.

We proceed as follows: First we assume that the initial condition x^0 has a finite second order moment, i.e., $\mathbb{E}\{|x^0|^2\} < \infty$, and that the constant M in assumption (5.7) is independent of r (as far as the restrictions $|x| \leq r$ and $|x'| \leq r$ are concerned), i.e., the coefficients satisfy a uniform (global) Lipschitz condition in the variable x and the region $[t_0, t_0 + r] \times \mathbb{R}^d \times \mathbb{R}^q$.

It i clear that existence and uniqueness should be proved only within any arbitrary time interval $[t_0, t_1]$. Thus, for a fixed time $t_1 > t_0$, denote by $\mathcal{Y}$ the vector space of measurable functions y from $[t_0, t_1] \times \Omega$ into $\mathbb{R}^d$ satisfying:

(a) the maps $\omega \mapsto y(t, \omega)$ is $\mathcal{F}(t)$-measurable for every t in $[t_0, t_1]$,

(b) the process x is stochastically (actually, either right- or left- is sufficient) continuous, i.e., for every t in $[t_0, t_1]$ and $\varepsilon > 0$ there exists a positive $\delta = \delta(t, \varepsilon)$ (actually, δ may take independent of t in any compact interval $[t_0, t_1]$) such that $P\{|x(t) - x(s)| > \varepsilon\} < \varepsilon$ for every s satisfying $|t - s| < \delta$,

(c) for a (positive) constant α (to be selected later on) the following quantity

$$\|y\|_{\mathcal{Y}} := \big(\mathbb{E}\{\sup_{t_0 \leq t \leq t_1} \mathrm{e}^{-2\alpha(t-t_0)}|y(t)|^2\}\big)^{1/2}$$

is finite and therefore defines a semi-norm on $\mathcal{Y}$, which is denoted by $\|\cdot\|$ when possible.

By taking equivalence classes, we obtain the (in general, non-separable) Banach space $\mathcal{Y}$ with the sup-norm $\|\cdot\| = \|\cdot\|_{\mathcal{Y}}$, and because $t_1 - t_0$ is a finite value, the space $\mathcal{Y}$ is independent of the constant α. Condition (b) is not really necessary, but it is very convenient to have the concept of *version* the same as the concept of *undistinguishable* (e.g., another good choice could be to replace continuity in probability with almost surely cad-lag processes). Certainly, this is the space of adapted processes with finite sup-norm second moment. Moreover, condition (a) may be replaced by the predictability condition (since there is not deterministic jumps, i.e., the continuity in probability holds). As seen below, an essential tool is the martingale-type estimates previously mentioned for the stochastic Wiener and Poisson integrals. In any way, a fixed point can be found first in the closed subspace of optional processes of the Hilbert $L^2(]t_0, t_1 \times \Omega)$ with the product

[1] i.e., a predictable version of an adapted process

[2] which are implicitly supposed to be Borel measurable

measure $\mathrm{d}t \times P(\delta\omega)$, and then the equality (5.3) shows that this fixed point has a version belonging to the space $\mathcal{Y}$.

Consider the non-linear mapping

$$T(x)(t) := x^0 + \int_{t_0}^t g(s, x(s), v(s))\mathrm{d}s + \int_{t_0}^t \sigma(s, x(s), v(s))\mathrm{d}w(s) + \\ + \int_{\mathbb{R}_*^m \times]t_0,t]} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \in [t_0, t_1].$$

Since each integral preserves the sup-norm, i.e., for the first integral the Hölder inequality implies

$$\Big\| \int_{t_0}^{\cdot} g(s, x(s), v(s))\mathrm{d}s \Big\|^2 = \\ = \mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \mathrm{e}^{-2\alpha(t-t_0)} \Big| \int_{t_0}^t g(s, x(s), v(s))\mathrm{d}s \Big|^2 \Big\} \le \\ \le \frac{1}{2\alpha}\mathbb{E}\Big\{ \int_{t_0}^{t_1} \mathrm{e}^{-2\alpha(s-t_0)} \Big| g(s, x(s), v(s)) \Big|^2 \mathrm{d}s \Big\} \le \\ \le \frac{t_1 - t_0}{2\alpha}\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \mathrm{e}^{-2\alpha(t-t_0)} |g(t, x(t), v(t))|^2 \Big\},$$

while for the other integrals using the martingale inequality (for stochastic integrals) we get

$$\Big\| \int_{t_0}^{\cdot} \sigma(s, x(s), v(s))\mathrm{d}w(s) \Big\|^2 = \\ = \mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \Big| \int_{t_0}^t \mathrm{e}^{-\alpha(t-t_0)} \sigma(s, x(s), v(s))\mathrm{d}w(s) \Big|^2 \Big\} \le \\ \le \frac{4}{2\alpha}\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \mathrm{e}^{-2\alpha(t-t_0)} |\sigma(t, x(t), v(t))|^2 \Big\}$$

and

$$\Big\| \int_{\mathbb{R}_*^m \times]t_0,\cdot]} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \Big\|^2 = \\ = \mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \Big| \int_{\mathbb{R}_*^m \times]t_0,t]} \mathrm{e}^{-\alpha(t-t_0)} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \Big|^2 \Big\} \le \\ \le \frac{4}{2\alpha}\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \mathrm{e}^{-2\alpha(t-t_0)} \int_{\mathbb{R}_*^d} |\gamma(\zeta, t, x(t), v(t))|^2 \pi(\mathrm{d}\zeta) \Big\}.$$

Thus, assumption (5.6) implies that T maps the Banach space $\mathcal{Y}$ into itself. Moreover, computations similar to the above yield

$$\|T(x) - T(y)\| \le \frac{K}{\sqrt{\alpha}}\|x - y\|, \quad \forall x, y \in \mathcal{Y},$$

for some constant K depending only on $t_1 - t_0$ and the constant M of hypothesis (5.7).

Therefore, if we take α sufficiently large then T becomes a contraction on $\mathcal{Y}$, and the result follows. Note that for any y in $\mathcal{Y}$, the image $T(y)$ is a cad-lag adapted process.

To extend these the arguments to the general case, we begin with the uniqueness. Let x and y be two solutions (cf. Definition 5.2) of (5.3). Because the second moments of x and y may not be finite, we use a truncation technique. For any $r > 0$, define the adapted process

$$\chi_r(t) := \begin{cases} 0 & \text{if } |x(s)| > r \text{ or } |y(s)| > r \text{ for some } s > t, \\ 1 & \text{otherwise,} \end{cases}$$

i.e., if $\tau_r(x(\cdot))$ denotes the first exit time from the closed ball of radius r for the process $x(\cdot)$, namely,

$$\tau_r(x(\cdot)) = \inf\{s \geq 0 : |x(s)| > r\},$$

which implies $|x(s)| \leq r$ for every $s < \tau_r(x(\cdot))$, and then

$$\chi_r(t) = \mathbb{1}_{\tau_r(x(\cdot))>t}\, \mathbb{1}_{\tau_r(y(\cdot))>t}, \quad \forall t \geq 0.$$

Since $\chi_r(t) = \chi_r(t)\,\chi_r(s)$ for every $s \leq t$, we have

$$\begin{aligned}[x(t) - y(t)]\chi_r(t) = \chi_r(t)\Big\{ &\int_{t_0}^t \chi_r(s)\big[g(s, x(s), v(s)) - g(s, y(s), v(s))\big]\mathrm{d}s \\ &+ \int_{t_0}^t \chi_r(s)\big[\sigma(s, x(s), v(s)) - \sigma(s, y(s), v(s))\big]\mathrm{d}w(s) + \\ &+ \int_{\mathbb{R}^m_* \times]t_0,t]} \chi_r(s)\big[\gamma(\zeta, s, x(s), v(s)) - \gamma(\zeta, s, y(s), v(s))\big]\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)\Big\}.\end{aligned}$$

By means of Hölder inequality and the martingale inequality for stochastic integrals, we deduce that

$$\mathbb{E}\{|x(t) - y(t)|^2\chi_r(t)\} \leq K \int_{t_0}^t \mathbb{E}\{|x(s) - y(s)|^2\chi_r(s)\}\mathrm{d}s,$$

for every t in $[t_0, t_1]$, and for some constant K depending only on $t_1 - t_0$, r and the constant $M = M(r)$ of hypothesis (5.7). Using Gronwall inequality, i.e., in this case notice that the derivative of

$$t \mapsto \mathrm{e}^{-K(t-t_0)} \int_{t_0}^t \mathbb{E}\{|x(s) - y(s)|^2\chi_r(s)\}\mathrm{d}s$$

is nonnegative, we deduce that for any t in $[t_0, t_1]$ and any $r > 0$ we have $|x(t) - y(t)|\chi_r(t) = 0$ with probability one. Since x and y are cad-lag, and

$$\begin{aligned}P\big\{\chi_r(t) = 0 \text{ in } [t_0, t_1]\big\} \leq\ & P\Big\{\sup_{t_0 \leq t \leq t_1} |x(t)| > r\Big\} + \\ &+ P\Big\{\sup_{t_0 \leq t \leq t_1} |y(t)| > r\Big\},\end{aligned}$$

there exists a set with probability one where $x(t) = y(t)$ for every t in $[t_0, t_1]$. Note that uniqueness holds under the sole assumption (5.7), i.e., (measurability and) a uniform locally Lipschitz condition in the variable x.

Finally, we need to extend the existence result to any initial condition and local Lipschitz coefficients. Given $r > 0$, consider the orthogonal projection π_r in $\mathbb{R}^d$ on the ball with center 0 and radius r, i.e., $\pi_r(x) := x$ if $|x| \le r$ and $\pi_r(x) = r\, x/|x|$ otherwise. Define $x^{0,r} := \pi_r(x^0)$,

$$
\begin{aligned}
g^r(t,x,v) &:= g(t, \pi_r(x), v) \\
\sigma^r(t,x,v) &:= \sigma(t, \pi_r(x), v), \\
\gamma^r(\zeta,t,x,v) &:= \gamma(\zeta, t, \pi_r(x), v),
\end{aligned}
$$

for every $r > 0$. Because now the initial condition $x^{0,r}$ has finite second moment and the coefficients $g^r(t,x,v)$, $\sigma^r(t,x,v)$ and $\gamma^r(\zeta,t,x,v)$ satisfy a uniform (global) Lipschitz condition in the variable x, there is a unique solution, denoted by $(x^r(t) : t \in [t_0, t_1])$, of the stochastic ordinary differential equation with above truncated data.

For any t in $[t_0, t_1]$, we have

$$
\begin{aligned}
|x^r(t)|^2 \le 4|x^{0,r}|^2 &+ 4(t_1 - t_0) \int_{t_0}^t |g^r(s, x(s), v(s))|^2 \mathrm{d}s + \\
&+ 4 \Big| \int_{t_0}^t \sigma^r(s, x(s), v(s)) \mathrm{d}w(s) \Big|^2 + \\
&+ 4 \Big| \int_{\mathbb{R}^m_* \times]t_0, t]} \gamma^r(\zeta, s, x(s), v(s)) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \Big|^2 .
\end{aligned}
$$

Since the truncated coefficients satisfy the growth condition (5.6) uniformly in $r > 0$, we can make use of the sup-estimate for stochastic integrals to estimate the product $\psi \sup_t |x^r(t)|^2$ with $\psi := 1/(1 + |x^0|^2)$, i.e.,

$$
\mathbb{E}\{\psi \sup_{t_0 \le s \le t} |x^r(s)|^2\} \le K_1 + K_1 \int_{t_0}^t \mathbb{E}\{\psi\, |x^r(s)|^2\} \mathrm{d}s, \quad \forall t,
$$

for some constant K_1 independent of r, actually, depending only on $t_1 - t_0$ and the bound C of assumption (5.6). Hence, Gronwall inequality yields

$$
\mathbb{E}\{\psi \sup_{t_0 \le t \le t_1} |x^r(t)|^2\} \le K,
$$

for another constant K independent of r. Therefore

$$
\begin{aligned}
P\{\sup_{t_0 \le t \le t_1} |x^r(t)| > r\} &\le P\{\psi \sup_{t_0 \le t \le t_1} |x^r(t)|^2 > \psi\, r^2\} \le \\
&\le \frac{\mathbb{E}\{\psi \sup_{t_0 \le t \le t_1} |x^r(t)|^2\}}{\varepsilon\, r^2} + P\{\psi \le \varepsilon\} \le \\
&\le \frac{K}{\varepsilon\, r^2} + P\{\psi \le \varepsilon\},
\end{aligned}
$$

for any $\varepsilon > 0$. This proves that

$$P\{\sup_{t_0 \le t \le t_1} |x^r(t)| > r\}$$

vanishes as $r \to \infty$.

On the other hand, the same arguments used for the uniqueness prove that for any $r' \ge r > 0$ we have

$$P\{\sup_{t_0 \le t \le t_1} |x^r(t) - x^{r'}(t)| > 0\} \le P\{\sup_{t_0 \le t \le t_1} |x^r(t)| > r\},$$

and we conclude that x^r converges to a solution of (5.3). Finally, because t_1 is arbitrary the proof is completed. □

General Comments

There is a little room for improvement in the above result regarding the assumption on the drift coefficient g. To see this point, we apply Itô formula to the function $(x,t) \mapsto |x|^2$ and the solution process $x(t)$ to obtain

$$\begin{aligned}
\mathrm{d}|x(t)|^2 = \Big\{ & 2\sum_i x_i(t)\, g_i(t,x(t),v(t)) + \sum_{i,k} \big[\sigma_{ik}(t,x(t),v(t))\big]^2 + \\
& + \sum_i \int_{\mathbb{R}^m_*} \big[\gamma_i(\zeta,t,x(t),v(t))\big]^2 \pi(\mathrm{d}\zeta)\Big\}\mathrm{d}t + \\
& + 2\sum_k \Big[\sum_i x_i(t)\, \sigma_{ik}(t,x(t),v(t))\Big]\mathrm{d}w_k(t) + \\
& + \int_{\mathbb{R}^m_*} \Big[|x(t) + \gamma(\zeta,t,x(t),v(t))|^2 - |x(t)|^2\Big]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t).
\end{aligned}$$

We can revise the previous existence proof to see that condition (5.6) can be replaced by the following growth assumption: there exists a constant $C > 0$ such that

$$2\, x\, g(t,x,v) + |\sigma(t,x,v)|^2 + \int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^2 \pi(\mathrm{d}\zeta) \le C(1+|x|^2), \quad (5.8)$$

for every (t,x,v) in $[t_0,\infty \times \mathbb{R}^d \times \mathbb{R}^q$. In particular, from Gronwall inequality we deduce the estimate

$$\mathbb{E}\{|x(t)|^2\} \le \mathrm{e}^{C\,(t-t_0)}\mathbb{E}\{1+|x(t_0)|^2\}, \quad \forall t \ge t_0, \quad (5.9)$$

where the constant C is the same as in (5.8). The stochastic integral with respect to the Wiener process, which is temporarily denoted by

$$\int_{t_0}^t a(s)\, b(s)\mathrm{d}w(s),$$

can be bounded as follows

$$\mathbb{E}\Big\{ \sup_{t_0\le r\le t} \Big| \int_{t_0}^{r} a(s)\, b(s) \mathrm{d}w(s) \Big| \Big\} \le$$
$$\le 3 \Big\{ \mathbb{E}\big[\sup_{t_0\le s\le t} |a(s)|^2 \big] \Big\}^{1/2} \Big\{ \mathbb{E}\Big[\int_{t_0}^{t} |b(s)|^2 \mathrm{d}s \Big] \Big\}^{1/2} \le$$
$$\le \frac{1}{3} \mathbb{E}\Big\{ \sup_{t_0\le s\le t} |a(s)|^2 \Big\} + \frac{27}{4} \mathbb{E}\Big\{ \int_{t_0}^{t} |b(s)|^2 \mathrm{d}s \Big\},$$

and similarly for the stochastic integral with respect to the Poisson measure. After using Hölder and Gronwall inequalities, this yields the bound

$$\mathbb{E}\{ \sup_{t_0\le s\le t} |x(s)|^2 \} \le K \mathrm{e}^{K\,(t-t_0)} \mathbb{E}\{1 + |x(t_0)|^2\}, \quad \forall t \ge t_0, \tag{5.10}$$

for another constant K depending only on the constant C appearing in condition (5.8).

Similarly, using the process $y(t) - x(t)$ with two solutions $x(t)$ and $y(t)$, we conclude that the uniqueness proof in Theorem 5.3 holds true if the uniform locally Lipschitz condition in x, i.e., (5.6), is replaced by the following weaker assumption: for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\begin{cases} (x - x')\,[g(t,x,v) - g(t,x',v)] + |\sigma(t,x,v) - \sigma(t,x',v)|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v) - \gamma(\zeta,t,x',v)|^2 \pi(\mathrm{d}\zeta) \le M|x - x|^2, \end{cases} \tag{5.11}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty) \times \mathbb{R}^d \times \mathbb{R}^q$ with $t \le t_0 + r$, $|x| \le r$ and $|x'| \le r$.

On the other hand, under assumption (5.6) we also have

$$|x(t) - x^0|^2 \le C \int_{t_0}^{t} \big[\big(1 + |x(t)|^2\big) + \big(1 + |x(t) - x^0|^2\big) \big] \mathrm{d}t +$$
$$+2 \int_{t_0}^{t} \sum_k \Big[\sum_i (x_i(s) - x_i^0)\, \sigma_{ik}(s, x(s), v(s)) \Big] \mathrm{d}w_k(s) +$$
$$+2 \int_{\mathbb{R}^m_* \times]t_0,t]} \Big[\sum_i (x_i(s) - x_i^0)\, \gamma_i(\zeta, s, x(s), v(s)) \Big] \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s),$$

which yields

$$\mathbb{E}\{ \sup_{t_0\le s\le t} |x(s) - x^0|^2 \} \le K \mathrm{e}^{K\,(t-t_0)} \mathbb{E}\{1 + |x_0|^2\}(t - t_0), \tag{5.12}$$

for some constant K and for every $t \ge t_0$, after using (5.10). This shows some type of time Hölder continuity with exponent $1/2$ as expected.

• *Remark* 5.4. Based on the a priori estimates we can obtain existence and uniqueness of the d-dimensional stochastic differential equation (5.5) under more general conditions. For instance, the coefficients g, σ and γ may be measurable and such that the bounds $\alpha(p,\lambda)$ and $\beta(p)$ given by (5.44) and (5.49) for $p =$

2 are finite, without satisfying neither the linear growth condition (5.6) nor the Lipschitz continuity condition (5.7). This is the case when the first order coefficient (or drift) g is expressed as two terms $g - h$, where g (and σ, γ) satisfy the linear growth and Lipschitz conditions (5.6), (5.7), but h satisfies the following monotonicity condition

$$(x - x') \cdot \big(h(t, x, v) - h(t, x', v)\big) \geq 0, \quad \forall t,\, x,\, x'\, v, \tag{5.13}$$

and some polynomial growth condition in x uniformly in t, v. Even a multi-valued monotone h is allowed. The method is essentially as follows. First the argument begins by using a modified version of Caratheodory method on the ordinary differential equation (for each given ω or in some $L^p(\Omega, \mathcal{F}, P)$ Lebesgue space)

$$z(t) + \int_{t_0}^{t} h(s, z(s) + M(s), v(s))\mathrm{d}s = x^0, \quad \forall t \geq t_0 \tag{5.14}$$

where $(M(t) : t \geq 0)$ is a stochastic plus Lebesgue integral of the form

$$M(t) = \int_{t_0}^{t} g(s)\mathrm{d}s + \int_{t_0}^{t} \sigma(s)\mathrm{d}w(s) + \int_{\mathbb{R}^m_* \times]t_0, t]} \gamma(\zeta, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s),$$

so that the stochastic process $y(t) := z(t) + M(t)$ solve the stochastic differential equation

$$\begin{aligned} y(t) = x^0 &+ \int_{t_0}^{t} [g - h](s, y(s), v(s))\mathrm{d}s + \\ &+ \int_{t_0}^{t} \sigma(s)\mathrm{d}w(s) + \int_{\mathbb{R}^m_* \times]t_0, t]} \gamma(\zeta, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{aligned}$$

for every $t \geq t_0$, where the stochastic processes $\sigma(t)$ and $\gamma(\zeta, t)$ are given. By means of the a priori estimates and the monotonicity argument (to identify the limit) we may establish the existence and uniqueness of the previous ordinary differential equation (5.14), including the fact that the solution $(z(t) : t \geq 0)$ is adapted. Next, replacing $\sigma(s)$ and $\gamma(\zeta, s)$ by $\sigma(s, \xi(s), v(t))$ and $\gamma(\zeta, s, \xi(s), v(s))$ we define the mapping $\xi \mapsto y$. Clearly, the a priori estimate (5.54) implies the uniqueness, and with arguments similar to those of Theorem 5.3, the monotonicity property yields also the existence of a fixed point, which is the solution. For instance, see Bensoussan and Lions [16, Section 2.3.4, pp. 51–68] for the case $\gamma = 0$ In particular, this allows us to solve an equation of the type

$$\mathrm{d}x(t) \in \mathrm{sgn}(x(t))\mathrm{d}t + \mathrm{d}w(t) + \int_{\mathbb{R}^d_*} z\tilde{\nu}(\mathrm{d}z, \mathrm{d}t),$$

where the sgn function is treated as a monotone multi-valued operator, with $\mathrm{sgn}(x) = \pm 1$ if x is positive or negative, and $\mathrm{sgn}(0) = [0, 1]$. The above example has indeed one and only one solution. □

• *Remark* 5.5. For measurable and locally bounded coefficients satisfying a uniform locally Lipschitz condition in x, i.e., (5.6), the previous proof will apply if an a priori estimate on the $x^r(t)$, corresponding to the truncated coefficients g^r, σ^r and γ^r, can be established, i.e., if it can be proved that

$$P\{\sup_{t_0\le t\le t_1}|x^r(t)|>r\}$$

vanishes as $r\to\infty$. Usually, this is obtained from the Liapunov function, i.e, under the assumption that there exists a nonnegative function $v(t,x)$ such that $v(t,x)$ goes to ∞ locally uniform in t as x goes to ∞ and satisfies

$$\partial_t v(t,x)+A_0(t,x)v(t,x)\le cv(t,x),$$

for some constant $c\ge 0$, where $A_0(t,x)v(t,x)$ represents the integro-differential term obtained from the application of Itô formula, see Remark 5.49. The interested reader may consult the books Khasminskii [130] and Skorokhod [230]. □

• *Remark* 5.6. It may useful to consider coefficients that are unbounded in the control variable and later to take suitable controls process $(v(t):t\ge t_0)$. For instance, instead of the conditions (5.6) and (5.7) we may assume there exists a constant $C>0$ and a nonnegative measurable function $c(t,v)$, increasing in t, such that

$$\begin{cases} 2\,x\,g(t,x,v)+|\sigma(t,x,v)|^2+\displaystyle\int_{\mathbb{R}^m_*}|\gamma(\zeta,t,x,v)|^2\pi(\mathrm{d}\zeta)\le \\ \qquad\qquad\le c(t,v)+C(1+|x|^2), \end{cases} \tag{5.15}$$

for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, and (5.7) on bounded set in v, namely, for any $r>0$ there exists a positive constant $M=M(r)$ such that

$$\begin{cases} |g(t,x,v)-g(t,x',v)|^2+|\sigma(t,x,v)-\sigma(t,x',v)|^2+ \\ \quad+\displaystyle\int_{\mathbb{R}^m_*}|\gamma(\zeta,t,x,v)-\gamma(\zeta,t,x',v)|^2\pi(\mathrm{d}\zeta)\le M|x-x|^2, \end{cases} \tag{5.16}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ with $t\le t_0+r$, $|x|\le r$, $|x'|\le r$, and $|v|\le r$. The function $c(t,v)$ handle the growth of the control processes, e.g., we can allow only controls satisfying

$$\int_{t_0}^{t_1}c(t,v(t))\mathrm{d}t<\infty,$$

for any $t_1>t_0$. For instance, we replace the control process $(v(t):t\ge t_0)$ with $(v_n(t):t\ge t_0)$, $v_n(t)=v(t)$ if $c(n,v(t))\le n$ and $v_n(t)=0$ otherwise, use Theorem 5.3 with v_n and take limit in n. □

• *Remark* 5.7. Sometimes, we need to consider stochastic differential equations with delay, i.e., the drift (perhaps the diffusion and jumps coefficients too) has the form

$$g(s):=g(s,x(s),x(s-\delta),v(s)),\quad \forall t\ge 0,$$

where δ is a given positive constant, or even a more complicate expression. In this case, the initial condition is given on the interval $[-\delta, 0]$ instead of just at time $t = 0$. In general, the coefficients are regarded as adapted functional, where the dependency on the control v becomes irrelevant. This is $g := g(t, x(\cdot), \omega)$, $\sigma := \sigma(t, x(\cdot), \omega)$ and $\gamma := \gamma(\zeta, t, x(\cdot), \omega)$, where $x(\cdot)$ denotes a cad-lag function $s \mapsto x(s)$. Clearly, $g(t, x(\cdot), \omega)$, $\sigma(t, x(\cdot), \omega)$ and $\gamma(\zeta, t, x(\cdot), \omega)$ are adapted and depend only on the path up to t, i.e., $x(s)$ with $0 \le s \le t$. Assume a linear growth condition and a uniform locally Lipschitz condition in the variable x, i.e. there exists a constant $C > 0$ such that

$$\begin{cases} \mathbb{E}\Big\{ |g(t, x(\cdot))|^2 + |\sigma(t, x(\cdot))|^2 + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta, t, x(\cdot))|^2 \pi(\mathrm{d}\zeta) \Big\} \le \\ \qquad\qquad \le C\, \mathbb{E}\Big\{ \sup_{0 \le s \le t} (1 + |x(s)|^2) \Big\} \end{cases} \tag{5.17}$$

for every $t \ge 0$ and any continuous function $s \mapsto x(s)$, and for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\begin{cases} \mathbb{E}\Big\{ |g(t, x(\cdot)) - g(t, x'(\cdot))|^2 + |\sigma(t, x(\cdot)) - \sigma(t, x'(\cdot))|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta, t, x(\cdot)) - \gamma(\zeta, t, x'(\cdot))|^2 \pi(\mathrm{d}\zeta) \Big\} \le \\ \qquad\qquad \le M\, \mathbb{E}\Big\{ \sup_{0 \le s \le t} |x(s) - x'(s)|^2 \Big\}, \end{cases} \tag{5.18}$$

for every $t \ge 0$ and any pair of continuous functions $s \mapsto x(s)$ and $s \mapsto x'(s)$, satisfying

$$\sup_{0 \le s \le t} |x(s)| \le r, \qquad \sup_{0 \le s \le t} |x'(s)| \le r.$$

Under these conditions, we can revise the proof of Theorem 5.3 to obtain an existence and uniqueness result. This includes the so-called *functional* differential equations, for instance, the reader is referred to the book Mao [165, Chapters 5 and 6, pp. 147–232], where the case $\gamma = 0$ (i.e., without jumps) is considered. □

• *Remark* 5.8. If the initial condition is given at a stopping time t_0 instead of a deterministic time t_0, then we can solve the d-dimensional stochastic differential equation (5.5) and Theorems 5.3 and 5.11 remain valid. Actually, we can reformulate (5.3) as

$$\begin{cases} x(t) = x^0 + \displaystyle\int_0^t g(s, x(s), v(s))\, \mathbb{1}_{s \ge t_0} \mathrm{d}s + \\ \qquad + \displaystyle\int_0^t \sigma(s, x(s), v(s))\, \mathbb{1}_{s \ge t_0} \mathrm{d}w(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta, s, x(s), v(s))\, \mathbb{1}_{s \ge t_0} \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{cases} \tag{5.19}$$

for any $t \ge 0$. This means that $x(t)$ is set to x^0 for any $t < t_0$. Moreover, the initial condition x^0 can be replaced by $\xi(t)$, where $\xi(t)$ is an adapted $\mathbb{R}^d$-valued process satisfying $\mathbb{E}\{|\xi(t)|^2\} < C$, for every t. □

Adding a Density

As mentioned early, a Poisson measure ν is a generalization of a compound Poisson process, where the 'distribution of jumps' is clearly defined. However, the coefficient γ in the above stochastic differential equations does not directly affect this distribution of jumps, e.g., to introduce a density with respect to the initial distribution of jumps (generated by ν) is not a simple problem as discussed below.

Let π be a bounded measure on $\mathbb{R}^m_*$ and ϑ a measurable function from $\mathbb{R}^m_*$ into $\mathbb{R}^d$ such that the pre-image preserves bounded-away sets (i.e., for every $b > a > 0$ there exist $\beta > \alpha > 0$ such that $\vartheta^{-1}(\{x \in \mathbb{R}^d_* : a \leq |x| < b\}) \subset \{z \in \mathbb{R}^m_* : \alpha \leq |z| < \beta\}$) and let consider the image measure $\pi_\vartheta(B) = \pi(\vartheta^{-1}(B))$, for any B in $\mathcal{B}(\mathbb{R}^d_*)$.

If $\{\tau_n : n = 1, 2, \dots\}$ is a sequence of independent exponentially distributed (with parameter $\pi(\mathbb{R}^m_*) = c$) random variables, and $\{\zeta_n : n = 1, 2, \dots\}$ is another sequence of independent identically distributed with distribution law $\gamma = \pi/c$ random variables, with $\{\zeta_n : n = 1, 2, \dots\}$ independent of $\{\tau_n : n = 1, 2, \dots\}$, then for $\theta_n := \tau_1 + \tau_2 + \cdots + \tau_n$ (which has a Gamma distribution with parameters c and n), the expressions

$$
\begin{aligned}
&Z(t) := \sum_{n=1}^{\infty} \vartheta(\zeta_n) \mathbb{1}_{t \geq \theta_n}, \quad \text{with} \quad \delta Z(t) := Z(t) - Z(t-) \\
&\delta Z(\theta_n) = \vartheta(\zeta(n)), \quad \text{and} \quad \delta Z(t) = 0 \ \text{ if } t \neq \theta_n, \ \forall n, \ \text{or equivalently} \\
&Z(t) := \zeta_1 + \zeta_2 + \cdots + \zeta_n \ \text{ if } \ \sum_{i=1}^{n} \tau_i = \theta_n \leq t < \theta_{n+1} = \sum_{i=1}^{n+1} \tau_i,
\end{aligned}
$$

are realizations of a compound Poisson process $\{Z(t) : t \geq 0\}$ with Lévy measure $\pi_\vartheta(\mathrm{d}z)$. Its associate point (or jump) process can be written as $\{\delta Z(t) : t \geq 0\}$ and its associate integer measure ν yields the integer-valued random variables

$$
\nu(B_i \times]s_i, t_i]) := \sum_{n=1}^{\infty} \mathbb{1}_{s_i < \theta_n \leq t_i} \, \mathbb{1}_{\vartheta(\zeta_n) \in B_i}, \quad \forall i = 1, 2, \dots, k,
$$

which are independent identically Poisson distributed, with $(t_i - s_i)\pi_\vartheta(B_i)$ parameter (or mean), as long as $\{B_i \times]s_i, t_i] : i\}$ is a family of disjoint Borel sets. More general, if p is a Poisson measure on $\mathbb{R}^m_*$ with Lévy measure π that integrates the function $z \mapsto |z|^2$, $\tilde{p}$ is its associated compensated Poisson measure and $\{\delta\tilde{p}(s) : s \geq 0\}$ is its associated point (or jump) process then the $\mathbb{R}^m$-valued compensated purely jump process

$$
Z(t) = \int_{\mathbb{R}^m_* \times]0,t]} \vartheta(z) \, \tilde{p}(\mathrm{d}z, \mathrm{d}s)
$$

has

$$
p_\vartheta(B \times]a, b]) = \sum_{n=1}^{\infty} \mathbb{1}_{a < s \leq b} \, \mathbb{1}_{\vartheta(\delta Z(s)) \in B},
$$

as its associate integer measure, which satisfies

$$\mathbb{E}\Big\{\int_{\mathbb{R}^m_*\times]0,\infty[} f(z,s)p_\vartheta(\mathrm{d}z,\mathrm{d}s)\Big\} = \mathbb{E}\Big\{\int_0^\infty \mathrm{d}s\int_{\mathbb{R}^m_*} f(z,s-)\pi_\vartheta(\mathrm{d}z)\Big\},$$

for any nonnegative cad-lag optional process f.

In other words, if $\{\delta p(s) : s \geq 0\}$ denotes the Poisson point (or jump) process of the Poisson integer measure $p(B\times]a,b]) = \sum_{s\geq 0}\mathbb{1}_{\delta p(s)\in B}\mathbb{1}_{a<s\leq b}$ with Lévy measure π on $\mathbb{R}^m_*$ then $p_\vartheta(B\times]a,b]) = \sum_{s\geq 0}\mathbb{1}_{\vartheta(\delta p(s))\in B}\mathbb{1}_{a<s\leq b}$ is a Poisson integer measure with Lévy measure $\pi_\vartheta(\mathrm{d}\zeta)$. Actually, $\vartheta = \{\vartheta(\zeta,s) : \zeta\in\mathbb{R}^m_*,\ s\geq 0\}$ could be taken to be a $\mathbb{R}^d_*$-valued predictable process and so

$$\nu_\vartheta(B\times]a,b]) = \sum_{s\geq 0}\mathbb{1}_{\vartheta(\delta p(s),s)\in B}\mathbb{1}_{a<s\leq b},\quad \forall 0\leq a<b,\ B\in\mathcal{B}(\mathbb{R}^d_*),$$

is an integer measure with predictable compensator measure

$$\nu^p_\vartheta(B\times]a,b]) = \int_a^b \pi\big(\{z\in\mathbb{R}^m_* : \vartheta(z,s)\in B\}\big)\mathrm{d}s,$$

for every $0\leq a<b$, B in $\mathcal{B}(\mathbb{R}^d_*)$; and if $m=d$ and a predictable density process $0\leq\rho(z,s)\leq 1$ exists, i.e.,

$$\int_a^b \pi\big(\{z\in\mathbb{R}^m_* : \vartheta(z,s)\in B\}\big)\mathrm{d}s = \int_a^b \mathrm{d}s\int_{\mathbb{R}^m_*}\rho(z,s)\pi(\mathrm{d}z), \tag{5.20}$$

then the predictable compensator measure $\nu^p_\vartheta(\mathrm{d}z,\mathrm{d}s) = \rho(z,s)\pi(\mathrm{d}z)\mathrm{d}s$. Hence, provided the predictable process ϑ can be constructed from a predictable density ρ, the above argument yields a density on the compensator measure.

Therefore, assuming the relation (5.20) between ϑ and ρ, a density can be taken as part of the jump coefficients, i.e., besides $\gamma(\zeta,s,x,v)$ another function $\rho(\zeta,s,x,v)$ with values in $[0,1]$ could be given. This means that the existence and uniqueness Theorem 5.3 holds true under a linear growth condition, namely, there is a constant $C>0$ such that

$$\int_{\mathbb{R}^m_*}|\gamma(\zeta,s,x,v)|^2\rho(\zeta,s,x,v)\pi(\mathrm{d}\zeta)\leq C(1+|x|^2), \tag{5.21}$$

for every (ζ,s,x,v) in $\mathbb{R}^m_*\times[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, and a uniform locally Lipschitz condition, namely, for every $r>0$ there exists a positive constant $M=M(r)$ such that

$$\begin{cases}\displaystyle\int_{\mathbb{R}^m_*}|\gamma(\zeta,s,x,v)\rho(\zeta,s,x,v)-\gamma(\zeta,s,x',v)\rho(\zeta,s,x',v)|^2\pi(\mathrm{d}\zeta)\leq\\ \qquad\qquad\leq M|x-x'|^2,\end{cases} \tag{5.22}$$

for every (s,x,v) and (s,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ with $t\leq t_0+r$, $|x|\leq r$ and $|x'|\leq r$. Indeed, the compensated integer measure

$$\tilde\nu_\rho(\mathrm{d}\zeta,\mathrm{d}s) = \nu_\rho(\mathrm{d}\zeta,\mathrm{d}s) - \rho(\zeta,s,x,v)\pi(\mathrm{d}\zeta)\mathrm{d}s$$

is used instead of the compensated (or martingale) Poisson measure $\tilde{\nu}$. Then remark that the predictable quadratic variation corresponding to the stochastic integral

$$\int_{\mathbb{R}^m\times]0,t]} \gamma(\zeta,s,x(s),v(s))\rho(\zeta,s,x(s),v(s))\tilde{\nu}_\rho(\mathrm{d}\zeta,\mathrm{d}s) = \\ = \int_{\mathbb{R}^m\times]0,t]} \gamma(\zeta,s,x(s-),v(s))\rho(\zeta,s,x(s-),v(s))\tilde{\nu}_\rho(\mathrm{d}\zeta,\mathrm{d}s)$$

is given by

$$\int_0^t \int_{\mathbb{R}^m_*} |\gamma(\zeta,s,x(s-),v(s))|^2\rho(\zeta,s,x(s-),v(s))\pi(\mathrm{d}\zeta),$$

which makes the calculations in Theorem 5.3 valid again.

5.1.2 Mainly Jumps

In this section we consider the case without diffusion, i.e., the stochastic ordinary differential equation (5.3) with the diffusion coefficient $\sigma = 0$, no Wiener process is involved,

$$\begin{cases} x(t) = x^0 + \displaystyle\int_{t_0}^t g(s,x(s),v(s))\mathrm{d}s + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*\times]t_0,t]} \gamma(\zeta,s,x(s),v(s))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall\, t \geq t_0, \end{cases} \tag{5.23}$$

or in differential form as

$$\mathrm{d}x(t) = g(t,x(t),v(t))\mathrm{d}t + \int_{\mathbb{R}^m_*} \gamma(\zeta,t,x(t),v(t))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \tag{5.24}$$

plus the initial condition $x(t_0) = x^0$.

Depending on the integrability of the Levy measure π we can distinguish the case of a finite measure $\pi(\mathbb{R}^m_*) < \infty$, the case of integrable jumps where π integrates the function $\zeta \mapsto |\zeta|$ and the general case. Note that no attention is payed to the "large" jumps, in this analysis, only the "small" jumps are of interest for us (i.e., the Levy measure integrates infinite at our convenience, e.g., π has a compact support in $\mathbb{R}^m_*$).

Finite Levy Measure

Beside assuming that $\pi(\mathbb{R}^m_*) < \infty$, we need a companion hypothesis on the coefficient γ, namely, there exists a constant $C > 0$ such that

$$|\gamma(\zeta,t,x,v)| \leq C|\zeta|(1+|x|^2)^{1/2} \quad \text{and} \quad \int_{\mathbb{R}^m_*} |\zeta|^p\pi(\zeta) < \infty, \tag{5.25}$$

for every ζ, t, x, t, and $p \geq 1$. This last condition ensures that the large jumps have finite moments, i.e., the underlying process associated with the Poisson measure $\nu(\mathrm{d}\zeta, \mathrm{d}t)$,

$$\ell(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta \nu(\mathrm{d}\zeta, \mathrm{d}s) \quad \text{and} \quad \tilde{\ell}(t) = \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \tag{5.26}$$

is a compound Poisson process with parameter $c = \pi(\mathbb{R}^m_*)$ and distribution $\gamma = \pi/c$, so that the characteristic function $\mathbb{E}\{\mathrm{e}^{iy\cdot\ell(t)}\} = \exp\big(-t(\hat{\pi}(y) - c)\big)$. As mentioned in early chapters, there is a canonical construction of the process ℓ, namely, take a sequence $\{\zeta_k : k \geq 1\}$ of independent identically distributed (with law π/c) of random variables, and take another sequence $\{\tau'_i : i \geq 1\}$ (independent from the first one) of independent identically exponentially (with parameter 1) distributed of random variables (so that $\tau_i = c\tau'_i$ is exponentially distributed with parameter $1/c$), to construct $\theta_k = \sum_{i \leq k} \tau_i$ (which has a Gamma distribution with parameters $1/c$ and k), and

$$\ell(t) = \sum_k \zeta_k \mathbb{1}_{t \geq \theta_k} \quad \text{and} \quad \ell(\theta_k) - \ell(\theta_k -) = \zeta_k, \quad \forall k$$

$N(t) = \sum_k \mathbb{1}_{t \geq \theta_k}$ is the counting (of jumps) process, $\mathbb{E}\{\ell(t)\} = t\pi(\mathbb{R}^m_*)$ and $\tilde{\ell}(t) = \ell(t) - t\pi(\mathbb{R}^m_*)$, where $t_0 = 0$ have been chosen, for simplicity.

By observing either ℓ or $\tilde{\ell}$ only at the points of discontinuity, and by means of the coefficient $\gamma(\zeta, t, x, v)$, construct the process

$$\begin{cases} y(\theta_n) = y(\theta_{n-1}) + \gamma\big(\zeta_n, \theta_{n-1}, y(\theta_{n-1}), v(\theta_{n-1})\big), \quad \forall n \geq 1, \\ y(t) = y(\theta_{n-1}), \quad \theta_{n-1} \leq t < \theta_n, \end{cases} \tag{5.27}$$

for a given initial $y(\theta_0)$ with $\theta_0 = 0$. This is a jump process satisfying

$$y(t) = y(0) + \int_{\mathbb{R}^m_* \times]0,t]} \gamma\big(\zeta, s, y(s-), v(s)\big)\nu(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall t \geq 0,$$

or equivalently

$$y(t) = y(0) + \int_{\mathbb{R}^m_* \times]0,t]} \gamma\big(\zeta, s, y(s-), v(s)\big)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \int_{]0,t]} \mathrm{d}s \int_{\mathbb{R}^m_*} \gamma\big(\zeta, s, y(s-), v(s)\big)\pi(\mathrm{d}\zeta), \quad \forall t \geq 0.$$

Note that the integral in ν could be pathwise interpreted (and in the Riemann-Stieltjes sense if $s \mapsto \gamma(\zeta, s, x, v(s))$ is left-hand continuous for any ζ and x), the integral in $\mathrm{d}s$ is in the Lebesgue sense, and the integral in $\pi(\mathrm{d}\zeta)$ is in the measure sense, but the integral in $\tilde{\nu}$ is a stochastic integral. Thus, the expression (5.27) makes sense when the function $s \mapsto \gamma(\zeta, s, x, v(s))$ has been taken left-hand continuous. Naturally, this is a piecewise deterministic process as in Section 3.9.

An interesting case is when the jump coefficient is proportional to ζ, i.e., $\gamma_i(\zeta,t,x,v)=\sum_{k=1}^m \bar\gamma_{ik}(t,x,v)\zeta_k$. The integrals with respect to ν and $\tilde\nu$ become integrals with respect to $\mathrm{d}\ell$ and $\mathrm{d}\tilde\ell$, e.g.,

$$y(t)=y(0)+\int_{]0,t]}\bar\gamma\big(s,y(s-),v(s)\big)\mathrm{d}\ell(s),\quad \forall t\ge 0,$$

which is a pathwise integral with respect to a compound Poisson process.

This procedure yields an algorithm to solve stochastic ordinary differential equation with only drift g and jump γ coefficients. Indeed, define

$$\tilde g(t,x,v)=g(t,x,v)-\int_{\mathbb{R}_*^m}\gamma(\zeta,t,x,v)\pi(\mathrm{d}\zeta)$$

and consider the evolution operator $\varphi(t;s,x)$ given by the solution of the ordinary differential equation with drift $\tilde g$, i.e., $\varphi(s;s,x)=x$ and

$$\dot\varphi(t;s,x)=\tilde g\big(t,\varphi(t;s,x),v(t)\big),\quad \forall t>s.$$

Now, set

$$\begin{aligned}&y(\theta_n)=y(\theta_{n-1})+\gamma\big(\zeta_n,\theta_{n-1},y(\theta_{n-1}),v(\theta_{n-1})\big),\quad \forall n\ge 1,\\&y(t)=\varphi\big(t;\theta_{n-1},y(\theta_{n-1})\big),\quad \forall\theta_{n-1}\le t<\theta_n\end{aligned}$$

or equivalently

$$\begin{cases}y(\theta_n)=y(\theta_{n-1})+\gamma\big(\zeta_n,\theta_{n-1},y(\theta_{n-1}),v(\theta_{n-1})\big),\quad \forall n\ge 1,\\ \dot y(t)=g\big(t,y(t),v(t)\big)+\\ \qquad\qquad+\displaystyle\int_{\mathbb{R}_*^m}\gamma\big(\zeta,t,y(t),v(t)\big)\pi(\mathrm{d}\zeta),\quad \forall\theta_{n-1}\le t<\theta_n.\end{cases}\tag{5.28}$$

with the initial $y(\theta_0)$ with $\theta_0=0$. This process $(y(t):t\ge 0)$ is the solution of the stochastic ordinary differential equation (5.23) with $t_0=0$.

It is clear that we can adjust (5.28) when the diffusion coefficient σ does not vanish, but in between two consecutive jumps we have to solve a stochastic differential equation with a drift and a diffusion coefficients.

Integrable Jumps

This is the case where we keep the condition (5.25), but we do not assume a finite measure, i.e., we may have $\pi(\mathbb{R}_*^m)=\infty$. The integral with respect to the random measure ν in $\mathbb{R}_*^m\times(0,\infty)$ is well defined, i.e.,

$$\int_{\mathbb{R}_*^m\times]0,t]}|\gamma(\zeta,s,x,v)|\nu(\mathrm{d}\zeta,\mathrm{d}s)<\infty,\quad \text{a.s.},$$

and if the processes $(x(t) : t \geq 0)$ and $(v(t) : t \geq 0)$ are predictable with respect to ν (or equivalently to the Levy process ℓ) then

$$\mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times]0,t]} |\gamma(\zeta, s, x(s), v(s))| \nu(\mathrm{d}\zeta, \mathrm{d}s) \Big\} = \\ = \int_0^t \mathrm{d}s \int_{\mathbb{R}^m_*} \mathbb{E}\big\{ |\gamma(\zeta, s, x(s), v(s))| \big\} \pi(\mathrm{d}\zeta).$$

As early, if ℓ is defined by (5.26) then $\gamma_i(\zeta, t, x, v) = \sum_{k=1}^m \bar\gamma_{ik}(t, x, v)\zeta_k$ then

$$\int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta, s, x(s), v(s)) \nu(\mathrm{d}\zeta, \mathrm{d}s) = \int_{]0,t]} \bar\gamma(s, x(s), v(s)) \mathrm{d}\ell(s), \quad \text{and}$$
$$\mathbb{E}\Big\{ \int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta, s, x(s), v(s)) \nu(\mathrm{d}\zeta, \mathrm{d}s) \Big\} = \sum_{k=1}^m \int_0^t \mathbb{E}\big\{ \bar\gamma_{\cdot k}(s, x(s), v(s)) \big\} \mathrm{d}s.$$

However, the pathwise solution to an ordinary differential equation of the form

$$y(t) = y(0) + \int_0^t g\big(s, y(s), v(s)\big) \mathrm{d}s + \int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta, s, y(s), v(s)) \nu(\mathrm{d}\zeta, \mathrm{d}s)$$

could present other difficulties, e.g. see papers Barron et al. [11], Dal Maso and Rampazzo [54], and Schmaedke [222], which do not apply here. For instance, assuming that $\gamma(\zeta, t, x, v)$ is uniformly Lipschitz continuous in x, i.e.,

$$|\gamma(\zeta, t, x, v) - \gamma(\zeta, t, x', v)| \leq M|\zeta|\,|x - x'|, \quad \forall x, x', \zeta, t, v,$$

and for some constant $M > 0$, the arguments of Theorem 5.3 can be applied to the operator

$$T(y)(t) = y(0) + \int_0^t g\big(s, y(s), v(s)\big) \mathrm{d}s + \int_{\mathbb{R}^m_* \times]0,t]} \gamma(\zeta, s, y(s), v(s)) \nu(\mathrm{d}\zeta, \mathrm{d}s),$$

on the Banach space $\mathcal{Y}$, but with a L^1-type (instead of a L^2-type) norm

$$\|y\|_{\mathcal{Y}} := \mathbb{E}\{ \sup_{0 \leq t \leq t_1} \mathrm{e}^{-\alpha t} |y(t)| \},$$

as in the case of an ordinary differential equation, i.e., essentially by means of the estimate

$$\mathbb{E}\big\{ |T(y)(t) - T(y')(t)| \big\} \leq M \int_0^t \mathbb{E}\big\{ |y(s) - y'(s)| \big\} \mathrm{d}s$$

the theory of existence and uniqueness (for integrable jumps) follows.

It is clear that the above argument get lost if we add a diffusion coefficient, however, the real gain is the fact that for integrable jumps we can consider stochastic ordinary differential equation like (5.3) with the initial Poisson measure ν instead of the martingale Poisson measure $\tilde\nu$.

General Jumps

In this case, the expression of ℓ given by (5.26) is useless, only the Levy process $\tilde{\ell}$ is meaningful, and the stochastic ordinary differential equation has to be considered with the martingale Poisson measure $\tilde{\nu}$ as in (5.3).

However, if $\gamma_i(\zeta,t,x,v)=\sum_{k=1}^m \bar{\gamma}_{ik}(t,x,v)\zeta_k$ and we assume given the Levy process $\tilde{\ell}$ with characteristic $(0,0,\pi)$, i.e., a 'purely' jump process in $\mathbb{R}^m_*$, then the stochastic ordinary differential equation (5.3) becomes

$$\begin{cases} x(t)=x^0+\displaystyle\int_{t_0}^t g(s,x(s),v(s))\mathrm{d}s+\int_{t_0}^t \sigma(s,x(s),v(s))\mathrm{d}w(s)+ \\ \qquad\qquad +\displaystyle\int_{]t_0,t]} \bar{\gamma}(s,x(s),v(s))\mathrm{d}\tilde{\ell}(s), \quad \forall\, t\ge t_0, \end{cases} \tag{5.29}$$

or in differential form as

$$\begin{cases} \mathrm{d}x(t)=g(t,x(t),v(t))\mathrm{d}t+\sigma(t,x(t),v(t))\mathrm{d}w(t)+ \\ \qquad\qquad +\bar{\gamma}(t,x(t),v(t))\mathrm{d}\tilde{\ell}(t), \end{cases} \tag{5.30}$$

plus the initial condition $x(t_0)=x^0$. The assumptions (5.6) and (5.7) becomes

$$|g(t,x,v)|+|\sigma(t,x,v)|+|\bar{\gamma}(t,x,v)|\le C(1+|x|), \tag{5.31}$$

$$\begin{cases} |g(t,x,v)-g(t,x',v)|+|\sigma(t,x,v)-\sigma(t,x',v)|+ \\ \qquad\qquad +|\bar{\gamma}(t,x,v)-\gamma(t,x',v)|\le M|x-x'|, \end{cases} \tag{5.32}$$

which are more natural in some way, but yet, particular cases.

In any case, approximate the measure ν with ν_ε corresponding to the finite Lévy measure $\pi_\varepsilon(\mathrm{d}\zeta)=\mathbb{1}_{|\zeta|\ge\varepsilon}\pi(\mathrm{d}\zeta)$ to be able to use the construction iterating the jumps (or interlacing) derived from piecewise deterministic processes. Moreover, as $\varepsilon\to 0$ good estimates can be obtained, see later Remark 5.14 and compare with the analysis in the book by Applebaum [6].

Discrete Jumps

Keeping $\sigma=0$ and $g=\pi(\gamma)$ and going back to the case of finite measure, we consider the particular case when the support of the Lévy measure π is $\mathbb{Z}^m_*$, with $\mathbb{Z}$ is the integer numbers. In this case, we are dealing with a continuous time Markov chain with states in $\mathbb{Z}^m$.

Use the notation of the subsection on finite measure to construct the jump process ℓ

$$\ell(t)=\int_{\mathbb{Z}^m_*\times]0,t]}\zeta\nu(\mathrm{d}\zeta,\mathrm{d}s)=\sum_k \zeta_k\mathbb{1}_{t\ge\theta_k},$$

and then the evolution

$$\begin{cases} y(t)=y(\theta_{n-1}), \quad \forall\theta_{n-1}\le t<\theta_n, \\ y(\theta_n)=y(\theta_{n-1})+\gamma\big(\zeta_n,\theta_{n-1},y(\theta_{n-1}),v(\theta_{n-1})\big), \quad \forall n\ge 1, \end{cases} \tag{5.33}$$

following the proper jumps of ℓ instead of the centered jumps of $\tilde{\ell}$. To make this recurrent equation equivalent to the differential equation (5.23) we should take

$$g(t,x,v) = \int_{\mathbb{Z}_*^m} \gamma(\zeta,t,x,v)\pi(\mathrm{d}\zeta),$$

$\theta_0 = t_0$ and $y(\theta_0) = x^0$.

Certainly, if the initial condition $y(0)$ belongs to $\mathbb{Z}$ then the whole evolution is in $\mathbb{Z}$, and the differential equation is not quite meaningful. The solution $(y(t) : t \geq 0)$ is a controlled Markov chain in continuous time with states in $\mathbb{Z}^m$.

The Poisson measure ν corresponding to a finite Levy measure π with support in $\mathbb{Z}_*^m$ is indeed a Markov chain in continuous time with states in $\mathbb{Z}^m$, satisfying $\mathbb{E}\{\nu(A,\mathrm{d}s)\} = \pi(A)\mathrm{d}s$, for any subset A of (the denumerable space) $\mathbb{Z}_*^m$. Perhaps, we should remark that the cad-lag jump process $\ell = (\ell_i : i = 1,\dots,m)$ satisfies (a) $\{\ell_i(t) : i = 1,\dots,m\}$ are independent Poisson random variables with parameters

$$t\int_{\mathbb{Z}_*^m} \zeta_i \pi(\mathrm{d}\zeta), \quad i = 1,\dots,m,$$

and (b) $\ell(s)$ is independent of $\ell(t) - \ell(s)$ for every $0 \leq s < t$.

It should be clear that the emphasis is on the jumps of a process when interpreting a Markov chain as a Poisson measure, and this requires the addition operation to be defined on the state space $\mathbb{Z}^m$. Sometimes, it is usually better to interpret the Markov chain in term of the transition probability infinite matrix $(p_{ij} : i,j \in \mathbb{Z}^m\}$, where p_{ij} is the probability of *jumping* from the state i to the state j, in this case, the "addition" of jumps is not necessary.

5.1.3 Modeling with Jumps

In this section we take a quick look at some alternative way of presenting the jumps. Particularly, first we discuss the "problem" with the exponential function when jumping and then we present other possible settings for modeling the jumps.

Exponential Expressions

In the same way that for a given constant a in $\mathbb{R}$, the exponential function $t \mapsto \mathrm{e}^{at}$ is the solution $x(\cdot)$ of the ODE $\dot{x}(t) = ax(t)$ with initial condition $x(0) = 1$, i.e.,

$$\mathrm{e}^{at} = 1 + \int_0^t a\mathrm{e}^{as}\mathrm{d}s, \quad \forall t \in \mathbb{R},$$

the function $t \mapsto \mathrm{e}^{a\alpha(t)}$ is the solution of the ODE $\dot{x}(t) = ax(t)\dot{\alpha}(t)$ with initial condition $x(0) = 1$, i.e.,

$$\mathrm{e}^{a\alpha(t)} = 1 + \int_0^t a\mathrm{e}^{a\alpha(s)}\mathrm{d}\alpha(s), \qquad \forall t \in \mathbb{R},$$

for a given differentiable function $t \mapsto \alpha(t)$ with $\alpha(0) = 0$. However, the exponential expression $\exp[a\alpha(t)] = \mathrm{e}^{a\alpha(t)}$ is meaningful when α is not necessarily continuous, and in particular, whenever α is a cad-lag (or cag-lad) function. Thus one may ask in which sense the ODE should be understood. To make sense of the ODE as a Riemann-Stieltjes integral, the first correction has to do with the jumps, i.e., the proposed ODE is as follows

$$y(t) = 1 + \int_{]0,t]} ay(s-)\mathrm{d}\alpha(s), \qquad \forall t > 0. \tag{5.34}$$

However, there are other ways of fixing this problem, e.g., see papers Barron et al. [11], Dal Maso and Rampazzo [54], and Schmaedke [222].

Therefore, if α is a continuous function of bounded variation on any bounded time-interval then the exponential function $t \mapsto \exp[a\alpha(t)]$ is the solution of the ODE (5.34), where the integral is understood in the Riemann-Stieltjes sense.

For the jumps themselves, take a function ℓ with a sequence of successive jumps $\zeta_0, \zeta_1, \dots, \zeta_k, \dots$, at the times $0 = \theta_0 < \theta_1 < \cdots < \theta_k < \cdots$, i.e.,

$$\ell(t) = \sum_k \zeta_k \mathbb{1}_{t \geq \theta_k} \quad \text{and} \quad \ell(\theta_k) - \ell(\theta_k-) = \zeta_k, \quad \forall k \geq 0,$$

and consider the recursive formula

$$y(\theta_0) = 1, \quad \text{and} \quad y(\theta_n) = y(\theta_{n-1}) + a\zeta_n y(\theta_{n-1}), \quad \forall n \geq 1,$$

where the jump at $t = 0$ does not intervene. This yields the close formula $y(t) = \prod_{k\geq 1}[1 + a\zeta_k]\mathbb{1}_{t\geq\theta_k}$. Note that (1) if $a\zeta_n = -1$ for some n then $y(\theta_k) = 0$ for any $k \geq n$ and (2) if $a\zeta_n < -1$ for some n then $y(\theta_n) < 0$. Hence, assuming $a\zeta_k > -1$ for any k, the expression

$$y(t) = \prod_k [1 + a\zeta_k]\mathbb{1}_{t\geq\theta_k} = \exp\Big[\sum_{k\geq 1} \ln(1 + a\zeta_k)\mathbb{1}_{t\geq\theta_k}\Big]$$

provides the solution of the ODE (5.34) when $\alpha = \ell$. Moreover,

$$y(t) = \prod_{0<s\leq t} [1 + a\delta\ell(t)] = \exp\Big[\sum_{0<s\leq t} \ln\big(1 + a\delta\ell(s)\big)\Big],$$

where δ is the jump-operator, i.e., $\delta\ell(t) = \ell(t) - \ell(t-)$. Furthermore, if ℓ is a jump-function with bounded variation over any bounded time interval, with jump ζ_k at time θ_k such that $a\zeta_k > -1$ and

$$\sum_k \ln(1 + a\zeta_k)\mathbb{1}_{t\geq\theta_k} < \infty, \quad \forall t > 0,$$

then the previous formula holds true.

Thus, if α is a cad-lag function with bounded variation over any bounded time interval, and α_c denotes its continuous part, i.e., $\delta\alpha(t) = \alpha(t) - \alpha(t-)$ and $\alpha_c(t) = \alpha(t) - \sum_{s\leq t} \delta\alpha(s)$, then the exponential expression

$$y(t) = \mathrm{e}^{a\alpha_c(t)} \prod_{0<s\leq t} [1 + a\delta\alpha(t)]$$

or equivalently

$$y(t) = \exp\big[a\alpha_c(t) + \sum_{0<s\le t} \ln\big(1 + a\delta\alpha(s)\big)\big], \tag{5.35}$$

is the solution of the ODE (5.34), provided the jumps satisfy

$$a\delta\alpha(t) > -1 \quad \text{and} \quad \sum_{0<s\le t} \ln\big(1 + a\delta\alpha(s)\big) < \infty, \quad \forall t > 0.$$

Note that since α has bounded variation and $\ln(1+a\delta\alpha(s)) \le a\delta\alpha(s)$, the above convergence is satisfied under the condition: for every $t > 0$ there exists $\varepsilon > 0$ such that $a\delta\alpha(s) > -1+\varepsilon$ for every $0 < s \le t$. Moreover, if $0 = t_0 < t_1 < \cdots < t_k < \cdots < t_{n-1} < t_n = t$ is a partition of the time-interval $[0, t]$ and $\alpha(t) = \alpha(t_i)$ for any t in $]t_{i-1}, t_i]$ so that $\delta\alpha(t_i) = \alpha(t_i) - \alpha(t_{i-1})$, then

$$\Big[1 + \sum_{k=1}^{n}\prod_{i=k}^{n}\big(\alpha(t_i) - \alpha(0)\big)\Big] = \prod_{k=1}^{n}\big[1 + \delta\alpha(t_k)\big] =$$
$$= \exp\Big[\sum_{k=1}^{n}\delta\alpha(t_k)\Big]\exp\Big[\sum_{k=1}^{n}\ln\big(1 + \delta\alpha(t_k)\big) - \delta\alpha(t_k)\Big],$$

which proves that the limit

$$\lim_{\max_i\{t_i - t_{i-1}\}\to 0}\Big[1 + \sum_{k=1}^{n}\prod_{i=k}^{n}\big(\alpha(t_i) - \alpha(0)\big)\Big],$$

is equal to the exponential expression $y(t)$ given by (5.35), since

$$\sum_{k=1}^{n}|\delta\alpha_c(t_k)|^2 \le \big(\max_k\{|\alpha(t_k) - \alpha(t_{k-1})|\}\big)\sum_{k=1}^{n}|\delta\alpha_c(t_k)|,$$

vanishes as the mesh of the partition goes to zero.

Therefore, if the integer measure $\nu(\mathrm{d}\zeta, \mathrm{d}s)$ in $\mathbb{R} \times (0, \infty)$ is only integrable (jumps with bounded variation), the processes

$$\ell(t) = \int_{\mathbb{R}_*\times]0,t]}\zeta\nu(\mathrm{d}\zeta, \mathrm{d}s) \quad \text{and} \quad \tilde{\ell}(t) = \int_{\mathbb{R}_*\times]0,t]}\zeta\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s),$$

are well defined, and the mean process $t \mapsto \alpha(t) = \mathbb{E}\{\ell(t)\}$ is a continuous function of bounded variation on any bounded time-interval (actually $\alpha(t) = t$ for an integrable or bounded variation Poisson random measure), and the previous argument can be used. However, besides $\nu(]-\infty, -1/a]\times]0, t]) = 0$, for every $t > 0$, the condition

$$\int_{\mathbb{R}_*\times]0,t]}\big[|\zeta| + \ln\big(1 + a|\zeta|\big)\big]\nu(\mathrm{d}\zeta, \mathrm{d}s) < \infty, \quad \forall t > 0, \tag{5.36}$$

is necessary to deduce that the exponential expression

$$\begin{cases} y(t) = \exp\Big[\displaystyle\int_{\mathbb{R}_*\times]0,t]}\ln(1 + a\zeta)\nu(\mathrm{d}\zeta, \mathrm{d}s)\Big], \quad \forall t > 0, \quad \text{or} \\ y(t) = \exp\Big[-a\mathbb{E}\{\ell(t)\} + \displaystyle\int_{\mathbb{R}_*\times]0,t]}\ln(1 + a\zeta)\nu(\mathrm{d}\zeta, \mathrm{d}s)\Big], \quad \forall t > 0, \end{cases} \tag{5.37}$$

is the pathwise-solution of the differential equation (5.34) for either $\alpha = \ell$ or $\alpha = \tilde{\ell}$, where the integral is considered in the Riemann-Stieltjes sense and agrees with the stochastic integral when the jumps are integrable.

Now, recall Itô formula with jumps (see Theorem 4.30) for a one-dimension standard Wiener process w, a Poisson measure $p(\mathrm{d}\zeta, \mathrm{d}s)$ in $\mathbb{R}\times(0,\infty)$ with Lévy measure π, and a drift g, diffusion σ and jumps γ, i.e., if

$$x(t) = x(0) + \int_0^t g(s)\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}w(s) + \int_{\mathbb{R}_*\times]0,t]} \gamma(\zeta, s)\nu(\mathrm{d}\zeta, \mathrm{d}s)$$

then the process $y(t) = \mathrm{e}^{ax(t)}$ has the stochastic differential

$$\begin{aligned}\mathrm{d}y(t) = y(t)\Big[ag(t) + \frac{a\sigma^2}{2} + \int_{\mathbb{R}_*} [\mathrm{e}^{a\gamma(\zeta,t)} - 1 - a\gamma(\zeta,t)]\pi(\mathrm{d}\zeta)\Big]\mathrm{d}t + \\ + y(t-)\Big[a\sigma\mathrm{d}w(t) + \int_{\mathbb{R}_*} [e^{a\gamma(\zeta,t)} - 1]\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t)\Big], \quad t > 0.\end{aligned}$$

Assuming that g, σ and γ are constant functions in t and that $\mathrm{e}^{a\gamma}$ is integrable as $\gamma \to \infty$ with respect to the Lévy measure π, the exponential expression

$$\begin{aligned}y(t) = \exp\Big[a\sigma w(t) + a\int_{\mathbb{R}_*\times]0,t]} \gamma(\zeta)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \Big(ag - \frac{a\sigma^2}{2} - \int_{\mathbb{R}_*} \big[\mathrm{e}^{a\gamma(\zeta)} - 1 - a\gamma(\zeta)]\pi(\mathrm{d}\zeta)\Big)t\Big]\end{aligned}$$

is the solution of the stochastic differential equation

$$dy(t) = y(t-)\Big[ag\mathrm{d}t + a\sigma\mathrm{d}w + \int_{\mathbb{R}_*} [\mathrm{e}^{a\gamma(\zeta)} - 1]\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t)\Big],$$

with the initial condition $y(0) = 1$.

Hence, making the change of jumps $\mathrm{e}^{a\gamma} - 1 = a\bar{\gamma}$ or equivalently $a\gamma = \ln(1 + a\bar{\gamma})$ the solution of the stochastic differential equation

$$dy(t) = y(t-)\Big[ag\mathrm{d}t + a\sigma\mathrm{d}w + a\int_{\mathbb{R}_*} \bar{\gamma}(\zeta)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}t)\Big], \tag{5.38}$$

with the initial condition $y(0) = 1$ is given by the exponential expression

$$\begin{cases} y(t) = \exp\Big[a\sigma w(t) + a\displaystyle\int_{\mathbb{R}_*\times]0,t]} \ln\big(1 + a\bar{\gamma}(\zeta)\big)\tilde{p}(\mathrm{d}\zeta, \mathrm{d}s) + \\ \qquad + a\Big(g - \dfrac{\sigma^2}{2} - \displaystyle\int_{\mathbb{R}_*} \big[\bar{\gamma}(\zeta) - \ln\big(1 + a\bar{\gamma}(\zeta)\big)\big]\pi(\mathrm{d}\zeta)\Big)t\Big], \end{cases} \tag{5.39}$$

as long as the integrals are meaningful, i.e.,

$$a\bar{\gamma}(\zeta) > -1 \quad \text{and} \quad \int_{\mathbb{R}_*} \big|\bar{\gamma}(\zeta) - \ln\big(1 + a\bar{\gamma}(\zeta)\big)\big|\pi(\mathrm{d}\zeta) < \infty$$

for some suitable constant a. Note that as $\bar{\gamma} \to 0$ the integrand is bounded by $|\bar{\gamma}|^2$, but the problem with the integral is as $\bar{\gamma} \to -1/a$.

Usually, (5.39) is called the stochastic (or martingale when $g = 0$) exponential of the Lévy process ℓ (which has drift g, diffusion σ and jumps $\bar{\gamma}$ with Lévy measure π), while the pathwise exponential (5.37) or (5.35) is referred to as the "Riemann-Stieltjes exponential" of the cad-lag bounded variation process ℓ (with integer measure ν) or α.

Centered Lévy Models

In most models, the continuous-type noise is represented by a standard Wiener process w in $\mathbb{R}^n$, which can be regarded as n-independent one-dimensional standard Wiener processes $w_1, \dots, w_n$. Similarly, the jump-type noise is represented by Poisson martingale $\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t)$ in $\mathbb{R}^m$ (constructed from a Poisson measure $\nu(\mathrm{d}\zeta, \mathrm{d}t)$ with Lévy measure π in $\mathbb{R}^m_*$), of which a particular situation is the case of m-independent one-dimensional Poisson martingales $\tilde{\nu}_i(\mathrm{d}\zeta_i, \mathrm{d}t)$ (constructed from a Poisson measure $\nu_i(\mathrm{d}\zeta_i, \mathrm{d}t)$ with Lévy measure π_i on $\mathbb{R}_*$).

Alternatively, the purely jump (Poisson) martingales obtained by the stochastic integral

$$\tilde{p}_i(t) = \int_{\mathbb{R}_* \times]0,t]} \zeta_i \tilde{\nu}_i(\mathrm{d}\zeta_i, \mathrm{d}t), \quad i = 1, 2, \dots, m,$$

could be used as the model of the Poisson noise. This means that on a probability space $(\Omega, \mathcal{F}, P)$, with a filtration $\mathbb{F} = (\mathcal{F}_t, t \geq 0)$ (satisfying the usual conditions) we assume that there are n-independent one-dimensional standard Wiener processes $w_1, \dots, w_n$, which are independent of m-independent one-dimensional purely jump (Poisson) martingales $\tilde{p}_1, \dots, \tilde{p}_m$, with predictable quadratic variation given in term of the Lévy measures π_i by the integral

$$\mathrm{d}\langle \tilde{p}_i \rangle = \Big(\int_{\mathbb{R}_*} |\zeta_i|^2 \pi_i(\mathrm{d}\zeta_i) \Big) \mathrm{d}t.$$

These processes $w_1, \dots, w_n$ and $\tilde{p}_1, \dots, \tilde{p}_m$ could model the Wiener-Poisson noise, and the coefficients for the jumps-part become a matrix $\gamma(t, x, v) := (\gamma_{ik}(t, x, v),\ i = 1, \dots, d,\ k = 1, \dots, m)$ and the stochastic ordinary differential equation (5.5) takes the integral form

$$\begin{cases} x_i(t) = x_i^0 + \displaystyle\int_{t_0}^t g_i(s, x(s), v(s))\mathrm{d}s + \\ \qquad + \displaystyle\sum_{k=1}^n \int_{t_0}^t \sigma_{ik}(s, x(s), v(s))\mathrm{d}w_k(s) + \\ \qquad + \displaystyle\sum_{k=1}^m \int_{]t_0,t]} \gamma_{ik}(s, x(s), v(s))\mathrm{d}\tilde{p}_k(s), \quad \forall\, t \geq t_0, \end{cases} \tag{5.40}$$

for $i = 1, \dots, d$, or a differential form as

$$\begin{cases} \mathrm{d}x(t) = g(t, x(t), v(t))\mathrm{d}t + \sigma(t, x(t), v(t))\mathrm{d}w(t) + \\ \qquad\qquad + \gamma(t, x(t), v(t))\mathrm{d}\tilde{p}(t), \end{cases} \tag{5.41}$$

plus the initial condition $x(t_0) = x^0$.

In this setting, the assumptions (5.6) and (5.6) used in Theorem 5.3 on the matrices σ and γ are exactly the same, i.e., a linear growth condition, namely, there exists a constant $C > 0$ such that

$$|g(t,x,v)| + |\sigma(t,x,v)| + |\gamma(t,x,v)| \le C(1+|x|), \tag{5.42}$$

for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, and a uniform locally Lipschitz condition in the variable x, namely, for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\begin{cases} |g(t,x,v) - g(t,x',v)| + |\sigma(t,x,v) - \sigma(t,x',v)| + \\ \qquad\qquad +|\gamma(t,x,v) - \gamma(t,x',v)| \le M|x-x'|, \end{cases} \tag{5.43}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ with $t \le t_0 + r$, $|x| \le r$ and $|x'| \le r$. However, the interpretation of the jumps takes a different form. Indeed, the source jumps for the noise are the jumps of the Poisson measure, i.e., the Poisson point-process $\delta p(t)$, which yields the Poisson measure

$$p(K\times]0,t]) = \sum_{0<s\le t} \mathbb{1}_{\delta p(s)\in K}$$

for any compact subset K of $\mathbb{R}^m_*$, and then, the purely jump (Poisson) martingale $\tilde p = (\tilde p_1,\dots,\tilde p_m)$ via the stochastic integral. Therefore, a coefficient of the vector-form $\gamma(z,t,x,v)$ transforms the Poisson point-process $\{\delta p(t),\, t\ge 0\}$ into the point process $\{\gamma(\delta p(t),t,x,v),\, t\ge 0\}$, while a coefficient of the matrix-form $\gamma(z,t,x,v)$ transforms the purely jump (Poisson) martingales $\tilde p_i$ with predictable quadratic (scalar) variation $\mathrm{d}\langle\tilde p_k\rangle(t)$ into the purely jump (Poisson) martingale with predictable quadratic matrix-variation

$$\Big(\sum_{k=1}^m \gamma_{ik}(t,x,v)\gamma_{ik}(t,x,v)\langle\tilde p_k\rangle\Big)\mathrm{d}t,$$

i.e., it is clear the relation $\gamma(z,t,x,v) = \gamma(t,x,v)z$ between the two expressions of the jump-coefficients.

Similarly, the noise could be represented directly as a Lévy process with characteristic $(0,Q,\pi)$, having a continuous-part w and a jump-part $\tilde p$, i.e., a cad-lag process, where $w = (w_1,\dots,w_n)$ is an n-dimensional Wiener process with drift zero and covariance matrix Q, namely,

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}y\cdot w(t)}\big\} = \exp\big[-t\frac{Qy\cdot y}{2}\big],\quad \forall t>0,\ y\in\mathbb{R}^n,$$

and $\tilde p = (\tilde p_1,\dots,\tilde p_m)$ is a compensated purely jump Poisson process with Lévy measure π (which the stochastic integral of the jumps of a Poisson measure ν in $\mathbb{R}^m_*$), namely,

$$\mathbb{E}\big\{\mathrm{e}^{\mathtt{i}z\cdot\tilde p(t)}\big\} = \exp\Big[-t\int_{\mathbb{R}^m_*}(1-\mathrm{e}^{\mathtt{i}z\cdot\zeta}+\mathtt{i}z\cdot\zeta)\pi(\mathrm{d}\zeta)\Big],\quad \forall t>0,\ z\in\mathbb{R}^m.$$

Note that as mentioned early, the Lévy measure π is assumed to integrate the function $\zeta \mapsto |\zeta| \wedge |\zeta|^2$ on $\mathbb{R}^m_*$, and moreover, $n = m$ can be imposed without any loss of generality.

In this context, there is not problem in considering $n = m = \infty$ and replacing the Euclidean norm $|\cdot|$ with the Hilbert norm in ℓ^2 (the space of sequences $\{a_i\}$ with $\sum_i |a_i|^2 < \infty$). Essentially, the same existence and uniqueness result can be shown for coefficients σ and γ regarded as continuous linear operator from ℓ^2 into itself, for each fixed t, x, v (and z with some adaptation). This effectively includes stochastic differential equation in Hilbert spaces, but, to include stochastic partial differential equation the coefficient becomes unbounded linear operator, and thus, the SPDE theory has substantial changes with respect to SODE. There is a vast bibliography on SPDE, for instance, the reader may check the books by Chow [42], Da Prato and Zabczyk [53], Kallianpur and Xiong [123], Peszat and Zabczyk [197], among many others.

Types of Jumps

First note that the initial measure π on $\mathbb{R}^m_*$ (or any other more general space) used in the definition (or construction) of Poisson measure relative could be supposed only to be σ-finite. We required that π be a Lévy measure to clarify (simplify or complicate) the setting, i.e., the condition that π integrate the function $|x|^2$ or more general $|x|^2 \wedge 1$. This condition is not necessary for the stochastic ODE, this is incorporated with the condition on the coefficient γ when requiring square-integrability with respect to π. The Lévy measure of the "controlled" diffusion process generated by the stochastic ODE is actually given by the image kernel $M(t, x, v, B) = \pi(\{\zeta : \gamma(\zeta, t, x, v) \in B\})$, for any Borel subset B of $\mathbb{R}^d_*$. The condition that π is a Lévy measure help to tract the dependency on the parameters t, x, v, in the sense that two types of jumps can be distinguished, i.e., when a positive and arbitrary constant $\varepsilon > 0$ is fixed, the condition $|\zeta| \geq \varepsilon > 0$ or $|\gamma| \geq \varepsilon > 0$ yields "large" jumps, while the opposite inequality defines the so-called "small" jumps. It is perhaps important to recall that the stochastic integral for jumps processes is designed only with the small jumps, and for the large jumps, the Riemann-Stieltjes sense is used (or Lebesgue-Stieltjes sense when the integrand is kept left-continuous), as discussed in previous sections.

Therefore, it could make sense to decompose the jump-coefficient γ into two pieces, the small jumps (still denoted by) $\gamma(\zeta, t, x, v)$ and the large jumps, which are further divided into two categories, the really large jumps $\alpha(\zeta, t, x, v)$ and the relative small jumps $\beta(\zeta, t, x, v)$ which are integrable. In other words, this means that for some $\varepsilon > 0$, (a) the new coefficient $\gamma(\zeta, t, x, v)$ is attached to a Lévy measure π_2 (also called jumps with unbounded variation) with support in $\{\zeta : |\zeta| < \varepsilon\}$ and such that π_2 integrates the function $|\zeta|^2$, (b) the new coefficient $\beta(\zeta, t, x, v)$ is attached to a Lévy measure π_1 (also called jumps with bounded variation) with support in $\{\zeta : |\zeta| < \varepsilon\}$ and such that integrate π_1 integrates the function $|\zeta|$, (c) the new coefficient $\alpha(\zeta, t, x, v)$ is attached to a finite Lévy measure π_0 (also called bounded jumps) with support in $\{\zeta : |\zeta| \geq \varepsilon\}$.

The Poisson measure ν has three independent pieces ν_i, $i = 0, 1, 2$, the first part ν_0 yields a compound Poisson process $\sum_k \zeta_k \mathbb{1}_{\theta_k \le t}$, the second part ν_1 yields a purely jump Lévy process with bounded variation on every compact time interval, and the last part ν_2 yields a "general" martingale measure $\tilde{\nu}_2$. To simplify a little, assume $t_0 = 0$ so that the stochastic ODE takes the form

$$\begin{aligned} \mathrm{d}x(t) &= g(t, x(t), v(t))\mathrm{d}t + \sigma(t, x(t), v(t))\mathrm{d}w(t) + \\ &+ \int_{\mathbb{R}^m_*} \gamma(\zeta, t, x(t), v(t))\tilde{\nu}_2(\mathrm{d}\zeta, \mathrm{d}t) + \int_{\mathbb{R}^m_*} \beta(\zeta, t, x(t), v(t))\nu_1(\mathrm{d}\zeta, \mathrm{d}t), \end{aligned}$$

between jumps, i.e., for t in $[\theta_{k-1}, \theta_k[$, $k = 1, 2, \dots$ and $\theta_0 = 0$, with the jump-condition

$$x(\theta_k) = x(\theta_{k-1} + \alpha(\zeta_k, \theta_{k-1}, x(\theta_{k-1}), v(\theta_{k-1})),$$

plus the initial condition $x(0) = x^0$.

To preserve the type of jumps, it seems natural to impose extra conditions such that for every $T > 0$ there exists $K = K(T) > 0$ satisfying

$$|\gamma(\zeta, t, x, v)| + |\beta(\zeta, t, x, v)| + |\alpha(\zeta, t, x, v)| \le K(1 + |x|), \quad \forall \zeta, x, v,$$

and t in $[0, T]$, and even to suppose that $|\gamma(\zeta, t, x, v)| + |\beta(\zeta, t, x, v)| \le 2\varepsilon$ and $|\alpha(\zeta, t, x, v)| \ge \varepsilon$, for every ζ, t, x, v and some positive constant $\varepsilon > 0$, which is initially chosen arbitrary. In this case, note that

$$\begin{aligned} \frac{\mathbb{E}\{\mathrm{d}x(t)\}}{\mathrm{d}t} &= \mathbb{E}\{g(t, x(t), v(t))\} + \int_{\mathbb{R}^m_*} \mathbb{E}\{\beta(\zeta, t, x(t), v(t))\}\pi_1(\mathrm{d}\zeta) + \\ &+ \sum_k \mathbb{E}\big\{\alpha(\zeta_k, \theta_{k-1}, x(\theta_{k-1}), v(\theta_{k-1}))\,\mathbb{1}_{\theta_k \le t}\big\}, \end{aligned}$$

i.e., the drift (average) term is changed. It should be clear that the assumptions on each jump-coefficient can be modified for a particular purpose, e.g., to combine discrete and continuous component of the description of the state of a stochastic system as in the case of the so-called hybrid processes, e.g., see Bensoussan and Menaldi [18], Menaldi[172] and references therein.

5.1.4 A Priori Estimates

A priori estimates are very important for stochastic differential equations, and usually, they are obtained by applying Itô formula to some appropriate (polynomial-type) function as shown in the next result. Most of the a priori estimates obtained in this section have some importance in stochastic optimal control and they are not directly used in this book.

Polynomial and Lipschitz Bounds

To estimate the p-moment of either the solution $x(t)$ in term of its initial value $x(t_0)$ or the different $x(t) - y(t)$ in term of $x(t_0) - y(t_0)$, we use Itô formula with smooth functions of the type $x \mapsto (1 + |x|^2)^{p/2}$ with $p > 0$.

For fixed positive constants p, λ we define

$$\alpha(p,\lambda) := p\,\alpha_g(\lambda) + p\,\alpha_\sigma(p,\lambda) + \alpha_\gamma(p,\lambda), \tag{5.44}$$

where

$$\alpha_g(\lambda) := \sup\Big\{\sum_i \frac{x_i\, g_i(t,x,v)}{\lambda+|x|^2}\Big\},$$

$$\alpha_\sigma(p,\lambda) := \sup\Big\{\sum_{i,k} \frac{\sigma_{ik}^2(t,x,v)}{\lambda+|x|^2} + \\ +(p-2)\sum_{i,j,k} \frac{x_i\,\sigma_{ik}(t,x,v)\,\sigma_{jk}(t,x,v)\,x_j}{(\lambda+|x|^2)^2}\Big\},$$

and

$$\alpha_\gamma(p,\lambda) := \sup\Big\{\int_{\mathbb{R}^m_*}\Big[(\lambda+|x+\gamma(\zeta,t,x,v)|^2)^{p/2} - (\lambda+|x|^2)^{p/2} - \\ -\sum_i p\,x_i\,\gamma_i(\zeta,t,x,v)\,(\lambda+|x|^2)^{p/2-1}\Big]\,(\lambda+|x|^2)^{-p/2}\pi(\mathrm{d}\zeta)\Big\},$$

and the suprema are taken for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$.

Lemma 5.9. *Assume that (5.8) holds and, only for $p>2$, that there exists a positive constant $C=C_p$ such that*

$$\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^p\pi(\mathrm{d}\zeta) \le C(1+|x|^2)^{p/2}, \quad p\ge 2, \tag{5.45}$$

for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$. Then the constant (5.44) is finite.

Proof. Clearly, under assumption (5.8) we can verify that the quantity $\alpha_g(\lambda)$ is finite. However, some details are necessary to check that $\alpha_\sigma(p,\lambda)$ and $\alpha_\gamma(p,\lambda)$ are also finite.

First recall the moment inequalities for stochastic integrals, see (4.75), (4.95), (4.103) and (4.104) of Chapter 4, namely, for any $p>0$ there is a constant $C_p>0$ (in particular, $C_1=3$ and $C_2=4$) such that

$$\mathbb{E}\Big\{\sup_{t_0\le r\le t}\Big|\int_{t_0}^r a(s)\mathrm{d}w(s)\Big|^p\Big\} \le C_p\,\mathbb{E}\Big\{\Big[\int_{t_0}^t |a(s)|^2\mathrm{d}s\Big]^{p/2}\Big\}, \tag{5.46}$$

and for the stochastic Poisson integral, if $0<p\le 2$ then

$$\begin{cases} \mathbb{E}\Big\{\sup_{t_0\le r\le t}\Big|\int_{\mathbb{R}^m_*\times]t_0,r]} b(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p\Big\} \le \\ \qquad\le C_p\,\mathbb{E}\Big\{\Big[\int_{t_0}^t \mathrm{d}s\int_{\mathbb{R}^m_*}|b(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\}, \end{cases} \tag{5.47}$$

and if $p > 2$ then

$$\left\{\begin{aligned}&\mathbb{E}\Big\{\sup_{t_0\le r\le t}\Big|\int_{\mathbb{R}^m_*\times]t_0,r]} b(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s)\Big|^p\Big\}\le\\&\qquad\le C_p\,\mathbb{E}\Big\{\int_{t_0}^t \mathrm{d}s\int_{\mathbb{R}^m_*}|b(\zeta,s)|^p\pi(\mathrm{d}\zeta)\Big\}+\\&\qquad\quad+\mathbb{E}\Big\{\Big[\int_{t_0}^t \mathrm{d}s\int_{\mathbb{R}^m_*}|b(\zeta,s)|^2\pi(\mathrm{d}\zeta)\Big]^{p/2}\Big\},\end{aligned}\right.\tag{5.48}$$

for every $t\ge t_0\ge 0$, any adapted processes (and measurable) $(a(t):t\ge 0)$ and $(b(\zeta,t):t\ge 0,\,\zeta\in\mathbb{R}^m_*)$.

Now, the integrand inside the expression of $\alpha_\gamma(p,\lambda)$ can be rewritten as

$$p\int_0^1(1-r)\Big(\frac{\lambda+|x+r\gamma|^2}{\lambda+|x|^2}\Big)^{p/2-1}\alpha_{p,\lambda,x,\gamma}(r)\mathrm{d}r,$$

where

$$\alpha_{p,\lambda,x,\gamma}(r)=\sum_{ij}\Big[(p-2)\frac{(x_i+r\gamma_i)(x_j+r\gamma_j)}{\lambda+|x+r\gamma|^2}+\delta_{ij}\Big]\Big[\frac{\gamma_i\,\gamma_j}{\lambda+|x|^2}\Big],$$

with $\gamma_i=\gamma_i(\zeta,t,x,v)$, $\gamma=\gamma(\zeta,t,x,v)$ and $\delta_{ij}=1$ only if $i=j$ and vanishes otherwise. Thus, taking note of assumption (5.45), the suprema $\alpha_\sigma(p,\lambda)$ and $\alpha_\gamma(p,\lambda)$ are finite. □

Similarly, we define

$$\beta(p):=\sup_{\lambda>0}\big\{p\,\beta_g(\lambda)+(p/2)\,\beta_\sigma(p,\lambda)+\beta_\gamma(p,\lambda)\big\},\tag{5.49}$$

where

$$\beta_g(\lambda):=\sup\Big\{\sum_i\frac{(x_i-x_i')\,[g_i(t,x,v)-g_i(t,x',v)]}{\lambda+|x-x'|^2}\Big\},$$

$$\begin{aligned}\beta_\sigma(p,\lambda):=\sup\Big\{&\sum_{i,k}\frac{\big[\sigma_{ik}(t,x,v)-\sigma_{ik}(t,x',v)\big]^2}{\lambda+|x-x'|^2}+\\&+(p-2)\sum_{i,j,k}\Big[\frac{(x_i-x_i')\,\big[\sigma_{ik}(t,x,v)-\sigma_{ik}(t,x',v)\big]}{\lambda+|x-x'|^2}\times\\&\times\frac{\big[\sigma_{jk}(t,x,v)-\sigma_{jk}(t,x',v)\big]\,(x_j-x_j')}{\lambda+|x-x'|^2}\Big]\Big\},\end{aligned}$$

and

$$\begin{aligned}\beta_\gamma(p,\lambda):=\sup\Big\{&\int_{\mathbb{R}^m_*}\Big[(\lambda+|x-x'+\gamma(\zeta,t,x,v)-\gamma(\zeta,t,x',v)|^2)^{p/2}-\\&-(\lambda+|x-x'|^2)^{p/2}-\sum_i p\,(x_i-x_i')[\gamma_i(\zeta,t,x,v)-\gamma_i(\zeta,t,x',v)]\times\\&\times(\lambda+|x-x'|^2)^{p/2-1}\Big]\,(\lambda+|x-x'|^2)^{-p/2}\pi(\mathrm{d}\zeta)\Big\},\end{aligned}$$

and the suprema are taken for any t in $[t_0,\infty)$, x, x' in $\mathbb{R}^d$ and v in $\mathbb{R}^q$.

Lemma 5.10. *Assume a global version of assumption (5.11), namely, there exists a positive constant M such that*

$$\begin{cases} (x-x')\,[g(t,x,v)-g(x',t,v)] + |\sigma(t,x,v)-\sigma(x',t,v)|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)-\gamma(x',t,\zeta,v)|^2 \pi(\mathrm{d}\zeta) \le M|x-x'|^2, \end{cases} \tag{5.50}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, and, only for $p>2$, that there exists a positive constant $M=M_p$ such that

$$\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)-\gamma(\zeta,t,x',v)|^p \pi(\mathrm{d}\zeta) \le M|x-x'|^p, \tag{5.51}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$. Then the supremum $\beta_\gamma(p)$ is finite.

Proof. Again, by means of estimates (5.46) and under (5.50), we can verify that the quantities β_g are finite. However, some details are necessary to check that $\beta_\sigma(p)$ and $\beta_\gamma(p)$ are actually finite.

The part concerning $\beta_\sigma(p)$ is simple, and bounded in term of the constant appearing in assumption (5.50). On the other hand, the integrand inside the expression of $\beta_\gamma(p,\lambda)$ can be rewritten as

$$p\int_0^1 (1-r)\Big(\frac{\lambda+|x-x'+r(\gamma-\gamma')|^2}{\lambda+|x-x'|^2}\Big)^{p/2-1}\beta_{p,\lambda,x-x',\gamma-\gamma'}(r)\mathrm{d}r,$$

where

$$\begin{aligned}\beta_{p,\lambda,x-x',\gamma-\gamma'}(r) = \sum_{ij}\Big[(p-2)\frac{(\gamma_i-\gamma_i')\,(\gamma_j-\gamma_j')}{\lambda+|x-x'|^2}\Big]\times \\ \times\Big[\frac{\big(x_i-x_i'+r(\gamma_i-\gamma_i')\big)\big(x_j-x_j'+r(\gamma_j-\gamma_j')\big)}{\lambda+|x-x'+r(\gamma-\gamma')|^2}+\delta_{ij}\Big],\end{aligned}$$

with $\gamma_i=\gamma_i(\zeta,t,x,v)$, $\gamma=\gamma(\zeta,t,x,v)$, $\gamma_i'=\gamma_i(\zeta,t,x',v)$, $\gamma'=\gamma(\zeta,t,x',v)$ and $\delta_{ij}=1$ only if $i=j$ and vanishes otherwise. Thus, by means of estimate (5.47), (5.50) and only for $p>2$, also (5.51), we may proceed as follows. If $p\ge 2$ then

$$\frac{\lambda+|x-x'+\gamma-\gamma'|^2}{\lambda+|x-x'|^2}\le 2+2\,\frac{|\gamma-\gamma'|^2}{|x-x'|^2},\quad \forall\lambda>0,$$

and estimate (5.51) plays a key role. Alternatively, if $0<p<2$ then split the integral over $\mathbb{R}^d_*$ in two parts, first over the region $|\gamma-\gamma'|\le(1/2)|x-x'|$, which is denoted by $\beta_\gamma(p,\lambda)_a$ and second over its complementary region $|\gamma-\gamma'|>(1/2)|x-x'|$, which is denoted by $\beta_\gamma(p,\lambda)_b$, where $\gamma:=\gamma(\zeta,t,x,v)$ and $\gamma':=\gamma(\zeta,t,x',v)$.

To estimate $\beta_\gamma(p,\lambda)_a$, we use the fact that

$$|\gamma-\gamma'|\le(1/2)|x-x'|\ \Rightarrow\ |x-x'+\gamma-\gamma'|\ge(1/2)|x-x'|,$$

to get the inequality

$$\frac{\lambda + |x - x' + \gamma - \gamma'|^2}{\lambda + |x - x'|^2} \geq \frac{1}{4}, \quad \forall \lambda > 0,$$

which control the integrand of $\beta_\gamma(p,\lambda)_a$. However, to estimate $\beta_\gamma(p,\lambda)_b$, we do not need to use the second derivatives, directly from the definition of $\beta_\gamma(p,\lambda)$ and by means of the inequality

$$|a^r - b^r| \leq |a - b|^r, \quad \forall a, b \geq 0, \quad \forall r \in (0,1],$$

we can bound the integrand by

$$|\gamma - \gamma'|^p \, (\lambda + |x - x'|^2)^{-p/2} + \frac{|\gamma - \gamma'|}{|x - x'|} \leq 6 \, \frac{|\gamma - \gamma'|^2}{|x - x'|^2}$$

over the region where $|\gamma - \gamma'| > (1/2)|x - x'|$, for every $\lambda > 0$, $0 < p < 2$ and $x \neq x'$ in $\mathbb{R}^d$. Hence, the extra condition (5.51) is only needed when $p > 2$. □

Note that $\alpha(p,\lambda)$ [respectively $\beta(p)$], defined by (5.44) [respectively (5.49)], depends only the constant C [respectively M] of conditions (5.6) and (5.8) [respectively (5.11) and (5.50)]. In particular the length of the time interval $t_1 - t_0$ does not appear explicitly.

Theorem 5.11 (estimates)**.** *Let $(x(t) : t \geq t_0)$ be a solution of the d-dimensional stochastic ordinary differential equation (5.5). If $\alpha \geq \alpha(p,\lambda)$, $p > 0$, $\lambda > 0$, as defined by (5.44), then under assumptions (5.8) and (5.45) we have*

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ [\alpha - \alpha(p,\lambda)] \displaystyle\int_{t_0}^{t} (\lambda + |x(s)|^2)^{p/2} \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s + \\ \qquad + (\lambda + |x(t)|^2)^{p/2} \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \mathbb{E}\big\{ (\lambda + |x(t_0)|^2)^{p/2} \big\}, \end{array} \right. \tag{5.52}$$

for every $t \geq t_0$. Moreover, there is a constant C, depending only on p and the bounds of σ and γ through conditions (5.8) and (5.45) (with $2p$ instead of p), such that for every $t \geq t_0$ we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \displaystyle\sup_{t_0 \leq s \leq t} (\lambda + |x(s)|^2)^{p/2} \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq \\ \qquad \leq C \, \big[1 + \dfrac{1}{\alpha - \alpha(p,\lambda)} \big] \, \mathbb{E}\big\{ (\lambda + |x(t_0)|^2)^{p/2} \big\}. \end{array} \right. \tag{5.53}$$

Now, denote by $(y(t) : t \geq t_0)$ another solution. If $\alpha \geq \beta(p)$, $p > 0$, as defined by (5.49), then under assumptions (5.50) and (5.51) we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ [\alpha - \beta(p)] \displaystyle\int_{t_0}^{t} |x(s) - y(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s + \\ \qquad + |x(t) - y(t)|^p \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \mathbb{E}\big\{ |x(t_0) - y(t_0)|^p \big\}, \end{array} \right. \tag{5.54}$$

for every $t \geq t_0$. Furthermore, we have

$$\begin{cases} \mathbb{E}\Big\{ \sup\limits_{t_0 \leq s \leq t} |x(s) - y(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq \\ \qquad \leq M \left[1 + \dfrac{1}{\alpha - \beta(p)} \right] \mathbb{E}\big\{ |x(t_0) - y(t_0)|^p \big\}, \end{cases} \tag{5.55}$$

for every $t \geq t_0$ and for some constant M depending only on p and the bounds of σ and γ through conditions (5.50) and (5.51) (with $2p$ instead of p).

Proof. First, for positive constants α, λ and p, let us apply Itô formula (see previous chapter) to the function $\varphi : (x,t) \mapsto \mathrm{e}^{-\alpha(t-t_0)}(\lambda + |x|^2)^{p/2}$ and the solution process $x(t)$ of (5.3), to obtain

$$\mathrm{d}\varphi(x(t),t) = a(t)\mathrm{d}t + \sum_{k=1}^{n} b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \tag{5.56}$$

where

$$\begin{aligned} a(t) &:= \partial_t\varphi(x(t),t) + \sum_{i=1}^{d} g_i(t,x(t),v(t))\,\partial_i\varphi(x(t),t) + \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sum_{k=1}^{n}\sigma_{ik}(t,x(t),v(t))\sigma_{jk}(t,x(t),v(t))\Big)\,\partial^2_{ij}\varphi(x(t),t) + \\ &\quad + \int_{\mathbb{R}^m_*}\big[\varphi(x(t)+\gamma(\zeta,t,x(t),v(t)),t) - \varphi(x(t),t) - \\ &\quad - \sum_{i=1}^{d}\gamma_i(\zeta,t,x(t),v(t))\,\partial_i\varphi(x(t),t)\big]\,\pi(\mathrm{d}\zeta), \\ b_k(t) &:= \sum_{i=1}^{d}\sigma_{ik}(t,x(t),v(t))\,\partial_i\varphi(x(t),t), \\ c(\zeta,t) &:= \varphi(x(t)+\gamma(\zeta,t,x(t),v(t)),t) - \varphi(x(t),t). \end{aligned}$$

Since

$$\begin{aligned} \partial_t\varphi(x,t) &= -\alpha\varphi(x,t), \\ \partial_i\varphi(x,t) &= p x_i(\lambda + |x|^2)^{-1}\varphi(x,t), \\ \partial_{i,j}\varphi(x,t) &= p\big\{(p-2)x_i x_j(\lambda+|x|^2)^{-1} + \\ &\qquad + \delta_{ij}\big\}(\lambda+|x|^2)^{-1}\varphi(x,t), \end{aligned}$$

where $\delta_{ij} = 1$ only if $i = j$ and vanishes otherwise, we verify that $a(t) \leq \alpha(p,\lambda)\,\varphi(x(t),t)$, given by (5.44). By taking the mathematical expectation we deduce estimate (5.52).

Now, we come back to the Itô equality (5.56) to take first the supremum and then the mathematical expectation, we deduce, after using the stochastic

integral inequalities (5.46) and (5.47), with $p = 1$, that

$$\mathbb{E}\Big\{ \sup_{t_0\le t\le t_1} (\lambda + |x(t)|^2)^{p/2}\, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \le \mathbb{E}\Big\{ (\lambda + |x(t_0)|^2)^{p/2} \Big\} + \\ +3\,\mathbb{E}\Big\{ \Big[\sum_k \int_{t_0}^{t_1} |b_k(s)|^2 \mathrm{d}s \Big]^{1/2} + \Big[\int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^m_*} |c(\zeta,s)|^2 \pi(\mathrm{d}\zeta) \Big]^{1/2} \Big\},$$

To estimate the last two integrals, first find a constant C, depending on the bounds of σ through (5.8) and (5.45) (regarding only σ and γ and with $2p$ instead of p),

$$\sum_k |b_k(s)|^2 + \int_{\mathbb{R}^m_*} |c(\zeta,t)|^2 \pi(\mathrm{d}\zeta) \le C\, |\varphi(s,x(s))|^2,$$

which is obvious for $b_k(t)$ but not so for $c(\zeta,t)$ (see below), and proceed as follows. The part corresponding to Wiener process, with b_k, can be bounded by

$$\mathbb{E}\Big\{ \sup_{t_0\le s\le t_1} |\varphi(s,x(s))|^{1/2} \Big[\int_{t_0}^{t_1} |\varphi(s,x(s))| \mathrm{d}s \Big]^{1/2} \Big\}.$$

Thus, by means of the inequality $2ab \le \varepsilon a^2 + b^2/\varepsilon$ and the Hölder inequality we deduce that

$$3\,\mathbb{E}\Big\{ \Big[\sum_k \int_{t_0}^{t_1} |b_k(s)|^2 \mathrm{d}s \Big]^{1/2} \Big\} \le \\ \le \frac{1}{3}\,\mathbb{E}\Big\{ \sup_{t_0\le s\le t_1} |\varphi(s,x(s))| \Big\} + C_1\,\mathbb{E}\Big\{ \int_{t_0}^{t_1} |\varphi(s,x(s))| \mathrm{d}s \Big\}.$$

The part corresponding to γ can be handled with the same technique, i.e., first the Poisson integral part is bounded by

$$\mathbb{E}\Big\{ \sup_{t_0\le s\le t_1} \Big[\int_{\mathbb{R}^d_*} |c(\zeta,s)|^2 \pi(\mathrm{d}\zeta) \Big]^{1/2} \Big[\int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^m_*} |c(\zeta,s)|^2 \pi(\mathrm{d}\zeta) \Big]^{1/2} \Big\},$$

and

$$|c(\zeta,s)|^2 \le \mathrm{e}^{-2\alpha(s-t_0)} |\gamma(\zeta,s,x(s),v(s))|^2 \times \\ \times \int_0^1 |\nabla\varphi\big(x(s) + \theta\gamma(\zeta,s,x(s),v(s)),s\big)|^2 \mathrm{d}\theta.$$

To estimate the gradient $\nabla\varphi$ consider two cases. If $p \ge 1$ then use the inequality $|a+b|^q \le 2^q(|a|^q + |b|^q))$ with $q = p-1 > 1$ to get

$$|\nabla\varphi(x+\theta\gamma,s)|^2 \le p^2 2^{(p-1)} \big[(\lambda + |x|^2)^{p-1} + |\gamma|^{2p-2} \big] \mathrm{e}^{-2\alpha(s-t_0)}.$$

However, if $0 < p < 1$ then the inequality $|a+b|^r \le |a|^r + |b|^r$ with $r = p$ yields

$$|\nabla\varphi(x+\theta\gamma,t)|^2 \le p^2 \big[(\lambda + |x|^2)^{p-1} + |\gamma|^{2p-2} \big] \mathrm{e}^{-2\alpha(s-t_0)},$$

which is fine when $|\gamma| \geq (1/2)(\lambda + |x|^2)^{1/2}$, but on its complement we use the fact that for every θ in $[0,1]$ we have

$$|\gamma| \leq (1/2)(\lambda + |x|^2)^{1/2} \quad \Rightarrow$$
$$\Rightarrow \quad (\lambda + |x + \theta\gamma|^2)^{1/2} \geq (1/2)(\lambda + |x|^2)^{1/2},$$

which implies

$$|\nabla\varphi(x + \theta\gamma, t)|^2 \leq 2p^2(\lambda + |x|^2)^{p-1}\mathrm{e}^{-2\alpha(s-t_0)},$$

i.e.,

$$|\nabla\varphi(x + \theta\gamma, t)|^2 \leq 3p^2(\lambda + |x|^2)^{p-1}\mathrm{e}^{-2\alpha(s-t_0)}, \quad \text{if} \quad 0 < p < 1.$$

Hence, use assumptions (5.8) (now, only the part regarding γ) and (5.45) (with $2p$ instead of p) to deduce

$$\int_{\mathbb{R}^d_*} |c(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \leq C|\varphi(x(s), s)|^2,$$

and to continue as in the case of the Wiener process part. Thus, we obtain the same estimates for the stochastic integrals which yield

$$\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1} (\lambda + |x(t)|^2)^{p/2}\, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq 3\,\mathbb{E}\Big\{ (\lambda + |x(t_0)|^2)^{p/2} \Big\} +$$
$$+ C\,\mathbb{E}\Big\{ \int_{t_0}^{t_1} \mathrm{e}^{-\alpha(t-t_0)}\, (\lambda + |x(t)|^2)^{p/2} \mathrm{d}t \Big\},$$

for some constant C depending only on p and the bounds of σ and γ through the conditions (5.8) and (5.45) (with $2p$ instead of p). Hence, (5.53) follows from the estimate (5.52).

Similarly, using the process $x(t) - y(t)$ instead of $x(t)$, we can establish estimates (5.54) and (5.55), i.e., Itô formula is used for the function $\varphi : (z, t) \mapsto \mathrm{e}^{-\alpha(t-t_0)}(\lambda + |z|^2)^{p/2}$ and the difference $z(t) = x(t) - y(t)$, where $x(t)$ and $y(t)$ are two solutions of (5.3) with initial conditions $x(t_0)$ and $y(t_0)$. □

It is clear that Itô formula plays a fundamental role to establish a priori estimates, particularly the quadratic norm for the solution $x(t)$ of the d-dimensional stochastic ordinary differential equation (5.3), namely

$$\begin{cases} \mathrm{d}|x(t) - a|^2 = \mathrm{d}M(t) + \big\{2[x(t) - a] \cdot g(t, x(t), v(t)) + \\ \qquad + \mathrm{Tr}[\sigma^2(t, x(t), v(t))] + \displaystyle\int_{\mathbb{R}^d_*} |\gamma(\zeta, t, x(t), v(t))|^2 \pi(\mathrm{d}\zeta)\big\}\mathrm{d}t, \end{cases} \tag{5.57}$$

where the stochastic integral is given by

$$\mathrm{d}M(t) := 2\sum_k [x(t) - a] \cdot \sigma_k(t, x(t), v(t))\mathrm{d}w_k(t) +$$
$$+ \int_{\mathbb{R}^m_*} \big[|\gamma(\zeta, t, x(t), v(t))|^2 - 2[x(t) - a] \cdot \gamma(\zeta, t, x(t), v(t))\big]\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t),$$

Tr[·] is the trace operator, the dot · denotes the scalar product in $\mathbb{R}^d$ and a is a constant (which could be random, but measurable with respect to the initial time).

The case $p = 1$ and $\alpha = 0$ in estimate (5.54) is of a particular interest, since a more accurate bound can be established under weaker assumptions, this is the so-called Tanaka's formula, namely

$$\begin{cases} \mathrm{d}|x(t) - y(t)| = \mathrm{d}M(t) + \Big[\Sigma(t) + \displaystyle\int_{\mathbb{R}^m_*} \Gamma(t,\zeta)\pi(\mathrm{d}\zeta)\Big]\mathrm{d}t + \\ \qquad + sgn[x(t) - y(t)] \cdot \big[g(t, x(t), v(t)) - g(t, y(t), v(t))\big]\mathrm{d}t, \end{cases} \tag{5.58}$$

where the stochastic integral M is given by

$$\mathrm{d}M(t) = sgn[x(t) - y(t)] \cdot \Big[\big[\sigma(t, x(t), v(t)) - \sigma(t, y(t), v(t))\big]\mathrm{d}w(t) + \\ + \int_{\mathbb{R}^m_*} \big[\gamma(\zeta, t, x(t), v(t)) - \gamma(\zeta, t, y(t), v(t))\big])\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t)\Big],$$

and the drift term is

$$\Sigma(t) = \frac{1}{2}\sum_{ijk} \mathbb{1}_{x(t)\neq y(t)} \frac{|x(t) - y(t)|^2\delta_{ij} - [x_i(t) - y_i(t)][x_j(t) - y_j(t)]}{|x(t) - y(t)|^3} \times \\ \times[\sigma_{ik}(t, x(t), v(t)) - \sigma_{ik}(t, y(t), v(t))][\sigma_{jk}(t, x(t), v(t)) - \sigma_{jk}(t, y(t), v(t))],$$

and

$$\Gamma(t,\zeta) = \big|x(t) - y(t) + \gamma(\zeta, t, x(t), v(t)) - \gamma(\zeta, t, y(t), v(t))\big| - \\ -|x(t) - y(t)| - sgn[x(t) - y(t)] \cdot [\gamma(\zeta, t, x(t), v(t)) - \gamma(\zeta, t, y(t), v(t))],$$

where *sgn* denotes the sign function by coordinates. Note that a revision (by taken $\lambda \to 0$) of the arguments used to show estimate (5.54) provides a way to establish Tanaka's formula (5.58), similarly to Rong [215, Section 1.10, pp. 38–44].

• *Remark* 5.12. Using the fact that

$$|x(t) - x(t_0)|^p \le 3^{p\vee 1}\Big\{\Big|\int_{\mathbb{R}^m_*\times]t_0,t]} \gamma(\zeta, t, x(t), v(t))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t)\Big|^p + \\ +\Big|\sum_k \int_{t_0}^t \sigma_k(t, x(t), v(t))\mathrm{d}w_k(t)\Big|^p + \Big[\int_{t_0}^t |g(s, x(s), v(s))|\,\mathrm{d}s\Big]^p\Big\}$$

and the stochastic integral inequalities (5.46), (5.47) and (5.48), for any p in $(0, 2]$ and $t \ge t_0$, we obtain

$$\mathbb{E}\Big\{\sup_{t_0\le s\le t} |x(s) - x(t_0)|^p\Big\} \le \tilde{C}_p\,(t - t_0)^{p/2} \times \\ \times \mathbb{E}\Big\{\sup_{t_0\le s\le t}\Big[(t - t_0)^{p/2}\,\big|g(s, x(s), v(s))\big|^p + \\ + \big|\operatorname{Tr}[\sigma^2(s, x(s), v(s))]\big|^{p/2} + \\ + \Big(\int_{\mathbb{R}^m_*} \big|\gamma(\zeta, s, x(s), v(s))\big|^2\pi(\mathrm{d}\zeta)\Big)^{p/2}\Big]\Big\},$$

while for $p > 2$ we get

$$\begin{aligned}\mathbb{E}\Big\{ \sup_{t_0\le s\le t} |x(s)-x(t_0)|^p\Big\} \le \tilde{C}_p\,(t-t_0)\times \\ \times\,\mathbb{E}\Big\{ \sup_{t_0\le s\le t}\Big[(t-t_0)^{p-1}\,\big|g(s,x(s),v(s))\big|^p + \\ + (t-t_0)^{p/2-1}\Big(\big|\operatorname{Tr}[\sigma^2(s,x(s),v(s))]\big|^{p/2} + \\ + \Big[\int_{\mathbb{R}^m_*}\big|\gamma(\zeta,s,x(s),v(s))\big|^2\pi(\mathrm{d}\zeta)\Big]^{p/2}\Big) + \\ + \int_{\mathbb{R}^m_*}\big|\gamma(\zeta,s,x(s),v(s))\big|^p\pi(\mathrm{d}\zeta)\Big]\Big\},\end{aligned}$$

which combined with (5.53) yield

$$\begin{cases}\mathbb{E}\Big\{ \sup_{t_0\le s\le t} |x(s)-x(t_0)|^p\Big\} \le \\ \qquad \le C_{p,T}\,(t-t_0)^{(p/2)\wedge 1}\,\mathbb{E}\big\{(1+|x(t_0)|^2)^{p/2}\big\},\end{cases} \tag{5.59}$$

for any t in $[t_0,T]$ and some constant $C_{p,T}$ depending only on T and on the bounds of conditions (5.6) and (5.45). Reviewing the above arguments, it is clear that assuming $\gamma = 0$, i.e., that there is only the stochastic integral with respect to the Wiener process, we can improve (5.59) as follows:

$$\mathbb{E}\Big\{ \sup_{t_0\le s\le t} |x(s)-x(t_0)|^p\Big\} \le C_{p,T}\,(t-t_0)^{p/2}\,\mathbb{E}\big\{(1+|x(t_0)|^2)^{p/2}\big\}, \tag{5.60}$$

even for $p > 2$. Note that T is arbitrary. □

Coefficients Bounds

As in (5.55) of Theorem 5.11, we may study how the solution changes when the coefficients are changed. There are two possible norms to measure the change of the coefficients, namely,

$$\begin{cases}\|h-h'\|_0 = \sup_{t,x,v}\big\{|h(t,x,v)-h'(t,x,v)|\big\}, \\ \|h-h'\|_{0,p} = \sup_{t,x,v}\Big\{\Big(\int_{\mathbb{R}^m_*}|h(\zeta,t,x,v)-h'(\zeta,t,x,v)|^p\nu(\mathrm{d}\zeta)\Big)^{1/p}\Big\}\end{cases} \tag{5.61}$$

and

$$\begin{cases}\|h-h'\|_1 = \sup_{t,x,v}\big\{|h(t,x,v)-h'(t,x,v)|(1+|x|^2)^{-1/2}\big\}, \\ \|h-h'\|_{1,p} = \sup_{t,x,v}\Big\{(1+|x|^2)^{-1/2}\times \\ \qquad \times\Big(\int_{\mathbb{R}^m_*}|h(\zeta,t,x,v)-h'(\zeta,t,x,v)|^p\nu(\mathrm{d}\zeta)\Big)^{1/p}\Big\},\end{cases} \tag{5.62}$$

for $p \ge 2$. Define, for $i = 0, 1$,

$$\begin{aligned}\Lambda_{i,p}(g-g',\sigma-\sigma',\gamma-\gamma') = \\ = \|g-g'\|_i + \|\sigma-\sigma'\|_i + \|\gamma-\gamma'\|_{i,p} + \|\gamma-\gamma'\|_{i,2}\end{aligned}$$

to have

Proposition 5.13. *Under assumptions (5.8), (5.45), (5.50) and (5.51), if* $(y(t) : t \geq t_0)$ *and* $(y'(t) : t \geq t_0)$ *denote the solution corresponding to coefficients* g, σ, γ *and* g', σ', γ'*, respectively, satisfying* $y(t_0) = y'(t_0)$*, then we have*

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \sup\limits_{t_0 \leq s \leq t} |y(s) - y'(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq \\ \qquad\qquad \leq M \Lambda^p_{1,2p}(g - g', \sigma - \sigma', \gamma - \gamma') \mathbb{E}\big\{ (1 + |y(t_0)|^2)^{p/2} \big\}, \end{array} \right. \tag{5.63}$$

and

$$\mathbb{E}\Big\{ \sup_{t_0 \leq s \leq t} |y(s) - y'(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq M \Lambda^p_{0,2p}(g - g', \sigma - \sigma', \gamma - \gamma'), \tag{5.64}$$

for every $t \geq t_0$ *and for some constant* M *depending only on* $p \geq 2$, $\alpha > \beta(p)$ *(as defined by (5.49)), and the bounds of* g, σ *and* γ *through conditions (5.50) and (5.51) (with* $2p$ *instead of* p*).*

Proof. First, note that the process $x(t) = y(t) - y'(t)$ has drift, diffusion and jump coefficients $g - g'$, $\sigma - \sigma'$ and $\gamma - \gamma'$, respectively. Thus, for positive constants α, λ and p, let us apply Itô formula (see previous chapter) to the function $\varphi : (x, t) \mapsto \mathrm{e}^{-\alpha(t-t_0)}(\lambda + |x|^2)^{p/2}$ and the process $x(t) = y(t) - y'(t)$, where $y(t)$ and $y'(t)$ are the solution (with obvious change of notation!) of (5.3), to obtain

$$\mathrm{d}\varphi(x(t), t) = a(t)\mathrm{d}t + \sum_{k=1}^{n} b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta, t)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t), \tag{5.65}$$

where

$$\begin{aligned} a(t) &:= \partial_t \varphi(x, t) + \sum_i \tilde{g}_i(t)\, \partial_i \varphi(x, t) + \frac{1}{2} \sum_{i,j,k} \tilde{\sigma}_{ik}(t)\tilde{\sigma}_{jk}(t)\, \partial^2_{ij}\varphi(x, t) + \\ &\quad + \int_{\mathbb{R}^m_*} \big[\varphi(x + \tilde{\gamma}(\zeta, t), t) - \varphi(x, t) - \sum_i \tilde{\gamma}_i(\zeta, t)\, \partial_i \varphi(x, t) \big]\, \pi(\mathrm{d}\zeta), \\ b_k(t) &:= \sum_i \tilde{\sigma}_{ik}(t)\, \partial_i \varphi(x, t), \\ c(\zeta, t) &:= \varphi(x + \tilde{\gamma}(\zeta, t), t) - \varphi(x, t), \end{aligned}$$

and $x = x(t) = y(t) - y'(t)$,

$$\begin{aligned} \tilde{g}_i(t) &= g_i(t, y(t), v(t)) - g'_i(t, y'(t), v(t)), \\ \tilde{\sigma}_{ik}(t) &= \sigma_{ik}(t, y(t), v(t)) - \sigma'_{ik}(t, y'(t), v(t)), \\ \tilde{\gamma}_i(\zeta, t) &= \gamma_i(\zeta, t, y(t), v(t)) - \gamma'_i(\zeta, t, y'(t), v(t)), \end{aligned}$$

and certainly, $\tilde{\gamma}(\zeta, t) = \gamma(\zeta, t, y(t), v(t)) - \gamma'(\zeta, t, y'(t), v(t))$. Also we have

$$\begin{aligned} \partial_t \varphi(x, t) &= -\alpha \varphi(x, t), \\ \partial_i \varphi(x, t) &= p x_i (\lambda + |x|^2)^{-1} \varphi(x, t), \\ \partial_{ij} \varphi(x, t) &= p\big\{ (p - 2) x_i x_j (\lambda + |x|^2)^{-1} + \delta_{ij} \big\} (\lambda + |x|^2)^{-1} \varphi(x, t), \end{aligned}$$

where $\delta_{ij} = 1$ only if $i = j$ and vanishes otherwise. Essentially, the whole argument is to show that for every $\varepsilon > 0$ $p \geq 2$, $\lambda = 0$ and $i = 0, 1$ there exists a constant $C_{\varepsilon,p} > 0$, which also depends only on the dimension d, and the various bounds in the assumptions (5.8), (5.45), (5.50) and (5.51), such that

$$a(t) \leq \big(\beta(p) + \varepsilon - \alpha\big)\,\varphi\big(y(t) - y'(t), t\big) + \\ + C_{\varepsilon,p}\Lambda_{i,p}^{p}(g - g', \sigma - \sigma', \gamma - \gamma')(1 + |y'(t)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}.$$

Moreover, we also have

$$\sum_k |b_k(t)|^2 + \int_{\mathbb{R}^d_*} |c(\zeta, t)|^2 \pi(\mathrm{d}\zeta) \leq C\big[|\varphi\big(y(t) - y'(t), t\big)|^2 + \\ + \Lambda_{i,2p}^{2p}(g - g', \sigma - \sigma', \gamma - \gamma')(1 + |y'(t)|^2)^{p}\mathrm{e}^{-2\alpha(t-t_0)}\big],$$

for some other constant $C > 0$, depending on (5.50) (regarding only σ and γ) and (5.51) (with $2p$ instead of p).

Let us consider the simplest case when $\sigma = \sigma'$ and $\gamma = \gamma'$. In this case, the $\tilde{g}_i(t)$ is bounded as follows:

$$(x - x') \cdot [g(t, x, v) - g'(t, x', v)] \leq (x - x') \cdot [g(t, x, v) - g(t, x', v)] + \\ + (x - x') \cdot [g(t, x', v) - g'(t, x', v)] \leq \\ \leq \beta_g(\lambda)(\lambda + |x - x'|^2) + \Lambda_2(\lambda, g - g')(\lambda + |x - x'|^2)^{1/2}(1 + |x'|^2)^{1/2},$$

with $\beta_g(\lambda)$ as in (5.49) and

$$\Lambda_2(\lambda, g - g') = \sup_{x, x' \in \mathbb{R}^d} \Big\{ \sum_i \frac{(x_i - x_i')\,[g_i(t, x', v) - g_i'(t, x', v)]}{(\lambda + |x - x'|^2)^{1/2}(1 + |x'|^2)^{1/2}} \Big\} \leq \\ \leq \sup_{x' \in \mathbb{R}^d} \big\{ |g(t, x', v) - g'(t, x', v)|\,(1 + |x'|^2)^{-1/2} \big\} \leq \|g - g'\|_1.$$

Hence, we bound

$$a(t) \leq p\beta_g(\lambda)\varphi(x(t), t) + \\ + p\|g - g'\|_1(1 + |y'(t)|^2)^{1/2}\mathrm{e}^{-\alpha(t-t_0)}(\lambda + |x(t)|^2)^{(p-1)/2}$$

to deduce an expression similar to (5.54), but with an extra term with $p\|g - g'\|_1$ multiplying the integral

$$\mathbb{E}\Big\{ \int_{t_0}^{t} (1 + |y'(s)|^2)^{1/2}\mathrm{e}^{-\alpha(t-t_0)}(\lambda + |x(s)|^2)^{(p-1)/2}\mathrm{d}s \Big\}.$$

This term can be bounded by (using Hölder inequality and the fact that, for each $\varepsilon > 0$ there exists a constant $C_\varepsilon > 0$ such that $AB \leq \varepsilon A^p + C_\varepsilon B^{p'}$, with $1/p + 1/p' = 1$)

$$\varepsilon\mathbb{E}\Big\{ \int_{t_0}^{t} (\lambda + |x(s)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}\mathrm{d}s \Big\} + \\ + C_\varepsilon p\|g - g'\|_1^p\mathbb{E}\Big\{ \int_{t_0}^{t} (1 + |y'(s)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}\mathrm{d}s \Big\},$$

proving (after letting $\lambda \to 0$)

$$\mathbb{E}\Big\{[\alpha-\beta(p)-\varepsilon]\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s + \\ + |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(t-t_0)}\Big\} \le \\ \le C_\varepsilon p\|g-g'\|_1^p\mathbb{E}\big\{(1+|y'(t_0)|^2)^{p/2}\big\},$$

with the implicit assumption that $\alpha > \alpha(p,1)$, the constant used in the estimate (5.52) with $\lambda = 1$. Similarly, we can define

$$\bar\Lambda_2(\lambda, g-g') = \sup_{x,x'\in\mathbb{R}^d}\Big\{\sum_i \frac{(x_i-x_i')\,[g_i(t,x',v)-g_i'(t,x',v)]}{(\lambda+|x-x'|^2)^{1/2}}\Big\} \le \\ \le \sup_{x'\in\mathbb{R}^d}\big\{|g(t,x',v)-g'(t,x',v)|\big\} \le \|g-g'\|_0$$

to have

$$\mathbb{E}\Big\{[\alpha-\beta(p)-\varepsilon]\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s + \\ + |y(t)-y'(t)|^p\,\mathrm{e}^{-\alpha(t-t_0)}\Big\} \le C_\varepsilon p\|g-g'\|_0^p.$$

with the norms (5.61) and (5.62).

Next, let us consider the case $g = g'$ and $\gamma = \gamma'$. In this case, using the inequality $(a+b)^2 \le (1+\varepsilon)a^2 + (1+C_\varepsilon)b^2$ the term $\sum_k \tilde\sigma_{ik}(t)\tilde\sigma_{jk}(t)$ is bounded by $A_{ij} + B_{ij}$, with

$$A_{ij} = (1+\varepsilon)\sum_k [\sigma_{ik}(t,x,v)-\sigma_{ik}(t,x',v)][\sigma_{jk}(t,x,v)-\sigma_{jk}(t,x',v)]$$
$$B_{ij} = (1+C_\varepsilon)\sum_k [\sigma_{ik}(t,x',v)-\sigma'_{ik}(t,x',v)][\sigma_{jk}(t,x',v)-\sigma'_{jk}(t,x',v)].$$

Now, $\sum_{ij} A_{ij}\partial_{ij}\varphi(x,t)$ is bounded by $(1+\varepsilon)\beta(p)\varphi(t,x-x')$, where $\beta(p) = \sup_{\lambda>0}\{(p/2)\beta_\sigma(\lambda,p)\}$ as in (5.49). The term B_{ij} can be bounded by either $\|\sigma-\sigma'\|_0^2$ or $\|\sigma-\sigma'\|_1^2(1+|x'|^2)$. Hence

$$a(t) \le (\beta(p)+\varepsilon)\varphi(x(t),t) + C\|\sigma-\sigma'\|_0^2(\lambda+|x(t)|^2)^{(p-2)/2}\mathrm{e}^{-\alpha(t-t_0)}$$

and

$$a(t) \le (\beta(p)+\varepsilon)\varphi(x(t),t) + C\|\sigma-\sigma'\|_1^2(1+|y'(t)|^2)^{1/2}\mathrm{e}^{-\alpha(t-t_0)}\times \\ \times(\lambda+|x(t)|^2)^{(p-2)/2}.$$

If $p > 2$ then use Hölder inequality with $q' = p/2$, so that $1/q' = 2/p = 1-(p-2)/p$ and $q = p/(p-2) > 1$, to bound the integral

$$\mathbb{E}\Big\{\int_{t_0}^t (1+|y'(s)|^2)\mathrm{e}^{-\alpha(t-t_0)}(\lambda+|x(s)|^2)^{(p-2)/2}\mathrm{d}s\Big\}.$$

with

$$\Big(\mathbb{E}\Big\{\int_{t_0}^t (1+|y'(s)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}\mathrm{d}s\Big\}\Big)^{1/q'}\Big(\mathbb{E}\Big\{\int_{t_0}^t (\lambda+|x(s)|^2)^{p/2}\mathrm{d}s\Big\}\Big)^{1/q},$$

Hence, the last term in the previous inequality for $a(t)$ (including the constant C) can be bounded by

$$\varepsilon\mathbb{E}\Big\{\int_{t_0}^t (\lambda+|x(s)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}\mathrm{d}s\Big\}+ \\ + C_\varepsilon\|\sigma-\sigma'\|_1^p\mathbb{E}\Big\{\int_{t_0}^t (1+|y'(s)|^2)^{p/2}\mathrm{e}^{-\alpha(t-t_0)}\mathrm{d}s\Big\},$$

proving (after letting $\lambda\to 0$)

$$\mathbb{E}\Big\{[\alpha-\beta(p)-2\varepsilon]\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s+ \\ + |y(t)-y'(t)|^p\,\mathrm{e}^{-\alpha(t-t_0)}\Big\}\le \\ \le C_\varepsilon\|\sigma-\sigma'\|_1^p\mathbb{E}\big\{(1+|y'(t_0)|^2)^{p/2}\big\},$$

with the implicit assumption that $\alpha>\alpha(p,1)$, the constant used in the estimate (5.52) with $\lambda=1$. Similarly, we have

$$\mathbb{E}\Big\{[\alpha-\beta(p)-2\varepsilon]\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s+ \\ + |y(t)-y'(t)|^p\,\mathrm{e}^{-\alpha(t-t_0)}\Big\}\le C_\varepsilon\|\sigma-\sigma'\|_0^p,$$

with the norms (5.61) and (5.62).

Finally, let us consider the case $g=g'$ and $\sigma=\sigma'$. In this case, we have to deal with the jump terms. Begin with $p=2$ and $\lambda=0$ to have $\partial_i\varphi(x,t)=2x_i\mathrm{e}^{-\alpha(t-t_0)}$ and

$$\big[\varphi(x+\tilde\gamma(t),t)-\varphi(x,t)-\sum_i\tilde\gamma_i(t)\,\partial_i\varphi(x,t)\big]=|\tilde\gamma_i(t)|^2\mathrm{e}^{-\alpha(t-t_0)},$$

and essentially the same argument used before for the diffusion term yields

$$\begin{cases}\mathbb{E}\Big\{[\alpha-\beta(2)-2\varepsilon]\displaystyle\int_{t_0}^t |y(s)-y'(s)|^2\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s+ \\ \qquad + |y(t)-y'(t)|^2\,\mathrm{e}^{-\alpha(t-t_0)}\Big\}\le \\ \qquad\qquad \le C_\varepsilon\|\gamma-\gamma'\|_{1,2}^2\mathbb{E}\big\{1+|y'(t_0)|^2\big\},\end{cases} \tag{5.66}$$

and

$$\begin{cases}\mathbb{E}\Big\{[\alpha-\beta(2)-2\varepsilon]\displaystyle\int_{t_0}^t |y(s)-y'(s)|^2\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s+ \\ \qquad + |y(t)-y'(t)|^2\,\mathrm{e}^{-\alpha(t-t_0)}\Big\}\le C_\varepsilon\|\gamma-\gamma'\|_{0,2}^2,\end{cases} \tag{5.67}$$

with the norms (5.61) and (5.62).

For the general case with $p > 2$, we can also take $\lambda = 0$, but the calculations are longer anyway, and perhaps tedious. First, define

$$\bar\gamma(\zeta,t) = \gamma(\zeta,t,y(t),v(t)) - \gamma(\zeta,t,y'(t),v(t)),$$
$$\bar{\bar\gamma}(\zeta,t) = \gamma(\zeta,t,y'(t),v(t)) - \gamma'(\zeta,t,y'(t),v(t)),$$

so that $\tilde\gamma(\zeta,t) = \bar\gamma(\zeta,t) + \bar{\bar\gamma}(\zeta,t)$, without writing the variables ζ, $y(t)$ and $v(t)$, and recall

$$\begin{aligned}\big[\varphi(x+\tilde\gamma(t),t) - \varphi(x,t) - \sum_i \tilde\gamma_i(t)\,\partial_i\varphi(x,t)\big] &= \\ = \int_0^1 \sum_{ij} \tilde\gamma_i(t)\big[\partial_i\varphi\big(x+\theta\tilde\gamma(t),t\big) - \partial_i\varphi\big(x,t\big)\big]\mathrm{d}\theta &= \\ = \big[\varphi(x+\bar\gamma(t),t) - \varphi(x,t) - \sum_i \bar\gamma_i(t)\,\partial_i\varphi(x,t)\big] + A.&\end{aligned}$$

Hence, use the inequalities

$$\begin{aligned}\big|\partial_i\varphi\big(x+\theta\tilde\gamma(t),t\big) - \partial_i\varphi\big(x+\theta\bar\gamma(t),t\big)\big| \le 4^{p-2}|\bar{\bar\gamma}(t)|\big[|\bar{\bar\gamma}(t)|^{p-2} + \\ + |\bar\gamma(t)|^{p-2} + (\lambda+|x|^2)^{(p-2)/2}\big]\mathrm{e}^{-\alpha(t-t_0)},\end{aligned}$$

to see that $A \le 4^{p-2}B\mathrm{e}^{-\alpha(t-t_0)}$, where B is dominated by a sum of homogeneous terms of the form $|\bar{\bar\gamma}(t)|^p$, $|\bar{\bar\gamma}(t)|^2|\bar\gamma(t)|^{p-2}$, $|\bar{\bar\gamma}(t)|^2(\lambda+|x|^2)^{(p-2)/2}$, $|\bar{\bar\gamma}(t)|^{p-1}|\bar\gamma(t)|$, $|\bar{\bar\gamma}(t)|\,|\bar\gamma(t)|^{p-1}$, and $|\bar\gamma(t)|\,|\bar{\bar\gamma}(t)|\,(\lambda+|x|^2)^{(p-2)/2}$. After integration and by means of assumption (5.51), each term is suitable bounded. For instance, the integral of $|\bar{\bar\gamma}(t)|^p$ is bounded by either $\|\gamma-\gamma'\|_{0,p}^p$ or $\|\gamma-\gamma'\|_{1,p}^p\mathbb{E}\big\{(1+|y'(t_0)|^2)^{p/2}\big\}$. For the other terms, use the inequality $ab \le \varepsilon a^q + C_\varepsilon b^{q'}$ with $1/q + 1/q' = 1$, where either $q' = p/2$ or $q' = p$, to obtain a bound (for each term) of the form

$$\varepsilon\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s + C_\varepsilon B$$

where B is either $\|\gamma-\gamma'\|_{0,p}^p$ or $\|\gamma-\gamma'\|_{1,p}^p\mathbb{E}\big\{(1+|y'(t_0)|^2)^{p/2}\big\}$. However, when dealing with the term involving the factor $(1+|x|^2)^{(p-2)/2}$, we use the norms norms (5.61) and (5.62) with index 2 instead of index p. This yields (5.66) and (5.67) for $p > 2$, with $\alpha - \beta(p) - 5\varepsilon$ instead of $\alpha - \beta(2) - 2\varepsilon$, and the norms $\|\gamma-\gamma'\|_{i,p}$ and $\|\gamma-\gamma'\|_{i,2}$ instead of just $\|\gamma-\gamma'\|_{i,2}$.

Collecting all the inequalities we obtain

$$\left\{\begin{aligned}&\mathbb{E}\Big\{[\alpha-\beta(p)]\int_{t_0}^t |y(s)-y'(s)|^p\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s + \\ &\qquad + |y(t)-y'(t)|^p\,\mathrm{e}^{-\alpha(t-t_0)}\Big\} \le \\ &\qquad \le C\Lambda_{i,p}^p(g-g',\sigma-\sigma',\gamma-\gamma')\mathbb{E}\big\{\big(1+|y(t_0))^2\big)^{ip/2}\big\},\end{aligned}\right. \tag{5.68}$$

with the norms (5.61) and (5.62).

Recalling that

$$b_k(t) = \sum_i \big(\sigma_{ik}(t, y(t), v(t)) - \sigma'_{ik}(t, y'(t), v(t))\big)\partial_i\varphi\big(y(t) - y'(t), t\big),$$

we can bound $\sum_k |b_k(t)|^2$ with

$$C\big(|y(t) - y'(t)|^2 + \|\sigma - \sigma'\|_i^2(1 + |y'(t)|^2)^{i/2}\big)|\partial_i\varphi\big(y(t) - y'(t), t\big)|^2,$$

for a constant C depending on the bounds of σ on assumption (5.50). Hence, because $|\partial_i\varphi(x,t)| \le \varphi(x,t)(1+|x|^2)^{-1/2}$ we deduce

$$\sum_k |b_k(t)|^2 \le C\mathrm{e}^{-2\alpha(t-t_0)}\big[|y(t) - y'(t)|^{2p} + \|\sigma - \sigma'\|_i^2(1 + |y'(t)|^2)^{ip}\big)\big],$$

for $i = 0, 2$, with the norms (5.61) and (5.62).

Similarly, write

$$c(\zeta, t) = \sum_i \tilde\gamma_i(\zeta, t)\int_0^1 \partial_i\varphi\big(y(t) - y'(t) + \theta\tilde\gamma(\zeta, t)\big)\mathrm{d}\theta$$

to obtain $|c(\zeta,t)| \le 2^{p-1}|\tilde\gamma(\zeta,t)|\big(|y(t) - y'(t)|^{p-1} + |\tilde\gamma(\zeta,t)|^{p-1}\big)$. This, together with

$$|\tilde\gamma(\zeta,t)| \le |\gamma(\zeta,t,y(t)) - \gamma(\zeta,t,y'(t))| + |\gamma(\zeta,t,y'(t)) - \gamma'(\zeta,t,y'(t))|,$$

yields

$$\begin{aligned}|c(\zeta,t)|^2 \le K\big[&|\gamma(\zeta,t,y(t)) - \gamma(\zeta,t,y'(t))|^2|y(t) - y'(t)|^{2p-2} + \\ &+ |\gamma(\zeta,t,y(t)) - \gamma(\zeta,t,y'(t))|^{2p} + |\gamma(\zeta,t,y'(t)) - \gamma'(\zeta,t,y'(t))|^{2p} + \\ &+ |\gamma(\zeta,t,y'(t)) - \gamma'(\zeta,t,y'(t))|^2|y(t) - y'(t)|^{2p-2}\big],\end{aligned}$$

for some constant K and the norm (5.61), and analogously for the norm with $i = 1$ (5.62). Hence, integrating in ζ we deduce

$$\begin{aligned}\int_{\mathbb{R}^d_*} |c(\zeta,t)|^2\pi(\mathrm{d}\zeta) \le C\mathrm{e}^{-2\alpha(t-t_0)}\big[&|y(t) - y'(t)|^{2p} + \\ &+ \big(\|\gamma - \gamma'\|_{i,2}^{2p} + \|\gamma - \gamma'\|_{i,2p}^{2p}\big)(1 + |y'(t)|^2)^{ip}\big],\end{aligned}$$

for a constant C depending on the bounds of γ on assumptions (5.50) and (5.51).

Finally, following the last argument (regarding the sup in $[t_0, t]$) in Theorem 5.11, we obtain (5.63) as desired. □

• *Remark* 5.14. Perhaps, an important application of Proposition 5.13 is when we take $g = g'$, $\sigma = \sigma'$ and $\gamma'(\zeta,t,x,v) = \gamma(\zeta,t,x,v)\mathbb{1}_{|\zeta|\ge\varepsilon}$, i.e., without the small jumps. In this case, with $p \ge 2$, $i = 0, 1$,

$$\begin{aligned}\Lambda_{i,p}^p(g - g', \sigma - \sigma', \gamma - \gamma') \le C\sup_{t,x,v}\Big\{&(1 + |x|^2)^{-ip/2} \times \\ &\times \Big[\int_{|\zeta|<\varepsilon} |\gamma(\zeta,t,x,v)|^p\pi(\mathrm{d}\zeta) + \Big(\int_{|\zeta|<\varepsilon} |\gamma(\zeta,t,x,v)|^2\pi(\mathrm{d}\zeta)\Big)^{p/2}\Big]\Big\},\end{aligned}$$

which goes to zero as $\varepsilon \to 0$. □

• *Remark* 5.15. Note that in the definition of the norms (5.61) and (5.62), we could take the supremum for $t = s$ within the time horizon, i.e., s in $[t_0, t]$, with obvious notation change. Moreover, we could redefine the norms (5.61) and (5.62), where supremum is taken only on x, keeping t and v constant, so that

$$\begin{aligned}\Lambda_{i,p}(g-g', \sigma-\sigma', \gamma-\gamma', t, v) = \\ = \|g(t,\cdot,v) - g'(t,\cdot,v)\|_i + \|\sigma(t,\cdot,v) - \sigma'(t,\cdot,v)\|_i + \\ + \|\gamma(\cdot,t,\cdot,v) - \gamma'(\cdot,t,\cdot,v)\|_{i,p} + \|\gamma(\cdot,t,\cdot,v) - \gamma'(\cdot,t,\cdot,v)\|_{i,2},\end{aligned}$$

i.e.,

$$\Lambda_{i,p}(g-g', \sigma-\sigma', \gamma-\gamma') = \sup_{t_0 \le s \le t, v} \Lambda_{i,p}(g-g', \sigma-\sigma', \gamma-\gamma', t, v),$$

agree with the previous definition. In this case, we obtain the estimate

$$\begin{aligned}\mathbb{E}\Big\{ \sup_{t_0 \le s \le t} |y(s) - y'(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \le M \mathbb{E}\Big\{ \int_{t_0}^t (1 + |y(s)|^2)^{ip/2} \times \\ \times \, \mathrm{e}^{-\alpha(s-t_0)} \Lambda^p_{i,2p}(g-g', \sigma-\sigma', \gamma-\gamma', s, v(s)) \, \mathrm{d}s \Big\},\end{aligned}$$

for $p \ge 2$, $i = 0, 1$ and some constant M as in Proposition 5.13. Note the distinct role of the space variable x and the time variable t. □

It also clear that we can consider the dependency of the solution with respect to the control parameter. The norms (5.61) and (5.62) have to be adjusted. Define the following weighted norms, for $f : \mathbb{R}^d \to \mathbb{R}^d$ we set, for $i = 0, 1$,

$$[\![f]\!]_i := \sup \big\{ |f(x)| \, (1 + |x|^2)^{-i/2} : x \in \mathbb{R}^d \big\} \tag{5.69}$$

and, for $p \ge 2$,

$$\left\{ \begin{aligned}[\![h]\!]_{i,p} := \sup \Big\{ \Big(\int_{\mathbb{R}^m_*} |h(\zeta, x)|^p \pi(\mathrm{d}\zeta) \Big)^{1/p} (1 + |x|^2)^{-i/2} + \\ + \Big(\int_{\mathbb{R}^m_*} |h(\zeta, x)|^2 \pi(\mathrm{d}\zeta) \Big)^{1/2} (1 + |x|^2)^{-i/2} : x \in \mathbb{R}^d \Big\},\end{aligned} \right. \tag{5.70}$$

for $h : \mathbb{R}^m_* \times \mathbb{R}^d \to \mathbb{R}$, and for coefficients g, σ, γ

$$\left\{ \begin{aligned}Q_{i,p}(g, \sigma, \gamma, t, x, v, v') = \big([\![g(t,\cdot,v) - g(t,\cdot,v')]\!]_i + \\ + [\![\sigma(t,\cdot,v) - \sigma(t,\cdot,v')]\!]_i + [\![\gamma(\cdot,t,\cdot,v) - \gamma(\cdot,t,\cdot,v')]\!]_{i,2} + \\ + [\![\gamma(\cdot,t,\cdot,v) - \gamma(\cdot,t,\cdot,v')]\!]_{i,p} \big) (1 + |x|^2)^{i/2},\end{aligned} \right. \tag{5.71}$$

Recall that the indexes $i = 0, 1$ and $p \ge 2$. Note that $Q_{0,p}(g, \sigma, \gamma, t, x, v, v')$ is constant in the variable x.

Essentially, based on the argument used in Proposition 5.13, we can deduce that if $\varphi(x,t) : (x,t) \mapsto \mathrm{e}^{-\alpha(t-t_0)}(\lambda + |x|^2)^{p/2}$ then there exist positive constants $\tilde\beta_g(p,\lambda)$, $\tilde\beta_\sigma(p,\lambda)$, $\tilde\beta_\gamma(p,\lambda)$ and C_p depending only through the bounds in assumptions (5.11) and (5.50) such that

$$\sum_i \big[g_i(t,x,v) - g_i(t,x',v')\big]\,\partial_i\varphi(x-x',t) \le \tilde\beta_g(p,\lambda)\,\varphi(x-x',t) + \\ + C_p\,[\![g(t,\cdot,v) - g(t,\cdot,v')]\!]_i^p (1+|x'|^2)^{ip/2}\mathrm{e}^{-\alpha(t-t_0)},$$

$$\sum_{ijk} \big[\sigma_{ik}(t,x,v) - \sigma_{ik}(t,x',v')\big]\,\partial_{ij}\varphi^2(x-x',t) \times \\ \times\big[\sigma_{sk}(t,x,v) - \sigma_{jk}(t,x',v')\big] \le \tilde\beta_\sigma(p,\lambda)\,\varphi(x-x',t) + \\ + C_p \sum_k [\![\sigma_k(t,\cdot,v) - \sigma_k(t,\cdot,v')]\!]_i^p (1+|x'|^2)^{ip/2}\mathrm{e}^{-\alpha(t-t_0)},$$

and

$$\int_{\mathbb{R}_*^m} \Big[\varphi\big(x-x'+\gamma(\zeta,t,x,v)-\gamma(\zeta,t,x',v)\big) - \varphi(x-x') - \\ - \sum_i [\gamma_i(\zeta,t,x,v) - \gamma_i(\zeta,t,x',v)]\,\varphi(x_i-x_i')\Big]\pi(\mathrm{d}\zeta) \le \\ \le \tilde\beta_\gamma(p,\lambda)\,\varphi(x-x',t) + \\ + C_p\,[\![\gamma(\cdot,t,\cdot,v) - \gamma(\cdot,t,\cdot,v')]\!]_{i,p}^p (1+|x'|^2)^{ip/2}\mathrm{e}^{-\alpha(t-t_0)},$$

for every t,x,x',v,v' and $p,\alpha,\lambda>0$, and

$$\tilde\beta(p) := \sup\big\{\tilde\beta_g(p,\lambda) + \tilde\beta_\sigma(p,\lambda) + \tilde\beta_\gamma(p,\lambda) : \lambda > 0\big\}. \tag{5.72}$$

Actually, for any $\tilde\beta(p) > \beta(p)$ given by (5.49) we can determine a constant $C_p > 0$ matching the above conditions.

Now, let $(x(t) : t \ge t_0)$ and $(x'(t) : t \ge t_0)$ be two solutions of the d-dimensional stochastic ordinary differential equation (5.3) corresponding to the control parameters $(v(s) : t \ge t_0))$ and $(v'(s) : t \ge t_0))$, respectively. Now, use the Itô formula with the function φ and the process $x(t) - x'(t)$ to get

$$\mathrm{d}\varphi(x(t),t) = a(t)\mathrm{d}t + \sum_{k=1}^n b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}_*^m} c(\zeta,t)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}t), \tag{5.73}$$

where

$$a(t) \le \big(\tilde\beta(p) - \alpha\big)\,\varphi\big(x(t)-x'(t),t\big) + \\ + C_p Q_{i,p}^p(g,\sigma,\gamma,t,x'(t),v(t),v'(t))\mathrm{e}^{-\alpha(t-t_0)},$$

$$b_k(t) := \sum_{i=1}^d [\sigma_{ik}(t,x(t),v(t)) - \sigma_{ik}(t,x'(t),v'(t))]\,\partial_i\varphi\big(x(t)-x'(t),t\big),$$

$$c(\zeta,t) := \varphi\big(x(t)+\gamma(\zeta,t,x(t),v(t)) - \gamma(\zeta,t,x'(t),v'(t)),t\big) - \\ - \varphi\big(x(t)-x'(t),t\big),$$

and the constant C_p depends only on the parameter $p \geq 2$, the dimension d, and $\tilde{\beta}(p)$. Hence, for $\alpha \geq \tilde{\beta}(p)$, we deduce the estimate

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ [\alpha - \tilde{\beta}(p)] \displaystyle\int_{t_0}^{t} |x(s) - x'(t)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s + \\ \qquad\qquad + |x(t) - x'(t)|^p \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \\ \qquad \leq C_p \, \mathbb{E}\Big\{ \displaystyle\int_{t_0}^{t} Q^p_{i,p}\big(g, \sigma, \gamma, s, x'(s), v(s), v'(s)\big) \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s \Big\}, \end{array} \right. \tag{5.74}$$

for every $t \geq t_0$. Similarly, going back to the above Itô equality, taking the supremum in t before the expectation, and using the inequalities

$$\sum_k |b_k(s)|^2 + \int_{\mathbb{R}^d_*} |c(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \leq C\big[|\varphi\big(x(s) - x'(s), s\big)|^2 + \\ + Q^{2p}_{i,2p}\big(g, \sigma, \gamma, s, x'(s), v(s), v'(s)\big) \, \mathrm{e}^{-2\alpha(s-t_0)} \big],$$

we obtain for $\alpha > \tilde{\beta}(p)$ the following estimate

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \displaystyle\sup_{t_0 \leq t \leq t_1} \big|x(t) - x'(t)\big|^p \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \\ \qquad \leq M_p \, \mathbb{E}\Big\{ \displaystyle\int_{t_0}^{t} Q^p_{i,2p}\big(g, \sigma, \gamma, s, x'(s), v(s), v'(s)\big) \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s \Big\}, \end{array} \right. \tag{5.75}$$

for some constant M_p depending only on the parameter $p \geq 2$, the dimension d, the constant $\tilde{\beta}(p)$, and the bounds of σ and γ through conditions (5.50) and (5.51) (with $2p$ instead of p). Recall that the expression $Q_{i,p}$ is given by (5.71).

Exponential Bounds

An important case is when the coefficients are bounded, i.e., assume that there exist constants C, $\beta > 0$ such that

$$\left\{ \begin{array}{l} |g(t,x,v)| + |\sigma(t,x,v)|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^2 \exp\big(2\beta|\gamma(\zeta,t,x,v)|\big) \pi(\mathrm{d}\zeta) \leq C, \end{array} \right. \tag{5.76}$$

for every (t,x,v) in $[t_0, \infty) \times \mathbb{R}^d \times \mathbb{R}^q$.

Proposition 5.16. *Let $(x(t) : t \geq t_0)$ be a solution of the d-dimensional stochastic ordinary differential equation (5.5) under the assumption (5.76). Then for every $\lambda > 0$ there exists $\alpha_\lambda > 0$, depending only on the constants C and β in (5.76), such that for every $\alpha > \alpha_\lambda$ we have*

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ (\alpha - \alpha_\lambda) \displaystyle\int_{t_0}^{t} \exp\big(-\alpha(s-t_0) + \beta\sqrt{\lambda + |x(s)|^2}\big) \mathrm{d}s + \\ \qquad + \exp\big(-\alpha(t-t_0) + \beta\sqrt{\lambda + |x(t)|^2}\big) \Big\} \leq \\ \qquad \leq \mathbb{E}\Big\{ \exp\big(\beta\sqrt{\lambda + |x(t_0)|^2}\big) \Big\}, \end{array} \right. \tag{5.77}$$

for every $t \geq t_0$. Moreover,

$$\begin{cases} \mathbb{E}\Big\{ \sup\limits_{t_0 \leq s \leq t} \exp\big(-\alpha(s-t_0) + \beta\sqrt{\lambda + |x(s)|^2}\big)\Big\} \leq \\ \qquad\qquad \leq K\,\big[1 + \dfrac{1}{\alpha - \alpha_\lambda}\big]\,\mathbb{E}\big\{ \exp\big(\beta\sqrt{\lambda + |x(t_0)|^2}\big)\big\}. \end{cases} \tag{5.78}$$

for some constant K, depending only on the bounds of σ and γ through conditions (5.76).

Proof. As in the previous theorem, we can apply Itô formula to the function

$$\varphi : (x,t) \mapsto \exp[-\alpha(t-t_0) + \beta(\lambda + |x|^2)^{1/2}]$$

with α, $\lambda > 0$ and $\beta > 0$, the same constant in hypothesis (5.76), and the solution process $x(t)$ of (5.5), to obtain

$$\mathrm{d}\varphi(x(t),t) = a(t)\mathrm{d}t + \sum_{k=1}^{n} b_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^m_*} c(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \tag{5.79}$$

where

$$\begin{aligned} a(t) :=\ & \partial_t\varphi(x(t),t) + \sum_{i=1}^{d} g_i(t,x(t),v(t))\,\partial_i\varphi(x(t),t) + \\ & + \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sum_{k=1}^{n}\sigma_{ik}(t,x(t),v(t))\sigma_{jk}(t,x(t),v(t))\Big)\,\partial^2_{ij}\varphi(x(t),t) + \\ & + \int_{\mathbb{R}^m_*}\big[\varphi(x(t) + \gamma(\zeta,t,x(t),v(t)),t) - \varphi(x(t),t) - \\ & - \sum_{i=1}^{d}\gamma_i(\zeta,t,x(t),v(t))\,\partial_i\varphi(x(t),t)\big]\,\pi(\mathrm{d}\zeta), \end{aligned}$$

$$\begin{aligned} b_k(t) &:= \sum_{i=1}^{d}\sigma_{ik}(t,x(t),v(t))\,\partial_i\varphi(x(t),t), \\ c(\zeta,t) &:= \varphi(x(t) + \gamma(\zeta,t,x(t),v(t)),t) - \varphi(x(t),t). \end{aligned}$$

Since

$$\begin{aligned} \partial_t\varphi(x,t) &= -\alpha\varphi(x,t), \\ \partial_i\varphi(x,t) &= \beta x_i(\lambda + |x|^2)^{-1/2}\varphi(x,t), \\ \partial_{ij}\varphi(x,t) &= \beta\big[\delta_{ij}(\lambda + |x|^2)^{-1/2} - x_i x_j(\lambda + |x|^2)^{-3/2} + \\ &\qquad + \beta x_i x_j(\lambda + |x|^2)^{-1}\big]\varphi(x,t), \end{aligned}$$

where $\delta_{ij} = 1$ only if $i = j$ and vanishes otherwise, we obtain

$$\sum_i |g_i\partial_i\varphi| \leq |g|\beta\varphi, \qquad\qquad \sum_{ij}|\sigma_{ik}\sigma_{jk}\partial_{ij}\varphi| \leq |\sigma_k|^2\beta(\lambda^{-1/2} + \beta)\varphi,$$

and

$$\big|\varphi(x+\gamma,t)-\varphi(x,t)-\gamma\cdot\nabla\varphi(x,t)\big|\le\big[\beta(\lambda^{-1/2}+\beta)|\gamma|^2\,\mathrm{e}^{\beta|\gamma|}\big]\varphi(x,t).$$

Hence, for each $\lambda>0$ we can choose $\alpha_\lambda>0$ such that $a(t)\le 0$ and (5.77) is established.

Similarly to Theorem 5.11, by means of the stochastic integral estimates (5.46) and (5.47), with $p=1$, and the inequality

$$\begin{aligned}&|b_k(t)|^2\le|\sigma_k(t,x(t),v(t))|^2\beta^2\big(\varphi(x(t),t)\big)^2,\\&|c(\zeta,t)|^2\le|\gamma(\zeta,t,x(t),v(t))|^2\mathrm{e}^{2\beta|\gamma(\zeta,t,x(t),v(t))|}\beta^2\big(\varphi(x(t),t)\big)^2\end{aligned}$$

and (5.76), we deduce the bound (5.78).

Note α_λ is equal to $C\beta\max\{1,(\lambda^{-1/2}+\beta)\}$, where C is the constant in assumption (5.76) with $\beta/2$ instead of β. Thus, as $\lambda\to 0$ we may have $\alpha_\lambda\to\infty$. However, as $\beta\to 0$ we do have $\alpha_\lambda\to 0$. □

Sometimes, we may need to use a more general setting, e.g., a stochastic ordinary differential equation of the form

$$\begin{cases}x_i(t)=x_i^0+\displaystyle\int_{t_0}^t g_i(s,x(s))\mathrm{d}s+\\\qquad+\displaystyle\sum_{k=1}^n\int_{t_0}^t\sigma_{ik}(s,x(s))\mathrm{d}w_k(s)+\\\qquad+\displaystyle\int_{\mathbb{R}_*^m\times]t_0,t]}\gamma_i(\zeta,s,x(s))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s),\quad\forall\,t\ge t_0,\end{cases}\tag{5.80}$$

where the coefficients $g=g(t,x,\omega)$, $\sigma=\sigma(t,x,\omega)$ and $\gamma=\gamma(\zeta,t,x,\omega)$ are predictable processes, continuous in x for every fixed (t,ω), for the functions γ this means that

$$\lim_{y\to x}\int_{\mathbb{R}_*^m}|\gamma(\zeta,t,y,\omega)-\gamma(\zeta,t,x,\omega)|^2\pi(\mathrm{d}\zeta)=0,$$

for every x, t and ω, and also, locally bounded in x, locally (square-, for σ and γ), integrable in t, for every ω in the following sense

$$\begin{cases}\displaystyle\int_{t_0}^T\sup_{|x|\le R}\Big\{|g(t,x,\omega)|+|\sigma(t,x,\omega)|^2+\\\qquad+\displaystyle\int_{\mathbb{R}_*^m}|\gamma(\zeta,t,x,\omega)|^2\pi(\mathrm{d}\zeta)\Big\}\mathrm{d}t<\infty,\end{cases}\tag{5.81}$$

for every $T>t_0$ and $R>0$. Certainly $x(t)=x^0(t)$ for every $t\le t_0$ so that all integrals in the right-hand side of (5.80) have an implicit coefficient of the form $\mathbb{1}_{t_0<s}$.

Under these conditions, the so-called Euler's method can be applied as in Krylov [141, Chapter V, pp. 165–220], where only the diffusion case is fully

treated. The Euler's method consists in setting $y_i^m(t_0) = x_i^0$, $i = 1, \dots, d$, and then, recursively,

$$\begin{cases} y_i^m(t) = y_i^m(t_{\ell,m}) + \displaystyle\int_{t_{\ell,m}}^t g_i(s, y^m(t_{\ell,m}))\mathrm{d}s + \\ \qquad + \displaystyle\sum_{k=1}^n \int_{t_{\ell,m}}^t \sigma_{ik}(s, y^m(t_{\ell,m}))\mathrm{d}w_k(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m \times]t_{\ell,m}, t]} \gamma_i(\zeta, s, y^m(t_{\ell,m}))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{cases} \tag{5.82}$$

for every $t_{\ell,m} < t \le t_{\ell+1,m}$, $k = 0, 1, \dots$, where $t_{\ell,m} := t_0 + \ell/m$. Based on the left-continuous processes

$$p^m(t) := \begin{cases} y^m(\tau^m(t)) - y^m(t) & \text{if } t_0 < t, \\ 0 & \text{otherwise,} \end{cases}$$

where $\tau^m(t) = t_{\ell,m}$ for $t_{\ell,m} < t \le t_{\ell+1,m}$, it can be proved that for every $\varepsilon > 0$ there exists a $M = M(\varepsilon)$ such that

$$P\Big\{ \sup_{t_0 \le t \le 1/\varepsilon} |y^m(t) - x(t)| > \varepsilon \Big\} \le \varepsilon, \quad \forall m \ge M,$$

i.e., $y^m \to x$ locally uniformly in probability, under a local monotonicity condition, i.e. a local version of the condition (5.50), and a growth condition of the form (5.8), i.e. for any $T > t_0$ there exists a constant M such that

$$x\, g(t,x,\omega) + |\sigma(t,x,\omega)|^2 + \int_{\mathbb{R}_*^m} |\gamma(\zeta,t,x,\omega)|^2 \pi(\mathrm{d}\zeta) \le M|x|^2,$$

for every t in $[t_0, T]$, x in $\mathbb{R}^d$ and any ω.

An important point here is the fact that the above convergence of the Euler's method holds under weaker assumptions, in particular a polynomial growth may be allowed for the coefficients and still existence and uniqueness results can be deduced. An interesting case is a system of the form $x = (x_1, x_2)$, $g = (g_1, g_2)$, and so on, with $g_1 = g_1(t, x_1)$ but $g_2 = g_2(t, x_1, x_2)$, which may be solved first in x_1 and then in x_2, allowing a polynomial growth of all coefficients g_2, σ_2, γ_2 in the variable x_1.

We cite the monograph by Cerrai [36, Chapter 1, pp. 21-64] where a self-contained treatment of the stochastic equations with coefficients with polynomial growth of the Itô type are considered. Actually, the expressions (5.44) and (5.49) defining the constants $\alpha(p, \lambda)$ and $\beta(p)$ can be combined so that the growth (in the space variable x) of the drift coefficient g compensates the growth of the diffusion and jump coefficients σ and γ in such a way that both constants, $\alpha(p, \lambda)$ and $\beta(p)$, are finite. These conditions are of the form: for every $T > t_0$, $\varepsilon > 0$ there exists $M(\varepsilon, T)$ such that

$$\begin{aligned} &\varepsilon\xi \cdot \nabla g(t,x,\omega)\xi + |\nabla\sigma(t,x,\omega)\xi|^2 + \\ &\quad + \int_{\mathbb{R}_*^m} |\nabla\gamma(\zeta,t,x,\omega)\xi|^2 \pi(\mathrm{d}\zeta) \le M(\varepsilon,T)|\xi|^2, \end{aligned}$$

for every t in $[t_0, T]$, x, ξ in $\mathbb{R}^d$ and ω. Here ∇ means either the vector or the matrix of all first order derivatives in x. Clearly, because of the time-dependence, we need also to add a condition like

$$|g(t,0,\omega)| + |\sigma(t,0,\omega)| + \int_{\mathbb{R}^m_*} |\gamma(\zeta,t,0,\omega)|^2 \pi(\mathrm{d}\zeta) \le C_T,$$

for every t in $[t_0, T]$ and any ω. Finally, we need also to make some assumption to handle the large jumps, that is a variation of (5.51) combined with the above conditions. Essentially, this means that if the larger growth of g is of the form $-cx^{2m+1}$, $c > 0$ then σ and γ may have a growth up to x^m.

• *Remark* 5.17. It is clear that we can consider changing the coefficients and keeping them bounded, i.e., under assumption (5.76). Following the steps in Proposition 5.13, we can obtain bounds in term of exponential weighted norms of the type

$$\sup_x \left\{ |h(t,x,v) - h'(t,x,v)| \, \exp\left(\beta\sqrt{1+|x|^2}\right)\right\}$$
$$\sup_x \left\{ |h(t,x,v) - h(t,x,v')| \, \exp\left(\beta\sqrt{1+|x|^2}\right)\right\}$$

similar to the estimates (5.63), (5.64), (5.74) and (5.75), e.g.,

$$\mathbb{E}\Big\{ \sup_{t_0\le s\le t} |y(s) - y'(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \le$$
$$\le M\Lambda^p_{\beta,p}(g-g', \sigma-\sigma', \gamma-\gamma')\mathbb{E}\big\{ \exp\left(p\beta\sqrt{1+|y(t_0)|^2}\right)\big\},$$

where $\Lambda_{\beta,p}$ is suitable defined in term of exponential weighted norms. □

Second-Order Bounds

Most of this section can be rewritten using the differentiability property, but we prefer to use semi-concave arguments. Some preliminaries are necessary.

A function h from $\mathbb{R}^d$ into $\mathbb{R}^n$ is called *semi-concave* if for every ball $B(0,r)$, $r > 0$ there exists a constant $C = C_r > 0$ such that the function $x \mapsto h(x) - C_r|x|^2$ is concave on $B(0,r)$, i.e., for every x, y in $\mathbb{R}^d$, $|x| < r$, $|y| < r$, we have

$$\theta h(x) + (1-\theta)h(y) - h\big(\theta x + (1-\theta)y\big) \le C_r\theta(1-\theta)|x-y|^2, \tag{5.83}$$

for any θ in $[0,1]$. If h is continuous, this is equivalent to the condition

$$h(x+z) - 2h(x) + h(x-z) \le C_r|z|^2$$

for all z sufficiently small. As expected, it is clear that any concave function is indeed semi-concave. Based on expression

$$\theta h(x) + (1-\theta)h(y) - h\big(\theta x + (1-\theta)y\big) =$$
$$= \theta(1-\theta)(x-y)\cdot \int_0^1 \Big[\nabla h\big(y + t\theta(y-x)\big) - \nabla h\big(x + t(1-\theta)(x-y)\big)\Big]\mathrm{d}t,$$

we see that any continuously differentiable function with a locally Lipschitz derivative is indeed semi-concave. Perhaps a more relevant class of non-differentiable semi-concave functions with values in $[-\infty, \infty)$ follows from the fact that the infimum of uniformly semi-concave functions is indeed semi-concave. Many more properties on semi-concave function are

well known, for instance see Bardi and Capuzzo-Dolcetta [9, Section II.4.2, pp. 65–76], Da Prato and Zabczyk [53, Appendix C.1].

Our interest is on functions such that h and $-h$ are semi-concave[3], i.e., for every $r > 0$ there exists a constant $C_r > 0$ such that for every x, y in $\mathbb{R}^d$, $|x| < r$, $|y| < r$, we have

$$\big|\theta h(x) + (1-\theta)h(y) - h\big(\theta x + (1-\theta)y\big)\big| \le C_r\theta(1-\theta)|x-y|^2, \tag{5.84}$$

for any θ in $[0,1]$. It is clear that any continuously differentiable function h with a locally Lipschitz derivative satisfies (5.84), moreover, the converse is also true.

Let $x(t)$, $y(t)$ and $z(t)$ be three solutions of the d-dimensional stochastic ordinary differential equation (5.5). We want to estimate

$$\begin{cases} \psi_\theta(x,y,z) := \theta^2(1-\theta)^2|x-y|^4 + |\theta x + (1-\theta)y - z|^2, \\ \psi_{\lambda,\theta,p}(x,y,z) := \big(\lambda + \psi_\theta(x,y,z)\big)^{p/2}, \end{cases} \tag{5.85}$$

for any θ in $[0,1]$, $p \ge 2$, and let $\lambda > 0$ vanish. Thus, if h denotes any of the coefficients g, σ or γ then the Itô process $\theta x(t)+(1-\theta)y(t)-z(t)$ has coefficients of the form $\theta h(x(t)) + (1-\theta)h(y(t)) - h(z(t))$, which can be estimated in term of

$$\big[\theta h(x(t)) + (1-\theta)h(y(t)) - h\big(\theta x(t) + (1-\theta)y(t)\big)\big]$$

and

$$\big[h\big(\theta x(t) + (1-\theta)y(t)\big) - h(z(t))\big].$$

By means of the Lipschitz constant, the second difference is bounded by $|\theta x(t)+(1-\theta)y(t) - z(t)|$ while the first difference can be controlled by $\theta(1-\theta)|x(t) - y(t)|^2$, if the first derivative $h' = (\partial_i h : i = 1, \dots, d)$ is Lipschitz continuous.

Hence, assume that the coefficients have linear growth and their first derivatives are bounded and Lipschitz continuous, i.e., for some $p \ge 2$ there exists a positive constant $C = C_p$ such that

$$\begin{cases} |g(t,x,v)|^p + |\sigma(t,x,v)|^p + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^p \pi(\mathrm{d}\zeta) \le \\ \qquad\qquad\qquad\qquad \le C(1+|x|^p), \end{cases} \tag{5.86}$$

and

$$|g'(t,x,v)|^p + |\sigma'(t,x,v)|^p + \int_{\mathbb{R}^m_*} |\gamma'(\zeta,t,x,v)|^p \pi(\mathrm{d}\zeta) \le C, \tag{5.87}$$

[3] if $-h$ is semi-concave then h is called semi-convex.

for every (t,x,v) in $[0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, and a constant $M>0$ such that

$$\begin{cases} |g'(t,x,v)-g'(t,x',v)|^p+|\sigma'(t,x,v)-\sigma'(t,x',v)|^p+ \\ \qquad +\displaystyle\int_{\mathbb{R}^m_*}|\gamma'(\zeta,t,x,v)-\gamma'(\zeta,t,x',v)|^p\pi(\mathrm{d}\zeta)\le M|x-x'|^p, \end{cases} \tag{5.88}$$

for every (t,x,v), (t,x',v) in $[0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$.

Similarly to (5.49), we define

$$\kappa(p):=\sup_{\lambda>0}\big\{p\,\kappa_g(\lambda)+(p/2)\,\kappa_\sigma(p,\lambda)+\kappa_\gamma(p,\lambda)\big\}, \tag{5.89}$$

where the expressions of $\kappa_g(\lambda)$, $\kappa_\sigma(p,\lambda)$ and $\kappa_\gamma(p,\lambda)$ are found by Itô formula, e.g.,

$$\kappa_g(\lambda):=\sup\Big\{\frac{2|x-y|^3(x-y)\,[g(t,x,v)-g(t,y,v)]}{\lambda+\theta^2(1-\theta)^2|x-y|^4+|\theta x+(1-\theta)y-z|^2}+ \\ +\frac{(\theta x+(1-\theta)y-z)\,[\theta g(t,x,v)+(1-\theta)g(t,y,v)-g(t,z,v)]}{\lambda+\theta^2(1-\theta)^2|x-y|^4+|\theta x+(1-\theta)y-z|^2}\Big\},$$

and the suprema are taken for any t in $[0,\infty)$, x, y, z in $\mathbb{R}^d$, v in $\mathbb{R}^q$ and θ in $[0,1]$. This means that under the assumptions (5.86), (5.87) and (5.88) the sumpremum $\kappa(p)$ is finite and satisfies $a(t)\le\kappa(p)\psi_{\lambda,\theta,p}(x(t),y(t),z(t))$, where

$$\mathrm{d}\psi_{\lambda,\theta,p}(x(t),y(t),z(t))=a(t)\mathrm{d}t+\sum_{k=1}^n b_k(t)\mathrm{d}w_k(t)+\int_{\mathbb{R}^m_*}c(\zeta,t)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}t),$$

for every $t\ge0$. Moreover, we also have

$$\sum_{k=1}^n|b_k(t)|^2+\int_{\mathbb{R}^m_*}|c(\zeta,t)|^2\pi(\mathrm{d}\zeta)\le C_p|\psi_{\lambda,\theta,p}(x(t),y(t),z(t))|^2,$$

for every $t\ge0$ and some constant C_p. Note that we have $\kappa(p)\ge\beta(p)$.

As in Theorem 5.11 we can prove

Proposition 5.18. *Let $x(t)$, $y(t)$ and $z(t)$ be three solutions of the d-dimensional stochastic ordinary differential equation (5.5) for $t\ge t_0$. If $\alpha\ge\kappa(p)$, $p\ge2$, as defined by (5.89), then for $\psi_{\lambda,\theta,p}(x,y,z)$ given by (5.85) and under assumptions (5.86), (5.87) and (5.88) we have*

$$\begin{cases} \mathbb{E}\Big\{[\alpha-\kappa(p)]\displaystyle\int_{t_0}^t\big[\psi_\theta(x(s),y(s),z(s))\big]^{p/2}\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s+ \\ \qquad +\big[\psi_\theta(x(t),y(t),z(t))\big]^{p/2}\,\mathrm{e}^{-\alpha(t-t_0)}\Big\}\le \\ \qquad\le\mathbb{E}\big\{\big[\psi_\theta(x(t_0),y(t_0),z(t_0))\big]^{p/2}\big\}, \end{cases} \tag{5.90}$$

for every $t\ge t_0$. Moreover, there exists a positive constant C, depending only on p and the bounds of σ and γ through conditions (5.86), (5.87) and (5.88), such that

$$\begin{cases} \mathbb{E}\Big\{\displaystyle\sup_{t_0\le s\le t}\big[\psi_\theta(x(s),y(s),z(s))\big]^{p/2}\,\mathrm{e}^{-\alpha(s-t_0)}\Big\}\le \\ \qquad\le C\,\big[1+\frac{1}{\alpha-\kappa(p)}\big]\,\mathbb{E}\big\{\big[\psi_\theta(x(t_0),y(t_0),z(t_0))\big]^{p/2}\big\}. \end{cases} \tag{5.91}$$

for every $t \geq t_0$. □

Clearly, we apply the above estimate as follows. Denote by $y(t,x)$ the solution of the d-dimensional stochastic ordinary differential equation (5.5) with the initial condition $y(t_0,x) = x$. Then we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ [\alpha - \kappa(p)] \displaystyle\int_{t_0}^{t} \big| \theta y(s,x) + (1-\theta) y(s,x') - \\ \quad - y(s, \theta x + (1-\theta)x') \big|^p \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s + \big| \theta y(t,x) + \\ \quad + (1-\theta) y(t,x') - y(t, \theta x + (1-\theta)x') \big|^p \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \\ \qquad\qquad \leq \big[\theta(1-\theta) |x - x'|^2) \big]^p, \end{array} \right. \tag{5.92}$$

and

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \displaystyle\sup_{t_0 \leq s \leq t} \big| \theta y(t,x) + (1-\theta) y(t,x') - \\ \quad - y(t, \theta x + (1-\theta)x') \big|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq \\ \qquad\qquad \leq C \left[1 + \dfrac{1}{\alpha - \kappa(p)} \right] \big[\theta(1-\theta) |x - x'|^2) \big]^p, \end{array} \right. \tag{5.93}$$

for every $t \geq t_0$. These estimates are used to show that an expression of the form

$$x \mapsto \inf_{v(\cdot)} \mathbb{E}\Big\{ \int_0^\infty f(y(t,x), v(t)) \mathrm{e}^{-\alpha t} \mathrm{d}t \Big\}$$

is semi-concave if the function $x \mapsto f(x,v)$ is Lipschitz continuous and semi-concave (uniformly with respect to v), provided $\alpha > \kappa(2)$.

5.1.5 Linear Equations

For the sake of notation simplicity, the initial time t_0 is taken to be 0 and the initial valued x^0 is assume deterministic. In general, t_0 could be a stopping time and x^0 an $\mathcal{F}(t_0)$-measurable random variable with as many finite moments as needed. Moreover, we drop the control parameter process $(v(t) : t \geq 0)$, understanding that it can be incorporated if necessary.

The particular case where the coefficients $g(t,x,v)$, $\sigma(t,x,v)$ and $\gamma(\zeta,t,x,v)$ are linear (or affine) in the variable x is of special interest. The d-dimensional stochastic ordinary differential equation (5.5) takes the general form

$$\left\{ \begin{array}{l} \mathrm{d}x(s) = \big[A(s)x(s) + a(s) \big] \mathrm{d}s + \\ \qquad + \displaystyle\sum_{k=1}^{n} \big[B_k(s)x(s) + b_k(s) \big] \mathrm{d}w_k(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m} \big[C(\zeta,s)x(s) + c(\zeta,s) \big] \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall s \geq 0, \end{array} \right. \tag{5.94}$$

where $A(s)$, $B_k(s)$ and $C(\zeta, s)$ are $d \times d$-matrix-valued functions, and $a(s)$, $b_k(s)$ and $c(\zeta, s)$ are d-valued adapted processes[4] such that for each $T > 0$ and $p \geq 2$ there exist $K = K(T, p) > 0$ satisfying

$$|A(t)| + |B_k(t)| + \int_{\mathbb{R}_*^m} |C(\zeta, t)|^p \pi(\mathrm{d}\zeta) \leq K, \quad \forall t \in [0, T], \tag{5.95}$$

and

$$\begin{cases} \mathbb{E}\Big\{ \Big[\displaystyle\int_0^T |a(s)|\mathrm{d}s \Big]^p + \sum_{k=1}^n \Big[\int_0^T |b_k(s)|^2 \mathrm{d}s \Big]^{p/2} + \\ \qquad + \Big[\displaystyle\int_0^T \mathrm{d}s \int_{R_*^m} |c(\zeta, s)|^2 \mathrm{d}s \Big]^{p/2} \Big\} \leq K. \end{cases} \tag{5.96}$$

It is clear that we have a unique solution of (5.94) with a given initial condition. Moreover, the a priori bounds on the moments established in Theorem 5.11 hold true.

In the homogeneous case, i.e., $a = b_k = c = 0$, the solution $x(t; s, \xi)$ depends linearly on the initial datum ξ; writing $x(t; s, \xi) = U(t, s)\xi$ we have that

$$\begin{cases} U(t, s) = \mathbb{1} + \displaystyle\int_s^t A(r)U(r, s)\mathrm{d}r + \sum_{k=1}^n \int_s^t B_k(s)U(r, s)\mathrm{d}w_k(r) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m \times]s,t]} C(\zeta, r)U(r, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}r), \quad \forall t \geq s, \end{cases} \tag{5.97}$$

where $\mathbb{1}$ is the identity matrix and the solution $(U(t, s) : t \geq s)$ is a matrix-valued process, i.e., the column j of the matrix $U(t, s)$ is the solution $x(t; s, \xi)$ with $\xi = e_j$, where the j-component of e_j is 1 and zero all others. Its determinant $\mathfrak{w}(t, s) := \det(U(t, s))$ is usually referred to as the stochastic Wronskian determinant. It is the solution of the following stochastic differential equation

$$\begin{aligned} \mathrm{d}\mathfrak{w}(t, s) = \Big\{ \operatorname{Tr} A(t) - \frac{1}{2} \sum_{k=1}^n \Big[\operatorname{Tr} B_k^2(t) - (\operatorname{Tr} B_k(t))^2 \Big] + \\ + \int_{\mathbb{R}_*^m} \big[\det[\mathbb{1} + C(\zeta, t)] - 1 - \operatorname{Tr} C(\zeta, t) \big] \pi(\mathrm{d}\zeta) \Big\} \mathfrak{w}(t, s)\mathrm{d}t + \\ + \sum_{k=1}^n \operatorname{Tr}[B_k(t)] \mathfrak{w}(t, s)\, \mathrm{d}w_k(t) + \\ + \int_{\mathbb{R}_*^m} \big(\det[\mathbb{1} + C(\zeta, t)] - 1 \big) \mathfrak{w}(t, s) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t), \end{aligned}$$

which is known as the stochastic Liouville formula. It is easy to check that for any matrix C we have

$$\begin{aligned} \big| \det(\mathbb{1} + C) - 1 \big| &\leq (1 + |C|)^d - 1 \leq d\,|C| \big(1 + |C|\big)^{d-1} \\ \big| \det(\mathbb{1} + C) - 1 - \operatorname{Tr}(C) \big| &\leq (1 + |C|)^d - 1 - d\,|C| \leq \\ &\leq \frac{d(d-1)}{2} |C|^2 (1 + |C|)^{d-2}, \end{aligned}$$

[4] just adapted is enough, but the predictable projection have to be used anyway.

where d is the dimension and $|A|$ denotes the operator norm $|A| = \sup_{|x|\leq 1} |Ax|$, so that the integral in $\pi(\mathrm{d}z)$ is convergent.

For the proof we have only to apply the Itô formula, recalling that for the determinant function det we have

$$\begin{aligned}
\big(D\det(X)\big)(Y) &= \lim_{t\to 0} \frac{\det(X+tY)-\det(X)}{t} = \\
&= \sum_{h=1}^{d} (-1)^{i+h} \det(X_{ih})\, y_{ih} = \det(X)\,\mathrm{Tr}(X^{-1}Y), \\
\big(D^2\det(X)\big)(Y,Z) &= \lim_{t\to 0} \frac{\big(D\det(X+tZ)\big)(Y) - \big(D\det(X)\big)(Y)}{t} = \\
&= \sum_{i<j}\sum_{h<k} (-1)^{i+j+h+k} \det(X_{ih,jk})\,(y_{ih}z_{jk} - y_{ik}z_{jh}) = \\
&= \det(X)\,\mathrm{Tr}(X^{-1}Y)\,\mathrm{Tr}(X^{-1}Z) - \det(X)\,\mathrm{Tr}(X^{-2}YZ),
\end{aligned}$$

where the $d-1$ dimensional matrix X_{ih} is the matrix X without the i-th row and the h-th column, and the $d-2$ dimensional matrix $X_{ih,jk}$ is the matrix X without the i-th and j-th rows and the h-th and k-th columns. Clearly, the last equalities holds only when $\det(X) \neq 0$.

The explicit solution is given by

$$\begin{aligned}
\mathfrak{w}(t,s) = \exp\Big(\int_s^t \Big\{ &\mathrm{Tr}\big[A(r) - \frac{1}{2}\sum_{k=1}^{n} B_k^2(r)\big] + \\
&+ \int_{\mathbb{R}_*^m} \big[\ln\det[\mathbb{1} + C(\zeta,t)] - \mathrm{Tr}\, C(\zeta,r)\big]\pi(\mathrm{d}\zeta)\Big\}\mathrm{d}r + \\
&+ \sum_{k=1}^{n} \int_s^t \mathrm{Tr}[B_k(r)]\mathrm{d}w_k(r) + \int_{\mathbb{R}_*^m} \ln\det[\mathbb{1} + C(\zeta,t)]\;\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}r)\Big),
\end{aligned}$$

provided that foe any $T > 0$ there exist $a = a_T >$ such that

$$\det[\mathbb{1} + C(\zeta,t)] \geq a_T, \quad \forall (\zeta,t) \in \mathbb{R}_*^m \times [0,T]. \tag{5.98}$$

For instance, we refer to Mao [165, Chapter 3, pp. 91–106], for the case without jumps.

This explicit expression of the stochastic Wronskian process $\mathfrak{w}(t,s)$ proves that the matrix-valued solution process $U(t,s)$ is invertible. Moreover

Lemma 5.19. *Under the assumptions (5.95), (5.96) and (5.98) the evolution matrix-operator $U(t,s)$ is invertible and its inverse $V(t,s) = U^{-1}(t,s)$ is the solution to the following linear stochastic differential equation*

$$\begin{aligned}
V(t,s) = \mathbb{1} - \int_s^t V(r,s)\tilde{A}(r)\mathrm{d}r - \sum_{k=1}^{n}\int_s^t V(r,s)B_k(r)\mathrm{d}w_k(r) + \\
- \int_{\mathbb{R}_*^m \times]s,t]} V(r,s)C(\zeta,r)\big(\mathbb{1} + C(\zeta,r)\big)^{-1}\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}r),
\end{aligned}$$

where

$$\tilde{A}(t) := A(t) - \sum_k B_k^2(t) - \int_{\mathbb{R}_*^m} C^2(\zeta,t)\big(\mathbb{1} + C(\zeta,t)\big)^{-1}\pi(\mathrm{d}\zeta),$$

for every $t \geq 0$.

Proof. Use following version of Itô formula: if we consider the matrix-valued Itô processes

$$\begin{aligned}
\mathrm{d}X(t) &= \alpha(t)\mathrm{d}t + \sum_k \beta_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^n_*} \gamma(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t),\\
\mathrm{d}\bar{X}(t) &= \bar{\alpha}(t)\mathrm{d}t + \sum_k \bar{\beta}_k(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}^n_*} \bar{\gamma}(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t),
\end{aligned}$$

then the Itô formula yields

$$\mathrm{d}\big(\bar{X}(t)X(t)\big) = \big(\mathrm{d}\bar{X}(t)\big)X(t) + \bar{X}(t)\big(\mathrm{d}X(t)\big) + \big(\mathrm{d}\bar{X}(t)\big)\big(\mathrm{d}X(t)\big),$$

where

$$\big(\mathrm{d}\bar{X}(t)\big)\big(\mathrm{d}X(t)\big) = \mathrm{d}\langle\bar{X}(t)X(t)\rangle + \int_{\mathbb{R}^n_*} \bar{\gamma}(\zeta,t)\gamma(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t),$$

and

$$\mathrm{d}\langle\bar{X}(t)X(t)\rangle = \Big[\sum_k \bar{\beta}_k(t)\beta_k(t) + \int_{\mathbb{R}^n_*} \bar{\gamma}(\zeta,t)\gamma(\zeta,t)\pi(\mathrm{d}\zeta)\Big]\mathrm{d}t,$$

so that

$$\begin{aligned}
\mathrm{d}\big(\bar{X}(t)X(t)\big) = \Big[&\bar{\alpha}(t)X(t) + \bar{X}(t)\alpha(t) + \sum_k \bar{\beta}_k(t)\beta_k(t) +\\
&+ \int_{\mathbb{R}^n_*} \bar{\gamma}(\zeta,t)\gamma(\zeta,t)\pi(\mathrm{d}\zeta)\Big]\mathrm{d}t + \sum_k \big[\bar{\beta}_k(t)X(t) + \bar{X}(t)\beta_k(t)\big]\mathrm{d}w_k(t) +\\
&+ \int_{\mathbb{R}^n_*} \big[\bar{\gamma}(\zeta,t)X(t) + \bar{X}(t)\gamma(\zeta,t) + \bar{\gamma}(\zeta,t)\gamma(\zeta,t)\big]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t),
\end{aligned}$$

and in particular, if α, $\bar{\alpha}$, β, $\bar{\beta}$, γ, $\bar{\gamma}$ are replaced by αX, $\bar{X}\bar{\alpha}$, etc, we deduce

$$\begin{aligned}
\mathrm{d}\big(\bar{X}(t)X(t)\big) = \bar{X}(t)P(t)X(t)\mathrm{d}t + \sum_k \bar{X}(t)Q_k(t)X(t)\mathrm{d}w_k(t) +\\
+ \int_{\mathbb{R}^n_*} \bar{X}(t)R(\zeta,t)X(t)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t).
\end{aligned}$$

with

$$\begin{aligned}
P(t) &= \alpha(t) + \bar{\alpha}(t) + \sum_k \bar{\beta}_k(t)\beta_k(t) + \int_{\mathbb{R}^n_*} \bar{\gamma}(\zeta,t)\gamma(\zeta,t)\pi(\mathrm{d}\zeta),\\
Q(t) &= \sum_k \big[\bar{\beta}_k(t) + \beta_k(t)\big], \qquad R(\zeta,t) = \bar{\gamma}(z,t) + \gamma(\zeta,t) + \bar{\gamma}(\zeta,t)\gamma(\zeta,t).
\end{aligned}$$

Hence, if we choose

$$\begin{aligned}
\bar{\alpha}(t) &= -\alpha(t) - \sum_k \beta_k^2(t) - \int_{\mathbb{R}^m_*} \gamma(\zeta,t)\big(\mathbb{1} + \gamma(\zeta,t)\big)^{-1}\pi(\zeta),\\
\bar{\beta}_k &= -\beta_k, \qquad \bar{\gamma} = -\gamma(\mathbb{1}+\gamma)^{-1},
\end{aligned}$$

then all coefficients P, Q, R vanish, $\bar{X}(t)X(t)$ is constant. Finally, apply this argument to $X(t) = U(t,s)$ and $\bar{X}(t) = V(t,s)$ to deduce that

$$V(t,s)U(t,s) = V(s,s)U(s,s) = \mathbb{1}$$

as desired. □

Now, essentially as in the case of ordinary differential equations, the homogeneous equation (5.97) provides the fundamental matrix solution, which yields also a representation for the solution of the non-homogeneous equation (5.94), i.e., a stochastic version of the variation-of-constants formula

Theorem 5.20. *Under the conditions* (5.95), (5.96) *and* (5.98), *the fundamental solution matrix* $(\Phi(t) = U(t,0) : t \geq 0)$ *satisfies* $U(t,s) = \Phi(t)\Phi^{-1}(s)$, *and*

$$\begin{aligned} x(t) := \Phi(t)\xi + \Phi(t)\int_0^t \Phi^{-1}(s)\Big[a(s) - \sum_{k=1}^n B_k(s)b_k(s) - \\ - \int_{\mathbb{R}_*^m} C(\zeta,s)\big(\mathbb{1} + C(\zeta,s)\big)^{-1}c(\zeta,s)\pi(\mathrm{d}\zeta)\Big]\mathrm{d}s + \\ + \sum_{k=1}^n \Phi(t)\int_0^t \Phi^{-1}(s)b_k(s)\mathrm{d}w_k(s) + \\ + \Phi(t)\int_{\mathbb{R}_*^m\times]0,t]} \Phi^{-1}(s)\big(\mathbb{1} + C(\zeta,s)\big)^{-1}c(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s) \end{aligned}$$

is the solution of the linear d-dimensional stochastic ordinary differential equation (5.94) *with the initial condition* $x(0) = \xi$.

Proof. The uniqueness of the stochastic ordinary differential equation (5.94) implies the semi-group property for the composition evolution matrix-operator, i.e., the identity $U(t,s)U(s,r) = U(t,r)$, for any $t \geq s \geq r \geq 0$, and in particular $U(t,0) = U(t,s)U(s,0)$. On the other hand, in view of the previous Lemma 5.19, $U(s,0)$ is invertible. Thus define $\Phi(t) = U(t,0)$ to see that $\Phi(t) = U(t,s)\Phi(s)$, i.e., $U(t,s) = \Phi(t)\Phi^{-1}(s)$, which complete the first part of the proof.

Next, define $y(t) = \Phi^{-1}(t)x(t)$ and apply Itô formula to get

$$\begin{aligned} y(t) := \xi + \int_0^t \Phi^{-1}(s)\Big[a(s) - \sum_{k=1}^n B_k(s)b_k(s) - \\ - \int_{\mathbb{R}_*^m} C(\zeta,s)\big(\mathbb{1} + C(\zeta,s)\big)^{-1}c(\zeta,s)\pi(\mathrm{d}\zeta)\Big]\mathrm{d}s + \\ + \sum_{k=1}^n \int_0^t \Phi^{-1}(s)b_k(s)\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}_*^m\times]0,t]} \Phi^{-1}(s)\big(\mathbb{1} + C(\zeta,s)\big)^{-1}c(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t) \end{aligned}$$

and from $x(t) = \Phi(t)y(t)$ we conclude. □

• *Remark* 5.21. It should be clear that (5.95) is only used for $p = 2$ and $p = d$ to make sense for the stochastic Wronskian. Also, the condition (5.98) is a simplification of the assumption

$$\det[\mathbb{1} + C(\zeta, t)] > 0 \quad \text{and} \quad \int_{\mathbb{R}^m_*} |C(\zeta, t)|^2 \big|[\mathbb{1} + C(\zeta, t)]^{-1}\big| \pi(\mathrm{d}\zeta) \le C_T,$$

for every t in $[0, T]$, any $T > 0$. Note that the singularity is as $\det[\mathbb{1}+C(\zeta, t)] \to 0$, which is not as $C(\zeta, t) \to 0$. □

In some particular case there exist explicit representations of the solution. The simplest case is when all the matrices commute with each other, i.e.,

$$\begin{aligned}
&A(t)A(s) = A(s)A(t), \quad B_k(t)B_\ell(s) = B_\ell(s)B_k(t),\\
&C(\zeta, t)C(\zeta, s) = C(\zeta, s)C(\zeta, t),\\
&A(t)B_k(s) = B_k(s)A(t), \quad A(t)C(\zeta, s) = C(\zeta, s)A(t),\\
&B_k(t)C(\zeta, s) = C(\zeta, s)B_k(t), \quad \forall k, \ell, \zeta, t, s,
\end{aligned}$$

then the fundamental solution $\Phi(t)$ can be expressed as an exponential semi-martingale, namely,

$$\begin{aligned}
\Phi(t) = \exp\Big[\int_0^t \Big(A(r) - \frac{1}{2}\sum_k B_k^2(r) + \\
+ \int_{\mathbb{R}^m_*} [\ln(\mathbb{1} + C(\zeta, r)) - C(\zeta, r)]\pi(\mathrm{d}\zeta)\Big)\mathrm{d}r + \\
+ \sum_k \int_0^t B_k(r)\mathrm{d}w_k(r) + \int_{\mathbb{R}^m_* \times]0,t]} \ln(\mathbb{1} + C(\zeta, r))\, \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}r)\Big],
\end{aligned}$$

for every $t \ge 0$. This follows immediately from Itô formula applied to the exponential function.

Some Deterministic Coefficients

In the case where the coefficients a, b_k and c are deterministic, the formula of Theorem 5.20 simplifies to

$$\begin{aligned}
x(t) := U(t, 0)\xi + \int_0^t U(t, s)a(s)\,\mathrm{d}s + \sum_{k=1}^n \int_0^t U(t, s)b_k(s)\,\hat{\mathrm{d}}w_k(s) + \\
+ \int_{\mathbb{R}^m_* \times]0,t]} U(t, s)\, c(\zeta, s)\, \tilde{\nu}(\mathrm{d}\zeta, \hat{\mathrm{d}}s)
\end{aligned}$$

where the stochastic integration is in the backward sense. In fact we have the following relations between the standard stochastic Itô integrals and the backward Itô integrals:

$$\begin{aligned}
\int_0^t U(t, s)b_k(s)\,\hat{\mathrm{d}}w_k(s) = \Phi(t) \int_0^t \Phi^{-1}(s)b_k(s)\,\mathrm{d}w_k(s) - \\
- \int_0^t U(t, s)B_k(s)\, b_k(s)\,\mathrm{d}s
\end{aligned}$$

and

$$\int_{\mathbb{R}^m_*\times]0,t]} U(t,s)\,c(\zeta,s)\,\tilde\nu(\mathrm{d}\zeta,\hat{\mathrm{d}}s) = \\ = \Phi(t)\int_{\mathbb{R}^m_*\times]0,t]} \Phi^{-1}(s)\big(\mathbb{1}+C(\zeta,s)\big)^{-1}c(\zeta,s)\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s) - \\ -\int_0^t\int_{\mathbb{R}^m_*} U(t,s)C(\zeta,s)\big(\mathbb{1}+C(\zeta,s)\big)^{-1}c(\zeta,s)\pi(\mathrm{d}\zeta)\mathrm{d}s.$$

Mean and covariance matrix

On the other hand, if $(x(t) : t \geq 0)$ is a solution of the linear d-dimensional stochastic ordinary differential equation (5.94) then its first moment or mean value $m(t) := \mathbb{E}\{x(t)\}$ satisfies the ordinary differential equation

$$\dot m(t) = A(t)m(t) + a(t), \quad \forall t \geq 0. \tag{5.99}$$

The variance matrix $R(t) := \mathbb{E}\{[x(t) - m(t)][x(t) - m(t)]^*\}$, i.e., by components $r_{ij}(t) := \mathbb{E}\{[x_i(t) - m_i(t)][x_j(t) - m_j(t)]\}$, solve the following $d(d+1)/2$-dimensional (since $R(t)$ is a symmetric non-negative definite matrix) ordinary differential equation

$$\begin{cases} \dot R(t) = A(t)R(t) + R(t)A^*(t) + \\ \qquad + \sum_k [B_k(t)R(t)B_k^*(t) + \tilde b_k(t)\tilde b_k(t)^*] + \\ \qquad + \int_{\mathbb{R}^m_*} [C(\zeta,t)R(t)C^*(\zeta,t) + \tilde c(\zeta,t)\tilde c^*(\zeta,t)]\pi(\mathrm{d}\zeta), \end{cases} \tag{5.100}$$

for every $t \geq 0$, where

$$\tilde b_k(t) := [B_k(t)m(t) + b_k(t)]$$

$$\tilde c(\zeta,t) := [C(\zeta,t)m(t) + c(\zeta,t)].$$

Indeed, set $y(t) := x(t) - m(t)$ to have

$$\mathrm{d}y(t) = A(t)y + \sum_k [B_k(t)y(t) + \tilde b_k(t)]\mathrm{d}w_k(t) + \\ + \int_{\mathbb{R}^m_*} [C(\zeta,t)y(t) + \tilde c(\zeta,t)]\tilde\nu(\mathrm{d}\zeta,\mathrm{d}t).$$

and applying the previous Itô formula for products we easily get (5.100). In this way, once the mean $m(t)$ have been computed, we can compute the variance matrix $R(t)$.

Actually, we may want to compute the covariance matrix, namely, $R(t,s) := \mathbb{E}\{[x(t)-m(t)][x(s)-m(s)]^*\}$, which can be expressed in terms of the preceding variance matrix $R(t)$ and the evolution operator $\mathcal{A}(t,s)$ solution to

$$\dot{\mathcal{A}}(t,s) = A(t)\mathcal{A}(t,s), \quad \mathcal{A}(s,s) = \mathbb{1},$$

as $R(t,s) = \mathcal{A}(t,s)R(s)$ if $s < t$ and $R(t,s) = R(t)\mathcal{A}^*(s,t)$ if $s > t$.

Gaussian case

If $B_k(t) = 0$, $C(\zeta, t) = 0$, $c(\zeta, t) = 0$ and the initial condition x^0 is normal distributed then the solution of the linear d-dimensional stochastic ordinary differential equation (5.94), with the initial condition $x(0) = \xi$, is a Gaussian process. Notice that in this case the equation (5.97) simplifies and $\Phi(t) = \mathcal{A}(t, 0)$ is the (deterministic) fundamental matrix solution of ordinary differential equation $\dot{\Phi}(t) = A(t)\Phi(t), \quad \Phi(0) = \mathbb{1}$, $x(t)$ takes the form

$$x(t) = \Phi(t)\Big[x^0 + \int_0^t \Phi^{-1}(s)a(s)\mathrm{d}s + \sum_{k=1}^n \int_0^t \Phi^{-1}(s)b_k(s)\mathrm{d}w_k(s)\Big],$$

This yields the following expressions for the mean $m(t)$ function

$$m(t) = \Phi(t)\Big[\mathbb{E}\{x^0\} + \int_0^t \Phi^{-1}(s)a(s)\mathrm{d}s\Big],$$

and the covariance $R(t, s)$ function

$$R(t, s) = \Phi(t)\Big[\mathbb{V}\mathrm{ar}\{x^0\} + \sum_{k=1}^n \int_0^{t\wedge s} \Phi^{-1}(r)b_k(r)[\Phi^{-1}(r)b_k(r)]^*\mathrm{d}r\Big]\Phi^*(s),$$

which completely determined the Gaussian process.

The particular case of the so-called Ornstein-Uhlenbeck process

$$\mathrm{d}x(t) = -Ax(t) + \sum_k b_k \mathrm{d}w_k(t)$$

yields $\Phi(t) = \exp(-tA)$ and so the above formula for the mean $m(t)$ and covariance $R(t, s)$ functions are simple. For instance, the reader is referred to Arnold [7, Sections 8.2–8.5, pp. 128–144].

All Random Coefficients

Consider the case where the coefficients $A(s, \omega)$, $B_k(s, \omega)$, $C(\zeta, s, \omega)$, $a(s, \omega)$, $b_k(s, \omega)$ and $c(\zeta, s, \omega)$ are all predictable processes satisfying (almost surely)

$$\sup_{0\le s\le t}\Big\{|A(s)| + \sum_{k=1}^n |B_k(s)|^2 + \int_{R_*^m} |C(\zeta, s)|^2\pi(\mathrm{d}\zeta)\Big\} < \infty, \tag{5.101}$$

and

$$\int_0^t |a(s)|\mathrm{d}s + \sum_{k=1}^n \int_0^t |b_k(s)|^2\mathrm{d}s + \int_0^t \mathrm{d}s \int_{R_*^m} |c(\zeta, s)|^2\pi(\mathrm{d}\zeta) < \infty, \tag{5.102}$$

for every $t > 0$. Thus, for each $r \geq 0$, define the stopping time

$$\tau_r = \inf\Big\{ s \geq 0 : |A(s)| \geq r, \ \text{ or } \ \sum_{k=1}^{n} |B_k(s)|^2 \geq r^2 \ \text{ or } \int_{R_*^m} |C(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \geq r^2 \ \text{ or } \ \int_0^T |a(s)| \mathrm{d}s \geq r \ \text{ or } \sum_{k=1}^{n} \int_0^T |b_k(s)|^2 \mathrm{d}s \geq r^2 \ \text{ or } \ \int_0^T \mathrm{d}s \int_{R_*^m} |c(\zeta, s)|^2 \pi(\mathrm{d}\zeta) \geq r^2 \Big\}$$

to get $\tau_r \to \infty$ as $r \to \infty$.

Let us revise the existence and uniqueness proof. Indeed, with the notation of Theorem 5.3 consider the (now) affine operator $T(x)(t)$ defined as the integral of the right-hand side in the stochastic differential equation (5.94), for processes $x(\cdot)$ in the Banach space $\mathcal{Y}_r$ with the sup-norm

$$\|y\|_r := \mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \mathrm{e}^{-2\alpha(t-t_0)} |y(t)|^2 \Big\}.$$

Note that in this case $g(t, x, \omega) = A(t, \omega)x + a(t, \omega)$, $\sigma_k(t, x, \omega) = B_k(t, \omega)x + b_k(t, \omega)$ and $\gamma(\zeta, t, x) = C(\zeta, t, \omega)x + c(\zeta, t, \omega)$. Thus $\|T(x) - T(y)\|_r$ is bounded by three pieces, namely, I_1 involving g, I_2 involving σ and I_3 involving γ, i.e.,

$$\|T(x) - T(y)\|_r \leq I_1 + I_2 + I_3.$$

Thus

$$I_1 = \mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \mathrm{e}^{-2\alpha(t-t_0)} \Big| \int_{t_0}^{t} g(s, x(s)) - g(s, y(s)) \mathrm{d}s \Big|^2 \Big\} \leq$$
$$\leq \mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \big[\mathrm{e}^{-2\alpha(s-t_0)} |x(s) - y(s)|^2 \big] \Big[\int_{t_0}^{t} \mathrm{e}^{-\alpha(t-s)} |A(s)| \mathrm{d}s \Big]^2 \Big\} \leq$$
$$\leq \frac{r^2}{\alpha^2} \mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \mathrm{e}^{-2\alpha(s-t_0)} |x(s) - y(s)|^2 \Big\},$$

while

$$I_2 \leq 4\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \int_{t_0}^{t} \mathrm{e}^{-2\alpha(t-t_0)} |\sigma(s, x(s)) - \sigma(s, y(s))|^2 \mathrm{d}s \Big\} \leq$$
$$\leq 4\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \big[\mathrm{e}^{-2\alpha(s-t_0)} |x(s) - y(s)|^2 \big] \int_{t_0}^{t} \mathrm{e}^{-2\alpha(t-s)} |B(s)|^2 \mathrm{d}s \Big\}$$

and similarly with I_3, so

$$I_2 + I_3 \leq \frac{4r^2}{\alpha} \mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1 \wedge \tau_r} \mathrm{e}^{-2\alpha(s-t_0)} |x(s) - y(s)|^2 \Big\},$$

i.e.,

$$\|T(x) - T(y)\|_r \leq c(r, \alpha) \|x - y\|_r,$$

with $c(r, \alpha)$ vanishes as α goes to ∞, for any fixed $r > 0$. Hence, the operator T is a contraction on $\mathcal{Y}_r$ for some α sufficiently large. Then, under the conditions

(5.101) and (5.102) the stochastic differential equation (5.94) has one and only one solution x_r, up to the stopping time τ_r, which is given by the expression in Theorem 5.20 in term of the fundamental matrix as in (5.97). Finally, since $x_r(t) = x_p(t)$ for any $t_0 \le t \le t_1 \wedge \tau_r$ if $r \le p$, we conclude as $r \to \infty$.

5.1.6 Differentiability

Now we consider the (strong) solution of the d-dimensional stochastic ordinary differential equation (5.3) as function of the initial data x^0 given at $t = t_0$. Recall that the initial time t_0 may be considered as part of the initial state x^0, by doing a translation in time and by adding one more dimension to the equation. However, the solution $x(t) = x(t, x^0, t_0)$ of the SDE (5.3) cannot be differentiable with respect to t_0, because this involves the paths of the martingale. Hence, for simplicity, we take $t_0 = 0$ and the solution is denoted by $y(t) = y(t, x)$, with the initial condition $x^0 = x$, i.e., we have

$$\left\{ \begin{aligned} y_i(t) = x_i &+ \int_0^t g_i(s, y(s), v(s))\mathrm{d}s + \\ &+ \sum_{k=1}^n \int_0^t \sigma_{ik}(s, y(s), v(s))\mathrm{d}w_k(s) + \\ &+ \int_{\mathbb{R}^m_* \times]0,t]} \gamma_i(\zeta, s, y(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \ge 0, \end{aligned} \right. \tag{5.103}$$

where the parameter $x = (x_1, \dots, x_d)$ belongs to $\mathbb{R}^d$. Our interest is differentiability of the solution process $y(t, x)$ with respect to the initial data $y(0, x) = x$.

Assume the linear growth condition: for some $p \ge 2$ there exists a positive constant $C = C_p$ such that

$$\left\{ \begin{aligned} |g(t,x,v)|^p + |\sigma(t,x,v)|^p + \int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^p \pi(\mathrm{d}\zeta) \le \\ \le C(1 + |x|^p), \end{aligned} \right. \tag{5.104}$$

for every (t, x, v) in $[0, \infty) \times \mathbb{R}^d \times \mathbb{R}^q$. This is equivalent to conditions (5.6) and (5.45). Then, if the strong solution is unique (e.g., adding a locally Lipschitz condition in x, see Theorem 5.3) then for every fixed $t > 0$, the mapping $x \mapsto y(t, x)$ can be regarded as a function form $\mathbb{R}^d$ into the p-integrable random variables. Thus, differentiability in x is understood in the L^p-sense, i.e., a vector-valued process $y_i'(t, x) = (\partial_j y_i(t, x) : j = 1, \dots, d)$ is the derivative of real-valued process $x \mapsto y_i(t, x)$ if

$$\lim_{|\xi| \to 0} \mathbb{E}\left\{ \frac{\left| y_i(t, x + \xi) - y_i(t, x) - \xi \cdot y_i'(t, x) \right|^p}{|\xi|^p} \right\} = 0,$$

for every $t > 0$ and $i = 1, \dots, d$. Note that $\xi \cdot y_i'(t, x) = \sum_j \xi_j \, \partial_j y_i(t, x)$.

Assume that the coefficients $g_i(t, x, v)$, $\sigma_{ik}(t, x, v)$ and $\gamma_i(\zeta, t, x, v)$ are continuously differentiable in the variable x, and for some constant $C > 0$ satisfy

(recall $p \geq 2$)

$$|g_i'(t,x,v)|^p + |\sigma_{ik}'(t,x,v)|^p + \int_{\mathbb{R}_*^m} |\gamma_i'(\zeta,t,x,v)|^p \pi(\mathrm{d}\zeta) \leq C, \tag{5.105}$$

for every (t,x,v) in $[0,\infty) \times \mathbb{R}^d \times \mathbb{R}^q$, any i and k, where

$$\begin{aligned} g_i'(t,x,v) &:= (\partial_j g_i(t,x,v) : j = 1,\dots,d), \\ \sigma_{ik}'(t,x,v) &:= (\partial_j \sigma_{ik}(t,x,v) : j = 1,\dots,d), \\ \gamma_i'(\zeta,t,x,v) &:= (\partial_j \gamma_i(\zeta,t,x,v) : j = 1,\dots,d), \end{aligned}$$

are the gradient in x. Now, consider the following linear stochastic differential equation for the $d \times d$-dimensional (matrix) process $\partial_j y_i(t) = \partial_j y_i(t,x)$, $i,j = 1,\dots,d$,

$$\left\{ \begin{aligned} \partial_j y_i(t) = \delta_{ij} &+ \sum_\ell \int_0^t \partial_\ell g_i(s)\, \partial_j y_\ell(s) \mathrm{d}s + \\ &+ \sum_{\ell,k} \int_0^t \partial_\ell \sigma_{ik}(s)\, \partial_j y_\ell(s) \mathrm{d}w_k(s) + \\ &+ \sum_\ell \int_{\mathbb{R}_*^m \times]0,t]} \partial_\ell \gamma_i(\zeta,s)\, \partial_j y_\ell(s) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{aligned} \right. \tag{5.106}$$

for every $t \geq 0$, where $\delta_{ij} := 1$ if $i = j$ and $\delta_{ij} := 0$ otherwise, and

$$\begin{aligned} \partial_\ell g_i(s) &:= \partial_\ell g_i(s, y(s), v(s)), \\ \partial_\ell \sigma_{ik}(s) &:= \partial_\ell \sigma_{ik}(s, y(s), v(s)), \\ \partial_\ell \gamma_i(\zeta, s) &:= \partial_\ell \gamma_i(\zeta, s, y(s), v(s)). \end{aligned}$$

It is clear that (5.106) is a d-dimensional system of d-dimensional (linear) stochastic ordinary differential equation of the form (5.3), where the existence and uniqueness Theorem 5.3 can be used. Hence, under the differentiability assumption (5.105) on the coefficients, the derivative process $\partial_j y_i(t) = \partial_j y_i(t,x)$, $i,j = 1,\dots,d$ is uniquely defined as the solution of (5.106) with $y(t) = y(t,x)$.

Proposition 5.22. *Let conditions* (5.104) *and* (5.105) *be satisfied for some* $p \geq 2$. *Then the unique solution of the* $d \times d$*-dimensional linear stochastic differential equation* $\partial_j y_i(t) = \partial_j y_i(t,x)$, $i,j = 1,\dots,d$, *is the derivative in the* $L^{p'}$ *sense, for any* $p' < p$, *of the solution* $y(t,x)$ *of the* d*-dimensional stochastic ordinary differential equation* (5.103).

Proof. First, calculate the Itô differential as follows

$$\begin{aligned} \mathrm{d}\big[y_i(t,x+\xi) - y_i(t,x) - \xi \cdot y_i'(t,x)\big] &= a_i(t)\mathrm{d}t + \\ &+ \sum_k b_{ik}(t)\mathrm{d}w_k(t) + \int_{\mathbb{R}_*^m} c_i(\zeta,t)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t), \end{aligned}$$

where

$$a_i(t) := g_i(t, y(t, x+\xi), v(t)) - g_i(t, y(t,x), v(t)) - \\ - \sum_{\ell,j} \xi_j \partial_\ell g_i(t, y(t,x), v(t)) \partial_j y_\ell(t,x),$$

$$b_{ik}(t) := b_{ik}(t, y(t, x+\xi), v(t)) - b_{ik}(t, y(t,x), v(t)) - \\ - \sum_{\ell,j} \xi_j \partial_\ell b_{ik}(t, y(t,x), v(t)) \partial_j y_\ell(t,x),$$

and

$$c_i(\zeta, t) := c_i(\zeta, t, y(t, x+\xi), v(t)) - c_i(\zeta, t, y(t,x), v(t)) - \\ - \sum_{\ell,j} \xi_j \partial_\ell c_i(\zeta, t, y(t,x), v(t)) \partial_j y_\ell(t,x).$$

Define the modulus of differentiability

$$\rho(r,t,v) := \sup_{0<|y-x|\le r} \Big\{ \frac{\rho_g(x,y,t,v) + \rho_\sigma(x,y,t,v) + \rho_\gamma(x,y,t,v)}{|y-x|} \Big\},$$

where

$$\rho_g(x,y,t,v) := |g(t,y,v) - g(t,x,v) - \sum_j (y_j - x_j)\partial_j g(t,x,v)|,$$

$$\rho_\sigma(x,y,t,v) := |\sigma(t,y,v) - \sigma(t,x,v) - \sum_j (y_j - x_j)\partial_j \sigma(t,x,v)|,$$

$$\rho_\gamma(x,y,t,v) := \Big(\int_{\mathbb{R}^m_*} \big[|\gamma(\zeta,t,y,v) - \gamma(\zeta,t,x,v) - \\ - \sum_j (y_j - x_j)\partial_j \gamma(\zeta,t,x,v)|^p \big] \pi(\mathrm{d}\zeta) \Big)^{1/p}.$$

In view of the assumption (5.105), there is a constant $C > 0$ such that

$$|a_i(t)|^p + |b_{ik}(t)|^p + \int_{\mathbb{R}^m_*} |c_i(\zeta,t)|^p \pi(\mathrm{d}\zeta) \le \\ \le C \big[|y(t,x+\xi) - y(t,x)|^p \big(\rho(|y(t,x+\xi) - y(t,x)|, t, v(t)) \big)^p + \\ + |y(t,x+\xi) - y(t,x) - \xi \cdot y'(t,x)|^p \big].$$

Hence, an application of Itô formula or Gronwall inequalities yields the estimate

$$\mathbb{E}\big\{ |y(t,x+\xi) - y(t,x) - \xi \cdot y'(t,x)|^p \big\} \le \\ \le C_T \mathbb{E}\Big\{ \int_0^T |y(s,x+\xi) - y(s,x)|^p \times \\ \times \big(\rho(|y(s,x+\xi) - y(s,x)|, s, v(s)) \big)^p \mathrm{d}s \Big\},$$

for every t in $[0,T]$. By means of the a priori estimates in Theorem 5.11 we conclude. □

Notice that if we assume that the modulus of differentiability $\rho(r,t,v)$ is uniform in v and locally uniform in t then the differentiability is valid for the same L^p.

It is clear that we can iterate this result, so that if the coefficients satisfy the growth condition (5.104) and a uniform global Lipschitz condition in the variable x, i.e., (5.50) and (5.51), then the solution process $y(t,x)$ is n-times differentiable in x as long as the coefficients are also n-times differentiable in x with continuous bounded derivatives.

Now, assume that $g'_i(t,x,v)$, $\sigma'_{ik}(t,x,v)$ and $\gamma'_i(\zeta,t,x,v)$ are uniformly (in t,v) Lipschitz continuous in x, i.e., for some $p \geq 2$ and every $T > 0$ there exists a constant $M > 0$ such that

$$\begin{cases} |g'(t,x,v) - g'(t,x',v)|^p + |\sigma'(t,x,v) - \sigma'(t,x',v)|^p + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma'(\zeta,t,x,v) - \gamma'(\zeta,t,x',v)|^p \pi(\mathrm{d}\zeta) \leq M|x-x'|^p, \end{cases} \tag{5.107}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty) \times \mathbb{R}^d \times \mathbb{R}^q$ with $t \leq t_0 + T$.

Proposition 5.23. *Let conditions (5.104), (5.105) and (5.107) be satisfied for some $p \geq 2$ and denote by $y(t,x)$ the solution of the d-dimensional stochastic ordinary differential equation (5.103). Then for every $T > 0$ there is a positive constant $C = C_{p,T}$, depending only on bounds in assumptions, such that*

$$\begin{cases} \mathbb{E}\Big\{ \displaystyle\sup_{t_0 \leq t \leq T} |[\theta y(t,x) + (1-\theta)y(t,x')] - y(t,\theta x + (1-\theta)x')| \Big\} \leq \\ \qquad \leq C_{p,T}\theta(1-\theta)|x-x'|^{2p}, \end{cases} \tag{5.108}$$

for every θ in $[0,1]$ and x, x' in $\mathbb{R}^d$.

Proof. Clearly, under the above assumptions, the derivative process $y'_i(t,x)$ is Lipschitz continuous in x, so that the relation

$$[\theta y_i(t,x) + (1-\theta)y_i(t,x')] - y_i(t,\theta x + (1-\theta)x') = \theta(1-\theta) \times$$
$$\times (x-x') \cdot \int_0^1 \big[y'_i\big(t, x' + r\theta(x-x')\big) - y'_i\big(t, x + r(1-\theta)(x'-x)\big) \big] \mathrm{d}r,$$

yields the desired result. □

At this point we should notice that the solution $y(t,x)$ of the d-dimensional stochastic ordinary differential equation (5.103) can be considered as a (controlled) *random field*, i.e., for each t in $[0,\infty)$ we look at the paths $x \mapsto y(t,x,\omega)$ with x in R^d and ω in Ω. The previous Proposition 5.22 does ensure that except for a set of measure zero, the paths are differentiable in x, locally uniformly in t. This involves a modification, i.e., to choose a suitable version of the random field $y(t,x)$. Kolmogorov–Chentsov criterium, e.g., see Kunita [143, Section 1.4, pp. 31–42] or Da Prato and Zabczyk [52, Appendix B, pp. 311–316], affirms that any random field $(\Phi(x) : x \in \mathbb{R}^n)$, with values in some Banach space with norm $\|\cdot\|$, satisfying the following a priori estimate

$$\mathbb{E}\{\|\Phi(x) - \Phi(y)\|^\alpha\} \leq C|x-y|^{n+\beta}, \quad \forall x, y \in \mathbb{R}^n,$$

for some positive constants α, β and C, has obtain a continuous version. This a priori estimate can be established for the solution and its derivative processes, under the same assumptions of Proposition 5.22. Hence, there exists a version of $y(t,x,\omega)$ which is continuously differentiable in x for every t in $[0,\infty)$ and ω not in N, with $P(N)=0$.

Now, let us assume that the coefficients satisfy (5.104) and they are continuously differentiable and bounded in the variable x up to the order $r \geq 1$, i.e., for some $p > d$, for any $T > 0$ and any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = \alpha_1 + \cdots + \alpha_d \leq r$, there exists a constant $C = C_{\alpha,T,p}$ such that

$$|D_x^\alpha g(t,x,v)|^p + |D_x^\alpha \sigma(t,x,v)|^p + \int_{\mathbb{R}_*^m} |D_x^\alpha \gamma(\zeta,t,x,v)|^p \pi(\mathrm{d}\zeta) \leq C, \quad (5.109)$$

for every (t,x,v) in $[0,T] \times \mathbb{R}^d \times \mathbb{R}^q$. It is not hard to prove the following result

Theorem 5.24. *Under the assumptions* (5.104) *and* (5.109) *we denote by* $y(t,x)$ *the solution of the d-dimensional stochastic ordinary differential equation* (5.103). *Then* $y(t,x)$ *has a modification which is continuously differentiable in* x *up to the order* r *and Taylor development holds in probability, i.e.,*

$$\begin{cases} y(t,x) = \displaystyle\sum_{|\alpha| \leq r} \frac{(x-x_0)^\alpha}{\alpha!} D_x^\alpha y(t,x_0) + R_r(t,x,x_0), \\ \displaystyle\lim_{x \to x_0} \sup_{0 \leq t \leq T} |R_r(t,x,x_0)| \, |x-x_0|^{-r} = 0, \quad \forall t > 0,\ x, x_0 \in \mathbb{R}^d, \end{cases}$$

where the limit is in probability. □

Similarly, we have a continuity with respect to the small jumps, i.e., for any $\varepsilon > 0$ we set

$$\pi_\varepsilon(B) = \pi\big(\{\zeta \in B \,:\, \varepsilon \leq |\zeta| \leq \varepsilon^{-1}\}\big), \quad \forall B \in \mathcal{B}(\mathbb{R}_*^m),$$

and denote by $y_\varepsilon(t,x)$ the solution (5.103) with π_ε instead of π. We have

Theorem 5.25. *Under the assumptions and notation of* Theorem 5.24 *the solution* $y_\varepsilon(t,x)$ *converges in probability to* $y(t,x)$ *locally uniformly in* (t,x) *together with all derivative in* x *up to the order* r, *i.e., for any compact subset* K *of* $[0,\infty) \times \mathbb{R}^d$ *and any multi-index* α *with* $|\alpha| \leq r$ *we have*

$$\lim_{\varepsilon \to 0} \sup_{(t,x) \in K} |D_x^\alpha y(t,x) - D_x^\alpha y_\varepsilon(t,x)| = 0,$$

where the limit is in probability. □

• *Remark* 5.26. It is clear that the above result holds true under a much wider assumptions, namely, if the coefficients g, σ and γ depend on a parameter ε destined to vanish and satisfy conditions (5.104) and (5.109) uniformly in ε, then the solution $y_\varepsilon(t,x)$ converges (in probability) to the limiting solution $y(t,x)$, together with their derivatives in x up to the order r, locally uniformly in (t,x), provided the same convergence holds for the coefficients. □

5.2 Local and Global Solutions

Let us consider a complete filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, with a n-dimensional (standard) Wiener process $(w(t) : t \geq 0)$ and an independent (standard) Poisson measure $(\nu(B,]0,t]) : B \in \mathbb{R}^m_*,\ t \geq 0)$ with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{\nu(B,]0,t])\}/t$, satisfying

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1+|\zeta|}\pi(\mathrm{d}\zeta) < \infty,$$

and with martingale measure $\tilde{\nu}(B,]0,t]) := \nu(B,]0,t]) - t\pi(B)$.

5.2.1 Local Existence and Uniqueness

Again we deal with problem (5.3), that we rewrite

$$\begin{cases} x(t) = x^0 + \displaystyle\int_{t_0}^t g(s, x(s))\mathrm{d}s + \int_{t_0}^t \sigma(s, x(s))\mathrm{d}w(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]t_0,t]} \gamma(\zeta, s, x(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0, \end{cases} \tag{5.110}$$

but this time the initial time t_0 may be a stopping time, the initial value x^0 may be a $F(t_0)$-measurable random variable, and the coefficients g, σ and γ are locally bounded and Lipschitz predictable processes, i.e., for every $r > 0$ there exist non-negative predictable processes $C(t,r)$ and $M(t,r)$ such that we have

$$\begin{cases} \displaystyle\sup_{|x|\leq r} \Big[|g(t,x)| + |\sigma(t,x)|^2 + \int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x)|^2\pi(\mathrm{d}\zeta)\Big] \leq C(t,r), \\ \text{with} \quad P\Big\{\displaystyle\int_{t_0}^{t_0+r} C(t,r)\mathrm{d}t < \infty\Big\} = 1, \end{cases} \tag{5.111}$$

and such that for every (t,x), (t,x') in $[t_0,\infty) \times \mathbb{R}^d$ with $|x| \leq r$ and $|x'| \leq r$ we have

$$\begin{cases} |g(t,x) - g(t,x')|^2 + |\sigma(t,x) - \sigma(t,x')|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x) - \gamma(\zeta,t,x')|^2\pi(\mathrm{d}\zeta) \leq M(t,r)|x-x'|^2, \\ \text{with} \quad P\Big\{\displaystyle\sup_{t_0\leq t\leq t_0+r} M(t,r) < \infty\Big\} = 1. \end{cases} \tag{5.112}$$

In this case we introduce the concept of *local* solution:

Definition 5.27 (local solution)**.** A process $x(\cdot)$, defined in a right-open stochastic interval $[\![t_0, \theta[\![$, is a local solution of the stochastic differential equations if $\theta > t_0$ is a stopping time and the equality (5.110) holds almost surely for any t in $[\![t_0, \theta[\![$. The local solution $x(\cdot)$ is defined in a maximal interval of existence $[\![t_0, \tau[\![$ if $x(\cdot)$ cannot be extended further than τ, i.e., for any local solution $y(\cdot)$ defined on $[\![t_0, \theta[\![$, with $x(t) = y(t)$ for any t in $[\![t_0, \tau \wedge \theta[\![$, we have $\tau \geq \theta$. In this case, $\tau = \varsigma$ is called the lifetime of the solution $x(\cdot)$. □

We may extend the definition of local solution $x(\cdot)$ to a closed stochastic interval $[\![t_0, \theta]\!]$ whenever the equality (5.110) holds almost surely for any t in the stochastic interval $[\![t_0, \theta]\!]$, or equivalently, if

$$\begin{cases} x(t \wedge \theta) = x^0 + \displaystyle\int_{t_0}^{t\wedge\theta} g(s, x(s))\mathrm{d}s + \int_{t_0}^{t\wedge\theta} \sigma(s, x(s))\mathrm{d}w(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]\!]t_0, t\wedge\theta]\!]} \gamma(\zeta, s, x(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) \end{cases} \tag{5.113}$$

holds for any $t \geq t_0$. Note that if $[\![t_0, \tau[\![$ is the maximal interval of existence $[\![t_0, \varsigma[\![$ of $x(\cdot)$ then the left-hand limit $x(\varsigma-)$ is infinite almost surely on the set where $\varsigma < \infty$, i.e., if $x(\varsigma-)$ is finite then we may re-start the stochastic equation with initial condition

$$x(\varsigma) = x(\varsigma-) + \int_{\mathbb{R}^m_*} \gamma(\zeta, \varsigma-, x(\varsigma-))\tilde{\nu}(\mathrm{d}\zeta, \varsigma-),$$

which contradicts the definition of $\varsigma = \tau$.

Theorem 5.28. *Let (5.111) and (5.112) be verified, then the process x is the unique solution to (5.110) on the stochastic interval $[\![t_0, \tau[\![$. Actually $[\![t_0, \tau[\![$ is the maximal interval of existence $[\![t_0, \varsigma[\![$ of the solution, in other words $\tau = \varsigma$ is the* lifetime *of x. Moreover, if*

$$P\Big\{ \sup_{t \in [\![t_0, \varsigma[\![} |x(t)| = \infty \Big\} = 0 \tag{5.114}$$

holds then $\varsigma = \infty$, in other words x is the unique global solution to (5.110).

Proof. Let us consider

$$\Omega_{r,N}(s) = \{\omega \in \Omega \,:\, C(s, r) + M(s, r) \leq N\}$$

and the auxiliary approximating problems for $r = 1, 2, \ldots$, and $N = 1, 2, \ldots$,

$$\begin{cases} x_{r,N}(t) = x^0 + \displaystyle\int_{t_0}^{t} \mathbb{1}_{\Omega_{r,N}(s)} g_r(s, x_{r,N}(s))\mathrm{d}s + \\ \qquad\qquad + \displaystyle\int_{t_0}^{t} \mathbb{1}_{\Omega_{r,N}(s)} \sigma_r(s, x_{r,N}(s))\mathrm{d}w(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]t_0, t]} \mathbb{1}_{\Omega_{r,N}(s)} \gamma_r(\zeta, s, x_{r,N}(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{cases} \tag{5.115}$$

for $t \geq t_0$, where the coefficients are defined by $g_r(t, x, \omega) = g(t, \psi_r(x), \omega)$, $\sigma_r(t, x, \omega) = \sigma(t, \psi_r(x), \omega)$ and $\gamma_r(t, x, \omega) = \gamma(t, \psi_r(x), \omega)$, with

$$\psi_r(x) = x \quad \text{for } |x| \leq r \qquad \text{and} \qquad \psi_r(x) = r\frac{x}{|x|} \quad \text{for} \quad |x| > r. \tag{5.116}$$

Let $x_{r,N}(t)$ be the corresponding solution assured by the usual existence and uniqueness Theorem 5.3 and define

$$\tau_r = \inf\{t \geq t_0 \ \text{ such that } \ |x_{r,N}(t)| > r\}.$$

Clearly, $\tau_{r,N} \le \tau_{r,N+1}$ and by the uniqueness we have $x_{r,N}(t) = x_{r,N+1}(t)$ on $[\![t_0, \tau_{r,N}]\!]$. This allows us to define the process $x_r(t)$ on $[\![t_0, \tau_r[\![$ by setting

$$\tau_r = \sup_N \tau_{r,N} = \lim_{N\to\infty} \tau_{r,N} \quad \text{and} \quad x_r(t) = x_{r,N}(t) \ \text{ for } \ t \in [\![t_0, \tau_{r,N}]\!].$$

Hence, as r increases, we have $x_r(t) = x_{r+1}(t)$ on $[\![t_0, \tau_r]\!]$, and $\tau_r \le \tau_{r+1}$.

Next, defining the stopping time and the process

$$\tau = \sup_r \tau_r = \lim_{r\to\infty} \tau_r \quad \text{and} \quad x(t) = x_r(t) \ \text{ for } t \in [\![t_0, \tau_r]\!],$$

we deduce that the process $x(t)$ is a local solution of (5.110) in $[\![t_0, \tau[\![$.

Let us show now that the local solution just constructed is defined on a maximal interval, i.e., if $y(t)$ is any local solution in $[\![t_0, \theta[\![$, then we have

$$\theta \le \tau \qquad \text{and} \qquad y(t) = x(t), \quad t_0 \le t < \theta,$$

almost surely. Indeed, for every integer $r \ge 1$ let us define

$$\theta_r = \inf\{t \in [t_0, \theta) \ \text{ such that } \ |y(t)| > r\},$$

which satisfies $\theta = \lim_{r\to\infty} \theta_r$. Now, observe that for any θ' in $[\![t_0, \theta_r[\![$ and $r' \ge r$ we have

$$y(\theta') = x^0 + \int_{t_0}^{\theta'} g_{r'}(s, y(s))\mathrm{d}s + \int_{t_0}^{\theta'} \sigma_{r'}(s, y(s))\mathrm{d}w(s) + \\ + \int_{\mathbb{R}^m_* \times]\!]t_0, \theta'[\![} \gamma_{r'}(\zeta, s, y(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s),$$

i.e., y solves the problem (5.115) in the stochastic interval $[\![t_0, \theta_r[\![$ for any $r' \ge r$. Hence, the uniqueness property implies that $y = x_{r'}$ in the interval $[\![t_0, \theta_r[\![$ for any $r' \ge r$. Moreover $\theta_r \le \tau_{r'} \le \tau$, i.e., $\tau = \varsigma$ the lifetime of the local solution process x.

To show that (5.114) yields $\tau = \varsigma = \infty$ we note that

$$\{\tau_r < \tau\} = \Big\{ \sup_{t_0 \le t \le \tau} |x_r(t)| > r \Big\} \subset \Big\{ \sup_{t_0 \le t \le \tau} |x(t)| > r \Big\}$$

and (5.114) imply

$$P\Big(\bigcap_r \{\tau_r < \tau\} \Big) = 0 \quad \text{i.e.} \quad P\Big(\bigcap_r \{\tau_r = \tau\} \Big) = 1.$$

Therefore, we can extend the definition of the process x to $[\![t_0, \tau]\!]$. If $\tilde{x}$ denote this continuation then $\tilde{x}$ is an adapted process with continuous trajectories. Hence by continuity we have

$$\tilde{x}(\tau) = x_0 + \int_{t_0}^{\tau} g(t, \tilde{x}(t))\mathrm{d}t + \int_{t_0}^{\tau} \sigma(t, \tilde{x}(t))\mathrm{d}w(t) + \\ + \int_{]\!]t_0, \tau]\!] \times \mathbb{R}^m_*} \gamma(\zeta, t, \tilde{x}(t))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t),$$

so that $\tilde{x}$ is a solution of (5.110) in $[\![t_0, \tau]\!]$. Finally, it must be $\varsigma = \infty$ (or $\tau = t_1$, if the equation was initially set up to the final time t_1) almost surely. Indeed, if $P\{\tau < t_1\} > 0$ then we can extend further $\tilde{x}$ with the solution of the problem

$$y(t) = \tilde{x}(\tau) + \int_\tau^t g(s, y(s))\mathrm{d}s + \int_\tau^t \sigma(s, y(s))\mathrm{d}w(s) + \\ + \int_{]\!]\tau, t]\!]\times\mathbb{R}^m_*} \gamma(\zeta, s, y(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t),$$

whose existence is assured by the Theorem 5.28. This contradicts the maximality of x and concludes the proof. □

5.2.2 A Priori Bounds

In analogy with the deterministic case, if (5.114) is verified then we say that there exists an *a priori* bound for the solution to (5.110). We have the following sufficient condition to have an a priori bound. Recall the definition of the operator $\mathcal{L}$:

$$\mathcal{L}\Phi(t,x) = \frac{\partial\Phi(t,x)}{\partial t} + \tfrac{1}{2}\mathrm{Tr}\big[\sigma(t,x)\sigma^*(t,x)D^2\Phi(t,x)\big] + (g(t,x), D\Phi(t,x)) + \\ + \int_{\mathbb{R}^m_*}\big[\Phi(t, x+\gamma(\zeta,t,x)) - \Phi(t,x) - (\gamma(\zeta,t,x), D\Phi(t,x))\big]\pi(\mathrm{d}\zeta).$$

Theorem 5.29. *Let V be in $C^{1,2}([t_0,t_1]\times\mathbb{R}^d, \mathbb{R})$ such that*

(i) *$V(t,x) \geq 0$ and $DV(t,x)$ bounded on bounded sets*

(ii) *$\mathcal{L}V(t,x) \leq K(t)(1+V(t,x))$*

(iii) *$\lim_{|x|\to+\infty} V(t,x) = +\infty$ ([5]), uniformly in t*

Then under the assumptions (5.111) and (5.112) there exists an a priori bound for the problem (5.110).

Proof. Let $x\colon [\![t_0, \tau[\![\to \mathbb{R}^d$ be the maximal solution of (5.110).

Step 1.- There exists a constant $C > 0$ such that, maintaining the notation of the previous proof in Theorem 5.28, for every integer $r \geq 1$ it follows

$$\mathbb{E}\big(V((t\wedge\tau_r)\vee t_0, x((t\wedge\tau_r)\vee t_0))\big) \leq C. \tag{5.117}$$

In fact the Itô formula gives

$$\begin{cases} V(t,x(t)) = V(t_0,x^0) + \displaystyle\int_{t_0}^t (\mathcal{L}V)(s,x(s))\ \mathrm{d}s + \\ \qquad + \displaystyle\int_{t_0}^t \big(\sigma(s,x(s)), DV(s,x(s))\big)\mathrm{d}w(s) + \\ \qquad + \displaystyle\int_{]t_0,t]\times\mathbb{R}^m_*} \big[V\big(s,x(s)+\gamma(\zeta,s,x(s))\big) - V(s,x(s))\big]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s) \end{cases} \tag{5.118}$$

[5] It is possible to prove the theorem by assuming $\limsup_{|x|\to+\infty} V(t,x) = \infty$.

for $t_0 \le t < \tau$. Then, for every t, we have

$$V((t\wedge\tau_r)\vee t_0, x((t\wedge\tau_r)\vee t_0))) \le V(x^0)+ \\ + \int_{t_0}^{(t\wedge\tau_r)\vee t_0} K(s)\,(1+V(x(s)))\mathrm{d}s+ \\ + \int_{t_0}^{(t\wedge\tau_r)\vee t_0} \big(\sigma(s,x(s)), DV(s,x(s))\big)\mathrm{d}w(s)+ \\ + \int_{]\!]t_0,(t\wedge\tau_r)\vee t_0]\!]\times\mathbb{R}^m_*} \big[V(s,x(s)+\gamma(\zeta,s,x(s)))-V(s,x(s))\big]\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s),$$

from which, since $\big(g(s,x(s)), DV(s,x(s))\big)$ is bounded in $[\![t_0,\tau_r[\![$, we obtain

$$\mathbb{E}\big(V((t\wedge\tau_r)\vee t_0, x((t\wedge\tau_r)\vee t_0))\big) \le \\ \le \mathbb{E}(V(x^0)) + \mathbb{E}\int_{t_0}^{(t\wedge\tau_r)\vee t_0} K(s)\,(1+V(x(s)))\mathrm{d}s \\ \le \mathbb{E}(V(x^0)) + \int_{t_0}^{(t\wedge\tau_r)\vee t_0} K(s)\,[1+\mathbb{E}(V(x((s\wedge\tau_N)\vee t_0)))]\mathrm{d}s.$$

for any t in $[t_0,t_1]$. Therefore the first step is proved by the Gronwall's lemma.

Step 2.- $M^r(t)$ is a supermartingale, where

$$M^r(t) = V\big((t\wedge\tau_r)\vee t_0, x((t\wedge\tau_r)\vee t_0)\big) - \int_{t_0}^{(t\wedge\tau_r)\vee t_0} K(s)\,(1+V(s,x(s)))\mathrm{d}s.$$

In fact we have, from (5.118), that

$$M^r(t)-M^r(s) \le \int_{(s\wedge\tau_r)\vee t_0}^{(t\wedge\tau_r)\vee t_0} \big(\sigma(s,x(s)), DV(s,x(s))\big)\mathrm{d}w(s)+ \\ + \int_{]\!](s\wedge\tau_r)\vee t_0,(t\wedge\tau_r)\vee t_0]\!]\times\mathbb{R}^m_*} \big[V(s,x(s)+\gamma(\zeta,s,x(s)))-V(s,x(s))\big]\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s).$$

Hence $\mathbb{E}(M^r(t)-M^r(s)\mid\mathcal{F}_s)\le 0$ and (5.117) imply the desired result.

Step 3.- We have that $P\{\sup_{t_0\le t<\tau} V(t,x(t)) = +\infty\} = 0$. Indeed, the martingale's inequality implies ([6])

$$P\{\sup_{t_0\le t<\le\tau_r} M^r(t) > \lambda\} \le \frac{C_2}{\lambda} \qquad \lambda>0,$$

and the definition of $M^r(t)$ yields, after taking the limit as $r\to\infty$,

$$P\{\sup_{t_0\le t<\tau} M^\tau(t) > \lambda\} \le \frac{C_2}{\lambda}.$$

[6] See Arnold [7] for an analogous utilization of the supermartingale property of the Lyapunov function V in the chapter of stability.

Therefore, as λ tends to infinity, almost surely we have

$$\sup_{t_0\le t<\tau} V(t,x(t)) - \int_{t_0}^{t} K(s)\,(1+\sup_{t_0\le r<s} V(r,x(r)))\,\mathrm{d}s \le C_3$$

for $t<\tau$ and with C_3 depending on ω. Finally the Gronwall's lemma gives

$$\sup_{t_0\le t<\tau} V(t,x(t)) \le C_4,$$

where C_4 is depending on ω. Hence the theorem is proved. □

Theorem 5.29 easily applies to the following situation: $\gamma = 0$,

$$(g(x),x) \le K_1(1+|x|^2) \qquad \text{and} \qquad |(\sigma(x),x)| \ge K_2|\sigma(x)|\,|x|, \tag{5.119}$$

for every x in $\mathbb{R}^d$, with $1/\sqrt{2} < K_2 \le 1$. In fact, let $\varphi \in C^2(\mathbb{R}_+,\mathbb{R}_+)$ be such that $\varphi(r) = r^\alpha$ in $[r_0,+\infty)$ with $0<\alpha\le 1-1/(2K_2^2)$, then $V(x) = \varphi(|x|^2)$ verifies the assumption of the theorem as we have, for $|x|^2 > r_0$,

$$\begin{aligned} LV(x) &= 2\alpha|x|^{2\alpha-2}(g(x),x) + 2\alpha(\alpha-1)|x|^{2\alpha-4}(\sigma(x),x)^2 + \\ &\qquad + \alpha|x|^{2\alpha-2}|\sigma(x)|^2 \le \\ &\le 2\alpha K_1|x|^{2\alpha-2}(1+|x|^2) + \alpha|x|^{2\alpha-2}|\sigma(x)|^2(2K_2^2(\alpha-1)+1) \le \\ &\le 2\alpha K_1\, r_0^{\alpha-1} + 2\alpha K_1 V(x). \end{aligned}$$

We observe that (5.119) is verified in the case $d=1$ by $g=0$ and any σ (see McKean [167]).

5.2.3 Continuous Dependency on Data

Let us rephrase (5.6) and (5.7) in the following form: there exists a constant $C>0$ such that

$$|g(t,x)|^2 + |\sigma(t,x)|^2 + \int_{\mathbb{R}_*^m} |\gamma(\zeta,t,x)|^2\pi(\mathrm{d}\zeta) \le C(1+|x|^2), \tag{5.120}$$

for every (t,x) in $[t_0,\infty)\times\mathbb{R}^d$; and, for any $r>0$, there exists a positive constant $M=M(r)$ such that

$$\begin{cases} |g(t,x)-g(t,x')|^2 + |\sigma(t,x)-\sigma(t,x')|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m} |\gamma(\zeta,t,x)-\gamma(\zeta,t,x')|^2\pi(\mathrm{d}\zeta) \le M|x-x'|^2, \end{cases} \tag{5.121}$$

for every (t,x), (t,x') in $[t_0,\infty)\times\mathbb{R}^d$ with $t\le t_0+r$, $|x|\le r$ and $|x'|\le r$. Here g, σ and γ are adapted (predictable) processes, and the assumptions are uniformly in ω, for simplicity (see Remark 5.7).

Theorem 5.30. *Let $g^{(n)}$, $\sigma^{(n)}$ and $\gamma^{(n)}$ with (5.120) and (5.121) be verified uniformly with respect to n. Assume that as $n \to \infty$, we have $\mathbb{E}\{|x^{(n)}(t_0) - x(t_0)|^2\} \to 0$, where $x^{(n)}(t_0)$ and $x(t_0)$ are the initial data, and for every x in $\mathbb{R}^d$ also $g^{(n)}(t,x,\omega) \to g(t,x,\omega)$, $\sigma^{(n)}(t,x,\omega) \to \sigma(t,x,\omega)$,*

$$\int_{\mathbb{R}^m_*} |\gamma^{(n)}(\zeta,t,x,\omega) - \gamma(\zeta,t,x,\omega)|^2 \pi(\mathrm{d}\zeta) \to 0,$$

almost everywhere in (t,ω) with respect to $\mathrm{d}t \times P(\mathrm{d}\omega)$, then

$$\mathbb{E}\Big\{ \sup_{t_0 \le s \le t_1} |x(t) - x^{(n)}(t)| \Big\} \to 0, \quad \textit{as} \quad n \to \infty. \tag{5.122}$$

Proof. This follows from the uniqueness arguments used in Theorem 5.3. Note that assumptions (5.120) and (5.121), which are satisfied uniformly with respect to n, can be combined to obtain that for any $r > 0$, there exists a positive constant $M' = M'(r)$ such that

$$\begin{cases} |g^{(n)}(t,x) - g^{(n)}(t,x')|^2 + |\sigma^{(n)}(t,x) - \sigma^{(n)}(t,x')|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma^{(n)}(\zeta,t,x) - \gamma^{(n)}(\zeta,t,x')|^2 \pi(\mathrm{d}\zeta) \le M|x - x'|^2, \end{cases} \tag{5.123}$$

for every (t,x), (t,x') in $[t_0,\infty) \times \mathbb{R}^d$ with $t \le t_0 + r$, $|x| \le r$ and *any* x' in $\mathbb{R}^d$. Therefore, for each $r > 0$ use the notation

$$\chi_r(t) := \begin{cases} 0 & \text{if } |x(s)| > r \text{ for some } s > t, \\ 1 & \text{otherwise,} \end{cases}$$

to check that $\chi_r(t) = \chi_r(t)\,\chi_r(s)$ for every $s \le t$, and

$$\begin{aligned} [x(t) - x^{(n)}(t)] &= \int_{t_0}^t \chi_r(s)\big[g(s,x(s)) - g^{(n)}(s,x^{(n)}(s))\big]\mathrm{d}s + \\ &+ \int_{t_0}^t \chi_r(s)\big[\sigma(s,x(s)) - \sigma^{(n)}(s,x^{(n)}(s))\big]\mathrm{d}w(s) + \\ &+ \int_{\mathbb{R}^m_* \times]t_0,t]} \chi_r(s)\big[\gamma(\zeta,s,x(s)) - \gamma^{(n)}(\zeta,s,x^{(n)}(s))\big]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s), \end{aligned}$$

for any t (and ω) with $\chi_r(t) = 1$. Thus, by means of Hölder inequality and the martingale inequality for stochastic integrals, we deduce that

$$\begin{aligned} \mathbb{E}\big\{|x(t) - x^{(n)}(t)|^2 \chi_r(t)\big\} &\le A + C\int_{t_0}^t \mathbb{E}\{|x(s) - x^{(n)}(s)|^2 \chi_r(s)\}\mathrm{d}s, \\ A = A_r^{(n)}(x(\cdot)) &= 4\mathbb{E}\Big\{\Big[\int_{t_0}^t \chi_r(s)\big|g(s,x(s)) - g^{(n)}(s,x(s))\big|\mathrm{d}s\Big]^2 + \\ &\qquad + \int_{t_0}^t \chi_r(s)\big|\sigma(s,x(s)) - \sigma^{(n)}(s,x(s))\big|^2\mathrm{d}s + \\ &+ \int_{t_0}^t \chi_r(s)\mathrm{d}s \int_{\mathbb{R}^m_*} \big|\gamma(\zeta,s,x(s)) - \gamma^{(n)}(\zeta,s,x(s))\big|^2 \pi(\mathrm{d}\zeta)\Big\}, \end{aligned}$$

for every t in $[t_0, t_1]$, and for some constant C. Next, use Gronwall inequality to deduce that there is a constant $K > 0$ depending only on $t_1 - t_0$, r and the constants C, $M(r)$ of hypotheses (5.120), (5.121) such that

$$\mathbb{E}\big\{|x(t) - x^{(n)}(t)|^2 \chi_r(t)\big\} \le K A_r^{(n)}(x(\cdot)), \quad \forall n,$$

for any t in $[t_0, t_1]$. Hence, the convergence of $g^{(n)}$, $\sigma^{(n)}$ and $\gamma^{(n)}$ implies

$$\mathbb{E}\big\{|x(t) - x^{(n)}(t)|^2 \chi_r(t)\big\} \to 0, \quad \text{as} \quad n \to \infty,$$

for any fixed $r > 0$. Actually, a variation of the previous argument allows the introduction of $\sup_{t_0 \le s \le t_1}$ insides the expectation, and Hölder inequality yields

$$\mathbb{E}\Big\{ \sup_{t_0 \le s \le t_1} |x(t) - x^{(n)}(t)| \Big\} \le \big[\mathbb{E}\big\{ \sup_{t_0 \le s \le t_1} |x(t) - x^{(n)}(t)|^2 \chi_r(t) \big\}\big]^{1/2} + \\ + \big[\mathbb{E}\big\{[1 - \chi_r(t)]\big\}\big]^{1/2} \big[\mathbb{E}\big\{ \sup_{t_0 \le s \le t_1} |x(t) - x^{(n)}(t)|^2 \big\}\big]^{1/2}.$$

Now, recall estimate (5.10), namely,

$$\mathbb{E}\Big\{ \sup_{t_0 \le s \le t_1} |x^{(n)}(s)|^2 \Big\} \le C\mathbb{E}\{1 + |x^{(n)}(t_0)|^2\},$$

and an analogous estimate with x replacing $x^{(n)}$, for some constant C independent of n, i.e., depending only $t_1 - t_0$ and the constant appearing in assumption (5.120), and note that

$$\mathbb{E}\big\{[1 - \chi_r(t)]\big\} = P\Big(\sup_{t_0 \le s \le t_1} |x(s)| > r \Big).$$

to deduce that

$$\lim_n \mathbb{E}\Big\{ \sup_{t_0 \le s \le t_1} |x(t) - x^{(n)}(t)| \Big\} \le CP\Big(\sup_{t_0 \le s \le t_1} |x(s)| > r \Big), \quad \forall r > 0,$$

for a suitable constant C independent of $r > 0$, but depending on $\mathbb{E}\{|x(t_0)|^2\}$ and $\sup_n \mathbb{E}\{|x^{(n)}(t_0)|^2\}$. This proves that (5.122) as desired. □

• *Remark* 5.31. If we assume all conditions of Theorem 5.30, except that only $x^{(n)}(t_0) \to x(t_0)$ in probability then we have

$$P\Big(\sup_{t_0 \le t \le t_1} |x_r^{(n)}(t) - x_r(t)| \ge \varepsilon \Big) \to 0, \quad \text{as} \quad n \to \infty, \quad \forall \varepsilon > 0.$$

Indeed, define $x_r^{(n)}(t_0) = \psi_r(x^{(n)}(t_0))$ and $x_r(t_0) = \psi_r(x(t_0))$, where ψ_r is given by (5.116) to check that (because ψ_r is Lipschitz continuous), $\mathbb{E}(|x_r^{(n)}(t_0) - x_r(t_0)|^2) \to 0$ as $n \to \infty$, for every fixed $r > 0$. Therefore, if $x_r^{(n)}(t)$ and $x_r(t)$ denote the solutions with initial conditions $x_r^{(n)}(t_0)$ and $x_r(t_0)$, then Theorem 5.30 implies that

$$\mathbb{E}\Big(\sup_{t_0 \le t \le t_1} |x_r^{(n)}(t) - x_r(t)|^2 \Big) \to 0, \quad \text{as} \quad n \to \infty,$$

for any fixed $r > 0$. Since the uniqueness implies

$$P\big(x_r^{(n)}(\cdot) \neq x^{(n)}(\cdot)\big) \leq P\big(x_r^{(n)}(t_0) \neq x^{(n)}(t_0)\big) = P\big(|x_r^{(n)}(t_0)| > r\big),$$
$$P\big(x_r(\cdot) \neq x(\cdot)\big) \leq P\big(x_r(t_0) \neq x(t_0)\big) = P\big(|x_r(t_0)| > r\big),$$

the triangular inequality

$$|x^{(n)}(t_0)| \leq |x^{(n)}(t_0) - x(t_0)| + |x(t_0)|$$

yields

$$|x^{(n)}(t_0)| > r \quad \text{implies} \quad |x^{(n)}(t_0) - x(t_0)| > r/2 \quad \text{or} \quad |x(t_0)| > r/2,$$

we deduce

$$P\Big(\sup_{t_0\leq t\leq t_1} |x^{(n)}(t) - x(t)| > \varepsilon\Big) \leq P\Big(\sup_{t_0\leq t\leq t_1} |x_r^{(n)}(t) - x_r(t)| > \varepsilon\Big) + \\ + P(|x^{(n)}(t_0) - x(t_0)| > r/2) + 2P(|x(t_0)| > r/2).$$

Hence, take limit as $n \to \infty$ and then as $r \to \infty$ to complete the argument. □

Consider now the case of local existence: let the sequences $g^{(n)}$, $\sigma^{(n)}$ and $\gamma^{(n)}$ verify (5.111) and (5.112), with the sequence of initial data $x^{(n)}(0)$ of $\mathcal{F}_{t_0}$-measurable random variables.

Let $x^{(n)} \colon [\![t_0, \varsigma^{(n)}[\![\to \mathbb{R}^d$ be the maximal solutions of the problems

$$\begin{cases} x^{(n)}(t) = x_0^{(n)} + \displaystyle\int_{t_0}^t g^{(n)}(s, x^{(n)}(s))\mathrm{d}s + \int_{t_0}^t \sigma^{(n)}(s, x^{(n)}(s))\mathrm{d}w(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m \times]t_0, t]} \gamma^{(n)}(\zeta, s, x^{(n)}(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0, \end{cases} \tag{5.124}$$

then we have

Theorem 5.32. *Let $g^{(n)}$, $\sigma^{(n)}$ and $\gamma^{(n)}$ with (5.111) and (5.112) be satisfied uniformly with respect to n. If as $n \to \infty$, we have $x^{(n)}(t_0) \to x(t_0)$ in probability, and for every x in $\mathbb{R}^d$ also $g^{(n)}(t, x, \omega) \to g(t, x, \omega)$, $\sigma^{(n)}(t, x, \omega) \to \sigma(t, x, \omega)$,*

$$\int_{\mathbb{R}_*^m} |\gamma^{(n)}(\zeta, t, x, \omega) - \gamma(\zeta, t, x, \omega)|^2 \pi(\mathrm{d}\zeta) \to 0$$

almost everywhere in (t, ω) with respect to $\mathrm{d}t \times P(\mathrm{d}\omega)$, then, for every θ in $[\![t_0, \varsigma[\![$, we have

$$\lim_{n\to\infty} P(\varsigma^{(n)} \leq \theta) = 0,$$

and

$$\lim_{n\to\infty} P\Big(\varsigma^{(n)} > \theta, \sup_{t_0\leq t\leq\theta} |x^{(n)}(t) - x(t)| > \varepsilon\Big) = 0,$$

where x is the maximal solution corresponding to g, σ and γ, with lifetime ς.

Proof. Consider for each n the approximating solution $x_r^{(n)}$ relative to (5.124). By Theorem 5.30 and Remark 5.31, for every r, we have

$$P\{\sup_{t_0\le t\le t_1}|x_r^{(n)}(t)-x_r(t)|>\varepsilon\}\to 0 \qquad \text{as } n\to\infty. \tag{5.125}$$

Define

$$\tau_r^{(n)}=\inf\big\{t\in[t_0,\varsigma^{(n)}]:\ |x^{(n)}(t)|>r\big\}$$

and

$$\tau_r=\inf\big\{t\in[t_0,\varsigma]:\ |x(t)|>r\big\},$$

to get

$$\{\tau_r^{(n)}<\tau_{r-1}\}\subset\big\{\sup_{t_0\le t\le t_1}|x_r^{(n)}(t)-x_r(t)|\ge 1\big\}.$$

From (5.125), we have

$$\lim_{n\to\infty}P\{\tau_r^{(n)}<\tau_{r-1}\}=0,\quad \forall r.$$

Now, for any θ in $[\![t_0,\varsigma^{(n)}[\![$, we have

$$\{\tau_r^{(n)}\le\theta\}\subset\{\tau_r^{(n)}\le\tau_{r-1}\}\cup\{\tau_{r-1}\le\theta\},$$

so that

$$\lim_{r\to\infty}\lim_{n\to\infty}P\{\tau_r^{(n)}\le\varsigma\}=0. \tag{5.126}$$

Finally, taking in account the following inequalities,

$$P\{\varsigma^{(n)}\le\theta\})\le P\{\tau_r^{(n)}\le\theta\}$$

and

$$\begin{aligned}
P\big\{\varsigma^{(n)}>\theta\,,\ \sup_{t_0\le t\le\theta}|x^{(n)}(t)-x(t)|>\varepsilon\big\}\le\\
\le P\big\{\varsigma^{(n)}>\theta\,,\ \sup_{t_0\le t\le\theta}|x^{(n)}(t)-x_r^{(n)}(t)|>0\big\}+\\
+P\big\{\sup_{t_0\le t\le t_1}|x_r^{(n)}(t)-x_r(t)|>\varepsilon\big\})+\\
+P\big\{\sup_{t_0\le t\le\theta}|x_r(t)-x(t)|>0\big\}\le\\
\le P\big\{\sup_{t_0\le t\le t_1}|x_r^{(n)}(t)-x_r(t)|>\varepsilon\big\}+P\big\{\tau_r^{(n)}\le\theta\big\}+P\{\tau_r\le\theta\},
\end{aligned}$$

the thesis follows from (5.126). □

5.3 Special Semi-Martingale

It is also clear that the fixed-point technique used in the existence and uniqueness Theorem 5.3 can also be applied if we replace the Wiener process with a (local) square-integrable martingale and the time with a (continuous and) increasing predictable process, i.e.,

$$\mathrm{d}x(s) = g(s, x(s-), v(s))\mathrm{d}V(s) + \sigma(s, x(s-), v(s)\mathrm{d}M(s),$$

where $V = (V(s) : s \geq 0)$ is an adapted process with integrable bounded variation and $M = (M(s) : s \geq 0)$ is a (local) square-integrable martingale.

Semi-martingale are the combination of martingales, bounded variation processes and localization procedures. Of particular interest for us is the class where an *integrability* condition is imposed on the adapted bounded variation processes to be used. This class, so-called *special semi-martingale* (that we refer to as just *semi-martingale*, for brevity) could be defined in several equivalent forms, we adopt the following: On a complete filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, a (special) semi-martingale X is a cad-lag adapted $\mathbb{R}^d$-valued process decomposed as $X = X_0 + M + V$, where (a) X_0 is a random variable measurable with respect to $\mathcal{F}_0$, (b) M is a local martingale with $M_0 = 0$, and (c) V is a predictable cad-lag process with finite variation (by coordinates) on each bounded time interval. The localization procedure used on the martingale part M means that there exists a sequence of stopping times $(\tau_1 \leq \tau_2 \leq \cdots)$ such that $\tau_i \to \infty$ and $\tau_i < \tau_{i+1}$ if $\tau_i < \infty$, and $t \mapsto M(t \wedge \tau_i)$ is a martingale relative to the filtration $(\mathcal{F}_t : t \geq 0)$, for any i. The fact that V has local bounded variation yields two predictable increasing processes V^+ and V^- such that $V = V^+ - V^-$. It can be proved (see Jacod and Shiryaev [117, Corollary II.2.28, p. 85]) that the above (canonical) decomposition is unique and that the martingale part M can be written as

$$M(t) = X^c(t) + \int_{\mathbb{R}^d_* \times]0,t]} z\tilde{\nu}_X(\mathrm{d}z, \mathrm{d}t),$$

where X^c is a continuous local martingale and $\tilde{\nu}_X$ is the local martingale associated with the jumps of X, which is a purely jumps local martingale. Thus, (1) X is continuous if and only if $\tilde{\nu}_X = 0$ and V^+, V^- are continuous, (2) X is quasi-continuous if and only if $\nu^p_X(\mathbb{R}^d_*, \{t\}) = 0$ for every $t > 0$ and V^+ and V^- are continuous. The *characteristics* of X are (a) the predictable increasing (vector-valued) processes V^+ and V^-, (b) the predictable quadratic variation (matrix-valued) process $\langle M^c, M^c \rangle$ and (c) the predictable compensator random measure ν^p_X of the integer measure ν_X associated with the jumps of X. Certainly we have $\tilde{\nu}_X := \nu_X - \nu^p_X$. Moreover, there exist a predictable increasing locally integrable process v^0, which is continuous if X is quasi-continuous, such that $\mathrm{d}V = v\mathrm{d}v^0$, $\mathrm{d}\langle M^c, M^c \rangle = m^c \mathrm{d}v^0$ and $\nu^p(\mathrm{d}\zeta, \mathrm{d}t) = K(\mathrm{d}\zeta, t)\,\mathrm{d}v^0(t)$, where v is a $\mathbb{R}^d$-valued predictable process, m^c is a symmetric matrix-valued predictable process and $K(\mathrm{d}\zeta, t, \omega)$ is a transition kernel (i.e., $K(\mathrm{d}\zeta, \cdot, \cdot)$ is a predictable

process and $K(\cdot,t,\omega)$ is a positive measure in $\mathbb{R}^m_*$) such that

$$[v^0(t,\omega)-v^0(t-,\omega)]K(\mathbb{R}^m_*,t,\omega)\le 1 \quad \text{and} \quad \int_{\mathbb{R}^m_*}(1\wedge|\zeta|^2)K(\mathrm{d}\zeta,t,\omega)\le 1,$$

for every t and ω, see Jacod and Shiryaev [117, Proposition II.2.9, p. 77–78].

Thus, given a cad-lag adapted process $h(t,\omega)$ and predictable random fields $g(t,x,\omega)$, $\sigma(t,x,\omega)$ and $\gamma(z,t,x,\omega)$, a stochastic ODE in $\mathbb{R}^d$ for quasi-continuous semi-martingales has the following form

$$\left\{\begin{aligned} x(t) &= h(t)+\int_{t_0}^t g(s,x(s))\mathrm{d}V^c(s)+\int_{t_0}^t \sigma(s,x(s))\mathrm{d}M^c(s)+\\ &\quad+\int_{\mathbb{R}^m_*\times]t_0,t]}\gamma(\zeta,s,x(s))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s),\quad \forall\, t\ge t_0,\end{aligned}\right. \tag{5.127}$$

which can be written by components, $i=1,\dots,d$, as

$$\left\{\begin{aligned} x_i(t) &= h_i(t)+\sum_{j=1}^d\int_{t_0}^t g_{ij}(s,x(s))\mathrm{d}V^c_j(s)+\\ &\quad+\sum_{k=1}^n\int_{t_0}^t \sigma_{ik}(s,x(s))\mathrm{d}M^c_k(s)+\\ &\quad+\int_{\mathbb{R}^m_*\times]t_0,t]}\gamma_i(\zeta,s,x(s))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s),\quad \forall\, t\ge t_0.\end{aligned}\right. \tag{5.128}$$

The initial condition $x(0)=x^0$ has been replaced by a more general condition given by the process h, the process $(V^c(t):t\ge 0)$ is an adapted continuous process with local bounded variation (so, also predictable), and $(M^c(t):t\ge 0)$ is a continuous local martingale and $\tilde\nu:=\nu-\nu^p$ is a purely jumps local martingale, i.e., ν is an integer measure with its compensator ν^p satisfying $\nu^p(\mathbb{R}^m_*,\{t\})=0$ for every $t\ge 0$. By convenience, all integrals above are implicitly extended by 0 for $t\le t_0$ so that $x(t)=h(t)$ for every $t\le t_0$. Recall that the jumps compensator can be disintegrated as $\nu^p(\mathrm{d}\zeta,\mathrm{d}t)=K(\mathrm{d}\zeta,t)\,\mathrm{d}\nu^0(t)$, where ν^0 is an adapted increasing continuous process and the integral of $(1\wedge|\zeta|^2)$ with respect to $K(\mathrm{d}\zeta,t)$ is bounded by 1. Furthermore, we may write $\mathrm{d}V^c_j(t)=v^c_j(t)\mathrm{d}v^0(t)$, $\mathrm{d}\langle M^c_k,M^c_\ell\rangle(t)=m^c_{k\ell}(t)\mathrm{d}v^0(t)$ and take $v^0=\nu^0$, after a redefinition of K.

The terminal time t_1 may be infinite and the initial time t_0 may be a stopping time. Clearly, as mentioned above the process h is cad-lag and adapted, and the (coefficients) processes g, σ and γ are assumed to be progressively measurable, i.e., for every $t_1>t_0$,

$$\left\{\begin{aligned} &g,\sigma:[t_0,t_1]\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d\times\mathbb{R}^n,\\ &\gamma:\mathbb{R}^m_*\times[t_0,t_1]\times\mathbb{R}^d\times\Omega\to\mathbb{R},\end{aligned}\right. \tag{5.129}$$

are measurable relative to the product Borel σ-algebras and $\mathcal{F}(t_1)$ on Ω. Moreover, the coefficients are locally bounded, i.e., for any $r>0$ there exists a real

valued predictable process $C(t,r)$ such that for every $t \geq t_0$ we have

$$\begin{cases} \sup\limits_{|x|\leq r} \Big\{ |g(t,x,\omega)| + |\sigma(t,x,\omega)|^2 + \\ \qquad + \int\limits_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,\omega)|^2 K(\mathrm{d}\zeta,t,\omega) \Big\} \leq C(t,r,\omega), \\ \text{with} \quad P\Big\{ \int\limits_{t_0}^{t_0+r} C(t,r)\mathrm{d}v^0(t,\omega) < \infty \Big\} = 1, \end{cases} \tag{5.130}$$

and satisfy a uniform locally Lipschitz condition in the variable x, namely, there exists a real valued predictable process $M(r,t) > 0$ such that for every (t,x), (t,x') in $[t_0,\infty) \times \mathbb{R}^d$ with $|x| \leq r$ and $|x'| \leq r$ we have

$$\begin{cases} |g(t,x) - g(t,x')|^2 + |\sigma(t,x) - \sigma(t,x')|^2 + \\ \qquad + \int\limits_{\mathbb{R}^m_*} |\gamma(\zeta,t,x) - \gamma(\zeta,t,x')|^2 K(\mathrm{d}\zeta,t) \leq M(t,r)|x-x'|^2, \\ \text{with} \quad P\Big\{ \sup\limits_{t_0\leq t\leq t_0+r} M(t,r) < \infty \Big\} = 1, \end{cases} \tag{5.131}$$

i.e., an adaptation of (5.111) and (5.112).

It is necessary to localize the concept of solution, which makes sense for any progressively measurable locally bounded process. Indeed, for such a process $x(\cdot)$, the measurability assumption (5.129) and the linear growth condition (5.130) ensure the (locally) integrability of the terms appearing in the equation (5.127), so that a posteriori, a cad-lag adapted version, again denoted by $x(\cdot)$ satisfies the same equation. Thus, a solution is a cad-lag adapted process $x(\cdot)$ satisfying locally the ODE. Hence, for a sequence of stopping times $(\tau_1 \leq \tau_1 \leq \cdots)$ such that $\tau_i \to \infty$ and $\tau_i < \tau_{i+1}$ if $\tau_i < \infty$, and

$$|\mathrm{d}V^c(t)|^2 + |\mathrm{d}\langle M^c, M^c\rangle(t)| + |\mathrm{d}\nu^0(t)| \leq i, \quad \forall t \leq \tau_i, \tag{5.132}$$

we may say that the stochastic equation makes sense in L^2 on the stochastic time interval $[t_0,\tau_i]$, for every $i \geq 1$, essentially as in the previous sections. Note that $|\mathrm{d}V^c(t)|$ means the variation process.

We rephrase Definition 5.27. An adapted cad-lag process $x(\cdot)$ is a local solution of the d-dimensional stochastic ordinary differential equation on the time interval $[t_0,t_1]$ if there exists a stopping time $\tau_x = \tau > t_0$ such that (5.127) is satisfied for every t in the stochastic interval $[\![t_0, t_1 \wedge \tau[\![$. Actually, a solution means an equivalence class of processes represented by a cad-lag element. Thus, we say that the local (strong) *uniqueness* holds whenever two solutions are each one version of each other, i.e., if x and y are two solutions then

$$P\{x(t) = y(t) \,:\, t_0 \leq t \leq \tau_x \wedge \tau_y\} = 0.$$

Moreover, the stopping time $\tau = \varsigma$ is maximal if

$$\text{either} \quad P\big\{\varsigma \geq t_1\big\} = 1 \quad \text{or} \quad P\big\{\varsigma < \infty,\ \lim_{t\to\varsigma} |x(t)| = \infty\big\} > 0.$$

and called *lifetime* of the solution, usually denoted by ς or ς_x if necessary.

Theorem 5.33 (existence and uniqueness)**.** *Consider the stochastic ordinary differential equation* (5.127) *with* h, V^c, M^c *and* $\tilde{\nu}$ *defined as above on a complete filtered probability space* $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, *and with coefficients* g, σ *and* γ *satisfying assumptions* (5.129), (5.130) *and* (5.131). *Then there exists one and only one solution* x *with lifetime* ς *to the stochastic ODE* (5.127). *Moreover, if*

$$P\Big\{ \sup_{t_0 \leq t < t_1 \wedge \varsigma} |x(t)| < \infty \Big\} = 1 \tag{5.133}$$

then $\varsigma = \infty$, *i.e., the solution is global on* $[t_0, t_1]$.

Proof. First, let us add two extra assumption, namely, the coefficients have linear growth, i.e., there exists a constant $C > 0$ such that

$$\begin{cases} |g(t,x,\omega)|^2 + |\sigma(t,x,\omega)|^2 + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,\omega)|^2 K(\mathrm{d}\zeta, t, \omega) \leq \\ \qquad\qquad\qquad\qquad \leq C(1+|x|^2), \end{cases} \tag{5.134}$$

for every (t,x,ω) in $[t_0,\infty) \times \mathbb{R}^d \times \Omega$, and that the processes $M(t,r)$ of (5.131) is bounded in ω.

Then, the existence and uniqueness follows from a *localization* argument based on the techniques shown in Theorem 5.3. Indeed, for any constant $\varepsilon > 0$ there exists a stochastic partition of stopping times $(t_0 = \tau_0 < \tau_1 \leq \tau_2 \leq \cdots)$ satisfying $\tau_i \to \infty$, $\tau_i < \tau_{i+1}$ if $\tau_i < \infty$, (5.132) and

$$|\mathrm{d}V^c(t,\omega)|^2 + |\mathrm{d}\langle M^c, M^c\rangle(t,\omega)| + |\mathrm{d}\nu^0(t,\omega)| \leq \varepsilon, \quad \forall\, \tau_i \leq t \leq \tau_{i+1}.$$

Note that for each $t_1 > t_0$ and ω there exists $n = n(t_1 - t_0, \omega)$ such that $\tau_n(\omega) = t_0$. Now, using the Banach space $\mathcal{Y}$ of cad-lag adapted processes with the L^2-sup-norm

$$\|x\| := \sqrt{\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1} |x(t)|^2 \Big\}},$$

we check that the nonlinear operator

$$\begin{aligned} T_i(x)(t) := h_i(t) &+ \int_{t \vee \tau_i}^{t \wedge \tau_{i+1}} g(s-, x(s-), \omega) \mathrm{d}V^c(s) + \\ &+ \int_{t \vee \tau_i}^{t \wedge \tau_{i+1}} \sigma(s-, x(s-), \omega) \mathrm{d}M^c(s) + \\ &+ \int_{\mathbb{R}^m_* \times]t \vee \tau_i, t \wedge \tau_{i+1}]} \gamma(\zeta, s-, x(s-), \omega) \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \in [t_0, t_1], \end{aligned}$$

defined for any square integrable cad-lag adapted process $h_i(t)$, maps $\mathcal{Y}$ into itself.

By making used of the supermartingale inequality, a direct estimate shows that

$$\|T_i(x) - T_i(y)\| \leq C_0 \sqrt{\varepsilon} \|x - y\|, \quad \forall\, x, y \in \mathcal{Y},$$

where C_0 is a constant depending on the *global* Lipschitz property of the coefficients. Thus, if the process h is also square integrable and the coefficients satisfy a global Lipschitz condition in the variable x then T_i is a contraction map, for ε sufficiently small. Therefore, by successively taking $x_0(\cdot) = 0$ and

$$h_i(t) := [h(t) - x_{i-1}(\tau_i)]\mathbb{1}_{t>\tau_i} + x_{i-1}(t)\mathbb{1}_{t\le\tau_i},$$

we find a fixed point $T_i(x_i) = x_i$, which is a solution to the stochastic ODE on the stochastic interval $[t_0, \tau_{i+1}]$. Hence, a solution is defined on the whole interval $[t_0, t_1]$, with a finite number of iterations, for any $t_1 > t_0$ and each ω.

The proof of the existence of a solution can be completed, by removing the extra global Lipschitz condition (just uniformly locally Lipschitz condition suffices) and the square integrability on the process $h(\cdot)$ as in Theorem 5.3.

To prove the uniqueness, let $x(\cdot)$ and $y(\cdot)$ be two cad-lag adapted solutions. Because x and y are bounded almost surely on any bounded interval, we can modify the above sequence of stopping time $(t_0 = \tau_0 < \tau_1 \le \tau_2 \le \cdots)$ to accommodate also the condition

$$|x(t)| + |y(t)| + |h(t)| \le i, \quad \forall t \le \tau_i,$$

without losing the property $\tau_i \to \infty$. By iteration, the contraction property above shows that

$$\|[x - y]\mathbb{1}_{[\tau_i,\tau_{i+1}[}\| = 0, \quad i = 0, 1, \dots.$$

Hence $x(\cdot) = y(\cdot)$ on the stochastic interval $[t_0, \tau_{i+1}]$, and therefore $P\{x(\cdot) = y(\cdot)\} = 0$.

Finally, we complete the proof (without the extra assumptions) as in Theorem 5.28. □

Note that if the temporal (initial) condition is stochastically continuous then so is the solution. Moreover, under the condition (5.134) the solution is global, i.e., $\varsigma = \infty$.

A more challenging problem is to include *predictable jumps* (i.e., to allow discontinuities for the process V) in the above stochastic ODE. Even the deterministic case (i.e., $M^c = 0$ and $\tilde{\nu} = 0$) yields several difficulties as applied to control theory, for instance, see the papers Barron et al. [11], Motta and Rampazzo [185] and Schmaedke [222], among others.

For instance, once a cad-lag version $x(t)$ of the solution has been chosen, and assuming that the coefficients σ and γ are cad-lag in t, a predictable version needed for the integrands in stochastic integrals is simply $\sigma(t-, x(t-))$ and $\gamma(\zeta, t-, x(t-))$. Similarly, if the coefficients g is cad-lag in t then $g(t, x(t))$ and $g(t-, x(t-))$ differ (for every ω) in a countable time set, and because the bounded variation process V is also continuous, we deduce that their pathwise

integrals coincide. Therefore, equality (5.127) can be written as

$$\begin{cases} x(t) = h(t) + \displaystyle\int_{t_0}^{t} g(s-, x(s-))\mathrm{d}V(s) + \int_{t_0}^{t} \sigma(s-, x(s-))\mathrm{d}M^c(s) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_* \times]t_0,t]} \gamma(\zeta, s-, x(s-))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0. \end{cases} \tag{5.135}$$

Thus, the above condition (5.135) is the *proper* or *normalized* expression of the stochastic ODE for cad-lag processes. Moreover, this normalized form makes sense even when the driven system (i.e., V, M^c and $\tilde{\nu}$) is not quasi-continuous, i.e., V is only an adapted cad-lag bounded variation process, not necessarily continuous. Clearly, we may write the stochastic ODE (5.135) in the form

$$\begin{cases} x(t) = h(t) + \displaystyle\int_{t_0}^{t} g(s-, x(s-))\mathrm{d}V(s) + \\ \qquad + \displaystyle\int_{t_0}^{t} \sigma(s-, x(s-))\mathrm{d}M(s), \quad \forall\, t \geq t_0, \end{cases} \tag{5.136}$$

where M is a (quasi-continuous) local martingale, not necessarily continuous.

Sometimes, the formulation (5.136) is complemented with coefficients g and σ of a functional type, i.e., we take $g(x(\cdot), s)$ and $\sigma(x(\cdot), s)$ instead of having $g(s, x(t))$ and $\sigma(s, x(s))$. This means for a (cad-lag) local martingale M and an adapted cad-lag process with bounded variation relative to a filtration $\{\mathcal{F}_t : t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, P)$ we assume:

(a) the processes $s \mapsto g(x(\cdot), s)$ and $s \mapsto \sigma(x(\cdot), s)$ are adapted cad-lag processes, for every adapted cad-lag process $x(\cdot)$,

(b) for every stopping time T, if $x(s) = y(s)$ for every s in $[0, T)$ then $g(x(\cdot), s) = g(y(\cdot), s)$ and $\sigma(x(\cdot), s) = \sigma(y(\cdot), s)$ for every s in $[0, T)$,

(c) for every $r > 0$ there exist two increasing (real-valued) adapted processes $C_r = C_r(t, \omega)$ and $K_r = K_r(t, \omega)$ such that

$$\sup_{0 \leq s \leq t} \left\{ \left| g(x(\cdot), s) \right|^2 + \left| \sigma(x(\cdot), s) \right|^2 \right\} \leq C_r(t) \sup_{0 \leq s \leq t} \left\{ \left| x(s) \right|^2 \right\},$$

and

$$\sup_{0 \leq s \leq t} \left\{ \left| g(x(\cdot), s) - g(y(\cdot), s) \right|^2 + \left| \sigma(x(\cdot), s) - \sigma(y(\cdot), s) \right|^2 \right\} \leq$$
$$\leq K_r(t) \sup_{0 \leq s \leq t} \left\{ \left| x(s) - y(s) \right|^2 \right\},$$

for every adapted cad-lag processes x and y satisfying $|x(s)| \leq r$ and $|y(s)| \leq r$, for every s in $[0, t]$, any $t \geq 0$.

Now, if some conditions are imposed on the variation $t \mapsto \mathrm{var}(V, [0, t])$, the optimal quadratic variation $t \mapsto [M, M](t)$ and the processes C_r and K_r in condition (c) above, then existence and uniqueness o is established for a SDE with functional coefficients. For instance, if C_r and K_r are bounded (case of global Lipschitz condition) then full details (and several other interesting

properties) can be found in Protter [206, Chapter V, pp. 187–284]. However, to include a random measure in a formulation like (5.136) with a semi-martingale, we need to consider local martingales with values in a Hilbert space.

The same technique shown in Theorem 5.33 can be extended directly to this non-continuous case.

Corollary 5.34 (existence and uniqueness)**.** *Consider the stochastic ordinary differential equation* (5.135) *relative to the processes* h, V, M^c *and* $\tilde{\nu}$ *defined as above on a complete filtered probability space* $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, *and with coefficients* g, σ *and* γ *satisfying assumptions* (5.129), (5.130), (5.131), *which are cad-lag in the variable* t. *Then there exists one and only one local solution* x *to the stochastic ODE* (5.135), *with lifetime* ς. *Moreover, under the condition* (5.134) *the solution is global, i.e.,* $\varsigma = \infty$. □

Note that usually the process V is taken predictable, however adapted suffices. Thus, assuming that the local martingale M in (5.136) is quasi-continuous does not limit the settings. Clearly, the normalized stochastic ODE makes sense only for a cad-lag version of the solution $x(\cdot)$ and the form (5.136) do have some limitation on the way the jumps can be changed by the coefficients σ, instead of γ as in (5.135).

5.4 Measure and Time Changes

It is clear that if $w = (w_1, \dots, w_n)$ is a n-dimensional standard Wiener process defined in $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : 0 \leq t \leq T)$ with $\mathcal{F} = \mathcal{F}(T)$ then for any constants $\sigma = (\sigma_{ik} : i = 1, \dots, d,\ k = 1, \dots, n)$ the expression $\tilde{w}(t) := \sigma w(t)$ is a new d-dimensional Wiener process with covariance matrix $a = \sigma^* \sigma$, i.e., $a_{ij} := \sum_k \sigma_{ik} \sigma_{jk}$ (with vanishing drift, so a martingale). On the other hand, to add a constant drift $b = (b_1, \dots, b_n)$, instead of changing the process $\tilde{w}$, we are going to change the probability measure P, namely, $\tilde{P}(\mathrm{d}\omega) := Z(T, \omega)\, P(\mathrm{d}\omega)$, where

$$Z(t) := \exp\Big[\sum_k b_k\, w_k(t) - \frac{1}{2} \sum_k (b_k)^2\, t \Big], \quad \forall t \in [0, T].$$

We claim that

$$\begin{cases} \text{the process } \tilde{w} = \sigma w \text{ is a } d\text{-dimensional Wiener process} \\ \text{with covariance } a = \sigma^* \sigma \text{ and drift } \sigma b \text{ under } \tilde{P}. \end{cases} \tag{5.137}$$

Indeed, first set $\bar{w}_i(t) := \tilde{w}_i(t) - \sum_k \sigma_{ik} b_k\, t$ and notice that a simple application of Itô formula shows that the (density) process Z satisfies the stochastic differential equation

$$Z(t) = 1 + \sum_k b_k \int_0^t Z(s)\, \mathrm{d}w_k(s), \quad \forall t \in [0, T],$$

proving that $(Z(t) : 0 \leq t \leq T)$ is a non-negative continuous martingale. It is clear that the new probability measure $\tilde{P}(\mathrm{d}\omega)$ is defined on $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : 0 \leq$

$t \le T$), the mathematical expectation relative to $\tilde{P}$ is denoted by $\tilde{\mathbb{E}}\{\cdot\}$ and in view of the martingale property we have

$$\tilde{P}(A) := \int_A Z(t,\omega)\, P(\mathrm{d}\omega) \quad \forall A \in \mathcal{F}(t),\ t \in [0,T].$$

Moreover, for any $\mathcal{F}(t)$-measurable random variable y with $\tilde{\mathbb{E}}\{|y|\} < \infty$, the conditional expectation satisfies

$$\tilde{\mathbb{E}}\{y \mid \mathcal{F}(s)\} = \frac{1}{Z(s)}\mathbb{E}\{y\, Z(t) \mid \mathcal{F}(s)\},$$

almost surely relative to P and $\tilde{P}$. Since $\langle \tilde{w}_i, Z\rangle(t) = Z(t)\sum_k \sigma_{ik} b_k\, t$, by means of Itô (or the integration-by-parts) formula we have

$$\bar{w}(t)\, Z(t) = \bar{w}(s)\, Z(s) + \int_s^t \bar{w}(s)\mathrm{d}Z(t) + \int_s^t Z(t)\mathrm{d}\tilde{w}(s)$$

and because $\bar{w}$ and Z are both martingale relative to P, we deduce the equality $\mathbb{E}\{\bar{w}(t)\, Z(t) \mid \mathcal{F}(s)\} = \bar{w}(s)\, Z(s)$, which implies $\tilde{\mathbb{E}}\{\tilde{w}(t) \mid \mathcal{F}(s)\} = \bar{w}(s)$, i.e., $(\bar{w}(t) : t \ge 0)$ is a martingale relative to $\tilde{P}$. Similarly, by means of Itô formula we have

$$\bar{w}_i(t)\bar{w}_j(t)\, Z(t) = \bar{w}_i(s)\, \bar{w}_j(s)\, Z(s) + \int_s^t \bar{w}_i(s)\, \bar{w}_j(s)\mathrm{d}Z(t) + \\ + \int_s^t Z(t)\mathrm{d}(\tilde{w}_i(s)\tilde{w}_j(s)),$$

$$\mathrm{d}(\tilde{w}_i(s)\tilde{w}_j(s)) = \tilde{w}_i(s)\mathrm{d}\tilde{w}_j(s) + \tilde{w}_j(s)\mathrm{d}\tilde{w}_i(s) + \sigma_{ik}\,\sigma_{jk}\mathrm{d}s,$$

which yields

$$\mathbb{E}\{\bar{w}_i(t)\bar{w}_j(t)\, Z(t) \mid \mathcal{F}(s)\} = \bar{w}_i(s)\, \bar{w}_j(s)\, Z(s) + \sigma_{ik}\,\sigma_{jk}\, Z(s)\,(t-s).$$

Hence

$$\tilde{\mathbb{E}}\{\bar{w}_i(t)\bar{w}_j(t) \mid \mathcal{F}(s)\} = \bar{w}_i(s)\, \bar{w}_j(s) + \sigma_{ik}\,\sigma_{jk}\,(t-s),$$

proving that

$$\bar{w}_i(t) = \sum_k \sigma_{ik} w_k(t) - \sum_k \sigma_{ik} b_k\, t,$$

is a d-dimensional Wiener process with covariance a under $\tilde{P}$, i.e., our claim (5.137). In particular, if $n = d$ and σ is the identity then $w(t) - bt$ is a standard Wiener process under the probability measure $\tilde{P}$.

A general setting for a given n-dimensional standard Wiener process $w = (w_1, \dots, w_n)$ defined in $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \ge 0)$ with $\mathcal{F} = \cup_{t\ge 0}\mathcal{F}(t)$ is as follows. Let $b = (b_1, \dots, b_n)$ be an integrand with respect to w, i.e., b belongs to $L^2_{\mathrm{loc}}(w)$, namely

$$P\Big\{\int_0^T |b_k(t)|^2 \mathrm{d}t < \infty\Big\} = 1, \quad \forall T > 0,\ k = 1, \dots, n.$$

Define

$$Z(t) := \exp\Big[\sum_{k=1}^{n}\int_0^t b_k(s)\mathrm{d}w_k(t) - \frac{1}{2}\sum_{k=1}^{n}\int_0^t |b_k(s)|^2\mathrm{d}s\Big], \tag{5.138}$$

which is a (nonnegative) continuous local martingale satisfying the equation

$$Z(t) = 1 + \sum_{k=1}^{n}\int_0^t Z(s)\, b_k(s)\mathrm{d}w_k(s), \quad \forall t \geq 0.$$

Usually, the process $Z = Z_{b,w}$ is referred to as an exponential local martingale. A priori, Z is a super-martingale which becomes a martingale if $\mathbb{E}\{Z(t)\} = 1$ for every $t \geq 0$. This critical point can be overcame by means of Novikov sufficient condition which states that for any continuous local martingale M satisfying

$$\mathbb{E}\Big\{\exp\big[\frac{1}{2}\langle M\rangle_t\big]\Big\} < \infty, \quad \forall t \geq 0,$$

we have $\mathbb{E}\{\exp(M_t - \langle M\rangle_t/2)\} = 1$, for every $t \geq 0$. Thus, Z is a martingale whenever

$$\mathbb{E}\Big\{\exp\big[\frac{1}{2}\int_0^t |b_k(s)|^2\mathrm{d}s\big]\Big\} < \infty, \quad \forall t \geq 0$$

holds.

Now, consider the following Girsanov transformation $M \mapsto \tilde{M}$ defined for any continuous local martingale M by the formula

$$\tilde{M}(t) := M(t) - \sum_{k=1}^{n}\int_0^t b_k(s)\mathrm{d}\langle M, w_k\rangle(s), \quad \forall t \geq 0. \tag{5.139}$$

Since $\mathcal{F} = \cup_{t\geq 0}\mathcal{F}(t)$, we can define a unique probability measure $\tilde{P}$ on $\mathcal{F}$ satisfying

$$\tilde{P}(A) = \mathbb{E}\{\mathbb{1}_A Z(t)\}, \quad \forall A \in \mathcal{F}(t),$$

for every $t \geq 0$, where $Z = Z_{b,w}$ is given by (5.138). Denote by $\tilde{\mathbb{E}}$ the mathematical expectation relative to $\tilde{P}$.

Theorem 5.35 (Girsanov)**.** *Assume that $Z_{b,w}$ defined by (5.138) satisfies the condition $\mathbb{E}\{Z_{b,w}(t)\} = 1$ for every $t \geq 0$. Then for any two processes M and N which are continuous local martingales under P, the transformed processes via (5.139) $\tilde{M}$ and $\tilde{N}$ are continuous local martingales under $\tilde{P}$ with the same predictable quadratic covariation process $\langle\tilde{M}, \tilde{N}\rangle = \langle M, N\rangle$ computed under the appropriate probability measure. In particular, $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_n)$, with*

$$\tilde{w}_k = w_k(t) - \int_0^t b_k(s)\mathrm{d}s, \quad \forall t \geq 0,\ k = 1, \ldots, n$$

is a n-dimensional standard Wiener process under $\tilde{P}$. □

A proof with some discussion can be found in Ikeda and Watanabe [110, Section IV.4, pp. 190–202] or Karatzas and Shreve [124, Section 3.5, pp. 190–201]. On the other hand, for discontinuous local martingales the result can be rephrased as follows.

Theorem 5.36. *Let Z be an exponential martingale, i.e., a non-negative martingale in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ with $\mathcal{F} = \cup_{t\geq 0}\mathcal{F}(t)$ satisfying the initial condition $Z(0) = 1$. Assume M a local martingale with $M(0) = 0$ such that the optional quadratic covariation $[M, Z]$ has a locally integrable variation, and denote by $\langle M, Z\rangle$ its compensator (under P). Then, the process*

$$\tilde{M}(t) := M(t) - \int_{]0,t]} \frac{1}{Z(s-)} \mathrm{d}\langle M, Z\rangle(s), \quad \forall t \geq 0$$

is a local martingale under $\tilde{P}$, where $\tilde{P}$ is defined as the unique probability measure on $\mathcal{F}$ satisfying $\tilde{P}(A) = \mathbb{E}\{\mathbb{1}_A Z(t)\}$ for every A in $\mathcal{F}(t)$ and for every $t \geq 0$. Moreover, the predictable quadratic variation $\langle M^c\rangle$ of the continuous part M^c of M (under P) is also a version of the predictable quadratic variation $\langle \tilde{M}^c\rangle$ of the continuous part $\tilde{M}^c$ of $\tilde{M}$ (under $\tilde{P}$). □

Notice that if Z and M are two local martingales such that the jumps of M are bounded by a (deterministic) constant, i.e., $|\delta M| \leq r$ for some r in $(0, \infty)$, then the optional quadratic covariation $[M, Z]$ has a locally integrable variation. The point here is that $\delta[M, Z] = \delta M\, \delta Z$ and $|\delta Z(t)| \leq \sqrt{\sum_{s\leq t} |\delta Z(s)|^2}$, which is locally integrable. Clearly, this applies to the particular case where M is continuous.

It is also clear that the above statement includes the fact that such a probability $\tilde{P}$ exists uniquely and that $\tilde{M}$ is $\tilde{P}$-almost surely defined as a real-valued process, actually, we have $\tilde{P}\{\inf_t Z(t) > 0\} = 1$, which implies that $\tilde{M}$ is locally integrable.

This change of probability measures, known as Cameron-Martin change of measure or Girsanov transformation, is very useful. First, we focus our interest on a d-dimensional continuous local martingale M^c with predictable quadratic covariation (matrix) process $\langle M^c\rangle$. Let $(b_1(s), \ldots, b_d(s))$ be a progressively measurable process such that for every $i = 1, \ldots, d$ and any $r = 1, 2, \ldots$, we have

$$\mathbb{E}\Big\{ \sum_{ij=1}^{d} \int_{]0,\tau_r]} b_i(t) b_j(t) \mathrm{d}\langle M_i^c, M_j^c\rangle \Big\} < \infty, \tag{5.140}$$

where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. Define

$$Z_b(t) := \exp\Big\{ \int_{]0,t]} b(s) \mathrm{d}M^c(s) - \frac{1}{2} \int_{]0,t]} b^2(s) \mathrm{d}\langle M^c\rangle(s) \Big\}, \tag{5.141}$$

where

$$\int_{]0,t]} b(s)\mathrm{d}M^c(s) = \sum_{i=1}^{d} \int_{]0,t]} b_i(s)\mathrm{d}M_i^c(s)$$
$$\int_{]0,t]} b^2(s)\mathrm{d}\langle M^c\rangle(s) = \sum_{i,j=1}^{d} \int_{]0,t]} b_i(s)b_j(s)\mathrm{d}\langle M_i^c, M_j^c\rangle(s),$$

for every $t \geq 0$. As above, by means of Itô formula we deduce that the process Z_b satisfies the equation

$$Z_b(t) = 1 + \sum_{i=1}^{d} \int_{]0,t]} Z_b(s)b_i(s)\mathrm{d}M_i^c(s), \quad \forall t \geq 0,$$

which implies that Z_b is a local martingale and a super-martingale. Assume that Z_b is indeed a martingale, i.e., $\mathbb{E}\{Z_b(t)\} = 1$, for every $t > 0$, which is true under the Novikov condition, namely,

$$\mathbb{E}\Big\{ \exp\big[\frac{1}{2} \sum_{i,j=1}^{d} \int_{]0,t]} b_i(s)b_j(s)\mathrm{d}\langle M_i^c, M_j^c\rangle(s)\big]\Big\} < \infty, \quad \forall t > 0 \tag{5.142}$$

holds. Then, the exponential continuous martingale Z_b defines a unique probability measure P_b, which is absolutely continuous with respect to P and satisfies $P_b(A) = \mathbb{E}\{\mathbb{1}_A Z(t)\}$ for every A in $\mathcal{F}(t)$ and for every $t \geq 0$. Girsanov Theorem implies that the process $\tilde{M}^c = (\tilde{M}_1^c, \ldots, \tilde{M}_d^c)$, given by

$$\tilde{M}_i^c(t) := M_i^c(t) - \int_{]0,t]} \sum_{j=1}^{d} b_j(s)\mathrm{d}\langle M_i^c, M_j^c\rangle(s), \quad \forall t \geq 0,$$

is a d-dimensional continuous local martingale with predictable quadratic covariation (matrix) process $\langle \tilde{M}^c\rangle = \langle M^c\rangle$.

Now, we focus our attention on a quasi-left continuous m-dimensional purely discontinuous martingale. Let ν be a quasi-left continuous integer-valued random measure in $\mathbb{R}_*^d$, with local martingale measure $\tilde{\nu}$ and compensator ν^p, and let $\delta(z,s)$ be a positive predictable process such that

$$\begin{cases} \mathbb{E}\Big\{ \displaystyle\int_{\mathbb{R}_*^d \times]0,\tau_r]} [\delta(z,t) - 1]^2 \nu^p(\mathrm{d}z, \mathrm{d}t)\Big\} < \infty, \quad \forall r \geq 1, \\ \mathbb{E}\Big\{ \displaystyle\int_{\mathbb{R}_*^d \times]0,\tau_r]} \big[\delta(z,t) - 1 - \ln\big(\delta(z,t)\big)\big] \nu^p(\mathrm{d}z, \mathrm{d}t)\Big\} < \infty, \end{cases} \tag{5.143}$$

where $\{\tau_r : r \geq 1\}$ is a non-decreasing sequence of stopping times satisfying $\tau_r \to \infty$ almost surely. Note the inequalities $\delta - 1 - \ln(\delta) \geq 0$ for every $\delta > 0$, and $\delta - 1 - \ln(\delta) \leq [\delta - 1]^2/2$ if $\delta \geq 1$. So, if the process $\delta(z,s) \geq 1$ then only the first condition in (5.143) is relevant. The second condition in (5.143) is equivalent to

$$\mathbb{E}\Big\{ \int_{\mathbb{R}_*^d \times]0,\tau_r]} \big[\delta(z,t) - 1 - \ln\big(\delta(z,t)\big)\big] \nu(\mathrm{d}z, \mathrm{d}t)\Big\} < \infty,$$

which is understood as a (classic) Lebesgue integral as long as the predictable version of the integrand is used.

Now set

$$\begin{cases} Z_\delta(t) := \exp\Big\{ \displaystyle\int_{\mathbb{R}^d_*\times]0,t]} [\delta(z,s)-1]\tilde\nu(\mathrm{d}z,\mathrm{d}s) - \\ \qquad\qquad - \displaystyle\int_{\mathbb{R}^d_*\times]0,t]} \big[\delta(z,s)-1-\ln\big(\delta(z,s)\big)\big]\nu(\mathrm{d}z,\mathrm{d}s)\Big\}, \end{cases} \tag{5.144}$$

for every $t \geq 0$. Let ε be a positive constant and define $Z^\varepsilon_\delta(t)$ as above where the integer measure ν has been replaced by ν_ε,

$$\nu_\varepsilon(B,[a,b[) := \nu(B\cap\{|z|\geq\varepsilon\},[a,b[), \quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\geq 0,$$

with its corresponding compensator ν^p_ε and its local martingale measure $\tilde\nu_\varepsilon$. Recalling that $\tilde\nu_\varepsilon = \nu_\varepsilon - \nu^p_\varepsilon$, we have

$$\begin{aligned} Z^\varepsilon_\delta(t) = \exp\Big\{ &\int_{\mathbb{R}^d_*\times]0,t]} \ln\big(\delta(z,s)\big)\nu_\varepsilon(\mathrm{d}z,\mathrm{d}s) - \\ &- \int_{\mathbb{R}^d_*\times]0,t]} \big[\delta(z,s)-1\big]\nu^p_\varepsilon(\mathrm{d}z,\mathrm{d}s)\Big\}, \end{aligned}$$

as long as $\delta>0$ almost surely. Moreover, if $\{z^\varepsilon_i,\tau^\varepsilon_i : i\geq 1\}$ are the atoms of ν^ε, with $\tau_{i-1}<\tau_i$, then

$$\begin{aligned} Z^\varepsilon_\delta(\tau^\varepsilon_i) &= Z^\varepsilon_\delta(\tau^\varepsilon_i-)\delta(z^\varepsilon_i,\tau^\varepsilon_i), \\ Z^\varepsilon_\delta(\tau^\varepsilon_i-) &= \int_{\tau^\varepsilon_{i-1}}^{\tau^\varepsilon_i} Z^\varepsilon_\delta(s)\big[\delta(z,s)-1\big]\nu^p_\varepsilon(\mathrm{d}z,\mathrm{d}s), \end{aligned}$$

which proves that the density process Z_δ satisfies the stochastic differential equation

$$Z^\varepsilon_\delta(t) = 1 + \int_{\mathbb{R}^d_*\times]0,t]} Z^\varepsilon_\delta(s-)[\delta(z,s)-1]\tilde\nu_\varepsilon(\mathrm{d}z,\mathrm{d}s),$$

for every $t\geq 0$. Replace δ with $\delta+(1/n)$ and as ε vanishes and later as n becomes infinite, the same equation remains valid for Z_δ with $\tilde\nu$ in lieu of $\tilde\nu_\varepsilon$. This proves that Z_δ is a local martingale. Note that $\delta=0$ means no jump, so that we may allow a nonnegative density processes $\delta(z,s)\geq 0$ as long as we interpret $\ln\delta(z,s)=0$ when $\delta(z,s)=0$, in (5.144) and also in the assumption (5.143). Also, we may use either $Z_\delta(s-)$ or $Z_\delta(s)$ inside the stochastic integral, since the compensator satisfies $\nu^p(\mathbb{R}^d_*\times\{t\})=0$ for every $t\geq 0$. Now, assuming[7] that Z_δ is indeed a martingale, i.e., $\mathbb{E}\{Z_\delta(t)\}=1$, for every $t>0$, the process Z_δ is an exponential martingale which defines an unique probability measure P_δ, which is absolutely continuous with respect to P and satisfies $P_\delta(A)=\mathbb{E}\{\mathbb{1}_A Z_\delta(t)\}$, for every A in $\mathcal{F}(t)$ and for every $t\geq 0$. Set

$$M^\nu_i(t) := \int_{\mathbb{R}^d_*\times]0,t]} \gamma_i(z,s)\tilde\nu(\mathrm{d}z,\mathrm{d}s), \quad \forall t\geq 0,$$

[7]Kazamaki criterium affirms that it suffices to know that Z_δ is a uniformly integrable sub-martingale, e.g., δ bounded by a deterministic constant and $\nu^p(\mathbb{R}^d_*,\cdot)$ is finite.

where $\gamma = (\gamma_1, \dots, \gamma_d)$ is a predictable process bounded by a (deterministic) constant. Under the probability P_δ, the process $M^\nu = (M_1^\nu, \dots, M_d^\nu)$ is a d-dimensional quasi-left continuous local martingale purely discontinuous with predictable random measure compensator ν_M^p on $\mathbb{R}_*^d$ given by

$$\nu_M^p(B\times]0,t]) = \int_{\mathbb{R}_*^d\times]0,t]} \mathbb{1}_{\{\gamma(z,s)\in B\}}\nu^p(\mathrm{d}\zeta,\mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}_*^d)$ and $t>0$, i.e., in short $\nu_M^p = \nu^p(\gamma)$. By means of Girsanov Theorem we deduce that, under the probability measure P_δ, the process

$$\begin{aligned}\tilde{M}_i^\nu(t) := &\int_{\mathbb{R}_*^d\times]0,t]} \gamma_i(z,s)\tilde{\nu}(\mathrm{d}z,\mathrm{d}s) - \\ &- \int_{\mathbb{R}_*^d\times]0,t]} \gamma_i(z,s)[\delta(z,s)-1]\nu^p(\mathrm{d}z,\mathrm{d}s),\end{aligned}$$

or, equivalently,

$$\tilde{M}_i^\nu(t) := \int_{\mathbb{R}_*^d\times]0,t]} \gamma_i(z,s)\tilde{\nu}_\delta(\mathrm{d}z,\mathrm{d}s), \quad \forall t\ge 0,$$

is a quasi-left continuous local martingale purely discontinuous with predictable random measure compensator $\nu_{\tilde{M}}^p$ on $\mathbb{R}_*^d$ given by

$$\nu_{\tilde{M}}^p(B\times]0,t]) = \int_{\mathbb{R}_*^d\times]0,t]} \mathbb{1}_{\{\gamma(z,s)\in B\}}\nu_\delta^p(\mathrm{d}z,\mathrm{d}s),$$

for every B in $\mathcal{B}(\mathbb{R}_*^d)$ and $t>0$, where $\nu_\delta^p := \delta\nu^p$ and $\tilde{\nu}_\delta := \tilde{\nu} - (\delta-1)\nu^p$. In short $\nu_{\tilde{M}}^p = \delta\,\nu^p(\gamma)$. When γ is not necessarily bounded by (deterministic) constant, we can be replace γ by $\gamma\,\mathbb{1}_{|\gamma|\le r}$. As r becomes infinity, the previous transformation remains valid for coefficients satisfying only a local integrability condition with respect to the compensator ν^p.

It is clear that we can combine expressions (5.141) and (5.144). Besides the local integrability conditions (5.140) and (5.143) on the coefficients b and δ, we need to impose Novikov condition (5.142) on b and something else on δ. For instance, if δ is bounded by a deterministic constant then Z_δ satisfies $\mathbb{E}\{Z_\delta(t)\} = 1$, for every $t>0$.

A full proof of Theorem 5.36 as well as more details on Cameron-Martin-Girsanov transformation can be found in Jacod and Shiryaev [117, Section III.3, pp. 152–166] or Rogers and Williams [214, Section IV.6.38, pp. 79–83].

For the case of Itô processes with jumps we have the following results. Let $(\sigma_{ik}(t) : t\ge 0,\ i=1,\dots,d,\ k=1,\dots,n)$ and $(\gamma_i(\zeta,t) : t\ge 0,\ \zeta\in\mathbb{R}_0^m)$ be predictable (adapted suffices) processes such that for every $i=1,\dots,d$ and any $t>0$ we have

$$\begin{aligned}&P\Big\{\int_0^t \sum_{k=1}^n |\sigma_{ik}(s)|^2\mathrm{d}s + \int_{\mathbb{R}_*^m\times]0,t]} |\gamma_i(\zeta,s)|^2\pi(\mathrm{d}\zeta,\mathrm{d}s) < \infty\Big\} = 1,\\ &M_i(t) = \sum_{k=1}^n \int_0^t \sigma_{ik}(s)\mathrm{d}w_k(s) + \int_{\mathbb{R}_*^m\times]0,t]} \gamma_i(\zeta,s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s).\end{aligned}$$

in some Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, with Lévy measure $\pi(\cdot)$. On the other hand, let $(b_1(s), \ldots, b_k(s))$ and $c = \delta - 1$ be predictable (adapted suffices) processes satisfying

$$P\Big\{ \sum_{k=1}^{n} \int_0^t |b_k(s)|^2 \mathrm{d}s < \infty \Big\} = 1,$$
$$P\Big\{ \int_{\mathbb{R}^m_* \times]0,t]} c^2(\gamma(\zeta, s)), s) \pi(\mathrm{d}\zeta, \mathrm{d}s) < \infty \Big\} = 1,$$

and $c \geq 0$, for every $t > 0$. Define the density process

$$Z(t) := \exp\Big\{ \sum_k \int_0^t b_k(s) \mathrm{d}w_k(s) - \frac{1}{2} \sum_{ik} \int_0^t |b_k(s)|^2 \mathrm{d}s + \\ + \int_{\mathbb{R}^m_* \times]0,t]} \ln[1 + c(\gamma(\zeta, s)), s)] \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) - \\ - \int_{\mathbb{R}^m_* \times]0,t]} \big[c(\gamma(\zeta, s)), s) - \ln[1 + c(\gamma(\zeta, s)), s)] \big] \pi(\mathrm{d}\zeta, \mathrm{d}s) \Big\},$$

assume $\mathbb{E}\{Z(t)\} = 1$ for every $t > 0$, and construct the unique probability measure Q satisfying $Q(A) := \mathbb{E}\{Z(t) \mathbb{1}_A\}$, for every A in $\mathcal{F}(t)$ and for every $t \geq 0$.

Then, there is Wiener process $\tilde{w}$ (possible in extension of the given Wiener-Poisson space if σ does not has constant rank) such that under the probability measure Q we have

$$M_i(t) = \sum_{k=1}^{n} \int_0^t \sigma_{ik}(s) \mathrm{d}\tilde{w}_k(s) + \int_{\mathbb{R}^m_* \times]0,t]} \gamma_i(\zeta, s) \tilde{\nu}_c(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \sum_{k=1}^{n} \int_0^t \sigma_{ik}(s) b_k(s) \mathrm{d}s,$$

for every $t \geq 0$. The local martingale measure is given by $\tilde{\nu}_c := \tilde{\nu} - c\pi$ with its predictable compensator $\nu_c^p = (1 + c)\pi$, which is not necessarily deterministic. Recall that in this context, an extension of a given Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}(t), w(t), \tilde{\nu}(\mathrm{d}z, \mathrm{d}t) : t \geq 0,\ z \in \mathbb{R}^m_*),$$

with Lévy measure $\pi(\cdot)$, $\mathcal{F} = \mathcal{F}(\infty)$ and w being d-dimensional, is another Wiener-Poisson space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}, \tilde{\mathcal{F}}(t), \tilde{w}(t), \tilde{\nu}(\mathrm{d}z, \mathrm{d}t) : t \geq 0,\ z \in \mathbb{R}^m_*)$, where $\tilde{\Omega} = \Omega \times \Omega_0$, $\tilde{P} = P \times P_0$, $\tilde{w} = (w, w_0)$, $\tilde{\nu}$ is (possible another) martingale measure with predictable quadratic variation $\pi(\cdot)$, $\tilde{\mathcal{F}}(t)$ is the complete right-continuous filtration generated by $\mathcal{F}(t) \times \mathcal{F}_0(t)$, $\tilde{\mathcal{F}} = \tilde{\mathcal{F}}(\infty)$, and $(\Omega_0, \mathcal{F}_0, P_0, \mathcal{F}_0(t), w_0(t) : t \geq 0)$ is a $\tilde{d}$-dimensional Wiener space.

In particular, let $\nu(\mathrm{d}\zeta, \mathrm{d}t)$ be a Poisson random measure in a filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ with Lévy measure $\pi(\cdot)$ in $\mathbb{R}^m_*$ and $(\delta(\zeta, t) : t \geq 0,\ \zeta \in \mathbb{R}^m_*)$ be a predictable (adapted is sufficient, but we take its predictable version) nonnegative process bounded by a deterministic constant and

satisfying

$$\int_0^t \mathrm{d}s \int_{R_*^m} [\delta(\zeta,s)-1]^2 \pi(\mathrm{d}\zeta) < \infty,$$
$$\int_0^t \mathrm{d}s \int_{R_*^m} \big[\delta(\zeta,s)-1-\ln\big(\delta(\zeta,s)\big)\big]\pi(\mathrm{d}\zeta) < \infty,$$

P-almost surely for every $t \geq 0$, with $\ln\big(\delta(\zeta,s)\big) = 0$ whenever $\delta(\zeta,s) = 0$. Consider the density process

$$Z_\delta(t) := \exp\Big\{ \int_{\mathbb{R}_*^m \times]0,t]} [\delta(\zeta,s)-1]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s) - \\ - \int_{\mathbb{R}_*^m \times]0,t]} \big[\delta(\zeta,s)-1-\ln\big(\delta(\zeta,s)\big)\big]\chi_\delta(\zeta,s)\nu(\mathrm{d}\zeta,\mathrm{d}s)\Big\}$$

for every $t \geq 0$, where $\chi_\delta(\zeta,s) = 1$ if $\delta(\zeta,s) > 0$ and $\chi_\delta(\zeta,s) = 0$ otherwise. As before, Z_δ solves the equation

$$Z_\delta(t) = 1 + \int_{\mathbb{R}_*^m \times]0,t]} Z_\delta(s-)[\delta(\zeta,s)-1]\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s),$$

for every $t \geq 0$. Now, construct a new probability measure P_δ, which is absolutely continuous with respect to P and satisfies $P_\delta(A) = \mathbb{E}\{\mathbb{1}_A Z_\delta(t)\}$, for every A in $\mathcal{F}(t)$ and $t \geq 0$. Then the process

$$M_\delta(t) := \int_{\mathbb{R}_*^m \times]0,t]} \zeta\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s) - \int_0^t \zeta\delta(\zeta,s)\pi(\mathrm{d}\zeta)$$

is a quasi-left continuous purely discontinuous martingale process which associated local martingale random measure is given by

$$\tilde{\nu}_\delta(B,]0,t]) = \nu(B,]0,t]) - \int_0^t \mathrm{d}s \int_B \delta(\zeta,s)\pi(\mathrm{d}\zeta), \tag{5.145}$$

for every B in $\mathcal{B}(\mathbb{R}_*^m)$ and $t > 0$, with (jump) compensator

$$\nu_\delta^p(B,]0,t]) = \int_0^t \mathrm{d}s \int_B \delta(\zeta,s)\pi(\mathrm{d}\zeta), \quad \forall B \in \mathcal{B}(\mathbb{R}_*^m),\ t > 0 \tag{5.146}$$

under the probability measure P_δ.

Summing up, the Lévy measure π is changed into a new Lévy measure π_δ with a density δ, namely, $\pi_\delta(\mathrm{d}\zeta)\mathrm{d}t = \delta(\zeta,t)\pi(\mathrm{d}\zeta)\mathrm{d}t$, or equivalently $\pi_\delta(\mathrm{d}\zeta)\mathrm{d}t = \big(1+c(\zeta,t)\big)\pi(\mathrm{d}\zeta)\mathrm{d}t$, where the function c is square-integrable with respect to π, in other words, if π integrates $|\zeta|^2 \wedge 1$ then the added part $c\pi$ will integrate $|\zeta| \wedge 1$, i.e., what we are able to add have a weaker singularity.

Beside this change of measure, there are other transformations that can be performed with a Wiener process and Poisson measures. For instance, given a n-dimensional Wiener process $w = (w_1, \ldots, w_n)$ and an integrable (with respect

to w) matrix-valued process $p = (p_{hk} : h, k = 1, \dots, n)$ we can define the new process $\mathrm{d}w^p = p\,\mathrm{d}w$,

$$w_h^p(t) := \sum_{k=1}^{n} \int_0^t p_{hk}(s)\mathrm{d}w_k(s), \quad \forall t, h,$$

which is a continuous martingale with predictable quadratic variation

$$\langle w_h^p, w_k^p \rangle(t) = \sum_{l=1}^{n} \int_0^t p_{hl}(s)p_{lk}(s)\mathrm{d}s, \quad \forall t, h, k.$$

In particular, if $p(s)$ is an orthogonal matrix for each s (i.e., a rotation of coordinates) then $\langle w_h^p, w_k^p \rangle(t) = t\delta_{hk}$, so that w^p is again a Wiener process.

Similarly, if $(c(t) : t \geq 0)$ is an adapted real-valued process with $C_1 \leq c(t) \leq C_2$, for every $t \geq 0$ and some constants $C_2 \geq C_2 > 0$, define

$$\tau(t) := \int_0^t c(s)\mathrm{d}s, \qquad \tau^{-1}(s) := \inf\{t \geq 0 : \tau(t) > s\},$$

then $s \mapsto w_k(\tau^{-1}(s))$ is a martingale and the process

$$w_{c,k}(t) := \int_0^t \sqrt{c(s)}\mathrm{d}w_k(\tau^{-1}(s)), \quad \forall t \geq 0,\ k$$

is a d-dimensional Wiener process with respect to its natural filtration. The random variables $\tau(t)$ are stopping time for each t, and they generate a filtration $\mathcal{F}^c(t) := \mathcal{F}(\tau(t))$, for every $t \geq 0$. This is call a *time change*, which can be generalized to continuous local martingales, for instance, more details can be found in Chung and Williams [45, Section 9.3, pp. 187–196] or in Kallenberg [121, in various pages].

These transformations applied to Poisson measure produce changes in the jump compensator. For instance, if $\nu(\mathrm{d}z, \mathrm{d}t)$ is an integer measure with jump compensator $\nu^p(\mathrm{d}z, \mathrm{d}t)$, then (continuous) rotation or dilation in the variable z on $\nu(\mathrm{d}z, \mathrm{d}t)$ (or time changes) are moved to the jump compensator $\nu^p(\mathrm{d}z, \mathrm{d}t)$. Thus, unless the jump compensator processes have some *homogeneity properties* in the variable (z, t), we may not expect to preserve Poisson measures. Therefore, if $\nu(\mathrm{d}z, \mathrm{d}t)$ is the integer measure corresponding to a typical Cauchy process, i.e., with Lévy kernel $\Pi(\mathrm{d}z, \mathrm{d}t) = |z|^{-m-1}dz\,\mathrm{d}t$ in $\mathbb{R}_*^m \times (0, \infty)$ (which is invariant under the operation $(z, t) \mapsto (\lambda z, \lambda t)$, for every $\lambda \neq 0$), and c is a process as above, but predictable, then the random measure defined for any compact subset K of $\mathbb{R}_*^m$ and $t \geq 0$ by the expression

$$\tilde{\nu}_c(K,]0, t]) := \int_0^t c(s)\tilde{\nu}(K, \mathrm{d}[\tau^{-1}(s)]),$$

with τ^{-1} as above, is indeed a martingale measure, which yields a Poisson measure ν_c with the same Lévy kernel $\Pi(\mathrm{d}z, \mathrm{d}t) = |z|^{-m-1}dz\,\mathrm{d}t$.

In general, a *random change of time* $(\tau(t) : t \geq 0)$ is a cad-lag (not necessarily continuous) increasing process such that each $\tau(t)$ is a stopping time relative

to the filtration $(\mathcal{F}(r) : r \geq 0)$, for every $t \geq 0$. The inverse τ^{-1} is defined as $\tau^{-1}(s) := \infty$ if $\tau(t) \leq s$ for every $t \geq 0$ and $\tau^{-1}(s) := \inf\{t \geq 0 : \tau(t) > s\}$ otherwise. Also, $\tau(\infty) := \lim_{t\to\infty} \tau(t)$ is set by definition. Certainly τ^{-1} is cad-lag, $\tau(\tau^{-1}(s)) \geq s$ and $\tau(t) := \inf\{s \geq 0 : \tau^{-1}(s) > t\}$, for every $s, t \geq 0$. The process τ^{-1} is strictly increasing continuous if τ is so, but τ^{-1} may not be continuous, whenever τ is only increasing and continuous. The family $(\mathcal{F}^\tau(t) : t \geq 0)$ with $\mathcal{F}^\tau(t) := \mathcal{F}(\tau(t))$,

the σ-algebra generated by the stopping time $\tau(t)$, is indeed a filtration, i.e., it is right-continuous and complete if $(\mathcal{F}(r) : r \geq 0)$ is so. Any adapted (or optional) process X is transformed into an adapted (or optional) process $X^\tau(t) := X(\tau(t))$, relative to the filtration $(\mathcal{F}^\tau(t) : t \geq 0)$. If X is a martingale then the Doob optional sampling theorem shows that $X^\tau(t) := X(\tau(t))$ is also a martingale relative to the filtration $(\mathcal{F}^\tau(t) : t \geq 0)$. Also the maximal martingale or p-estimates proves X^τ is p-integrable if X is so. It is simple to establish that if X has (locally and/or integrable) bounded variation then the same is valid for X^τ. A little more delicate is the fact that if X is a semi-martingale then so is X^τ. Given a purely discontinuous (cad-lag) local martingale X, a typical absolutely continuous change of times takes the form

$$\tau^{-1}(s) = \int_0^s c(r)\mathrm{d}r \quad \text{and} \quad \tau(t) = \inf\{s \geq 0 : \tau^{-1}(s) > t\},$$

where $\tau(t) = \infty$ if $\tau^{-1}(s) \leq t$ for every $s \geq 0$, and $c(s)$ is a (usually bounded) strictly positive predictable process. In this case, $s \mapsto \tau^{-1}(s)$ and $t \mapsto \tau(t)$ are both strictly increasing and continuous.

The reader is referred to the book Liptser and Shiryayev [158, Section 4.7, pp. 246–248]. It is clear that we cannot expect to preserve predictable processes when τ is not continuous. The change of time τ may introduce some extra discontinuities to the process. If τ is continuous and X is a continuous local martingale with predictable quadratic covariation matrix process $\langle X_i, X_j\rangle(t)$ then X^τ is also a continuous local martingale with predictable quadratic covariation matrix process $\langle X_i^\tau, X_j^\tau\rangle(t) = \langle X_i, X_j\rangle(\tau(t))$. Similarly, if X is a purely discontinuous local martingale with jump compensator $\nu_X^p(K,]0, t])$ then X^τ is also a purely discontinuous local martingale with jump compensator $\nu_{X^\tau}^p(K,]\tau(0), \tau(t)])$, for every $t \geq 0$ and any compact subset K of $\mathbb{R}_*^d$.

What follows (until the end of this subsection) should be taken as a *heuristic* subsection, which is not completely understood (or at least, not in way needed in this manuscript) and certain assertions should be taken as incomplete or unclear. Perhaps, this will be clarified later, see Barndorff-Nielsen and Shiryaev [10] for a comprehensive discussion.

Now consider change of times in an integer-valued measure (see Definition 4.6) ν on $\mathbb{R}_*^d \times [0, \infty)$, i.e., besides ν being a measure for every ω, the random process $t \mapsto \nu(K \times [0, t[)$ takes only integer values and is locally integrable, for any compact set K of $\mathbb{R}_*^d$ and any $t > 0$. Its jump compensator ν^p (or dual predictable projection, see Definition 3.5) is the random measure satisfying $\mathbb{E}\{\nu(K \times [0, t \wedge \theta) - \nu^p(K \times [0, t \wedge \theta])\} = 0$ or equivalently $t \mapsto \nu(K\times]0, t]) - \nu^p(K\times]0, t])$ is a martingale, for every compact K of $\mathbb{R}_*^d$,

any $t > 0$, and any finite stopping time θ, all this relative to the filtration $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$, generated by the family of random variables $\nu(K\times]0,t])$.

It is clear that $h : \mathbb{R}^d_* \to \mathbb{R}^n_*$ is a Borel measurable function such that the closure of $h^{-1}(K)$ is a compact set in $\mathbb{R}^n_*$, for every compact K in $\mathbb{R}^d_*$), then $\nu_h(K\times]0,t]) = \nu(h^{-1}(K)\times]0,t])$ defines an integer ν_h measure with jump compensator $\nu^p_h(K\times]0,t]) = \nu^p(h^{-1}(K)\times]0,t])$. Actually, we may take h to be a random variable independent of ν, or even to be a predictable process with respect to ν, i.e., $\nu^p_h(K\times]0,t]) = \nu^p(\{(z,s) : (h(z,s),s) \in K\times]0,t]\})$ is a predictable process for any compact set K of $\mathbb{R}^n_*$ and any $t > 0$.

Continuing with the discussed above, a change of times is more delicate. Indeed, let $\tau(z,t,\omega)$ be Borel measurable random variable, (z,t) in $\mathbb{R}^d_* \times [0,\infty]$, with values in $[0,\infty]$, such that (a) $t \mapsto \tau(z,t)$ is a non-decreasing cad-lag process for every z and (b) $\tau(z,t)$ is a stopping time for every (z,t). Therefore, the random mapping $\tau\colon (z,t) \mapsto \big(z,\tau(z,t)\big)$ from $\mathbb{R}^d_* \times [0,\infty]$ into itself transforms an integer measure ν on into another integer measure

$$\nu_\tau(K\times]0,s]) = \nu\big(\{(z,t) : z \in K,\, 0 < \tau(z,t) \leq s\}\big), \quad \forall s \in (0,\infty)$$

with jumps compensator given by

$$\nu^p_\tau(K\times]0,s]) = \nu^p\big(\{(z,t) : z \in K,\, 0 < \tau(z,t) \leq s\}\big), \quad \forall s \in (0,\infty).$$

Certainly, both transformations (h in $\mathbb{R}^d$ and τ in $[0,\infty)$) can be combined, i.e., considering the random mapping $\vartheta : (z,t) \mapsto \big(h(z,t),\tau(z,t)\big)$ the integer measure becomes $\nu_{h,\tau}(K\times[0,s[) = \nu\big(\vartheta^{-1}(K\times[0,s[)\big)$, with jumps compensator given by $\nu^p_{h,\tau}(K\times[0,s[) = \nu^p\big(\vartheta^{-1}(K\times[0,s[)\big)$, Moreover, if ν^p is absolutely continuous in t and can be expressed as

$$\nu^p(K\times[0,t[) = \int_K \pi(\mathrm{d}z)\int_0^t m(z,r)\mathrm{d}r, \quad m(z,r) \geq 0,$$

and the inverse time $\tau^{-1}(z,s)$ has the representation

$$\tau^{-1}(z,s) = \int_0^s c(z,r)\mathrm{d}r \;\text{ and }\; \tau(z,t) = \inf\{s \geq 0 : \tau^{-1}(z,s) > t\}, \quad (5.147)$$

with $\tau(z,t) = +\infty$ if $\tau^{-1}(z,s) \leq t$ for every $s \geq 0$, and some nonnegative locally integrable $\mathbb{F}$-predictable process c, then

$$\nu^p_\tau(K\times[0,s[) = \int_K \pi(\mathrm{d}z)\int_0^s c(z,r)m(z,r)\mathrm{d}r,$$

i.e., $t \mapsto \nu_\tau(K\times]0,s]) - \nu^p_\tau(K\times]0,s])$ is a martingale, for every compact K of $\mathbb{R}^d_*$, any $s > 0$, all this relative to the filtration $\mathbb{F}_\tau = \{\mathcal{F}_\tau(s) : s \geq 0\}$, generated by the family of random variables $\nu_\tau(K\times]0,s])$.

Actually, if $\mathbb{G} = \{\mathcal{G}(t) : t \geq 0\}$ is another filtration independent of $\mathbb{F}$ then the initial integer measure ν could be considered with respect to the filtration $\mathbb{G} \vee \mathbb{F}$, i.e., generated by the family of random variables $\nu(K\times]0,t])$ and the σ-algebras $\mathcal{G}(t)$, $t \geq 0$. In this case, the process c is $\mathbb{G} \vee \mathbb{F}$-predictable and the

process $t \mapsto \nu_\tau(K\times]0,s]) - \nu^p_\tau(K\times]0,s])$ is a martingale relative to $\mathbb{G} \vee \mathbb{F}_\tau$, i.e., generated by the family of random variables $\nu_\tau(K\times]0,s])$ and the σ-algebras $\mathcal{G}(s)$. In particular, if the density is independent of z, i.e., $c = c(s)$, then the filtration $\mathbb{F}_\tau = \{\mathcal{F}_\tau(t) : t \geq 0\}$ generated image integer measure ν_τ satisfies $\mathcal{F}_\tau(t) = \mathcal{F}(\tau(t))$, where $\mathtt{F} = \{\mathcal{F}(t) : t \geq 0\}$ is the filtration generated by the initial integer measure ν. Note that the existence of a jump compensator is used to obtain ν^p and the uniqueness of the jump compensator is necessary when ν^p_τ is identified as $\nu^p_\tau\big((K \times [0,s[)\big) = \nu^p\big(\vartheta^{-1}(K \times [0,s[)\big)$. Moreover, if

$$\tau^{-1}_\varepsilon(s) = \nu\big(\{(z,t) : |z| \geq \varepsilon,\ 0 < \tau(z,t) \leq s\}\big),$$
$$\text{and its inverse} \quad \tau_\varepsilon(t) = \inf\{s \geq 0 : \tau^{-1}_\varepsilon(s) > t\},$$

and if $f(z,t)$ and $c(z,s)$ with $0 \leq c(z,s) \leq 1$ are predictable processes with respect to the filtration $\mathbb{G} \vee \mathbb{F}$ then the stochastic integral

$$\int_{\{|z|\geq\varepsilon\}\times]0,\tau_\varepsilon(t)]} f(z,\tau^{-1}(z,s))\,\nu_\tau(\mathrm{d}z,\mathrm{d}s)$$

is defined and

$$\mathbb{E}\Big\{\Big|\int_{\{|z|\geq\varepsilon\}\times]0,\tau_\varepsilon(t)]} f(z,\tau^{-1}(z,s))\nu_\tau(\mathrm{d}z,\mathrm{d}t)\Big|^2\Big\} = \\ = \mathbb{E}\Big\{\int_{\{|z|\geq\varepsilon\}\times]0,t]} |f(z,t)|^2 c(z,t)\nu^p(\mathrm{d}z,\mathrm{d}t)\Big\},$$

as expected, and the argument is completed as $\varepsilon \to 0$.

In particular, if ν is a homogeneous Poisson measure (see Definition 4.9 with intensity measure $\pi(\mathrm{d}z)\mathrm{d}t$, i.e., ν^p is the deterministic Radon measure $\pi(\mathrm{d}z)\mathrm{d}t$ on $\mathbb{R}^d_* \times [0,\infty)$, and the transformation (5.147) is used with a nonnegative bounded Borel (deterministic) function $c(z,r)$, then the integer measure ν_τ defined as above is indeed a Poisson measure with intensity measure $\Pi_\tau(\mathrm{d}z,\mathrm{d}t) = c(z,t)\pi(\mathrm{d}z)\mathrm{d}t$. What is desired is to use the parameters $c(z,t,x)$ and π to construct a Poisson measure $\Pi(c,\mathrm{d}z,\mathrm{d}t)$ with intensity measure $c(z,t,x)\pi(\mathrm{d}z)\mathrm{d}t$ such that x can be replaced by a predictable process $x = y(t)$ later on.

Moreover, if π is a finite measure then the integer measure ν corresponds to a composed Poisson process. For instance, let $\{\zeta_n,\theta_n\}$ be a sequence of iid random variables such that ζ_n is $\mathbb{R}^d_*$-valued with distribution $\pi/\pi(\mathbb{R}^d_*)$ and θ_n has exponential distribution with mean $\mathbb{E}\{\theta_n\} = \pi(\mathbb{R}^d_*)$. Define $p(t) = \sum_n \zeta_n \mathbb{1}_{\{\sum_{i\leq n}\theta_i \leq t\}}$. If $c(z)$ is a strictly positive π-integrable function and the change of times is given by (5.147) then

$$p_\tau(t) = \sum_n \zeta_n \mathbb{1}_{\{\tau(\zeta_n,\sum_{i\leq n}\theta_i)\leq t\}} = \sum_n \zeta_n \mathbb{1}_{\{\sum_{i\leq n}\theta_i \leq \tau^{-1}(\zeta_n,t)\}}$$

is another composed Poisson process with respect to the measure $c(z)\pi(\mathrm{d}z)$. This is also valid if the density function c is only nonnegative, but the second expression of $p_\tau(t)$ with $\tau^{-1}(\zeta_n,t)$ may not hold true.

Regarding Lévy processes, consider a purely discontinuous Lévy process, i.e., a compensated Poisson process $\tilde{p} = \{\tilde{p}(t) : t \geq 0\}$ with valued in $\mathbb{R}^d$,

or equivalently a standard Poisson measure p in $\mathbb{R}^d_*$ with Lévy measure π on $\mathbb{R}^d_*$, $\pi(\mathrm{d}z)\,\mathrm{d}t = \mathbb{E}\{p(\mathrm{d}z,\mathrm{d}t)\}$, and martingale measure $\tilde{p}(\mathrm{d}z,\mathrm{d}t) = p(\mathrm{d}z,\mathrm{d}t) - \pi(\mathrm{d}z)\,\mathrm{d}t$, i.e., the Lévy process is defined by means of the stochastic integral

$$\tilde{p}(t) = \int_{\mathbb{R}^d_* \times]0,t]} z\tilde{p}(\mathrm{d}z,\mathrm{d}t).$$

Given a bounded predictable process $c(z,t) \geq 0$, indexed by z in $\mathbb{R}^d_*$, define the change of times

$$\tau^{-1}(z,s) = \int_0^s c(z,r)\mathrm{d}r \quad \text{and} \quad \tau(z,t) = \inf\{s \geq 0 : \tau^{-1}(z,s) > t\}.$$

where $\tau(z,t) = \infty$ if $\tau^{-1}(z,s) \leq t$ for every $s \geq 0$, as in (5.147). Therefore, the transformed purely discontinuous martingale or compensated Poisson process $\tilde{p}_\tau = \{\tilde{p}_\tau(t) : t \geq 0\}$ should be expressed in term of the Poisson martingale measure $\tilde{p}(\mathrm{d}z,\mathrm{d}t)$, i.e.,

$$\tilde{p}_\tau(t) = \int_{\mathbb{R}^d_* \times]0,\infty)} \mathbb{1}_{\{\tau(z,r) \leq t\}} \tilde{p}(\mathrm{d}z,\mathrm{d}r),$$

and not in term of the compensated Poisson process $\{\tilde{p}(t) : t \geq 0\}$. This may look non-anticipative,

$$\tilde{p}_\tau(t) = \int_{]0,\infty)} \mathbb{1}_{\{\tau(\tilde{p}(r)-\tilde{p}(r-),r) \leq t\}} \tilde{p}(\mathrm{d}r),$$

i.e., the stochastic integral cannot be defined? If a filtration $\mathbb{F}$ is attached to the compensated Poisson process $\tilde{p} = \{\tilde{p}(t) : t \geq 0\}$ then the new filtration $\mathbb{F}_\tau$ induced by the change of time τ has the σ-algebras $\mathcal{F}_\tau(t)$, which are generated by $\mathcal{F}\big(\tau(z,t)\big)$

To end this heuristic discussion and go back to actual assertions, if we begin with a Lévy process p with characteristic $(0, I, \pi)$ in $\mathbb{R}^d$ then the continuous part of p is a Wiener process w and the jumps yield a Poisson measure ν with Lévy measure π (both in $\mathbb{R}^d$), all relative to the filtration generated by p and satisfying the usual conditions, where the Wiener processes and the Poisson measures result independent. Equivalently, if we begin with a Wiener process w and a Poisson measure ν with Lévy measure π, both in $\mathbb{R}^d$ and independent of each other, then we can construct a Lévy process p (and the filtration satisfying the usual conditions) with characteristic $(0, I, \pi)$. Based on these arguments, we can apply the above random change of time to a Lévy process. The reader is referred to the book Barndorff-Nielsen and Shiryaev [10] for a comprehensive discussion.

5.5 Weak and Strong Solutions

The assumptions of Theorem 5.3 are rather neat, but sometime weaker conditions are needed. Note that in the previous sections we have assumed a fixed and

given Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, with Lévy measure $\pi(\cdot)$, i.e., a complete filtered probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$, the stochastic process $(w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener space and $(\nu(B,]0, t]) : B \in \mathbb{R}^m_*,\ t \geq 0)$ is an independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{\nu(B,]0, t])\}/t$, which satisfies

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1 + |\zeta|} \pi(\mathrm{d}\zeta) < \infty,$$

with martingale measure $\tilde{\nu}(B,]0, t]) := \nu(B,]0, t]) - t\pi(B)$. This martingale measure $\tilde{\nu}$ is identified with the $\mathbb{R}^m$-valued (Poisson) compensated-jump process

$$\tilde{p}(t) := \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0,$$

in the sense that given the Poisson integer measure ν we obtain the Poisson martingale measure $\tilde{\nu}$, which yields the Poisson compensated-jump process $\tilde{p}$, and conversely, starting from a Poisson compensated-jump process $\tilde{p}$ we may define a Poisson integer measure

$$\nu(B,]0, t]) = \sum_{0 < s \leq t} \mathbb{1}_{\{\tilde{p}(s) - \tilde{p}(s-) \in B\}},$$

which yields the Poisson martingale measure $\tilde{\nu}$.

• *Remark* 5.37. Using $\tilde{p}$ instead of $\tilde{\nu}$ in the setting of the stochastic ordinary differential equation correspond to taking a coefficient

$$\gamma_i(\zeta, t, x) = \sum_j \tilde{\gamma}_i(t, x)\zeta_j,$$

i.e., particular cases, but sufficiently general. □

A simple situation appears when the coefficients are time-independent, i.e., setting $x(t, x^0, t_0) = x(t)$, with initial condition $x(t_0) = x^0$, and differential

$$\mathrm{d}x(t) = g(x(t), v(t))\mathrm{d}t + \sigma(x(t), v(t))\mathrm{d}w(t) + \int_{\mathbb{R}^m_*} \gamma(\zeta, x(t), v(t))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t).$$

For the deterministic case, namely, when $\sigma = 0$ and $\gamma = 0$, it is clear that $x(\cdot, x^0, t_0) = x(\cdot - t_0, x^0, 0)$, but for the stochastic case, the equality holds only in law, i.e., in the same way that for a Wiener process $w(\cdot)$ we deduce that $w(\cdot + t_0) - w(t_0)$ is again a Wiener process, even if $w(\cdot) \neq w(\cdot + t_0) - w(t_0)$. Certainly, the statistics of $x(\cdot, x^0, t_0)$ and $x(\cdot - t_0, x^0, 0)$ are the same. A more delicate problem occurs when the coefficients are not so smooth, as we will see later. Thus, supposing that the Wiener-Poisson space is part of the unknown solution becomes essential.

If the filtration $\mathbb{F} = \{F_t : t \geq 0\}$ is generated by the processes w and $\tilde{p}$, i.e., $\mathbb{F}$ is the minimal completed filtration (right-continuous) such that w and $\tilde{p}$ are adapted, then

(i) we can always reduce to the canonical Wiener-Poisson space, and

(ii) the solution can be expressed as a deterministic functional of the processes w and $\tilde{p}$, i.e., $x = h(w, \tilde{p})$, almost surely, where h is a Borel measurable functions from $C([0,\infty[,\mathbb{R}^d) \times D([0,\infty[,\mathbb{R}^d)$ into $D([0,\infty[,\mathbb{R}^d)$, as seen later in details.

Actually, a key point is the fact that the given filtration $\mathbb{F} = \{F_t : t \geq 0\}$ is not necessarily the filtration generated by the processes w and $\tilde{p}$, so in general, to be able to immerse the problem in the canonical sample space $\Omega = C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)$ we need to consider all three processes w, $\tilde{p}$ and the solution x. This will be discuss in some detail later on this section, see also Section 4.3.5 of the previous Chapter.

The setting of the d-dimensional stochastic ordinary differential equation is slightly modified by including in the unknown elements not only the stochastic process $(x(t) : t \geq 0)$ but the whole $\mathbb{R}^n \times \mathbb{R}^m_*$ Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0),$$

and eventually the control process $(v(t) : t \geq 0)$. In the previous section, the Wiener-Poisson space was given a priori so that the solution process $(x(t) : t \geq 0)$ is adapted to the *initial* filtration $(\mathcal{F}_t : t \geq 0)$, i.e., the solution is found in term of *filtration* given on the the Wiener-Poisson space. Thus, we may use the canonical setting, i.e., the canonical sample space $C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)$ as the probability space (Ω, P), where $P = P_w \times P_{\tilde{\nu}}$ is the product of the Wiener and the Poisson measures as in (5.2) and (5.1). The projection maps $(\omega_1, \omega_2) \mapsto \big(\omega_1(t), \omega_2(t)\big)$, defined for every $t \geq 0$, from $C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)$ into $\mathbb{R}^n \times \mathbb{R}^m$, as $X_w(t,\omega) := \omega_1(t) = w(t,\omega)$ and $X_{\tilde{\nu}}(t,\omega) := \omega_2(t) = \tilde{p}(t,\omega)$, are a standard Wiener process $(w(t) : t \geq 0)$ and the compensated jump (Poisson) process

$$\tilde{p}(t) := \int_{\mathbb{R}^m_* \times]0,t]} \zeta \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad t \geq 0.$$

corresponding to a standard Poisson measure $(\nu(\mathrm{d}\zeta, \mathrm{d}s) : \zeta \in \mathbb{R}^m_*,\ s \geq 0)$, with martingale measure $\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s)$ and Lévy measure $\pi(\mathrm{d}\zeta)$ on $\mathbb{R}^m_*$. Clearly, the *minimal* filtration, i.e., the generated by (the null sets and) the process $(w, \tilde{p})$ can be used. In this context (and with coefficients independent of the control process), the solution process $x(\cdot)$ is written as a deterministic functional of the Wiener process and the Poisson measure, as discussed in detail later.

On the other hand, by letting the Wiener-Poisson space (and the control) be part of the unknown, we can enlarge the filtration (but keeping the martingale properties of the Wiener process w and the Poisson measure ν) so that the solution may not be a predictable functional on the canonical sample space. Sometime, a candidate for a solution is found and then the Wiener-Poisson process is constructed accordingly.

In stochastic control theory, the control process $(v(t) : t \geq 0)$ is a feedback of the state, i.e., $v(t) := k(t, x(t-))$, where $k(t,x)$ is only a (deterministic) Borel measurable function. Thus, the interest of this section is to study d-dimensional

stochastic ordinary differential equation of the form

$$\begin{cases} \mathrm{d}x(t) = g(t,x(t))\mathrm{d}t + \sum_{k=1}^{n} \sigma_k(t,x(t))\mathrm{d}w_k(t) + \\ \qquad\qquad + \int_{\mathbb{R}^m_*} \gamma(\zeta,t,x(t))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \end{cases} \tag{5.148}$$

without any Lipschitz condition in the x variable.

Definition 5.38 (weak or strong)**.** As we mention early, for a solution of the above SDE (5.148) corresponding to the coefficients g, σ, γ, with initial condition x_0 at t_0 and Lévy measure $\pi(\cdot)$, we mean an adapted cad-lag process $(x(t) : t \geq t_0)$ satisfying $x(t_0) = x^0$, and (5.148) for every $t \geq t_0$, which implies that the integrals

$$\int_{t_0}^{t} |g\big(s,x(s)\big)|\mathrm{d}s, \quad \int_{t_0}^{t} |\sigma\big(s,x(s)\big)|^2\mathrm{d}s, \quad \int_{t_0}^{t} \mathrm{d}s \int_{\mathbb{R}^m_*} |\gamma\big(\zeta,s,x(s)\big)|^2\pi(\mathrm{d}\zeta),$$

are almost surely finite, for any $t \geq t_0$.

(a) Then, by *weak existence* of a solution with initial distribution P_0 at t_0, we mean that for *some* Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, with Lévy measure $\pi(\cdot)$, there are a random variable x_0 with distribution P_0 and a solution $(x(t) : t \geq t_0)$ of the SDE (5.148) with initial condition $x(t_0) = x^0$. Thus a *weak solution* is a solution not necessarily adapted to the Wiener process and Poisson measure, in the sense that the filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ could be strictly larger than the canonical filtration generated by w and ν. In most cases, the weak solution is found in the canonical sample space $\mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)$, with the product probability $P_0 \times P$ and some filtration to which the solution process is adapted and the canonical Wiener process and Poisson measure are martingales.

(b) However, by *strong existence* of a solution with initial distribution P_0 at t_0, we mean that for *every* Wiener-Poisson space $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, with Lévy measure $m(\cdot)$ and any $\mathcal{F}_0$-measurable random variable x_0 with distribution P_0, there is a solution $(x(t) : t \geq t_0)$ of the SDE (5.148) with initial condition $x(t_0) = x^0$. Because the initial condition x^0 (with distribution P_0, at $t_0 = 0$ to simply notation) is independent of the Wiener process and the Poisson measure, a *strong solution* is a solution in the canonical sample space $\mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)$, with the product probability $P_0 \times P$, and the filtration $\mathbb{F}$ is the usually augmentation of the one generated by the initial condition x^0, the Wiener process w and the Poisson process, i.e., the strong solution is a functional of the data x^0, w and p.

(c) Thus, we say that the *uniqueness in law* holds whenever two solutions, in possible distinct Wiener-Poisson spaces but with the same Lévy measure and initial condition, have the same probability law, i.e., if x and y are two solutions then x and y induce the same probability law on the canonical space $D([t_0,\infty[,\mathbb{R}^d)$.

(d) On the other hand, we say the *pathwise uniqueness* holds whenever two solutions, in a common Wiener-Poisson space and with the same initial condition, are each one version of each other, i.e., if x and y are two solutions then $P\{x(t) = y(t)\} = 1$, for every $t \geq t_0$.

Sometimes, uniqueness in law and pathwise uniqueness are called *weak uniqueness* and *strong uniqueness*, respectively. □

The following Tanaka's example is used to clarify the two concepts of uniqueness. The one-dimensional stochastic ordinary differential equation

$$x(t) = \int_0^t \operatorname{sgn}(x(t))\mathrm{d}w(t), \quad t \geq 0,$$

where $\operatorname{sgn}(\cdot)$ is the sign function, has only uniqueness in law. Indeed, if (x, w) is a weak solution then x is a continuous square-integrable martingale with quadratic variation

$$\langle x(t)\rangle = \int_0^t \operatorname{sgn}^2(x(s))\mathrm{d}s = t.$$

Hence x is a standard Wiener process, and the weak uniqueness holds. However, the strong uniqueness can not hold since $(-x, w)$ is also a solution and x is not a version of $-x$. To make this clear, notice that if x is a standard Wiener process on a given filtered space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \geq 0)$ then defining

$$w(t) := \int_0^t \operatorname{sgn}(x(s))\mathrm{d}x(s), \quad \forall t \geq 0$$

we deduce that

$$x(t) = \int_0^t \operatorname{sgn}^2(x(s))\mathrm{d}x(s) = \int_0^t \operatorname{sgn}(x(s))\mathrm{d}w(s),$$

which shows the existence of a weak solution (x, w). Moreover there is no strong solution to this equation. Namely, if $x(t)$ were a strong solution, it should be adapted to the (completed) natural filtration of the driving Wiener w. Now, since x is a Wiener process, we can consider the Tanaka formula and write

$$w(t) = |x(t)| - \ell_t^x(0),$$

where $\ell_t^x(0)$ is the local time at 0 of the Wiener process x. This implies that w is adapted to the filtration of $|x|$. Hence, x is adapted to the filtration of $|x|$ which is a contradiction.

A key point in the study of weak solutions is that fact that pathwise uniqueness implies uniqueness in law. To simplify the notation, we take $t_0 = 0$ in most of this section.

Theorem 5.39 (Yamada-Watanabe)**.** *If the pathwise uniqueness holds for the stochastic ordinary differential equation* (5.148)*, as expressed by Definition 5.38, then uniqueness in law also holds.*

Proof. This is essentially based on disintegration of probability measures, or in other words, based on regular conditional probabilities. Indeed, a solution x of (5.148) in $[0,\infty[$ can be regarded as the element $(x^0, w, \tilde{p}, x - x^0)$ in the Polish space

$$\tilde{\Omega} := \mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m) \times D([0,\infty[,\mathbb{R}^d)$$

together with a probability measure (still denoted by) P, which is defined by the expression

$$\tilde{P}(A) := P\{(x^0, w, \tilde{p}, x - x^0) \in A\}, \quad \forall A \in \mathcal{B}(\tilde{\Omega})$$

in terms of the initial probability $P\{\cdot\}$. The marginal distribution in the variables $(x^0, w, \tilde{p})$ is $P_* = P_{x^0} \times P_w \times P_{\tilde{\nu}}$, see (5.1) and (5.2), where P_{x^0} is the distribution in $\mathbb{R}^d$ of the random variable x^0 used as the initial condition. The filtration $(\tilde{\mathcal{F}}(t) : t \geq 0)$ is generated by $(x^0, w, \tilde{p}, x - x^0))$, made right-continuous and completed with respect to the probability measure $\tilde{P}$.

Denote by $q = q(\xi, w, \tilde{p}, F)$ a regular conditional probability for Borel σ-algebra $\mathcal{B}\big(D([0,\infty[,\mathbb{R}^d)\big)$ given $(\xi, w, \tilde{p})$, i.e., a mapping q from

$$\mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m) \times \mathcal{B}\big(D([0,\infty[,\mathbb{R}^d)\big)$$

into $[0,1]$ such that:

(a) for each ξ in $\mathbb{R}^d$, w in $C([0,\infty[,\mathbb{R}^n)$ and $\tilde{p}$ in $D([0,\infty[,\mathbb{R}^m)$ the map $F \mapsto q(\xi, w, \tilde{p}, F)$ is a probability measure on $\mathcal{B}\big(D([0,\infty[,\mathbb{R}^d)\big)$,

(b) for each F in $\mathcal{B}\big(D([0,\infty[,\mathbb{R}^d)\big)$ the map $(\xi, w, \tilde{p}) \mapsto q(\xi, w, \tilde{p}, F)$ is measurable with respect to $\mathcal{B}\big(\mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)\big)$,

(c) for every B in $\mathcal{B}\big(\mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m)\big)$ we have

$$P(B \times F) = \int_B q(\xi, w, \tilde{p}, F) P_*(\mathrm{d}\xi, \mathrm{d}w, \mathrm{d}\tilde{p}),$$

for every F in $\mathcal{B}\big(D([0,\infty[,\mathbb{R}^d)\big)$.

At this point, if we have two weak solutions $(x_1(t) : t \geq t_0)$ and $(x_2(t) : t \geq t_0)$ with the same initial condition and probabilities P_1 and P_2, we can follow the above construction on $[t_0,\infty)$ to obtain two regular conditional probabilities $q_i = q_i(\xi, w, \tilde{p}, \mathrm{d}y^i)$, with $i = 1, 2$. Hence, on the Polish space

$$\Omega := \mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m) \times D([0,\infty[,\mathbb{R}^d) \times D([0,\infty[,\mathbb{R}^d)$$

we define the probability measure

$$P(\mathrm{d}\omega) := q_1(\xi, w, \tilde{p}, \mathrm{d}y_1)\, q_2(\xi, w, \tilde{p}, \mathrm{d}y_2) P_*(\mathrm{d}\xi, \mathrm{d}w, \mathrm{d}\tilde{p}),$$

where $\omega = (\xi, w, \tilde{p}, y_1, y_2)$ denotes a generic element in Ω. The filtration $(\mathcal{F}(t) : t \geq 0)$ is generated by the projection maps $(x^0, w(t), \tilde{p}(t), x_1(t) - x^0, x_2(t) - x^0)$, made right-continuous and completed with respect to the above probability

measure P. In view of the properties satisfied by the regular conditional we have

$$P\big((\xi, w, \tilde{p}, y_i) \in A\big) = P_i\{(x^0, w, \tilde{p}, x_i - x^0) \in A\}, \quad i = 1, 2,$$

for every A in $\mathcal{B}\big(\mathbb{R}^d \times C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m)\big)$. which shows that the distribution of $(\xi + y_i, w, \tilde{p})$ under P is the same as the initial distribution of $(x_i, w, \tilde{p})$ under P_i. Since the pathwise uniqueness means that

$$P\big(\xi + y_1(t) = \xi + y_2(t),\ \forall t \geq 0\big) = 1,$$

which can be written as

$$P(\omega = (\xi, w, \tilde{p}, y_1, y_2) \in \Omega \,:\, y_1 = y_2) = 1.$$

This proves that

$$\begin{aligned} P\big(\omega = (\xi, w, \tilde{p}, y_1, y_2) \in \Omega \,:\, (\xi, w, \tilde{p}, y_1) \in A\big) = \\ = P\big(\omega = (\xi, w, \tilde{p}, y_1, y_2) \in \Omega \,:\, (\xi, w, \tilde{p}, y_2) \in A\big), \end{aligned}$$

for every A in $\mathcal{B}\big(\mathbb{R}^d \times C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m) \times D([0,\infty[, \mathbb{R}^d))\big)$, i.e., the uniqueness in law holds true. □

The reader may check the book Rogers and Williams [214, Section V.3, pp. 140–157] for more detailed on the above proof.

Corollary 5.40. *The pathwise uniqueness and the weak existence imply the strong existence (and uniqueness).*

Proof. To be clear, set $S = C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m)$ and note that the diagonal

$$B = \big\{(y_1, y_2) \in D([0,\infty[, \mathbb{R}^d) \times D([0,\infty[, \mathbb{R}^d) \,:\, y_1 = y_2\big\}$$

is a Borel set. Now, we have proved above that

$$P(\mathbb{R}^d \times S \times B) = 1$$

and

$$\int_{\mathbb{R}^d \times S} P_*(\mathrm{d}\xi, \mathrm{d}w, \mathrm{d}\tilde{p}) \int_B q_1(\xi, w, \tilde{p}, \mathrm{d}y_1)\, q_2(\xi, w, \tilde{p}, \mathrm{d}y_2) = 1.$$

Hence

$$\int_B q_1(\xi, w, \tilde{p}, \mathrm{d}y_1)\, q_2(\xi, w, \tilde{p}, \mathrm{d}y_2) = 1$$

almost everywhere in $(\xi, w, \tilde{p})$ with respect to the measure $P_*(\mathrm{d}\xi, \mathrm{d}w, \mathrm{d}\tilde{p})$. Now, using the fact that a product measure with support on the diagonal is necessarily a Dirac measure, we deduce that there exists some function

$$y = k(\xi, w, \tilde{p})\colon \mathbb{R}^d \times S \to D([0,\infty[, \mathbb{R}^d)$$

such that

$$q_1(\xi, w, \tilde{p}, \{k(\xi, w, \tilde{p})\}) = q_2(\xi, w, \tilde{p}, \{k(\xi, w, \tilde{p})\}) = 1$$

almost everywhere with respect to $P_*(\mathrm{d}\xi, \mathrm{d}w, \mathrm{d}\tilde{p})$. Since the functions q_1 and q_2 are measurable, a monotone class argument yields the measurability of the function k with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d \times S)$. Moreover, this also proves that k is adapted. Note that this function h depends on the initial density P_0 of the initial condition x_0.

Hence, for any given weak solution (x, x_0, w, p), relative to the same initial distribution P_0, we consider the function

$$h\colon \mathbb{R}^d \times C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m) \to D([0,\infty[, \mathbb{R}^d),$$

defined by $h(\xi, w, \tilde{p}) = \xi + k(\xi, w, \tilde{p})$, which satisfies $x = h(x_0, w, \tilde{p})$. Now, we have also that starting from another probability base (x_0', w', p'), with x_0' distributed as x_0, the process $x' = h(x_0', w', \tilde{p}')$ verifies the equation, see Proposition 4.46. Hence there is strong existence of solution. □

Recall that the function h depends on the distribution P_0 of the initial condition and that the pathwise uniqueness holds under a linear growth and a locally Lipschitz assumptions like: there exists a constant $C > 0$ such that

$$x \cdot g(t,x) + |\sigma(t,x)|^2 + \int_{\mathbb{R}^m_*} |\gamma(\zeta, t, x)|^2 \pi(\mathrm{d}\zeta) \le C(1 + |x|^2), \tag{5.149}$$

for every (t, x) in $[t_0, \infty) \times \mathbb{R}^d$, and for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\begin{cases} (x - x') \cdot \big(g(t,x) - g(t,x')\big) + |\sigma(t,x) - \sigma(t,x')|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x) - \gamma(\zeta,t,x')|^2 \pi(\mathrm{d}\zeta) \le M|x - x'|^2, \end{cases} \tag{5.150}$$

and for some $q > d$ there exists a constant $M_q = M_q(r)$ such that

$$\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x) - \gamma(\zeta,t,x')|^q \pi(\mathrm{d}\zeta) \le M_q |x - x'|^q, \tag{5.151}$$

for every (t, x), (t, x') in $[t_0, \infty) \times \mathbb{R}^d$ with $t \le t_0 + r$, $|x| \le r$ and $|x'| \le r$.

Corollary 5.41. *Under the assumptions* (5.149) *and* (5.150) *there exists a non-anticipating functional*

$$F : \mathbb{R}^d \times C([0,\infty[, \mathbb{R}^n) \times D([0,\infty[, \mathbb{R}^m) \to D([0,\infty[, \mathbb{R}^d),$$

such that for any initial condition x_0 the expression $x = F(x_0, w, \tilde{p})$ yields the unique (strong) solution of the stochastic ordinary differential equation (5.148). *Moreover, if* (5.151) *also holds then the function F is continuous in the first variable.*

Proof. In this context, a (deterministic) non-anticipating functional means a Borel measurable function F with the property that the function $(\omega_1, \omega_2) \mapsto F(t, \omega_1, \omega_2)$ is measurable with respect to the σ-algebra generated by the random variables $X_s : (\omega_1, \omega_2) \mapsto (\omega_1(s), \omega_2(s))$ for $0 \le s \le t$, see Definition 4.44. Actually, the proposed non-anticipating functional F is the same constructed in the previous corollary under the name h. Moreover, by means of an truncation argument as in Theorem 5.3, we may assume that (5.150) and (5.151) are global, i.e., it is satisfied with a constant $M = M(r)$ independent of r.

First, we have to prove that h can be constructed independently of the initial distribution P_0. Indeed, let $F_0 = h$ be the above functional corresponding to a given initial distribution P_0, and denote by F_x the particular case where $P_0 = \delta_x$ is the delta distribution concentrated at x. A simple conditioning argument yields

$$P_0\big\{x \in \mathbb{R}^d : F_0(x, w, \tilde{p}) = F_x(x, w, \tilde{p})\big\} = 1,$$

which means that the function $F(x_0, w, \tilde{p}) = F_x(x, w, \tilde{p})\big|_{x=x_0}$ is the right candidate, i.e., $x = F(x_0, w, \tilde{p})$ almost surely.

Next, we should establish the (joint) measurability of F. To this purpose, we note that the polynomial estimate (5.54) of Theorem 5.11 yields an estimate of the type: there exist some positive constants α and β, such that for any $T > 0$ there is a constant $M = M_T > 0$ satisfying

$$\mathbb{E}\Big\{ \sup_{0\le t\le T} |x(t) - x'(t)|^\alpha \Big\} \le M\, |x_0 - x_0'|^{d+\beta},$$

for any initial conditions x_0 and x_0'. Hence, by means of this estimate, the Kolmogorov continuous extension technique can be applied (e.g., see Kunita[143, Section 1.4, pp. 31–38]) to get a continuous version of the process (or field) $x_0 \mapsto x$, i.e., $F(x_0, w, \tilde{p})$ can be redefined as a continuous function in the variable x_0, and hence F is jointly measurable.

Now, without the condition (5.151) the arguments are slightly different. Indeed, if we revise the proof of Theorem 5.3 then we realize that the fixed-point argument and Proposition 4.47 (in the previous Chapter) prove that the iteration $x^{n+1} = T(x^n)$ yields a non-anticipative functional, and so is the limit as $n \to \infty$, i.e., assumption (5.151) is not necessary to get a (joint measurable) non-anticipating functional. □

• *Remark* 5.42 (Carathéodory). In a measure space $(\Omega, \mathfrak{F}, \mu)$, consider a function $f\colon \mathbb{R}^d \times \Omega \to \mathbb{R}$ such that: (a) for any x the function $\omega \to f(x, \omega)$ is $\mathfrak{F}$-measurable, (b) for any ω the function $x \to f(x, \omega)$ is continuous. Then f is jointly measurable with respect to $\mathcal{B}^d \times \mathfrak{F}$. Indeed, to simplify the notation, assume $d = 1$ and, for a fixed n, define the functions q_n and f_n as follows:

$$q_n(x) = i/n \quad \text{if} \quad i/n \le x < (i+1)/n, \quad i = 0, \pm 1, \dots$$
$$f_n(x, \omega) = k/n \quad \text{if} \quad k/n \le f(q_n(x), \omega) < (k+1)/n, \quad i = 0, \pm 1, \dots,$$

which enjoy, for any x, the following properties: (1) $\omega \to f_n(x,\omega)$ is a simple function, (2) $f_n(x,\omega) = f_n(q_n(x),\omega)$, (3) $|f(q_n(x),\omega) - f_n(q_n(x),\omega)| \le 1/n$, (4) $|q_n(x) - x| \le 1/n$. Therefore, the function f_n is jointly measurable and, the continuity in x and the inequality

$$|f(x,\omega) - f_n(x,\omega)| \le |f(x,\omega) - f(q_n(x),\omega)| + |f(q_n(x),\omega) - f_n(x,\omega)|$$

imply that $f_n \to f$ pointwise, i.e., f is jointly measurable. On the other hand, if only (c) the function $x \to f(x,\omega)$ is continuous in measure, i.e., for any $\varepsilon > 0$ and for any sequence $x_n \to x$ we have $\mu\{|f(x_n,\omega) - f(x,\omega)| \ge \varepsilon\} \to 0$ as $n \to \infty$. Then the previous arguments show that $f_n(x,\cdot) \to f(x,\cdot)$ in measure, so by taken a subsequence, we may define the limit almost surely. This is, under the assumption (c), instead of (b), there is a version of f which is jointly measurable. This same argument is used to show that any adapted process which is continuous in probability has a progressively measurable version. □

For the Wiener case, the reader may find more details on the above proof in Roger and Williams [214, Theorem V.13.1, pp 136–138]. Clearly, to show the joint measurability it is not necessary to have continuity in x_0, i.e., the assumptions (5.149), (5.150) and (5.151) are not necessary. Indeed, in the case without jumps, the following result is found in Kallenberg [121, Theorem 21.14, pp 423–426]: Assume that weak existence and pathwise uniqueness hold for solutions starting at arbitrary fixed points. Then the strong existence and uniqueness in law hold for every initial distribution. Furthermore, there exists a Borel measurable and universally predictable function $h\colon \mathbb{R}^d \times C([0,\infty[,\mathbb{R}^n) \to C([0,\infty[,\mathbb{R}^d)$ such that any solution (x,w) satisfies $x = h(x(0),w)$ almost surely. Clearly, a result similar to the above seems to hold for the Wiener-Poisson case.

• *Remark* 5.43. Actually, the coefficients g, σ and γ may be considered predictable functional on the canonical space, see Section 4.3.5 of the previous Chapter. For some technical reasons, we consider the universally completed predictable σ-algebra, and hence, universally predictable functionals on the canonical sample space. Under these assumptions on the coefficients, the previous Theorem and Corollaries remain valid with obvious modifications on the conditions (5.149), (5.150) and (5.151). Furthermore, if the coefficients are predictable functional like $g(t,x,v)$, $\sigma(t,x,v)$ and $\gamma(\zeta,t,x,v)$, where v is a non-anticipative process relative to the canonical Wiener-Poisson process (see Definition 4.41) then Corollary 5.41 holds true with natural changes, the solution is a non-anticipating functional on w, $\tilde{p}$, v and the initial condition ξ, which may be regarded as another non-anticipative process, see Proposition 4.47. □

To emphasize this difference, let us discuss Tsirel'son example. Consider the equation

$$x(t) = \int_0^t b(s,x(\cdot))\,\mathrm{d}s + w(t) \tag{5.152}$$

where $b\colon [0,1] \times C([0,1],\mathbb{R}) \to [0,1]$ is given as follows: choose a decreasing sequence of times converging to 0

$$0 \leftarrow t_{-n} < \cdots < t_{-3} < t_{-2} < t_{-1} < t_0 = 1,$$

and defining

$$b(t,w) = f\Big(\frac{w(t_k) - w(t_{k-1})}{t_k - t_{k-1}}\Big), \qquad t_k \le t < t_{k+1}$$

with f the fractional part, i.e., $f(\xi) = \xi - [\xi]$, with $[\xi]$ being the integer part of a real number. Then, it is clear that by means of Girsanov theorem to reduce to the Wiener process, and so, there exists a unique weak solution to (5.152). Now, let us check that equation (5.152) has no strong existence of solution. Indeed, let us suppose that the solution x is adapted to the Wiener process and we will get a contradiction. First, from the equation (5.152) we can write

$$x_{k+1} = w_{k+1} + f(x_k)$$

where

$$x_k = \frac{x(t_k) - x(t_{k-1})}{t_k - t_{k-1}}, \qquad w_k = \frac{w(t_k) - w(t_{k-1})}{t_k - t_{k-1}}.$$

The particular definition of the function f implies that $e^{2\pi \mathrm{i} f(x)} = e^{2\pi \mathrm{i} x}$. Hence we can write, inductively,

$$e^{2n\pi \mathrm{i} x_{k+1}} = e^{2n\pi \mathrm{i} w_{k+1}} e^{2n\pi \mathrm{i} w_k} \dots e^{2n\pi \mathrm{i} w_{k-j+1}} e^{2n\pi \mathrm{i} x_{k-j}}.$$

Since the solution x is adapted to the Wiener process w, the random variable x_{k+1} must be independent of $x_k, x_{k-1}, \dots, w_k, w_{k-1}, \dots$, and we deduce $\mathbb{E}\{e^{2\pi \mathrm{i} x_{k+1}}\} = 0$. Moreover, for any j,

$$\mathbb{E}(e^{2\pi \mathrm{i} x_{k+1}} \mid \mathcal{B}_j^{k+1}) = e^{2\pi \mathrm{i} w_{k+1}} e^{2\pi \mathrm{i} w_k} \cdots e^{2\pi \mathrm{i} w_{k-j+1}} \mathbb{E}(e^{2\pi \mathrm{i} x_{k-j}}) = 0,$$

where $\mathcal{B}_j^{k+1}$ is the σ-algebra generated by all increments $w(t) - w(s)$ with $t_{k-j} \le s < t \le t_{k+1}$, and therefore $\mathcal{F}_{t_{k+1}} = \bigvee_j \mathcal{B}_j^{k+1}$. Then

$$e^{2\pi \mathrm{i} f(x_{k+1})} = \mathbb{E}(e^{2\pi \mathrm{i} x_{k+1}} \mid \mathcal{F}_{t_{k+1}}) = \lim_{j \to \infty} \mathbb{E}(e^{2n\pi \mathrm{i} x_{k+1}} \mid \mathcal{B}_j^{k+1}) = 0,$$

is uniformly distributed in $[0,1)$, which is a contradiction. The reader may consult the book Rogers and Williams [214, Section V.3.18, pp. 155–157] for more comments on the above example.

• *Remark* 5.44. In the above example, the solution $x(t)$ is *not* adapted to the Wiener process w. Compare this case with the simpler equation

$$x(t) = \int_0^t b(x(s))\, \mathrm{d}s + w(t).$$

where b is only bounded and measurable, and therefore there exists a strong solution, e.g., see Karatzas and Shreve [124, Proposition 5.17, pp. 341–342]. □

Returning to the control process, we can be regarded v as predictable functional, but for simplicity, it is preferred the above setting with a specific control v. Thus, to include the control process $(v(t) : t \ge 0)$ in the weak formulation we need to fix the probability space. To be more specific, the data are:

(a) the (deterministic) functions $g(t,x,v)$, $\sigma(t,x,v)$, $\gamma(\zeta,t,x,v)$ and a Lévy measure $\pi(\cdot)$ as in the previous section,

(b) the canonical (pre-) filtered (Polish) space

$$\Omega := C([0,\infty[,\mathbb{R}^n) \times D([0,\infty[,\mathbb{R}^m), \quad \mathcal{B}(\Omega), \quad (\mathcal{F}^o(t) : t \geq 0),$$

where $\mathcal{F}^o(t)$ is the intersection in $\varepsilon > 0$ of the σ-algebra generated by the projection maps $X_w(s,\omega_1,\omega_2) := \omega_1(s)$ and $X_{\tilde{\nu}}(s,\omega_1,\omega_2) := \omega_2(s)$ for $0 \leq s \leq t+\varepsilon$, with $\omega = (\omega_1,\omega_2)$, ω_1 in $C([0,\infty[,\mathbb{R}^n)$ and ω_2 in $D([0,\infty[,\mathbb{R}^m)$. The canonical filtration $(\mathcal{F}^o(t) : t \geq 0)$ is right-continuous, but it will be completed as soon as a probability measure is defined on the Polish space Ω.

(c) a distribution P_0 on $\mathbb{R}^d$ and a $\mathbb{R}^q$-valued adapted process $(v(t) : t \geq 0)$.

And the unknowns are:

(1) a probability $P = P_w \times P_{\tilde{\nu}}$ on $\big(\Omega, \mathcal{B}(\Omega)\big)$ such that P_w is the standard Wiener measure, i.e., (5.2) holds, and $P_{\tilde{\nu}}$ is the standard compensated Poisson measure with Lévy measure $\pi(\cdot)$, i.e., (5.1) holds,

(2) a $\mathbb{R}^d$-valued random variable x^0 with the given distribution P_0 and a $\mathbb{R}^d$-valued adapted cad-lag process $(x(t) : t \geq 0)$ on $(\Omega, \mathcal{B}(\Omega), P, \mathcal{F}(t) : t \geq 0)$ such that the equation (5.4) holds, where the filtration $(\mathcal{F}(t) : t \geq 0)$ is the completion of $(\mathcal{F}^o(t) : t \geq 0)$ with respect to null sets in $\big(P, \mathcal{B}(\Omega)\big)$.

Due to some technical issues, sometime we allow the control process $(v(t) : t \geq 0)$ to be universally predictable, the filtration $\{\mathcal{F}^o(t) : t \geq 0\}$ are universally completed to a new filtration $\{\mathcal{F}^u(t) : t \geq 0\}$ to which the process is assumed to be predictable.

Clearly, uniqueness in law implies that the law in $D([0,\infty[,\mathbb{R}^d)$ of the stochastic process $(x(t) : t \geq 0)$ is unique. With this setting we do have the key fact that pathwise uniqueness implies uniqueness in law.

It is interesting to realize that identifying the concept of a $\mathbb{R}^d$-valued cad-lag stochastic process with a $D([0,\infty[,\mathbb{R}^d)$-valued random variable, we can rephrase the meaning of a weak solution as follows. The data are (a) as above, and instead of (b) and (c) we require

(b') a distribution P_0 on $\mathbb{R}^d$ and a measurable functional v on $D([0,\infty[,\mathbb{R}^q)$ such that the $\mathbb{R}^q$-valued function $\omega(\cdot) \mapsto v(\omega(t))$, where $\omega(\cdot)$ denote a generic element in $D([0,\infty[,\mathbb{R}^q)$, is measurable with respect to $\mathcal{F}^o(t)$, defined as above, for every $t \geq 0$.

The unknowns are:

(1) a (standard) $n \times m$ Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*, \ t \geq 0)$$

with Lévy measure $\pi(\cdot)$,

(2) a $\mathbb{R}^d$-valued random variable x^0 with distribution P_0 and a $D([0,\infty[,\mathbb{R}^d)$-valued random variable x such that $\omega \mapsto x(t,\omega)$ is measurable with respect to $\mathcal{F}_t$, for every $t \geq 0$, and finally, the equation (5.4) holds.

Another important point to remark is the fact that as long as pathwise uniqueness holds, there is not differences between strong and weak solution.

Global existence of weak solutions is only discussed under (Borel measurability is implicitly assumed) the growth condition (5.8), i.e., linear growth in the variable x for the diffusion term $\sigma(t,x,v)$, the jumps term $\gamma(\zeta,t,x,v)$ and the non-monotone part of the drift term $g(t,x,v)$. This allows any growth for the drift in variable x as long as it is in a "good" direction. The moment estimates (5.52) and (5.53) are then valid.

Girsanov transformation is a good tool to establish the existence of a weak solution for measurable drift coefficients. Indeed, let M be a (real-valued) continuous local martingale in a filtered probability space

$$(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0).$$

Assume that the process

$$Z(t) := \exp\Big(M(t) - \frac{\langle M\rangle(t)}{2}\Big),$$

is a martingale, where $\langle\cdot\rangle$ denotes the optional quadratic variation. Since Z is always a local martingale and a sub-martingale this means that we assume $\mathbb{E}\{Z(t)\} = 1$ for every $t \geq 0$, which is true if

$$\mathbb{E}\Big\{\exp\Big(\frac{\langle M\rangle(t)}{2}\Big)\Big\} < \infty, \quad \forall t \geq 0,$$

in view of Novikov sufficient condition, see Section 5.4 for more details and references. If Ω is a good probability space (e.g., a Polish space as the canonical space D of cad-lag functions) then there exits a probability $\tilde{P}$ on Ω such that for each $t \geq 0$ we have

$$\tilde{P}(A) = \int_A Z(t) P(\mathrm{d}\omega), \quad \forall A \in \mathcal{F}(t),$$

in short we say either $\tilde{P}(\mathrm{d}\omega) = ZP(\mathrm{d}\omega)$ or $\tilde{P}(\mathrm{d}\omega)|\mathcal{F}(t) = Z(t)P(\mathrm{d}\omega)$, for every $t \geq 0$. The measure P and $\tilde{P}$ are absolutely continuous one to each other. This is know as Cameron-Martin change of measure or Girsanov transformation, which affirms that for any continuous local martingale X in $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$ the process

$$\tilde{X}(t) := X(t) - \langle X, M\rangle(t), \quad \forall t \geq 0$$

results a continuous local martingale in $(\Omega, \mathcal{F}, \tilde{P}, \mathcal{F}(t) : t \geq 0)$, where $\langle\cdot,\cdot\rangle$ is the optional quadratic covariation of local martingales in $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq 0)$.

To make use of the above result, let $(x(t) : t \geq t_0)$ be a solution of the previous d-dimensional stochastic ordinary differential equation (5.5) with a

initial condition at t_0 and let $b(t,x,v)$ be a $\mathbb{R}^n$-valued measurable function. Define the (density or exponential martingale) process

$$Z(t) := \exp\Big(\sum_k \int_{t_0}^t b_k(s,x(s),v(s))\mathrm{d}w_k(s) - \\ -\frac{1}{2}\sum_k \int_{t_0}^t |b_k(s,x(s),v(s))|^2\mathrm{d}s\Big),$$

for every $t \geq 0$ under the assumption that

$$\mathbb{E}\Big\{\exp\Big(\frac{1}{2}\sum_k \int_{t_0}^t |b_k(s,x(s),v(s))|^2\mathrm{d}s\Big)\Big\} < \infty, \quad \forall t \geq 0.$$

which is clearly satisfied if b is bounded. Calculating the optional quadratic covariation of continuous martingale part of the solution $(x(t) : t \geq 0)$ and the Wiener process as follow

$$y_i(t) := \sum_k \int_{t_0}^t \sigma_{ik}(s,x(s),v(s))\mathrm{d}w_k(s),$$
$$\langle y_i, w_k\rangle(t) = \int_{t_0}^t \sigma_{ik}(s,x(s),v(s))\, b_k(s,x(s),v(s))\mathrm{d}s,$$

the Girsanov transformation shows that $(x(t) : t \geq 0)$ is a solution of the following d-dimensional stochastic ordinary differential equation

$$\begin{cases} \mathrm{d}x(t) = \tilde{g}(t,x(t),v(t))\mathrm{d}t + \sigma(t,y(t),v(t))\mathrm{d}w(t) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_*} \gamma(\zeta,t,y(t),v(t))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \end{cases} \tag{5.153}$$

where $\tilde{g}_i := g_i + \sum_k \sigma_{ik}\, b_k$, in the filtered probability space $(\Omega, \mathcal{F}, \tilde{P}, \mathcal{F}(t) : t \geq 0)$. Since we can do the inverse change of variables, we deduce that both stochastic differential equation (5.5) and (5.153) are equivalent as long as we change the probability measure. Hence, we have existence and uniqueness for any drift of the form of $\tilde{g}$, where b is measurable and bounded, and g, σ and γ satisfy the conditions of Theorem 5.3. Moreover, we can add a monotone term to the drift coefficient g.

For instance, the reader may consult the books Ikeda and Watanabe [110, Chapter IV, pp. 159–246] and Karatzas and Shreve [124, Chapter 5, pp. 281–398] for a more detailed treatment.

5.6 Martingale Problems

After discussing weak solutions we realize that a solution is indeed a probability measure in the canonical space satisfying a certain number of conditions which

involve the given data. In particular, consider the following integro-differential operator

$$\begin{cases} A(t,v)\varphi(x) = \displaystyle\sum_{i,j=1}^{d} a_{ij}(t,x,v)\partial_{ij}\varphi(x) + \sum_{i=1}^{d} g_i(t,x,v)\partial_i\varphi(x) + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,t,x,v), \end{cases} \tag{5.154}$$

where ∂_i and ∂_{ij} denote the partial first x_i and second x_i, x_j derivatives and ∇ is the gradient in the variable x,

$$a_{ij}(t,x,v) := \frac{1}{2}\sum_{k=1}^{n} \sigma_{ik}(t,x,v)\,\sigma_{jk}(t,x,v), \tag{5.155}$$

and

$$\mathtt{M}(B,t,x,v) = \pi(\{\zeta : \gamma(\zeta,t,x,v)\in B\}), \quad \forall B\in\mathcal{B}(\mathbb{R}^d_*) \tag{5.156}$$

is the jumps or Lévy kernel in $\mathbb{R}^d_*$.

The proper integro-differential part of A, i.e., the last term or integral term in expression (5.154), can be rewritten as

$$\begin{cases} I(t,v)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} [\varphi(x+\gamma(\zeta,t,x,v)) - \varphi(x) - \\ \qquad -\gamma(\zeta,t,x,v)\cdot\nabla\varphi(x)]\pi(\mathrm{d}\zeta), \end{cases} \tag{5.157}$$

or, after using Taylor formula with $\gamma(\zeta) := \gamma(\zeta,t,x,v)$,

$$\begin{aligned} I(t,v)\varphi(x) &= \int_0^1 \mathrm{d}\theta \int_{\mathbb{R}^m_*} \gamma(\zeta)\cdot[\nabla\varphi(x+\theta\gamma(\zeta)) - \nabla\varphi(x)]\pi(\mathrm{d}\zeta) = \\ &= \int_0^1 (1-\theta)\mathrm{d}\theta \int_{\mathbb{R}^m_*} \gamma(\zeta)\cdot\big[\nabla^2\varphi\big(x+\theta\gamma(\zeta)\big)\gamma(\zeta)\big]\pi(\mathrm{d}\zeta), \end{aligned}$$

as long as φ is smooth. Moreover, for any function $\psi(z)$ satisfying $\psi(z) = z$ in a neighborhood of $z = 0$ and with a compact support in $\mathbb{R}^m$, we see that a term of the form

$$\int_{\mathbb{R}^m_*} \big[\gamma(\zeta,t,x,v) - \psi\big(\gamma(\zeta,t,x,v)\big)\big]\pi(\mathrm{d}\zeta)\cdot\nabla\varphi(x),$$

can be regarded as part of the drift coefficient g. Therefore, the proper integro-differential part I may take the form

$$\begin{cases} I(t,v)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} \big[\varphi(x+\gamma(\zeta,t,x,v)) - \varphi(x)]\pi(\mathrm{d}\zeta) - \\ \qquad -\psi\big(\gamma(\zeta,t,x,v)\big)\cdot\nabla\varphi(x)\big]\pi(\mathrm{d}\zeta), \end{cases} \tag{5.158}$$

whenever necessary. This proves that if the coefficients g, σ and γ are locally bounded, in particular if the growth condition (5.8) is satisfied then $A\varphi$ is

bounded (in t, x and v) for every twice-continuously differentiable function φ with a compact support in $\mathbb{R}^d$.

It is clear that the operator A is acting on the space variable x for smooth functions φ (e.g., C^2 with bounded second derivatives) and holds t and v as parameters. Itô formula shows that for any solution $(x(t) : t \geq t_0)$ on the canonical Wiener-Poisson space

$$(\Omega, \mathcal{F}, \tilde{P}, \mathcal{F}(t), w(t), \tilde{\nu}(\mathrm{d}t, \mathrm{d}\zeta) : t \geq 0,\, \zeta \in \mathbb{R}^m_*),$$
$$\Omega = C([0, \infty[, \mathbb{R}^n) \times D([0, \infty[, \mathbb{R}^m),$$

with Lévy measure $\pi(\cdot)$, the stochastic process for $t \geq 0$,

$$M_\varphi(t) := \varphi(x(t)) - \varphi(x(t_0)) - \int_{t_0}^t A(s, v(s))\varphi(x(s))\mathrm{d}s, \tag{5.159}$$

is a stochastic integral, and so a real-valued (quasi-left continuous, bounded, local) martingale for any smooth function φ with a compact support. The reader may check the books Stroock and Varadhan [241], where the so-called *martingale problem* was introduced for diffusion processes or Ethier and Kurtz [76] where the martingale problems are considered for more general Markov processes. For instance, a more recent approach is given in He et al. [105] or Kallenberg [121].

Without the growth assumption (5.8) on the coefficients of the (non-local) operator A we need to assume that the integral expression in (5.159) makes sense as a Lebesgue integral in $[t_0, t]$ almost surely, i.e., the local martingale condition for M_φ should include the following condition:

$$P\Big\{ \int_{t_0}^t |A(s, v(s))\varphi(x(s))|\mathrm{d}s < \infty \Big\} = 1, \quad \forall t.$$

If the coefficients of the integro-differential operator A are bounded, then the above integral is bounded for any φ function which is bounded together with its first and second derivatives.

The key point is the converse of the (local) martingale condition, namely,

Theorem 5.45. *Suppose that g, σ and γ are Borel measurable coefficients satisfying the growth condition (5.8), and that a control process $v(\cdot)$ is given as an adapted (predictable) mapping from $D([t_0, \infty), \mathbb{R}^d)$ into $\mathbb{R}^q$. Let $a(s, x, v)$ and $\mathtt{M}(B, s, x, v)$ be defined by (5.155) and (5.156), and let $(x(t) : t \geq t_0)$ be a an adapted cad-lag stochastic process in a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}(t) : t \geq t_0)$ such that the process M_φ in (5.159) is a bounded martingale, for every smooth (i.e., differentiable of any order) real valued function φ with a compact support in $\mathbb{R}^d$, Then there is a $\mathbb{R}^d$-valued (quasi-left continuous) local square-integrable martingale $M = (M_1, \ldots, M_d)$ with $M(t_0) = 0$ such that*

$$\begin{cases} x(t) = x(t_0) + \displaystyle\int_{t_0}^t g(s, x(s), v(s))\mathrm{d}s + M(t), \\ \langle M^c, M^c \rangle(t) = \displaystyle\int_{t_0}^t a(s, x(s-), v(s))\mathrm{d}s, \\ \nu^p_M(B,]t_0, t]) = \displaystyle\int_{t_0}^t \mathtt{M}(B, s, x(s-), v(s))\mathrm{d}s, \quad \forall B \in \mathcal{B}(\mathbb{R}^d_*), \end{cases} \tag{5.160}$$

for every $t \geq t_0$, where M^c is the continuous part of M and ν_M^p is the compensator of the integer measure ν_M associated with M.

Proof. First, since $\mathbb{R}^d$ is a locally compact space, by truncation on closed balls we deduce that the local martingale condition remains true for any φ function which is C^2 with bounded second derivatives. Moreover, M_φ is a local square-integrable martingale for any smooth function φ.

Next, based on the arguments used to prove the a priori estimate (5.52) in Theorem 5.11 and because any smooth function with bounded second derivatives φ produces a local martingale, we do have the a priori estimate (5.52) with $p = 2$. Thus, even if the coefficients have a linear growth, the process $x(\cdot)$ is locally square-integrable.

Then, take $\varphi(x) = x_i$ to deduce that

$$M(t) := x(t) - x(t_0) - \int_{t_0}^t g_i(s, x(s), v(s))\mathrm{d}s, \quad \forall t \geq t_0$$

is a $\mathbb{R}^d$-valued local square-integrable martingale with $M(t_0) = 0$.

By means of the integration by parts applied to the (special and quasi-left continuous) semi-martingale $x(\cdot)$ we have

$$\begin{aligned} x_i(t)\, x_j(t) = {} & x_i(t_0)\, x_j(t_0) + [M_i, M_j](t) + \\ & + \int_{t_0}^t \big[x_i(s)\, g_j(s, x(s), v(s)) + x_j(s)\, g_i(s, x(s), v(s))\big]\mathrm{d}s + \\ & + \int_{]t_0, t]} \big[x_i(s-)\mathrm{d}M_j(s) + x_j(s-)\mathrm{d}M_i(s)\big], \end{aligned}$$

and by taking $\varphi(x) := x_i\, x_j$ we deduce that

$$\begin{aligned} [M_i, M_j](t) = {} & q_{ij}(t) + \int_{t_0}^t a_{ij}(s, x(s), v(s))\mathrm{d}s + \\ & + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}_*^m} z_i z_j \mathtt{M}(\mathrm{d}z, s, x(s), v(s)), \end{aligned}$$

where q_{ij} is a martingale. This proves that the predictable quadratic covariation of M_i and M_j satisfies

$$\begin{aligned} \langle M_i, M_j\rangle(t) = {} & \int_{t_0}^t a_{ij}(s, x(s-), v(s))\mathrm{d}s + \\ & + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}_*^m} z_i z_j \mathtt{M}(\mathrm{d}z, s, x(s-), v(s)). \end{aligned}$$

In the case without jumps, i.e., when $\gamma = 0$, we have finished as long as M is assumed continuous. Thus, only the last statement in (5.160) remain to be verified.

In general, denote by $\nu_M = \nu$ the integer random measure in $\mathbb{R}^d_*$ associated with M, i.e., $\nu(B\times]a,b])$ is the number of jumps $M(t) - M(t-)$ in the interval $]a,b]$ that belong to B, a Borel subset of $\mathbb{R}^d_*$. Now, define the stochastic process

$$M^c(t) := M(t) - \int_{\mathbb{R}^d_*\times]t_0,t]} z\tilde{\nu}(\mathrm{d}z,\mathrm{d}s), \quad \forall t \geq t_0,$$

where $\tilde{\nu} := \nu - \nu^p$ and $\nu^p = \nu^p_M$ is the predicable compensator of $\nu = \nu_M$. It is clear that M^c is the continuous part of M.

To check that ν_M is quasi-left continuous, we take a smooth function φ with compact support in $\mathbb{R}^d_*$. The martingale property and Doob's optional sampling theorem show that

$$\mathbb{E}\Big\{\varphi\big(x(\tau) - x(\theta)\big) + \int_\theta^\tau A(s,v(s))\varphi(x(s))\mathrm{d}s \mid \mathcal{F}(\theta)\Big\} = 0,$$

for any bounded stopping times $\theta \leq \tau$. Because φ vanishes near the origin and $x(\cdot)$ has always a finite number of jumps (outside of neighborhood of the origin) on any bounded time interval, so that the expression $\varphi\big(x(\tau) - x(\theta)\big)$ is a finite sum for each ω, and

$$\begin{aligned}\mathbb{E}\{\varphi\big(x(\tau) - x(\theta)\big)\} &= \mathbb{E}\Big\{\int_{\mathbb{R}^d_*}\varphi(z)\big(\nu(\mathrm{d}z,\tau) - \nu(\mathrm{d}z,\theta)\big)\Big\} \leq \\ &\leq \mathbb{E}\Big\{\int_\theta^\tau |A(s,v(s))\varphi(x(s))|\mathrm{d}s\Big\} < \infty,\end{aligned}$$

the proof of the quasi-left continuity is completed.

From the integration by parts we deduce the following property: if X is a local square-integrable martingale and Y is an adapted cad-lag process with local bounded variation paths, such that $s \mapsto Y(s-)$ is locally integrable with respect to X, then the process

$$\begin{aligned}X(t)Y(t) - \int_{t_0}^t X(s-)\mathrm{d}Y(s) - \\ - \sum_{t_0<s\leq t}[X(s) - X(s-)][Y(s) - Y(s-)],\end{aligned}$$

is a stochastic integral and therefore a local martingale. Essentially, by means of the mentioned property and by using $\varphi(x) = \exp[\mathtt{i}\xi\cdot x + \psi(x)]$ (with $\mathtt{i} := \sqrt{-1}$ and ψ a smooth function having a compact support in $\mathbb{R}^d_*$) in the martingale condition (5.159), we deduce that the process

$$\left\{\begin{aligned}&\exp\Big\{\mathtt{i}\xi\cdot\Big[x(t) - x(t_0) - \int_{t_0}^t g(s,x(s),v(s))\mathrm{d}s\Big] + \\ &\quad + \int_{R^d_*}\psi(z)\nu(\mathrm{d}z,]t_0,t]) - \int_{t_0}^t \xi\cdot a(s,x(s),v(s))\xi\mathrm{d}s + \\ &\quad + \int_{t_0}^t \mathrm{d}s\int_{\mathbb{R}^m_*}\big[1 - \mathrm{e}^{\mathtt{i}\xi\cdot z + \psi(z)} + \mathtt{i}\xi\cdot z\big]\mathtt{M}(\mathrm{d}z,s,x(s),v(s))\Big\},\end{aligned}\right. \tag{5.161}$$

is a local (square-integrable) martingale. This remains valid for any bounded Borel measurable function ψ vanishing near the origin.

Now, set

$$R_\psi(t) := \int_{\mathbb{R}^d_*} \psi(z)\nu(\mathrm{d}z,]t_0, t]) - \int_{t_0}^t \psi(z)\mathtt{M}(\mathrm{d}z, s, x(s), v(s))$$

and take $\xi = 0$, $\theta\psi$ instead of ψ in (5.161) to show that the process

$$\exp\Big\{\theta R_\psi(t) + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^m_*} \big[1 - \mathrm{e}^{\theta\psi(z)} + \theta\psi(z)\big]\mathtt{M}(\mathrm{d}z, s, x(s), v(s))\Big\}$$

is a local martingale. Hence, differentiating with respect to θ and taking $\theta = 0$ we obtain that R_ψ is a local martingale. A second differentiation shows that

$$R^2_\psi(t) - \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^m_*} \psi^2(z)\mathtt{M}(\mathrm{d}z, s, x(s), v(s))$$

is also a local martingale. This proves that for any compact subset K of $\mathbb{R}^d_*$ the process

$$t \mapsto \nu_M(K,]t_0, t]) - \int_{t_0}^t \mathrm{d}s \int_K \mathtt{M}(\mathrm{d}z, s, x(s), v(s))$$

is a local martingale, i.e., $\mathtt{M}(\mathrm{d}z, s, x(s), v(s))$ is the compensator or ν_M and (5.160) is established. □

A more detailed proof can be found in the book Bensoussan and Lions [16, Section 3.7.2 pp. 252–259].

• *Remark* 5.46 (characteristics). The converse of Theorem 5.45 (i.e., if we have a semi-martingale with characteristics as in (5.160) then we have a solution of the martingale problem) follows from the Itô formula for semi-martingales, e.g. see Jacod and Shiryaev [117, Theorem 4.57, Chapter 1, pp. 57–58], Remark 4.32 and Theorem 4.31 of Chapter 4. □

It is clear now that, based on identification (5.160), the representation of martingales (see Section 4.2.6 in Chapter 7 for more detail) shows that martingale problem solution is equivalent to the weak solution of the stochastic differential equation. Essentially, if the matrix a of (5.160) is non degenerate then we construct easily a Wiener process and the Poisson to represent M, otherwise, we need to enlarge (standard extension) the probability space to obtain the desired processes.

To define the integro-differential operator A by (5.154) we only need the *drift* vector $g(t, x, v)$, the *diffusion* covariance matrix $a(t, x, v)$ and the *jump* or Lévy kernel $\mathtt{M}(\mathrm{d}z, t, x, v)$. It is clear that g takes values in $\mathbb{R}^d$, $a(t, x, v)$ is a symmetric non-negative d-dimensional square matrix, i.e., $\xi \cdot a(t, x, v)\xi \geq 0$ for every ξ in $\mathbb{R}^r$, and $\mathtt{M}(\cdot, t, x, v)$ is a Radon measure in $\mathbb{R}^d_*$, i.e., it is finite on every compact

(actually, any closed subset separated from the origin) subset of $\mathbb{R}^d_*$. Besides measurability, we impose a growth condition, namely,

$$\begin{cases} x \cdot g(t,x,v) + |a(t,x,v)| + \displaystyle\int_{\mathbb{R}^d_*} |z|^2 \mathtt{M}(\mathrm{d}z,t,x,v) \leq \\ \qquad\qquad \leq C\,(1+|x|^2), \end{cases} \tag{5.162}$$

for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ and some positive constant C. Under this condition, the a priori estimate (5.52) in Theorem 5.11 with $p=2$ as discussed above, and the following definition may be considered.

Definition 5.47. A probability measure on the canonical space $D([t_0,\infty),\mathbb{R}^d)$ or equivalently an adapted cad-lag stochastic process $x(\cdot)$ in a (completed) filtered probability space $(\Omega,\mathcal{F},P,\mathcal{F}(t):t\geq t_0)$ is called a solution of the martingale problem relative to the integro-differential operator (5.154) with measurable coefficients g, a and $\mathtt{M}(\mathrm{d}z,t,x,v)$ satisfying (5.162) and a control process $v(\cdot)$ on the time interval $[t_0,\infty)$ if the stochastic process (5.159) is a martingale for every test function. □

It is clear that under the growth condition (5.162) the martingale condition (5.159) is equivalent to the assertion (5.160), which can be used as the definition of the martingale problem.

As mentioned early, a local integrability condition for $A\varphi$ must be part of the definition of the property (5.159) if local martingale is used instead of martingale. This is actually a local square-integrable martingale for any C^2 function with bounded second derivatives. It is also clear that we consider the problem on a bounded interval $[t_0,t_1]$ instead of the semi-line $[t_0,\infty)$.

Note that the martingale condition (5.159) on the process M_φ can be rewritten as

$$\begin{cases} \mathbb{E}\Big\{\varphi\big(x(t)\big)\displaystyle\prod_{i=1}^n h_i\big(x(s_i)\big)\Big\} = \mathbb{E}\Big\{\varphi\big(x(s)\big)\displaystyle\prod_{i=1}^n h_i\big(x(s_i)\big)\Big\} + \\ \qquad +\mathbb{E}\Big\{\displaystyle\int_s^t A(r,v(r))\varphi\big(x(r)\big)\prod_{i=1}^n h_i\big(x(s_i)\big)\mathrm{d}r\Big\}, \end{cases} \tag{5.163}$$

for any integer n, for every $t_0\leq s_1<s_2\cdots\leq s_n\leq s<t$, any (real-valued) Borel and bounded functions h_i, $i=1,\dots,n$. It is clear know that the filtered probability space is a convenient tool, but it may be taken to the canonical space of cad-lag (or continuous, when no jumps are involved) functions from $[t_0,\infty)$ into $\mathbb{R}^d$. Moreover, $\mathbb{R}^d$ may be replaced by a Polish space $\mathcal{O}$. If $\mathcal{O}$ is not locally compact, then some extra difficulties appear and extra care should be taken to work with smooth functions. Essentially, when $\mathcal{O}$ is a Banach space then smooth functions with compact support should be replaced by smooth functions vanished outside of a ball (this produces some extra difficulties).

An initial condition takes the form

$$P\{x(t_0)\in B\} = P_0(B), \quad \forall B\in\mathcal{B}(\mathbb{R}^d), \tag{5.164}$$

where P_0 is a distribution in $\mathbb{R}^d$. The control process $v(\cdot)$ is regarded as an adapted (predictable) mapping from $D([t_0,\infty),\mathbb{R}^d)$ into $\mathbb{R}^q$. Certainly, the martingale condition (5.159) is equivalent to the condition (5.160) of Theorem 5.45, where the characteristics of the martingale are explicitly given.

By measurability for the Lévy kernel, we mean that for every compact subset K of $\mathbb{R}^d_*$ the function $(t,x,v)\mapsto \mathtt{M}(K,t,x,v)$ is Borel measurable. However, by a continuous Lévy kernel with respect to the variable x we mean that for any continuous and integrable function $\varphi(z)$, every t and v, the function

$$x\mapsto \int_{\mathbb{R}^d_*}\varphi(z)\mathtt{M}(\mathrm{d}z,t,x,v)$$

is continuous. If the kernel is given by an expression like (5.156) then this means that the jump coefficient $\gamma(\zeta,t,x,v)$ is continuous in x. The dependency on the variables t, x, and v of the Lévy kernel $\mathtt{M}(\mathrm{d}z,t,x,v)$ is hard to track. For instance, a uniform integrability condition of the type

$$\lim_{\varepsilon\to 0}\sup_{t,x,v}\Big\{(1+|x|^2)^{-1}\int_{\{|z|<\varepsilon\}}|z|^2\mathtt{M}(\mathrm{d}z,t,x,v)\Big\}=0. \tag{5.165}$$

is very general and useful. This is mainly used for bounded coefficients and without the term $(1+|x|^2)$.

As discussed in the books Stroock and Varadhan [241, Section 5.2, pp. 131–134] and Ethier and Kurtz [76, Section 5.3, pp. 290–301], if the symmetric non-negative matrix valued function $a(t,x,v)$ is uniformly (local) Lipschitz continuous in x with either a bounded inverse (matrix) $a^{-1}(t,x,v)$ or bounded second derivatives $\partial_{ij}a(t,x,v)$, then there exists a square root $a^{1/2}(t,x,v)$ uniformly (local) Lipschitz continuous in x.

Clearly, we say that the drift coefficient $g(t,x,v)$ and diffusion coefficients $a^{1/2}(t,x,v)$ (or simply $a(t,x,v)$ even if we really mean square root of a) have linear growth if

$$|g(t,x,v)|^2+|a(t,x,v)|\le C\,(1+|x|^2), \tag{5.166}$$

or even the following weaker growth condition

$$x\cdot g(t,x,v)+|a(t,x,v)|\le C\,(1+|x|^2), \tag{5.167}$$

for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ and some positive constant C.

Instead of given *structural* condition on the Lévy kernel $\mathtt{M}(\mathrm{d}z,t,x,v)$ we prefer to use the following concrete representation, see Garroni and Menaldi [93, 94],

$$\mathtt{M}(B,t,x,v)=\int_{\{\zeta\,:\,\mathtt{j}(\zeta,t,x,v)\in B\}}\mathtt{m}(\zeta,t,x,v)\pi(\mathrm{d}\zeta), \tag{5.168}$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$, where $\mathtt{j}(\cdot)$ and $\mathtt{m}(\cdot)$ are measurable functions taking values in $\mathbb{R}^d$ and $[0,1]$, respectively, and π is a Radon measure on $\mathbb{R}^m_*$ such that for

some measurable real-valued function $\bar\jmath(\zeta)$ we have

$$\begin{cases} |\mathtt{j}(\zeta,t,x,v)|\,\mathtt{m}(\zeta,t,x,v) \le (1+|x|)\,\bar\jmath(\zeta), \\ \displaystyle\int_{\mathbb{R}^m_*} \bar\jmath^2(\zeta)\pi(\mathrm{d}\zeta) \le C, \end{cases} \tag{5.169}$$

for every ζ, t, x, v and for some positive constant C. Thus, conditions (5.167) and (5.169) imply the growth assumption (5.162). In most of the cases, all coefficients are bounded and the linear growth term $(1+|x|)$ is not necessary. Usually, it is convenient to set $\mathtt{j}(\zeta,t,x,v) = 0$ when $\mathtt{m}(\zeta,t,x,v) = 0$. The fact that the density $\mathtt{m}$ takes values in $[0,1]$ is a normalization, always we may reduce (by changing the base measure π) to this case if we assume $\mathtt{m}$ bounded. Hence, the full integro-differential operator $A(t,v)$ is expressed as

$$\begin{cases} A(t,v)\varphi(x) = \displaystyle\sum_{i,j=1}^{d} a_{ij}(t,x,v)\partial_{ij}\varphi(x) + \\ \qquad\qquad + \displaystyle\sum_{i=1}^{d} g_i(t,x,v)\partial_i\varphi(x) + I(t,v)\varphi(x), \end{cases} \tag{5.170}$$

where the proper integro-differential operator $I(t,v)$ takes the form

$$\begin{cases} I(t,v)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} [\varphi(x+\mathtt{j}(\zeta,t,x,v)) - \varphi(x) - \\ \qquad\qquad -\mathtt{j}(\zeta,t,x,v)\cdot\nabla\varphi(x)]\mathtt{m}(\zeta,t,x,v)\pi(\mathrm{d}\zeta), \end{cases} \tag{5.171}$$

on which Taylor formula can be used when φ is twice continuously differentiable with bounded second derivative. Note that under the above conditions on $\mathtt{j}$ and $\mathtt{m}$, cf., (5.169), the non-local operator $I(t,v)$ maps twice continuously differentiable functions with compact supports into bounded functions, uniformly in t and v.

We say that the coefficients $g(t,x,v)$, $a(t,x,v)$, $\mathtt{j}(\zeta,t,x,v)$ and $\mathtt{m}(\zeta,t,x,v)$ are *uniformly local continuous* x if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every t, x, x', v satisfying $0 \le t - t_0 \le 1/\varepsilon$, $|x-x'| < \delta$, $|x| \le 1/\varepsilon$, $|x'| \le 1/\varepsilon$, we have

$$\begin{cases} |g(t,x,v) - g(t,x',v)| + |a(t,x,v) - a(t,x',v)| \le \varepsilon, \\ |\mathtt{j}(\zeta,t,x,v) - \mathtt{j}(\zeta,t,x',v)| \le \varepsilon\,\bar\jmath(\zeta) \\ |\mathtt{m}(\zeta,t,x,v) - \mathtt{m}(\zeta,t,x',v)| \le \varepsilon, \end{cases} \tag{5.172}$$

where $\bar\jmath(\zeta)$ is a square-integrable function as in (5.169), and a posteriori taken to be the same function.

As we may expect, Borel measurability on the coefficients and the bound (5.162) are not sufficient for the existence of a weak solution. In ODE, continuity and a priori bounds imply global existence of a solution. A quite similar result can be established for the weak solution or the martingale problem.

Theorem 5.48 (existence). *Let A be an integro-differential operator as given (5.170), (5.171), where the drift vector $g(t,x,v)$, the diffusion covariance matrix $a(t,x,v)$ and the jump or Lévy kernel $\mathtt{M}(\mathrm{d}z,t,x,v)$, written as in (5.168), are Borel measurable and satisfy the growth conditions (5.167) and (5.169). Assume also that the coefficients are uniformly local continuous x in the sense (5.172), that a $\mathbb{R}^q$-valued adapted process $v(\cdot)$ is given on the canonical space $D([t_0,\infty),\mathbb{R}^d)$ and that an initial distribution P_0 is prescribed in $\mathbb{R}^d$. Then there exists a solution to the martingale problem (5.159) relative to the integro-differential operator (5.170) satisfying the initial condition (5.164).*

Proof. The proof for the diffusion case, i.e., without the jumps, is due to Stroock and Varadhan [241, Section 6.1, pp. 139–145]. A more general case can be found in Jacod and Shiryaev [117, Sections III.1, III.2, pp. 129–152], among others. The arguments are as follows: first existence is established for regular coefficients and then a passage to the limit is applied. We proceed in several steps.

• Let $0 < p \le 2$ and $\lambda > 0$ be given and let $\alpha(p,\lambda)$ be defined by (5.44) with γ and $\pi(\mathrm{d}\zeta)$ replaced by $\mathtt{j}(\zeta,t,x,v)$ and $\mathtt{m}(\zeta,t,x,v)\,\pi(\mathrm{d}\zeta)$. If P is a martingale solution satisfying the initial condition then estimates (5.52) and (5.53) of Theorem 5.11 hold true. Moreover, to consider $p > 2$ we need to add the p-integrability condition

$$\int_{\mathbb{R}^m_*} \bar{\jmath}^p(\zeta)\pi(\mathrm{d}\zeta) \le C,$$

for some constant C and where $\bar{\jmath}(\zeta)$ is the same function in assumption (5.169). Indeed, the martingale condition (5.159) for a function $\varphi(x) = (1+|x|^2)^{p/2}$ for $|x| \le r$ and the time s in $[0, \tau_r \wedge t]$, where τ_r is the first exit time of the process $x(s)$ from the ball $\{|x| \le r\}$. As $r \to \infty$ we deduce the estimates by using Fatou's Lemma. Actually, because the initial condition may not have a finite p moment, we review the proof of Theorem 5.11 to obtain the estimates

$$\left\{\begin{array}{l} \mathbb{E}\Big\{(\lambda+|x(t_0)|^2)^{-p/2}\,(\lambda+|x(t)|^2)^{p/2}\,\mathrm{e}^{-\alpha(t-t_0)} + \\ \qquad +[\alpha-\alpha(p,\lambda)](\lambda+|x(t_0)|^2)^{-p/2}\times \\ \qquad\qquad \times \displaystyle\int_{t_0}^t (\lambda+|x(s)|^2)^{p/2}\,\mathrm{e}^{-\alpha(s-t_0)}\mathrm{d}s\Big\} \le 1, \end{array}\right. \tag{5.173}$$

and

$$\left\{\begin{array}{l} \mathbb{E}\Big\{(\lambda+|x(t_0)|^2)^{-p/2} \displaystyle\sup_{t_0\le s\le t} (\lambda+|x(s)|^2)^{p/2}\,\mathrm{e}^{-\alpha(s-t_0)}\Big\} \le \\ \qquad\qquad \le C\,\big[1+\dfrac{1}{\alpha-\alpha(p,\lambda)}\big]. \end{array}\right. \tag{5.174}$$

for every $t \ge t_0$, where the constant C depends only on p and the bounds of σ and γ through assumptions (5.167) and (5.169). This yields the property: for every $\varepsilon > 0$ there exists a sufficient constant N, depending only on ε, the initial

distribution P_0, and the bounds in the assumptions (5.167) and (5.169) such that

$$P\{\sup_{0\le t\le 1/\varepsilon}|x(t)|\ge N\}\le\varepsilon.$$

Now, remark that

$$|x(t)-x(s)|\le\int_s^t|g(r,x(r),v(r))|\mathrm{d}r+$$
$$+\Big|\int_s^t\mathrm{d}M^c(r)\Big|+\Big|\int_{]s,t]\times\mathbb{R}^m_*}z\tilde{\nu}_x(\mathrm{d}z,\mathrm{d}r)\Big|,$$

where $(M^c(t):t\ge 0)$ is a continuous (local) martingale with predictable quadratic variation $\langle M^c,M^c\rangle(t)=a(t,x(t),v(t)$, and $\mathtt{M}(\mathrm{d}z,t,x,v)$, given by (5.168) is the Lévy kernel associated with the local martingale measure $\tilde{\nu}_x$. Then, if τ_N denotes the first exit time of the process $x(\cdot)$ from the ball $\{|x|\le N\}$ we deduce

$$\mathbb{E}\{\sup_{s\le r\le t\le\tau_N}|x(r)-x(s)|\}\le C_N\left[(t-s)+(t-s)^{1/2}\right],$$

for every $0\le s,t\le N$ and some constant C_N depending only on N, and the bounds on the coefficients. Therefore, for every $\varepsilon>0$ there is a $\delta>0$ depending only on ε, the initial distribution P_0, and the bounds in the assumptions (5.166) and (5.169) such that

$$\begin{cases}P\{\sup_{0\le t\le 1/\varepsilon}|x(t)|\ge 1/\delta\}+\\ \qquad+P\{\sup_{0\le r\le\delta}|x(s+r)-x(s)|\ge\varepsilon\}\le\varepsilon.\end{cases}\tag{5.175}$$

This proves that the set of probability measures P that solve a martingale problem is tight or pre-compact. Moreover, because of the continuity assumption (5.172), the martingale condition (5.159) is preserved by weak convergence of probabilities. Hence, the martingale solutions form a compact set.

• Let assume that the coefficients have a linear growth, i.e., assumptions (5.166) and (5.169). By means of the uniform local continuity in x of the coefficients in the sense (5.172), we can use a mollification in x to show that there exist sequences of smooth functions in x (actually C^2 is sufficient) $\{g_n(t,x,v)\}$, $\{a_n(t,x,v)\}$, $\{\mathtt{j}_n(\zeta,t,x,v)\}$ and $\{\mathtt{m}_n(\zeta,t,x,v)\}$, $n\ge 1$ such that the linear growth conditions (5.166) and (5.169), and the uniformly local continuity are satisfied uniformly in n, and also the following convergence holds: for every $\varepsilon>0$ there exits N_ε such that for every $n\ge N_\varepsilon$, and for every t, x, v satisfying $0\le t-t_0\le 1/\varepsilon$, $|x|\le 1/\varepsilon$, we have

$$|g_n(t,x,v)-g(t,x,v)|+|a_n(t,x,v)-a(t,x,v)|\le\varepsilon$$
$$|\mathtt{j}_n(\zeta,t,x,v)-\mathtt{j}(\zeta,t,x,v)|\le\varepsilon\,\bar{\jmath}(\zeta),$$
$$|\mathtt{m}_n(\zeta,t,x,v)-\mathtt{m}(\zeta,t,x,v)|\le\varepsilon,$$

where $\bar{\jmath}(\zeta)$ is the same square-integrable function in assumption (5.169) and (5.172). Since $a_n(t,x,v)$ is C^2 in x and symmetric non-negative square matrix, the exits a (square-root) symmetric non-negative square matrix $\sigma_n(t,x,v)$ such that $2\,a_n = \sigma_n^2$. Also we define the finite Radon measure $\pi_n(\mathrm{d}\zeta)$ by

$$\pi_n(B) := \pi(B \cap \{(1/n) \le |\zeta| \le n\}), \quad \forall B \in \mathcal{B}(\mathbb{R}^m),$$

for every $n \ge 1$.

Since the coefficients $g_n(t,x,v)$, $a_n(t,x,v)$ and $\mathtt{j}_n(\zeta,t,x,v)$ are locally Lipschitz in x, the assumptions of Theorem (5.3) on the existence and uniqueness of strong solution are satisfied with g_n, σ_n, $\mathtt{j}_n$ and π_n (even π) in lieu of g, σ, γ and π_n. Let be given a d-dimensional Wiener process $w(t)$ and a Poisson random measure ν_n on $\mathbb{R}^m_*$ on the canonical product space of continuous and cad-lag functions $C([t_0,\infty),\mathbb{R}^d) \times D([t_0,\infty),\mathbb{R}^m)$ under a probability measure Q_n and the canonical filtration $(\mathcal{F}(t) : t \ge t_0)$. Then, there exists one and only one strong solution, denoted by x_n, of the stochastic differential solution

$$\begin{aligned}\mathrm{d}x_n(t) = {} & g_n(t,x_n(t),v(t))\mathrm{d}t + \sigma_n(t,x_n(t),v(t))\mathrm{d}w(t) + \\ & + \int_{\mathbb{R}^m_*} \mathtt{j}_n(\zeta,t,x_n(t),v(t))\tilde{\nu}_n(\mathrm{d}\zeta,\mathrm{d}t), \quad t \ge t_0,\end{aligned}$$

with the initial condition $x(t_0) = x^0$, where x^0 is a random variable having P_0 as its distribution and the Poisson martingale measure $\tilde{\nu}_n$ has π_n as its Lévy measure. On the other hand, without a lost of generality we may suppose that $(1/n) \le \mathtt{m}_n(\zeta,t,x,v) \le 2$ and because π_n is finite, the exponential martingale solution of stochastic differential solution

$$Z_n(t) = 1 + \int_{]t_0,\infty)\times\mathbb{R}^m_*} Z_n(s-)\big[\mathtt{m}_n(\zeta,s,x_n(s-),v(s)) - 1\big]\tilde{\nu}_n(\mathrm{d}\zeta,\mathrm{d}s),$$

is defined for every $t \ge t_0$. Moreover, if $\mathtt{m}_n(\zeta,s) := \mathtt{m}_n(\zeta,s,x_n(s-),v(s))$ then

$$\begin{aligned}Z_n(t) = \exp\Big\{ & \int_{]t_0,\infty)\times\mathbb{R}^m_*} \big[\mathtt{m}_n(\zeta,s) - 1\big]\tilde{\nu}_n(\mathrm{d}\zeta,\mathrm{d}s) - \\ & - \int_{]t_0,\infty)\times\mathbb{R}^m_*} \big[\mathtt{m}_n(\zeta,s) - 1 - \ln\big(\mathtt{m}_n(\zeta,s)\big)\big]\nu_n(\mathrm{d}\zeta,\mathrm{d}s)\Big\}.\end{aligned}$$

Now, construct a new probability measure $\bar{Q}_n$, which is absolutely continuous with respect to Q_n and satisfies $\bar{Q}_n(\mathrm{d}\omega)|\mathcal{F}(t) = Z_n(t,\omega)Q_n(\mathrm{d}\omega)$, for every $t \ge 0$. Therefore, under the probability measure $\bar{Q}_n$, the quasi-left continuous random measure

$$\begin{aligned}\tilde{\bar{\nu}}_n(B,]0,t]) := {} & \tilde{\nu}_n(B,]0,t]) - \\ & - \int_0^t \mathrm{d}s \int_B \mathtt{m}_n(\zeta,s,x_n(s-),v(s))\pi_n(\mathrm{d}\zeta)\end{aligned}$$

is a martingale measure with jump compensator

$$\bar{\nu}^p_n(B,]0,t]) := \int_B \mathtt{m}_n(\zeta,s,x_n(s-),v(s))\pi_n(\mathrm{d}\zeta)$$

and we have

$$\mathrm{d}x_n(t) = g_n(t, x_n(t), v(t))\mathrm{d}t + \sigma_n(t, x_n(t), v(t))\mathrm{d}w(t) + \\ + \int_{\mathbb{R}^m_*} \mathtt{j}_n(\zeta, t, x_n(t), v(t))\tilde{\bar{\nu}}_n(\mathrm{d}\zeta, \mathrm{d}t), \quad t \geq t_0.$$

In other words, if P_n is the probability law in $D([t_0, \infty), \mathbb{R}^d)$ of the process x_n under the probability $\bar{Q}_n$, i.e.,

$$P_n\{B\} := \bar{Q}_n\{\omega : x_n(\cdot, \omega) \in B\},$$

for every Borel subset B of $D([t_0, \infty), \mathbb{R}^d)$, is a solution of the martingale problem relative to the coefficients $g_n(t, x, v)$, $a_n(t, x, v)$, $\mathtt{j}_n(\zeta, t, x, v)$, $\mathtt{m}_n(\zeta, t, x, v)$ and π_n. Moreover, if necessary we may fix π and assume that $\mathtt{m}_n(\zeta, t, x, v) = 0$ for any $|\zeta| \leq 1/n$ or $|\zeta| \geq n$. Thus, the initial Poisson random measure $\tilde{\nu_n} = \tilde{\nu}$ is independent of n. However, the random measure $\tilde{\bar{\nu}}_n$ depends on n, in general. Notice that for any $t \geq 0$, we have

$$P_n\{B\} := \int_{\{\omega : x_n(\cdot, \omega) \in B\}} Z_n(t, \omega) Q_n(\mathrm{d}\omega),$$

for every Borel subset B of $D([t_0, t], \mathbb{R}^d)$.

• If A_n denotes the full integro-differential operator corresponding to the coefficients $g_n(t, x, v)$, $a_n(t, x, v)$, $\mathtt{j}_n(\zeta, t, x, v)$, $\mathtt{m}_n(\zeta, t, x, v)$ and π_n, then the local uniform convergence of the coefficients shows that for every test function with compact support φ in $\mathbb{R}^d$ we have the following convergence holds: for every $\varepsilon > 0$ there exists N_ε such that for every $n \geq N_\varepsilon$, and for every t, x, v satisfying $0 \leq t - t_0 \leq 1/\varepsilon$, $|x| \leq 1/\varepsilon$, we have

$$|A_n(t, v)\varphi(x) - A(t, v)\varphi(x)| \leq \varepsilon.$$

Thus, the process

$$M^n_\varphi(t) := \varphi(x(t)) - \varphi(x(t_0)) - \int_{t_0}^t A_n(s, v(s))\varphi(x(s))\mathrm{d}s, \quad \forall t \geq t_0,$$

is a P_n-martingale, for any test function with compact support φ in $\mathbb{R}^d$, where $x(t) := \omega(t)$ is the canonical process in $D([t_0, \infty), \mathbb{R}^d)$. In view of the continuity assumption on the variable x (5.172), the Lebesgue integral

$$\int_{t_0}^t A(s, v(s))\varphi(x(s))\mathrm{d}s, \quad \forall t \geq t_0$$

is a continuous functional on the Polish space $D([t_0, \infty), \mathbb{R}^d)$. Then, the a priori estimates (5.175) applied to P_n proves that there exist a subsequence, still denoted by $\{P_n\}$, and a probability P such that P_n converges weakly to P. Hence, P is a solution of the martingale problem relative to the integro-differential operator (5.170).

• To complete the proof, we need to pass from the linear growth condition (5.167) to the weaker growth condition (5.166). To this purpose, we approximate the drift coefficient g by truncation as follows

$$g_M(t,x,v) := \begin{cases} g(t,x,v) & \text{if } |g(t,x,v)| \le M, \\ g(t,x,v)\dfrac{M}{|g(t,x,v)|} & \text{otherwise.} \end{cases}$$

Thus, $g_M(t,x,v)$ is bounded (for each $M>0$), and the growth and continuity conditions (5.166) and (5.172) are satisfied uniformly in M. Hence, we proceed as above to pass to the limit as $M\to\infty$ and the theorem is established. □

Notice the use of Girsanov transformation. Let us assume that there exist $d\times n$ matrix-valued measurable functions $\sigma_k := (\sigma_{ik}(t,x,v) : i=1,\dots,d)$, $k=1,\dots,n$, and a constant $C>0$ such that

$$a_{ij}(t,x,v) := \frac{1}{2}\sum_k \sigma_{ik}(t,x,v)\,\sigma_{jk}(t,x,v),$$
$$\sum_k |\sigma_k(t,x,v)| \le C\,(1+|x|), \quad \forall t,x,v.$$

As in Theorem 5.45, for any solution P of the martingale problem relative to the coefficients $\mathtt{j}(\zeta,t,x,v)$, $\mathtt{m}(\zeta,t,x,v)=1$, $a(t,x,v)$ and $g(t,x,v)$, the canonical process $x(t):=\omega(t)$ in $D([t_0,\infty),\mathbb{R}^d)$ satisfies the equation

$$\mathrm{d}x(s) = g(s,x(s),v(s))\mathrm{d}s + \sum_k \sigma_k(s,x(s),v(s))\mathrm{d}w_k(s) + \\ + \int_{\mathbb{R}^m_*} \mathtt{j}(\zeta,s,x(s),v(s))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall s>t_0,$$

where $w=(w_1,\dots,w_n)$ is a n-dimensional Wiener process and $\tilde\nu$ is the Poisson martingale measure obtained from the (standard) Poisson random measure ν with Lévy measure π.

Given a measurable function $\mathtt{k}(\zeta,t,x,v)$ satisfying the same assumptions (5.169) as $\mathtt{j}(\zeta,t,x,v)$, let us consider the integer measure $\nu_{\mathtt{k}}$ associated with the purely jumps local martingale

$$\int_{\mathbb{R}^m_*\times]t_0,t]} \mathtt{k}(\zeta,s,x(s),v(s))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}s), \quad \forall t\ge t_0,$$

with local martingale measure $\tilde\nu_{\mathtt{k}}$ and jumps compensator

$$\nu^p_{\mathtt{k}}(B,]t_0,t]) := \int_{t_0}^t \pi\big(\{\zeta : \mathtt{k}(\zeta,s,x(s-),v(s))\in B\}\big)\,\mathrm{d}s,$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$ and $t>t_0$. Suppose that $b:=(b_k(t,x,v) : k=1,\dots,n)$ and $\mathtt{c}(z,t,x,v)$ are measurable functions such that for some constants $C,c_0>0$

and any ζ, t, x, v we have

$$\begin{cases} |b(t,x,v)| \le C, \\ -1 < -c_0 \le |\mathtt{c}(z,t,x,v)| \le C, \\ \displaystyle\int_{\mathbb{R}^d_*} |\mathtt{c}(\mathtt{k}(\zeta,t,x,v),t,x,v)|^2 \pi(\mathrm{d}\zeta) \le C, \end{cases} \tag{5.176}$$

and define the following density process

$$\begin{aligned} \delta(t) := \exp\Big\{ \sum_k \int_{t_0}^t b_k(s)\mathrm{d}w_k(s) - \frac{1}{2}\sum_k \int_{t_0}^t |b_k(s)|^2 \mathrm{d}s + \\ + \int_{\mathbb{R}^d_* \times]t_0,t]} c(z,s)\tilde{\nu}_{\mathtt{k}}(\mathrm{d}z,\mathrm{d}s) - \\ - \int_{\mathbb{R}^d_* \times]t_0,t]} \big[c(z,s) - \ln[c(z,s)+1]\big]\nu_{\mathtt{k}}(\mathrm{d}z,\mathrm{d}s)\Big\}, \end{aligned}$$

where $b_k(s) := b_k(s,x(s),v(s))$ and $c(z,s) := \mathtt{c}(z,s,x(s),v(s))$. Since $0 \le c - \ln(1+c) \le c/(1+c_0)$, for every $c \ge -c_0 > -1$, the above stochastic integrals are well defined and the process δ is an exponential martingale, i.e., $\mathbb{E}\{\delta(t)\} = 1$ for every $t > t_0$. Note that

$$\begin{aligned} \int_{\mathbb{R}^d_* \times]t_0,t]} c(z,s)\tilde{\nu}_{\mathtt{k}}(\mathrm{d}z,\mathrm{d}s) = \int_{\mathbb{R}^m_* \times]t_0,t]} c(\mathtt{k}(\zeta,s),s)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s), \\ \int_{\mathbb{R}^d_* \times]t_0,t]} \big[c(z,s) - \ln[c(z,s)+1]\big]\nu_{\mathtt{k}}(\mathrm{d}z,\mathrm{d}s) = \\ = \int_{\mathbb{R}^m_* \times]t_0,t]} \big[c(\mathtt{k}(\zeta,s),s) - \ln[c(\mathtt{k}(\zeta,s),s)+1]\big]\nu(\mathrm{d}\zeta,\mathrm{d}s), \end{aligned}$$

where $\mathtt{k}(\zeta,s) := \mathtt{k}(\zeta,s,x(s),v(s))$.

Then, the probability measure $\bar{P}$ satisfying $\bar{P}(\mathrm{d}\omega)|\mathcal{F}(t) = \delta(t,\omega)\,P(\mathrm{d}\omega)$, for every $t \ge t_0$, is a solution of the martingale problem relative to the coefficients $\bar{g}(t,x,v)$, $\sigma_k(t,x,v)$ and $\bar{\mathtt{M}}(\mathrm{d}z,t,x,v)$. where

$$\begin{cases} \bar{g}_i(t,x,v) := g_i(t,x,v) + \displaystyle\sum_k b_k(t,x,v)\,\sigma_{ik}(t,x,v), \\ \bar{\mathtt{M}}(B,t,x,v) := \pi(\{\zeta \,:\, \mathtt{j}(\zeta,t,x,v) \in B\}) + \\ \qquad + \displaystyle\int_{\{\mathtt{k}(\zeta,t,x,v)\in B\}} \mathtt{c}(\mathtt{k}(\zeta,t,x,v),t,x,v)\,\pi(\mathrm{d}\zeta), \end{cases} \tag{5.177}$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$, and any t, x, and v. It is clear that $\mathtt{j}(\zeta,t,x,v)$, $\mathtt{k}(\zeta,t,x,v)$, $\bar{\mathtt{M}}(\zeta,t,x,v)$, $a(t,x,v)$, $g(t,x,v)$ and $\bar{g}(t,x,v)$ are measurable and they satisfy the growth conditions in Theorem 5.48, namely, (5.167) and (5.169).

In particular, if the locally uniform continuity (in x) condition (5.172) holds then there is a solution to the martingale problem relative to the coefficients $\bar{\mathtt{M}}(\mathrm{d}\zeta,t,x,v)$, $a(t,x,v)$ and $\bar{g}(t,x,v)$. Thus, if the coefficients $g(t,x,v)$, $\sigma_k(t,x,v)$, $\mathtt{j}(\zeta,t,x,v)$ are locally Lipschitz in x (i.e., $\delta = K\varepsilon$ in (5.172) for some constant K) and $\mathtt{m}(\zeta,t,x,v) = 1$, then the solution is unique, i.e., a strong solution plus

a Girsanov transformation. Notice that the measure $\mathtt{c}(\mathtt{k}(\zeta,t,x,v),t,x,v)\,\pi(\mathrm{d}\zeta)$ is of first order in the sense that it integrates the jumps $\mathtt{k}(\zeta,t,x,v)$ instead of the square of the jumps $|\mathtt{k}(\zeta,t,x,v)|^2$ as it happens with just π.

Now, if P is a solution of the martingale measure relative to the coefficients $g(t,x,v)$, $a(t,x,v)$ and the Lévy kernel $\mathtt{M}(\mathrm{d}z,s,x,v)$ then we write

$$\mathrm{d}x(s) = g(s,x(s),v(s))\mathrm{d}s + \mathrm{d}M^c(s) + \int_{\mathbb{R}^d_*} z\nu_x(\mathrm{d}z,\mathrm{d}s),$$

where the continuous martingale $M^c(t)$ has a predictable compensator satisfying

$$\langle M^c\rangle(t) = \int_0^t a(s,x(s-),v(s))\mathrm{d}s,$$

the canonical integer measure ν_x associated with the process $x(\cdot)$, i.e., $\tilde\nu_x := \nu_x - \nu_x^p$ and ν_x^p is the jumps compensator which satisfies

$$\nu_x^p(B,]t_0,t]) = \int_0^t \mathrm{d}s \int_{\{\zeta\in\mathbb{R}^m_* \,:\, \mathtt{j}(\zeta,s,x(s-),v(s))\in B\}} \mathtt{m}(\zeta,s,x(s-),v(s))\pi(\mathrm{d}\zeta),$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$ and $t > t_0$, i.e., $\nu_x^p(\mathrm{d}z,ds) = \mathtt{M}(\mathrm{d}z,s,x(s-),v(s))\,\mathrm{d}s$, where the Lévy kernel is as in (5.168). If $c(z,s,x,v)$ and $b(s,x,v)$ are measurable such that for some constants $C, c_0 > 0$ and for every z,s,x,v we have

$$\begin{cases} |b(s,x,v)\,a(s,x,v)| \le C, \\ -1 < -c_0 \le c(z,s,x,v) \le C, \\ \displaystyle\int_{\mathbb{R}^d_*} |c(z,s,x,v)|^2 \mathtt{M}(\mathrm{d}z,s,x,v) \le C, \end{cases} \tag{5.178}$$

then going back to the definition of the density process $\delta(t,\omega)$ we can set

$$\begin{aligned}\delta(t) := \exp\Big\{ &\int_{t_0}^t b(s)\cdot \mathrm{d}M^c(s) + \int_{\mathbb{R}^d_*\times]t_0,t]} c(z,s)\tilde\nu_x(\mathrm{d}z,\mathrm{d}s) - \\ &- \int_{t_0}^t \big[b(s)\cdot a(s,x(s),v(s))\,b(s)\big]\mathrm{d}s - \\ &- \int_{\mathbb{R}^d_*\times]t_0,t]} \big[c(z,s) - \ln[c(z,s)+1]\big]\nu_x(\mathrm{d}z,\mathrm{d}s)\Big\},\end{aligned}$$

where $b(s) := b(s,x(s),v(s))$ and $c(z,s) := \mathtt{c}(z,s,x(s),v(s))$, which is again an exponential martingale.

Then, the probability measure $\bar P$ satisfying $\bar P(\mathrm{d}\omega)|\mathcal{F}(t) = \delta(t,\omega)\,P(\mathrm{d}\omega)$, for every $t \ge t_0$, is a solution of the martingale problem relative to the coefficients $\bar g(t,x,v)$, $a(t,x,v)$ and the Lévy Kernel $\bar{\mathtt{M}}(\mathrm{d}z,s,x,v)$, where

$$\begin{cases} \bar g(t,x,v) := g(t,x,v) + 2b(t,x,v)\,a(t,x,v), \\ \bar{\mathtt{M}}(\mathrm{d}z,t,x,v) := \big[1+\mathtt{c}(z,t,x,v)\big]\mathtt{M}(\mathrm{d}z,t,x,v). \end{cases} \tag{5.179}$$

Hence, we see that by means of Girsanov transformation we can add a bounded measurable drift $2ba$ (which is arbitrary if the covariance diffusion matrix a is non-singular) and a jump density $\mathtt{c}(z,t,x,v)$ which yields a first order Lévy kernel $\mathtt{c}(z,t,x,v)\mathtt{M}(\mathrm{d}z,t,x,v)$.

• *Remark* 5.49. The arguments used here are not valid when the linear growth conditions (5.166) or (5.167) and (5.169) hold only locally, i.e. for any x with $|x| \le r$ and constants $C = C(r)$. However, if a Lyapunov function can be constructed, then most of the technique can be used. A Lyapunov function is a nonnegative smooth function $v(t,x)$ such that $v(t,x)$ goes to ∞ locally uniform in t as x goes to ∞ and satisfies

$$\partial_t v(t,x) + A(t,x)v(t,x) \le cv(t,x),$$

for some constant $c \ge 0$, for every x and t, where $A(t,x)$ is the integro-differential operator (5.154). The reader is referred to Khasminskii [130], for a study of this technique for diffusion processes in the whole space, with focus on stationary distributions and invariant measures. □

5.7 Strong Markov Property

A family of controlled Markov processes (or a Markov process dynamic)

$$(X(t) : t \ge 0) = \{X(t,x,s,v(\cdot)) : t \ge s \ge 0,\ x \in \mathbb{R}^d\} \tag{5.180}$$

is naturally constructed by means of stochastic differential equations. Usually, X represents the state of the system at time t when the initial state was x at time s under the control $v(\cdot)$. After the coefficients have been estimated, i.e., the drift $g(t,x,v)$, the diffusion $\sigma(t,x,v)$ and the jumps $\gamma(\zeta,t,x,v)$ are known, the evolution of the system should be defined, i.e., the initial state x at the initial time t and the control $v(\cdot)$ on the interval $[s,t[$ should determine the state x at the time t. Thus, a natural model is a stochastic ordinary differential equation, namely,

$$\begin{cases} X(t) = x + \displaystyle\int_s^t g(r,X(r),v(r))\mathrm{d}r + \int_s^t \sigma(r,X(r),v(r))\mathrm{d}w(r) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]s,t]} \gamma(\zeta,r,X(r),v(r))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}r), \quad \forall\, t \ge s, \end{cases} \tag{5.181}$$

where the solution process is denoted by (5.180). The driven Wiener process and Poisson martingale measure

$$(w(t) : t \ge 0), \quad (\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \ge 0)$$

are statistically given through their characteristic functions (or Fourier transforms) as discussed at the beginning of this chapter. The particular realization of this Wiener-Poisson space is (of little importance and) mainly used to set up a filtered probability space $(\Omega, P, \mathcal{F}, \mathcal{F}(t) : t \ge 0)$ where the control process $(v(t) : t \ge 0)$ is realized as an adapted (or progressively measurable) process taking values in some closed subset of an Euclidean space. In particular, only the control $v(\cdot)$, and the increments of the Wiener process and Poisson martingale measure on the time interval $[s,t]$ are used to defined the solution $X(t,x,s,v(\cdot))$.

Moreover, since (5.181) has a unique (almost surely) solution we deduce that $(t,x,s) \mapsto X(t,x,s,v(\cdot))$ is continuous in probability and

$$X(t,x,s,v(\cdot)) = X(t, X(r,x,s,v(\cdot)), r, v(\cdot)), \tag{5.182}$$

for every $s \le r \le t$ and x in $\mathbb{R}^d$. Moreover, the time r and the state x can be replaced by a stopping time R and a $\mathcal{F}(R)$-measurable random variable. Equality (5.182) is usually refer to as a *flow condition*.

Because of the control $v(\cdot)$, it is clear that $(X(t) : t \ge 0)$ is not (unless $v(\cdot)$ is a deterministic constant) a Markov process. However, if another variable X_0 is added the couple $\bar{X} = (X, X_0)$ becomes a Markov process. Indeed, $(X_0(t) : t \ge 0$ is the constant process $v(\cdot)$, i.e., with values in the space $\mathtt{C}$ of adapted processes (the control space policies) endowed with the trivial σ-algebra $\mathcal{C} = \{\mathtt{C}, \emptyset\}$. The *Markov property* for $\bar{X}$ becomes

$$P\{\bar{X}(t) \in \bar{A} \,|\, \bar{\mathcal{F}}(s)\} = P\{\bar{X}(t) \in \bar{A} \,|\, \bar{X}(s)\}, \quad \text{a.s.}$$

for every $t > s$ and any $\bar{A}$ in $\mathcal{B}(\mathbb{R}^d) \times \mathcal{C}$, where $\bar{\mathcal{F}}(s) = \mathcal{F}(s) \times \mathcal{C}$. In short, this *controlled Markov property* can be written as

$$P\{X(t) \in B \,|\, \mathcal{F}(s)\} = P\{X(t) \in B \,|\, X(s),\, v(\cdot)\}, \quad \text{a.s.}$$

for every $t > s$ and any B in $\mathcal{B}(\mathbb{R}^d)$, or equivalently as

$$\mathbb{E}\{\varphi(X(t)) \,|\, \mathcal{F}(s)\} = \mathbb{E}\{\varphi(X(t)) \,|\, X(s),\, v(\cdot)\}, \quad \text{a.s.} \quad \forall t > s, \tag{5.183}$$

for any smooth function φ with compact support in $\mathbb{R}^d$. Because of the trivial σ-algebra $\mathcal{C}$, the conditioning with respect to $v(\cdot)$ means the evaluation for (or conditioning to a set) $X_0 = v(\cdot)$. Thus, the controlled transition function is defined as a temporal functional on the control (i.e., on the restriction $v[s,t)$ to the interval $[s,t)$ of $v(\cdot)$, namely,

$$P\big(s,t,x,B; v[s,t)\big) = P\{X(t,x,s,v(\cdot)) \in B\},$$

for every $t > s$ and any B in $\mathcal{B}(\mathbb{R}^d)$, provided a unique solution of the stochastic ordinary differential equation (5.181) exists and a continuous version in the variables t, x, s is chosen.

Actually, only measurability with respect to the variables s, x, t is necessary to show the validity of the controlled Markov property. Indeed, for every smooth function φ with compact support we can use Itô formula (i.e., the martingale property) to get

$$\mathbb{E}\{\varphi(X(t)) \,|\, \mathcal{F}(s)\} = \varphi(X(s)) + \mathbb{E}\Big\{ \int_s^t A(r, v(r))\varphi(X(r))\mathrm{d}r \;\Big|\; \mathcal{F}(s)\Big\},$$

almost surely, for every $t > s$, where $A(t,v)$ is the second-order integro-differential operator (5.154) associated with the coefficients. Hence, by means of equality (5.182) we obtain

$$\mathbb{E}\{\varphi(X(t,x^0,t_0,v(\cdot))) \,|\, \mathcal{F}(s)\} = \varphi(X(s)) + \Phi(X(s); v(\cdot)),$$
$$\Phi(x; v(\cdot)) = \mathbb{E}\Big\{ \int_s^t A(r, v(r))\varphi(X(r,x,s,v(\cdot)))\mathrm{d}r \;\Big|\; \mathcal{F}(s)\Big\},$$

for every $t_0 \le s \le t$ and x, x^0 in $\mathbb{R}^d$, which yields (5.183).

For the uncontrolled process, namely,

$$\begin{cases} X(t) = x + \displaystyle\int_s^t g(r, X(r))\mathrm{d}r + \int_s^t \sigma(r, X(r))\mathrm{d}w(r) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_* \times]s,t]} \gamma(\zeta, r, X(r))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}r), \quad \forall\, t \ge s, \end{cases} \tag{5.184}$$

where $X(t) = X(t, x, s)$, there are several way of proving the Markov property in the form

$$\mathbb{E}\{\varphi(X(t)) \mid X(s_1), \dots, X(s_n)\} = \mathbb{E}\{\varphi(X(t)) \mid X(s_n)\}, \text{ a.s.} \tag{5.185}$$

for every $s_1 \le \cdots \le s_n < t$ and for any smooth function φ with compact support in $\mathbb{R}^d$.

As mentioned early, the stochastic ordinary differential equation (5.184) or (5.181) can be used to define in a unique manner (under suitable assumptions) the solution process $(X(t) : t \ge s)$ with an initial condition $X(s) = x$ where s could be a stopping time and x a $\mathcal{F}(s)$-measurable function. Alternatively, if a continuous (measurable suffices) version in all variables of the solution $\{X(t, x, s) : t \ge s,\, x \in \mathbb{R}^d$ has been chosen for, then the substitution in the function

$(x, s) \mapsto X(t, x, s)$ for s a stopping time and x for a $\mathcal{F}(s)$-measurable random variable should give the unique solution. Since both, the Wiener process and the Poisson martingale measure have independent increment, the following result on conditional expectation, which is proved with the usual Fubini's and monotone class theorems,

Proposition 5.50. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $f(x, \omega)$ be a measurable function with respect to the product σ-algebra $\mathcal{B}(\mathbb{R}^d) \times \mathcal{F}$. Suppose ξ is a $\mathbb{R}^d$-valued random variable which is independent of the σ-algebra generated by the family of random variables $\{f(x, \omega) : x \in \mathbb{R}^d\}$. If $\mathbb{E}\{|f(\xi, \cdot)|\} < \infty$ and $g(x) = \mathbb{E}\{f(x, \cdot)\}$ then the function g is $\mathcal{B}(\mathbb{R}^d)$-measurable and $\mathbb{E}\{f(\xi, \cdot) \mid \xi\} = g(\xi)$, almost surely.* □

can be applied to verify the Markov property (5.185). Moreover, a stronger version

$$\mathbb{E}\{\varphi(X(t)) \mid \mathcal{F}(S)\} = \mathbb{E}\{\varphi(X(t, x, s))\}|_{x=X(S), s=S}, \text{ a.s.} \tag{5.186}$$

for every stopping time $S \le t$ and for any smooth function φ with compact support in $\mathbb{R}^d$.

This strong Markov property (5.186) can be proved in a more direct way and in several steps as in Krylov [142, Section 9.11, pp. 220-224]. Some details are given in what follows.

First, denote by $\mathcal{G}(t, s)$ the σ-algebra generated by the random variables $\{w(r) - w(s) : s < r \le t\}$, $\{\tilde{\nu}(B,]s, r]) : B \in \mathcal{B}(\mathbb{R}^m_*),\, s < r \le t\}$ and the null sets, i.e., the completed σ-algebra generated by the Wiener process and the Poisson martingale measure on the interval $(s, t]$. Now, note that $x_n \to x$

implies $X(t,x_n,s) \to X(t,x,s)$ in probability for any sequence $\{x_n\}$ of random variables independent of $\mathcal{G}(t,s)$ which converges in probability to x. Also, for any $t \geq s$ the random variable $X(t,x,s)$ is measurable with respect to $\mathcal{G}(s,t)$.

Since the Wiener process and the Poisson martingale measure have independent increments, the σ-algebras $\mathcal{F}(s)$ and $\mathcal{G}(s,t)$ are independent, proving that the random variable $X(t,x,s)$ is independent of $\mathcal{F}(s)$.

Next, for a smooth function φ with compact support in $\mathbb{R}^d$ (bounded continuous suffices) define

$$\Phi(t,x,s) = \mathbb{E}\{\varphi(X(t,x,s))\}, \quad \forall\, x \in \mathbb{R}^d,\, t \geq s.$$

If ξ is a $\mathbb{R}^d$-valued random variable $\mathcal{F}(s)$-measurable assuming a countable number of values, say N, and ζ is a bounded $\mathcal{F}(s)$-measurable set then

$$\begin{aligned}\mathbb{E}\{\zeta\varphi(X(t,\xi,s)\} &= \sum_{n\in N} \mathbb{E}\{\zeta\varphi(X(t,n,s)\mathbb{1}_{\xi=n}\} = \\ &= \sum_{n\in N} \mathbb{E}\{\zeta\Phi(t,n,s)\mathbb{1}_{\xi=n}\} = \mathbb{E}\{\zeta\Phi(t,\xi,s)\}.\end{aligned}$$

By means of the continuity in probability, this remains valid for any $\mathcal{F}(s)$-measurable random variable ξ. Now, consider $X(t) = X(t,x^0,s_0)$ for some x^0 and $s_0 \leq s_1 \leq \cdots \leq s_n \leq t$, so that the previous arguments show that

$$\mathbb{E}\{\varphi(X(t)) \mid \mathcal{F}(s_n)\} = \Phi(t,X(s_n),s_n), \quad \text{a.s.},$$

which yields the Markov property (5.185). Because both sides of (5.185) are measures, the equality holds for any Borel and bounded function φ.

Finally, the above arguments are also valid for a stopping time S in lieu of the deterministic time s, proving the validity of strong Markov property (5.186).

5.8 Diffusions with Jumps

It is clear by now that there are two ways of study (or starting) the construction of (standard) Markov processes or more specifically diffusion process with jumps. One way is to look for a family of cad-lag stochastic processes $X = \{X_x(t) : t \geq 0,\, x \in \mathbb{R}^d\}$ with values in $\mathbb{R}^d$ and defined in some abstract probability space or equivalently to look for a family of probability measures $P = \{P_x : x \in \mathbb{R}^d\}$ on the canonical (Polish) space $D([0,\infty),\mathbb{R}^d)$. Another way is to look for a family of probability measures (a transition function) $\{P(x,t,\mathrm{d}y) : t > 0,\, x \in \mathbb{R}^d\}$ on the Borel space $\mathcal{B}(\mathbb{R}^d)$. Actually, when talking about Markov processes we are implicitly referring to all this elements and more.

Stochastic differential equations (SDE) technique is a typical way to construct the family of stochastic processes X based on previously discussed (and better well known) spacial homogeneous processes such as the Wiener process or the Poisson process (in general the Lévy process). This approach requires some probability (in particular, the stochastic integral theory) to extend the classic techniques used in analysis to study ordinary differential equations.

Equivalently, the so-called Martingale problem (MP) is better adapted to the study of the family of probability measures $P = \{P_x : x \in \mathbb{R}^d\}$ on the canonical (Polish) space $D([0,\infty), \mathbb{R}^d)$. However, a deeper probability (and analysis) knowledge is necessary to understand this approach.

Besides the variety of assumptions needed for each method, SDE or MP, the comparison is relatively simple, stochastic differential equations handles efficiently degeneracy in the diffusion coefficients while martingale problems allows weaker regularity assumptions (with not equivalent for the deterministic case) on the coefficients.

Partial differential equations (and integro-differential equations) methods are necessary to discuss the construction of a transition function. we may go directly for the transition function, which in analysis is called *fundamental solution* or *Green function* associated with a *Cauchy* or (*parabolic problem*, or alternatively, study the Cauchy problem itself, meaning getting existence and uniqueness results as well as a priori estimates. Again, here we may attack the parabolic problem directly or as a product of a well develop techniques known as the *semigroup theory*. Essentially, the semigroup arguments extend the existence and uniqueness results and a priori estimates proven on elliptic equations (i.e., stationary problems) to parabolic equations. Certainly, there are plenty of comments about this point that are not mentioned here. All techniques (in this more analytic way) start with a previous analysis of easier problems, such as the heat equation, and growth all the way to more complete and complex cases.

The key connection between the probability and the analysis view points is the realization results (given the transition function a Markov process is constructed) and the conditional probability results (given the Markov process the transition function is defined). Clearly, the Kolmogorov (backward and forward) equations is at the heart of the discussion. Except for a few number of simple examples, such as the Wiener or Poisson process, there is not an explicit expression for the transition probability (or Green) function, so that the explicit characteristics of a diffusion process with jumps (or even a standard Markov process) is the second order integro-differential operator, which acts as its infinitesimal generator.

5.8.1 Probabilistic Viewpoint

A Wiener process, a Poisson process and a Lévy process are typical prototypes of a diffusion process, a jump process and a diffusion process with jumps, respectively. Let $x(\cdot)$ be a (cad-lag) Markov process with values in $\mathbb{R}^d$ such that for any $t \geq 0$ and any test function φ in $C_0^\infty(\mathbb{R}^d)$ (infinite differentiable with a compact support) we have

$$\lim_{r \downarrow 0} \frac{\mathbb{E}\{\varphi(x(t+r)) - \varphi(x(t)) \,\big|\, x(s),\, s \leq t\} - A(t)\varphi(x(t))}{r} = 0, \qquad (5.187)$$

where $A(t)$ is a linear operator from $C_0^\infty(\mathbb{R}^d)$ into $C_b^0(\mathbb{R}^d)$ (bounded continuous functions), moreover

$$\begin{cases} A(t)\varphi(x) = \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij}\varphi(x) + \sum_{i=1}^{d} a_i(t,x)\partial_i\varphi(x) + \\ \qquad + \int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,t,x), \end{cases} \tag{5.188}$$

where $a_{ij} = a_{ji}$, all coefficients a_{ij} and a_i are measurable and $\mathtt{M}(\mathrm{d}z,t,x)$ integrates the function $z \mapsto |z|^2 \wedge |z|$. Except for the measurability of the coefficients and the integrability at infinity of the function $z \mapsto |z|$ for the Lévy kernel, any linear operator as above must have the integro-differential form (5.188).

The above process $x(\cdot)$ is called a *diffusion with jumps* or just a *diffusion* if the Lévy kernel $\mathtt{M}(\mathrm{d}z,t,x)$ vanishes. Similarly, we can say that process $x(\cdot)$ is a (special) semi-martingale with characteristics $(a_i(s,x(s)))$ (drift-vector), $(a_{ij}(s,x(s)))$ (diffusion-matrix) and $\mathtt{M}(\mathrm{d}z,s,x(s))$ (jump-measure), i.e., the processes

$$\begin{cases} t \mapsto M_i(t) := x_i(t) - x_i(t_0) - \int_{t_0}^{t} a_i(s,x(s))\mathrm{d}s, \quad \forall t > t_0, \\ t \mapsto M_i^c(t)\, M_j^c(t) - \int_{t_0}^{t} a_{ij}(s,x(s))\mathrm{d}s, \\ t \mapsto \sum_{t_0<s\le t} \psi\big(M(s) - M(s-)\big) - \int_{t_0}^{t}\mathrm{d}s \int_{\mathbb{R}^d_*} \psi(z)\mathtt{M}(\mathrm{d}z,s,x(s)), \end{cases} \tag{5.189}$$

for every $t > t_0$, are local martingales, for any i,j, and for every test function ψ in $C_0^\infty(\mathbb{R}^d_*)$ (infinite differentiable with a compact support in $\mathbb{R}^d_*$), and where $M_i^c(\cdot)$ is the *continuous part* of the local martingale $M_i(\cdot)$. Actually,

$$M_i^c(t)\, M_j^c(t) = M_i(t)\, M_j(t) - \\ - \sum_{t_0<s\le t} \big(M_i(s) - M_i(s-)\big)\big(M_j(s) - M_j(s-)\big),$$

for every $t > t_0$, and $M_i(s) - M_i(s-) = x_i(s) - x_i(s-)$, for every $s > t_0$. This holds for every $t_0 \ge 0$.

In term of the characteristic function, these conditions means that the complex-valued process

$$\begin{cases} \mathbb{E}\Big\{ \exp\Big[\mathtt{i}\xi\cdot\Big[x(t) - x(0) - \int_{t_0}^{t} \mathtt{a}(s,x(s))\mathrm{d}s\Big] + \\ \qquad + \sum_{t_0<s\le t} \psi\big(x(s) - x(s-)\big) - \int_{t_0}^{t} \xi\cdot a(s,x(s))\xi\mathrm{d}s + \\ \qquad + \int_{t_0}^{t}\mathrm{d}s \int_{\mathbb{R}^m_*} \big[1 - \mathrm{e}^{\mathtt{i}\xi\cdot z + \psi(z)} + \mathtt{i}\xi\cdot z\big]\mathtt{M}(\mathrm{d}z,s,x(s))\Big\}, \end{cases} \tag{5.190}$$

is a local (square-integrable) martingale, for every test function ψ in $C_0^\infty(\mathbb{R}^d_*)$ and any ξ in $\mathbb{R}^d$, where $\mathtt{i} = \sqrt{-1}$, $\mathtt{a}(s,x(s)) = (a_i(s,x(s)))$ and $a(s,x(s)) =$

$(a_{ij}(s,x(s)))$. Perhaps a more elegant way is to use the martingale problem, i.e., to impose that the real-valued process

$$M_\varphi(t) := \varphi(x(t)) - \varphi(x(t_0)) - \int_{t_0}^t A(s)\varphi(s)\mathrm{d}s, \quad \forall t \geq t_0, \tag{5.191}$$

be a (bounded) martingale for any test function φ in $C_0^\infty(\mathbb{R}^d)$. This is translated into the condition

$$\left\{ \begin{array}{c} \mathbb{E}\Big\{\varphi\big(x(t)\big)\prod_{i=1}^n h_i\big(x(s_i)\big)\Big\} = \mathbb{E}\Big\{\varphi\big(x(s)\big)\prod_{i=1}^n h_i\big(x(s_i)\big)\Big\} + \\ +\mathbb{E}\Big\{\int_s^t A\varphi\big(x(r)\big)\prod_{i=1}^n h_i\big(x(s_i)\big)\mathrm{d}r\Big\}, \end{array} \right. \tag{5.192}$$

for any integer n, for every $t_0 \leq s_1 < s_2 \cdots \leq s_n \leq s < t$, any (real-valued) Borel and bounded functions h_i, $i = 1, \ldots, n$. Recall that the integro-differential operator $A(t)$ maps twice-continuously differentiable functions with compact support into Borel and bounded functions, locally uniformly in $t > t_0$, and any $t_0 \geq 0$.

As in Ikeda and Watanabe [110], we adopt the following framework;

Definition 5.51. A family of probability measures $(P_{s,x} : s \geq 0,\ x \in \mathbb{R}^d)$ on the canonical space $D([0,\infty),\mathbb{R}^d)$ is called a *Markov probability measures* if (1) $P_{s,x}\{\omega(s) = x\} = 1$ for every x in $\mathbb{R}^d$,

(2) the function $(s,x) \mapsto P_{s,x}(B)$, form $[0,\infty)\times\mathbb{R}^d$ into $[0,1]$, is Borel measurable for each Borel subset B of $D([0,\infty),\mathbb{R}^d)$,

(3) for every x in $\mathbb{R}^d$, $t > r > s \geq 0$, any subset B in $\mathcal{B}_{s,r}\big(D([0,\infty),\mathbb{R}^d)\big)$ and Borel subset A of $\mathbb{R}^d$ we have

$$P_{s,x}(B \cap \{\omega : \omega(t) \in A\}) = \int_B P_{r,\omega'(r)}(\omega : \omega(t) \in A)\, P_{s,x}(\mathrm{d}\omega'),$$

where $\mathcal{B}_{s,r}\big(D([0,\infty),\mathbb{R}^d)\big)$ is the $]s,r]$-Borel sets, i.e, the σ-algebra generated by cylinders of the forms $\{\omega(r_1) \in A_1,\ \ldots, \omega(r_n) \in A_n\}$, $s < r_1 \leq \cdots < r_n \leq r$, $A_i, \ldots, A_n$ are Borel subsets of $\mathbb{R}^d$, for any integer $n \geq 1$.

The family of measures $\{P(s,x,t,\mathrm{d}y),\ t > s \geq 0,\ x \in \mathbb{R}^d\}$ on $\mathbb{R}^d$ defined by

$$P(s,x,t,A) := P_{s,x}\{\omega : \omega(t) \in A\}, \quad \forall A \in \mathcal{B}(\mathbb{R}^d)$$

is called the *transition function.*

(4) if the transition function is time-invariant, i.e., $P(s,x,t,A) = P(s+r,x,t+r,A)$ for every $r \geq 0$, then either transition function $P(x,t,\mathrm{d}y) := P(0,x,t,\mathrm{d}y)$, $t > 0$, $x \in \mathbb{R}^d$ or the family of probability measures $P_x := P_{0,x}$, $x \in \mathbb{R}^d$) is relevant and called (time) homogeneous.

On the other hand, a Markov probability measures $(P_{s,x} : s \geq 0, x \in \mathbb{R}^d$ is called a *diffusion with jumps* determined by an integro-differential operator $(A(t) : t \geq 0)$ as in (5.188) if

(5) for every $s \geq 0$, x in $\mathbb{R}^d$ and any test function φ in $C_0^\infty(\mathbb{R}^d)$, the process

$$M_\varphi(t) := \varphi(\omega(t)) - \varphi(\omega(s)) - \int_s^t A(r)\varphi(\omega(r))\mathrm{d}r, \quad \forall t \geq s,$$

is a (bounded) martingale with respect to the probability measure $P_{s,x}$. If the coefficients are independent of t, i.e., $A(t) = A$, then $P_x := P_{0,x}$ is called *homogeneous* diffusion with jumps. □

A successive application of the Markov property (3) of Definition 5.51 yields

$$\begin{aligned} &P_{s,x}\big(\omega(t_1) \in A_1,\, \omega(t_2) \in A_2, \ldots, \omega(t_n) \in A_n,\big) = \\ &= \int_{A_1} P(s,x,t_1,\mathrm{d}y_1) \int_{A_2} P(t_1,y_1,t_2,\mathrm{d}y_2) \ldots \int_{A_n} P(t_{n-1},y_{n-1},t_n,\mathrm{d}y_n), \end{aligned}$$

for every $0 \leq t_1 < t_2 < \cdots < t_n$, A_i in $\mathcal{B}(\mathbb{R}^d)$, $1 \leq i \leq n$. It is then clear that two Markov measures with the same transition function coincide. However, an integro-differential operator determines a version of a diffusion with jumps, i.e., the uniqueness of $P_{s,x}$ is usually related with the continuity properties in (s,x). On the other hand, we also have

$$\lim_{t \downarrow s} P(s,x,t,\{y : |y-x| \geq \varepsilon\}) = 0,$$

for every $\varepsilon > 0$, any $s \geq 0$ and x in $\mathbb{R}^d$. Nevertheless, continuity in the variables (s,x) should be assumed if needed. Indeed, the transition function $P(s,x,t,\mathrm{d}y)$ is called a (*strong*) *Feller* transition if for any $t > s \geq 0$, any x in $\mathbb{R}^d$ and every bounded and continuous (bounded and Borel measurable) function h the mapping

$$(s,x) \mapsto \int_{\mathbb{R}^d} h(y) P(s,x,t,\mathrm{d}y) \tag{5.193}$$

is continuous. Setting $x(t,\omega) := \omega(t)$ the canonical process in $D([0,\infty),\mathbb{R}^d)$ we see that a diffusion with jumps $P_{s,x}$ solves the martingale problem relative to the integro-differential operator $A(t)$. If not jumps kernel is given, i.e., $\mathtt{M}(\mathrm{d}z,s,x(s)) = 0$ then $P_{s,x}$ is a probability measure in the canonical space $C([0,\infty),\mathbb{R}^d)$ of continuous processes.

A Markov probability measures $(P_{s,x} : s \geq 0,\, x \in \mathbb{R}^d$ on the canonical space of cad-lag functions $D([0,\infty),\mathbb{R}^d)$ or continuous functions $C([0,\infty),\mathbb{R}^d)$ is called *strong Markov probability measures* if the property (3) of Definition 5.51 holds true for any stopping times. To properly state this condition we need some notation. As above $\mathcal{B}_{s,t}\big(D([0,\infty),\mathbb{R}^d)\big)$ denotes the $]s,r]$-Borel sets, i.e, the σ-algebra generated by cylinders of the forms $\{\omega(r_1) \in A_1, \ldots, \omega(r_n) \in A_n\}$, $s < r_1 \leq \cdots < r_n \leq t$, $A_i, \ldots, A_n$ are Borel subsets of $\mathbb{R}^d$, for any integer $n \geq 1$. The universally complete filtration $\bar{\mathcal{B}}_{s,t}\big(D([0,\infty),\mathbb{R}^d)\big)$ with respect to a Markov probability measures $(P_{s,x} : s \geq 0,\, x \in \mathbb{R}^d$ is defined as the minimal filtration (i.e., right-continuous) which contains all negligible sets with respect to the probabilities $P_{r,x}$ for every $s \leq r \leq t$ and x in $\mathbb{R}^d$. A s-stopping time is a

measurable variable $\tau > s$ on $D([0,\infty),\mathbb{R}^d)$ such that $\{\omega : \tau(\omega) \le t\}$ belongs to $\bar{\mathcal{B}}_{s,t} = \bar{\mathcal{B}}_{s,t}\big(D([0,\infty),\mathbb{R}^d)\big)$ for every $t > s$. Then $\bar{\mathcal{B}}_{s,\tau} = \bar{\mathcal{B}}_{s,\tau}\big(D([0,\infty),\mathbb{R}^d)\big)$ is the σ-algebra generated by Borel sets A satisfying $A \cup \{\tau \le t\}$ belongs to $\bar{\mathcal{B}}_{s,t}$. Hence, the strong Markov property is written as

$$P_{s,x}(B \cap \{\omega : \omega(\tau(\omega)+t) \in A\}) = \\ = \int_B P_{\tau(\omega'),\omega'(\tau(\omega'))}(\omega : \omega(\tau(\omega)+t) \in A)\, P_{s,x}(\mathrm{d}\omega'),$$

for any subset B in $\bar{\mathcal{B}}_{s,\tau}\big(D([0,\infty),\mathbb{R}^d)\big)$, or equivalently (in term of the regular conditional probability)

$$P_{s,x}\big\{\omega : \omega(\tau(\omega)+t) \in A \,\big|\, \bar{\mathcal{B}}_{s,\tau}\big\}(\mathrm{d}\omega') = \\ = P_{\tau(\omega'),\omega'(\tau(\omega'))}\big(\omega : \omega(\tau(\omega)+t) \in A\big)\, P_{s,x}(\mathrm{d}\omega'),$$

for every Borel subset A of $\mathbb{R}^d$, every s-stopping time τ, any $t \ge 0$. This expresses the fact that the canonical process $x(t,\omega) := \omega(t)$ is a strong Markov process in $D([0,\infty),\mathbb{R}^d)$ with the transition function $P(s,x,t,\mathrm{d}y)$.

Proposition 5.52. *Let $(P_{s,x} : s \ge 0,\ x \in \mathbb{R}^d)$ be a diffusion with jumps determined by an integro-differential operator $(A(t) : t \ge 0)$ as in* Definition 5.51. *Assume that the martingale problem relative to $(A(t) : t \ge 0)$ has the uniqueness property, i.e., if $(P'_{s,x} : s \ge 0,\ x \in \mathbb{R}^d)$ is another Markov probability measures then for every real-valued smooth function h with a compact support in $\mathbb{R}^d$ we have either*

$$\mathbb{E}_{s,x}\{h(\omega(t))\} = \mathbb{E}'_{s,x}\{h(\omega(t))\}, \quad \forall t > s \ge 0,\ x \in \mathbb{R}^d,$$

or

$$\int_s^\infty \mathrm{e}^{-\lambda(t-s)}\mathbb{E}_{s,x}\{h(\omega(t))\}\mathrm{d}t = \int_s^\infty \mathrm{e}^{-\lambda(t-s)}\mathbb{E}'_{s,x}\{h(\omega(t))\}\mathrm{d}t,$$

for every $\lambda > 0$, $s \ge 0$ and x in $\mathbb{R}^d$, where $\mathbb{E}_{s,x}$ and $\mathbb{E}'_{s,x}$ denotes the mathematical expectation with respect to $P_{s,x}$ and $P'_{s,x}$. Then $(P_{s,x} : s \ge 0,\ x \in \mathbb{R}^d)$ is a strong *diffusion with jumps, i.e., the above strong Markov property is satisfied.*

Proof. It is clear that the above assumption yields the uniqueness in law of the martingale problem, i.e., $P_{s,x} = P'_{s,x}$, for every $s \ge 0$ and x in $\mathbb{R}^d$. Since the process $M_\varphi(t)$ is a $P_{s,x}$-martingale, it is also a martingale relative to the universally completed filtration $\bar{\mathcal{B}}_{s,\infty}\big(D([0,\infty),\mathbb{R}^d)\big)$ and by Doob's optimal sampling theorem, the process $t \mapsto M_\varphi(t+\tau)$ is also a martingale relative to $\bar{\mathcal{B}}_{\tau,\infty}\big(D([0,\infty)\big)$, for any bounded s-stopping time τ. In particular, for every $t > s \ge 0$ A in $\bar{\mathcal{B}}_{s,\tau}\big(D([0,\infty)\big)$ and B in $\bar{\mathcal{B}}_{\tau,\infty}\big(D([0,\infty)\big)$ we have

$$\mathbb{E}_{s,x}\big\{\big[M_\varphi(\tau+t) - M_\varphi(\tau+s)\big]\,\mathbb{1}_A\,\mathbb{1}_B\big\} = 0,$$

which remains true for any s-stopping time not necessarily bounded, e.g., replace τ by $\tau \wedge n$, with $n \to \infty$. Hence, if $P^{\omega'}$ is a regular conditional probability of

$P_{s,x}$ given $\bar{\mathcal{B}}_{\tau,\infty}\big(D([0,\infty)\big)$ then $P^{\omega'}\{\omega : \omega(\tau(\omega)) = \omega(\tau(\omega'))\} = 1$. By the uniqueness of the martingale problem, $P^{\omega'} = P_{\tau(\omega'),\omega'(\tau(\omega'))}$, which proves the desired result. □

It is clear that we may replace $\mathbb{R}^d$ with any locally compact Polish space $\mathcal{O}$ and the previous Proposition remains true. Moreover, if $\mathcal{O}$ is not locally compact, then we need to take a larger family of smooth functions φ, for instance, bounded functions with a bounded support if $\mathcal{O}$ is a Banach space.

The martingale problem can be considered for any continuous and linear operator $(A(t) : t \geq 0)$ from $C_0^\infty(\mathbb{R}^d)$ into $C_b^0(\mathcal{B}^d)$, not necessarily an integro-differential operator of the form (5.188). Moreover, $(A(t) : t \geq 0)$ may be a multi-valued operator as in Ethier and Kurtz [76, Chapter 4, pp. 155–274]. Thus, a direct consequence of the previous Proposition 5.52 is the following fact: if $P_{s,x}$ is the unique solution of a martingale problem with $P_{s,x}\{\omega(s) = x\} = 1$ and the mapping $(s,x) \mapsto P_{s,x}$ then $(P_{s,x} : s \geq 0,\, x \in \mathbb{R}^d)$ is a strong diffusion with jumps.

Suppose the coefficients $a_{ij}(t,x)$, $a_i(t,x)$ satisfy

$$a_{ij} := \frac{1}{2}\sum_{k=1}^{n} \sigma_{ik}\,\sigma_{jk}, \qquad a_i := g_i + \sum_{k=1}^{n} b_k\,\sigma_{ik}, \tag{5.194}$$

where g_i, b_k and σ_{ik} are (Borel) measurable and for some constant $C > 0$ and any t, x, we have

$$x \cdot g(t,x) + \sum_{k=1}^{n} |\sigma_k(t,x)|^2 \leq C\,(1 + |x|^2),$$
$$\sum_k |b_k(t,x)| \leq C,$$

and for every $r > 0$ there is a constant C_r such that

$$|g(t,x) - g(t,x')| + \sum_{k=1}^{n} |\sigma_k(t,x) - \sigma_k(t,x')| \leq C_r\,|x - x'|,$$

for any $0 \leq t \leq r$, $|x| \leq r$, $|x'| \leq r$. The Lévy or jump kernel $\mathtt{M}(\mathrm{d}z,t,x)$ has the form

$$\begin{cases} \mathtt{M}(B,t,x) := \pi(\{\zeta : \mathtt{j}(\zeta,t,x) \in B\}) + \\ \qquad\qquad + \displaystyle\int_{\{\zeta\,:\,\mathtt{k}(\zeta,t,x)\in B\}} \mathtt{c}(\mathtt{k}(\zeta,t,x),t,x)\,\pi(\mathrm{d}\zeta), \end{cases} \tag{5.195}$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$, $t \geq 0$, and x in $\mathbb{R}^d$, where all coefficients are measurable and π is a Radon measure on $\mathbb{R}^m_*$, such that there exit some positive constant C, c_0 and some measurable real-valued function $\bar{\jmath}(\zeta)$, such that for every ζ, t, x, z

we have

$$|\mathtt{j}(\zeta,t,x)| + |\mathtt{k}(\zeta,t,x)| \le (1+|x|)\,\bar\jmath(\zeta),$$
$$\int_{\mathbb{R}^m_*} \bar\jmath^2(\zeta)\pi(\mathrm{d}\zeta) \le C,$$
$$-1 < -c_0 \le \mathtt{c}(z,t,x) \le C,$$
$$|\mathtt{c}(\mathtt{k}(\zeta,t,x),t,x)| \le \bar\jmath(\zeta),$$

and for every $r>0$ there is a constant C_r such that

$$|\mathtt{j}(\zeta,t,x) - \mathtt{j}(\zeta,t,x')| + |\mathtt{k}(\zeta,t,x) - \mathtt{k}(\zeta,t,x')| \le C_r\,\bar\jmath(\zeta)\,|x-x'|,$$

for any $0\le t\le r$, $|x|\le r$, $|x'|\le r$, where $\bar\jmath(\zeta)$ is a square-integrable with respect to π as above, and a posteriori taken to be the same function.

Based on the strong existence and uniqueness of stochastic differential equations Theorem 5.3 and Girsanov transformation, we deduce

Theorem 5.53. *Let $(A(t) : t \ge 0)$ be an integro-differential operator as in (5.188) where the coefficients satisfy conditions (5.194) and (5.195). Then the corresponding diffusion with jumps can be defined as a strong Markov process. Moreover, if the coefficients $b(t,x)$, $\mathtt{k}(\zeta,t,x)$ and $\mathtt{c}(z,t,x)$ are locally uniform continuous in x then the diffusion with jumps satisfies the Feller property (5.193).*

Proof. Under these assumptions the martingale problem has one and only one solution, so that the previous Proposition 5.52 proves that the diffusion with jumps is indeed a strong Markov process. On the other hand, if Girsanov transformation is not used, i.e., the coefficients $b(t,x)$, $\mathtt{k}(\zeta,t,x)$ and $\mathtt{c}(z,t,x)$ vanish, then the estimates for the strong solution of stochastic differential equations Theorem 5.11 shows the Feller property (5.193). In general, if all coefficients are locally uniform continuous in x then the expression of the density process (used to applied Girsanov transformation) is continuous in $x(\cdot)$, and the Feller character is preserved. □

There is a case not included in the previous theorem, which is related to uniqueness question of the martingale problem with continuous coefficients, cf. Jacod and Shiryaev [117, Theorem III.2.34, p. 146].

Theorem 5.54. *Assume that the diffusion matrix $a_{ij}(t,x)$ is bounded, invertible and continuous everywhere, that the drift vector $a_i(t,x)$ is measurable and bounded, and that the Lévy kernel $\mathtt{M}(\mathrm{d}z,t,x)$ is such that for every B in $\mathcal{B}(\mathbb{R}^d_*)$ the function*

$$(t,x) \mapsto \int_B |z|^2 \wedge |z|\,\mathtt{M}(\mathrm{d}z,t,x)$$

is continuous, uniformly in B. Then the martingale problem has one and only one solution, which is a strong Markov process satisfying the strong Feller property. □

The proof for the diffusion case, i.e. without the jump kernel $\mathtt{M}(\mathrm{d}z, t, x)$, is very well know and several variations on the above assumptions are possible, e.g. see Stroock and Varadhan [241, Chapter 7, pp. 171–190]. The general case is very similar. The existence is handled a way similar to Theorem 5.48, but the uniqueness and the strong Feller property are obtained from analytic results on partial differential equations. The arguments are as follows: on one hand, for any real-valued smooth function φ with a compact support in $\mathbb{R}^d$ we consider the integro-partial differential equation

$$\begin{cases} \partial_s \psi(s,x) = -A(s)\varphi(x), \quad \forall s < t,\, x \in \mathbb{R}^d, \\ \psi(t,x) = \varphi(x), \quad \forall x \in \mathbb{R}^d, \end{cases}$$

and on the other hand, an integral version of Itô formula (or equivalently the so-called Dynkin formula) is applied to the function ψ to deduce

$$\psi(s,x) = \mathbb{E}_{s,x}\{\varphi(x(t))\} = \int_{\mathbb{R}^d} \varphi(y) P(s,x,t,\mathrm{d}y),$$

Hence, the uniqueness of the solution ψ yields the uniqueness of the transition function $P(s,x,t,\mathrm{d}y)$ and so the uniqueness of the martingale problem. The uniqueness property for the above integro-partial differential equation in ψ is usually obtained from the weak maximum principle, and the solution ψ should be found in a class of function where Dynkin formula can be used. A priori estimates on the solution ψ show the Feller character of the diffusion with jumps. The continuity condition on the Lévy kernel can be replaced by the condition that the the function

$$(t,x) \mapsto \int_B |z|^2 \wedge 1\, \mathtt{M}(\mathrm{d}z, t, x),$$

is continuous, but in this case, the canonical process $x(t) = \omega(t)$ may not have a first order moment, due to the *large jumps*. When the coefficients have a linear growth in x, the (integro-)partial differential equation has some extra difficulties, for instance see Bensoussan and Lions [16, Section III.1.13, pp. 202–207], Besala [24] and Cerrai [36] for the diffusion case, and Menaldi and Robin [174] for diffusion with jumps.

Even if no uniqueness is ensured, we may be able to find a measurable selection from the solutions of a martingale problem and then to construct a strong Markov process, this is the case under the assumptions of Theorem 5.48. For the sake of completeness we rewrite the integro-differential operator $A(t)$ as

$$\begin{cases} A(t)\varphi(x) = \displaystyle\sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij}\varphi(x) + \sum_{i=1}^{d} g_i(t,x)\partial_i\varphi(x) + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,t,x), \end{cases} \tag{5.196}$$

where $a_{ij} = a_{ji}$, all coefficients $a = (a_{ij})$ and $g = (g_i)$ are measurable satisfying the growth (linear) condition

$$x \cdot g(t,x) + |a(t,x)| \le C\,(1 + |x|^2),$$

for every (t,x) in $[0,\infty)\times\mathbb{R}^d$ and some positive constant C.

$$\mathtt{M}(B,t,x)=\int_{\{\zeta\,:\,\mathtt{j}(\zeta,t,x)\in B\}}\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta),\quad\forall B\in\mathcal{B}(\mathbb{R}^d_*),$$

where $\mathtt{j}(\cdot)$ and $\mathtt{m}(\cdot)$ are measurable functions taking values in $\mathbb{R}^d$ and $[0,1]$, respectively, and π is a Radon measure on $\mathbb{R}^m_*$ such that for some measurable real-valued function $\bar\jmath(\zeta)$ we have

$$\begin{aligned}&|\mathtt{j}(\zeta,t,x)|\,\mathtt{m}(\zeta,t,x)\le(1+|x|)\,\bar\jmath(\zeta),\\&\int_{\mathbb{R}^m_*}\bar\jmath^{\,2}(\zeta)\pi(\mathrm{d}\zeta)\le C,\end{aligned}$$

for every ζ,t,x and for some positive constant C. Again, these assumptions are referred to as measurability and a linear growth condition, even if the diffusion matrix $a(t,x)$ has at most a quadratic growth and the drift vector $g(t,x)$ may not has a linear growth in all directions.

Moreover, suppose that all coefficients $g(t,x)$, $a(t,x)$, $\mathtt{j}(\zeta,t,x)$ and $\mathtt{m}(\zeta,t,x)$ are *uniformly local continuous* x if for every $\varepsilon>0$ there exists a $\delta>0$ such that for every t, x, x' satisfying $0\le t-t_0\le 1/\varepsilon$, $|x-x'|<\delta$, $|x|\le 1/\varepsilon$, $|x'|\le 1/\varepsilon$, we have

$$\begin{cases}|g(t,x)-g(t,x')|+|a(t,x)-a(t,x')|\le\varepsilon,\\|\mathtt{j}(\zeta,t,x)-\mathtt{j}(\zeta,t,x')|\le\varepsilon\,\bar\jmath(\zeta),\\|\mathtt{m}(\zeta,t,x)-\mathtt{m}(\zeta,t,x')|\le\varepsilon,\end{cases}\tag{5.197}$$

where $\bar\jmath(\zeta)$ is the square-integrable function used above.

Theorem 5.55. *Let $(A(t)\,:\,t\ge0)$ be an integro-differential operator as in (5.196) satisfying (5.197). Then the solution of the martingale problem $P_{s,x}$ with initial condition $x(s)=x$ forms a non-empty compact set with respect to the weak convergence of probability measures defined on the canonical space $D([0,\infty),\mathbb{R}^d)$. Furthermore, there exists a measurable selection which is a diffusion with jumps satisfying the strong Markov property.* □

The diffusion case (which has almost no differences with the case of diffusion with jumps) follows from Krylov [137]. The fact that the solutions form a non-empty compact set is essentially proved in Theorem 5.48. The measurable selection is based on the following result. First, if Q is a given separable metric space with a metric $\mathrm{d}(\cdot,\cdot)$ then $\mathrm{C_p}(Q)$, the class of all compact subsets of Q, with the metric $\mathrm{d}_c(\cdot,\cdot)$ is also a separable metric space, the metric $\mathrm{d}_c(K_1,K_2)$ is defined as the infimum of all $\varepsilon>0$ such that $K_1\subset K_2^\varepsilon$ and $K_2\subset K_1^\varepsilon$, where

$$K^\varepsilon:=\{q\in Q\,:\,\mathrm{d}(q,p)<\varepsilon,\text{ for some }p\in K\},$$

i.e., K^ε is a ε-ball around K. It is clear that Q is embedded isometrically in $\mathrm{C_p}(Q)$.

Theorem 5.56. *Let $(E,\mathcal{E})$ a measurable space and consider $e \mapsto K_e$ a measurable map from E into $\mathrm{C_p}(Q)$. Then there is a measurable map $e \mapsto k(e)$ from E into Q such that $k(e)$ belongs to K_q for every e in E.* □

The proof of this result and a complete discussion can be found in Stroock and Varadhan [241, Chapter 12, pp. 285–303].

From the control point of view, a very important question is the existence of a solution under measurability assumption, without any condition of continuity on the variable x. This is due to the fact that the control process $v(t)$ is Markov feedback, i.e., $v(t) = k(x(t-))$ for some measurable function $k(\cdot)$. The optimal feedback $k(\cdot)$ is obtained as a minimizer of some equation, which in most of the cases is only measurable, no condition of continuity can be assumed, cf. Krylov [139].

A proof of the following result, for non-degenerate diffusion with jumps and measurable coefficients with $\mathtt{m} = 1$, can be found in the paper Anulova and Pragarauskas [5]. Thus, assume that the diffusion matrix is non-degenerate, i.e., there exists a constant $a_0 > 0$ such that

$$\xi \cdot a(t,x)\xi \geq a_0|\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \tag{5.198}$$

for every $t \geq 0$ and x in $\mathbb{R}^d$. Similar to Theorem 5.55 we have

Theorem 5.57. *Let $(A(t) : t \geq 0)$ be an integro-differential operator as in (5.196) satisfying only the linear growth condition and the non-degeneracy assumption (5.198). Then the solution of the martingale problem $P_{s,x}$ with initial condition $x(s) = x$ forms a non-empty compact set with respect to the weak convergence of probability measures defined on the canonical space $D([0,\infty),\mathbb{R}^d)$. Furthermore, there exists a measurable selection which is a diffusion with jumps satisfying the strong Markov property.*

Proof. We discuss only the existence of the martingale problem under non-degeneracy and measurability conditions. The arguments are the same as in Theorem 5.48, by means of the strong solutions and Girsanov transformation, we find a solution of the martingale problem $P^n = P^n_{s,x}$ (we take $s = 0$ for simplicity) with smooth coefficients $g_n(t,x)$, $a_n(t,x)$, $\mathtt{j}_n(\zeta,t,x)$ and $\mathtt{m}_n(\zeta,t,x)$, i.e., if $A_n(t)$ denotes the integro-differential operator with these smooth coefficients then the process for $t \geq s \geq 0$,

$$M^n_{\varphi,s}(t) := \varphi(x(t)) - \varphi(x(s)) - \int_s^t A_n(r)\varphi(x(r))\mathrm{d}r,$$

is a (bounded) martingale with respect to P^n, for any $n \geq 1$, for every test function φ, where $x(t) := \omega(t)$ is the canonical process defined on the canonical space $D([0,\infty),\mathbb{R}^d)$.

These coefficients satisfy a linear growth condition and a non-degeneracy assumption uniformly in n, with $\mathtt{m}_n(\zeta,t,x) = 0$ if $|\zeta| \leq 1/n$ or $|\zeta| \geq n$ and $1/n \leq \mathtt{m}_n(\zeta,t,x) \leq 2$, otherwise. Moreover, $g_n \to g$, $a_n \to a$, $\mathtt{j}_n(\zeta,\cdot,\cdot) \to \mathtt{j}(\zeta,\cdot,\cdot)$ and $\mathtt{m}_n(\zeta,\cdot,\cdot) \to \mathtt{m}(\zeta,\cdot,\cdot)$ in $L^p(]0,T[\times B)$ for every bounded measurable subset B of $\mathbb{R}^d$, any $p \geq d+1$ and any ζ in $\mathbb{R}^m_*$. Recall that we may also assume

that $|\mathtt{j}_n(\zeta,t,x)| \le (1+|x|)\,\bar\jmath(\zeta)$, where $\bar\jmath(\zeta)$ is square-integrable with respect to π. Thus, for every test function φ, any bounded subset B of $\mathbb{R}^d$ and $T>0$ we have

$$\lim_n \int_0^T \mathrm{d}t \int_B |A_n(t)\varphi(x) - A(t)\varphi(x)|^p \mathrm{d}x = 0,$$

for every $p \ge d+1$. Furthermore, because φ has a compact support, the assumptions on the coefficients imply that the sequence $\{A_n(t)\varphi(x) : n \ge 1\}$ is equi-bounded in $[0,\infty \times \mathbb{R}^d$.

Since the coefficients satisfy the linear growth condition uniformly in n, we also have the following property (referred to as bounded and continuous in probability): for every $\varepsilon > 0$ there is a $\delta > 0$ independent of n such that

$$P^n\{\sup_{0\le t\le 1/\varepsilon} |x(t)| \ge 1/\delta\} + \\ + P^n\{\sup_{0\le r\le \delta} |x(s+r) - x(s)| \ge \varepsilon\} \le \varepsilon.$$

This proves that the set of probability measures P^n is tight (or pre-compact) and that there exists a subsequence (still denoted by P^n) which converges weakly to some limit P.

Hence, the arguments in Theorem 4.69 of the previous chapter show that

$$\mathbb{E}^n\{M^n_{\varphi,s}(t) \mid \mathcal{F}(r)\} \to \mathbb{E}\{M_{\varphi,s}(t) \mid \mathcal{F}(r)\}, \quad \text{as} \quad n \to \infty,$$

for every $t \ge r \ge s \ge 0$, where

$$M_{\varphi,s}(t) := \varphi(x(t)) - \varphi(x(s)) - \int_s^t A(r)\varphi(x(r))\mathrm{d}r,$$

for every $t \ge s \ge 0$. This proves that the limiting probability P is a solution of the martingale problem relative to the integro-differential operator $A(t)$ given above. □

Notice that for any measurable and bounded Lévy measures, i.e., when $\mathtt{M}(\mathbb{R}^d_*, s, x) \le C < \infty$, for every x in $\mathbb{R}^d$ and $s \ge 0$, the diffusion with jumps can be constructed, without any particular structure on the kernel, e.g., see Bensoussan and Lions [17, Section 3.7.5, pp. 274–280]. Thus, assume that the integro-differential operator $A(t)$ can be approximated by a sequence $\{A^n(t) : n \ge 1\}$ satisfying uniformly the measurability and growth conditions, with bounded Lévy measures and such that $A^n(t)\varphi(x)$ is equi-bounded and converges to $A(t)\varphi(x)$ in $L^{d+1}_{\mathrm{loc}}(]0,\infty[\times\mathbb{R}^d)$, for every smooth function φ with a compact support in $\mathbb{R}^d$. Then the above arguments can be used to prove the existence of a strong diffusion with jumps.

5.8.2 Analytic Viewpoint

From the analysis (rather than probability) point of view, we are interested in the Markov semigroup or backward evolution operator

$$\Phi(s,t)h(x) := \mathbb{E}\{h\big(x(t)\big) \,\big|\, x(s) = x\}, \quad \forall t > s \geq 0,\, x \in \mathbb{R}^d,$$

defined on Borel and bounded function h. Its infinitesimal generator given by the expression

$$\begin{aligned} A(s)h &:= \lim_{t \to s^+} \frac{\Phi(s,t)h(x) - h(x)}{t-s}, \quad s \geq 0, \\ A(t)h &:= \lim_{s \to t^-} \frac{h(x) - \Phi(s,t)h(x)}{t-s}, \quad t > 0, \end{aligned}$$

which define the operator and its domain of definition. The homogeneous case, i.e., when $\Phi(s,t) = \Phi(t-s)$ and $A(s) = A$, is better well known. It is clear that by adding one more dimension, the variable t can be regarded as part of the state variable x, however, the assumptions on t are generally weaker than those on x. Using the transition function $P(s,x,t,\mathrm{d}y)$ we can rewrite the semigroup as

$$\Phi(s,t)h(x) := \int_{\mathbb{R}^d} h(y) P(s,x,t,\mathrm{d}y), \quad \forall t > s \geq 0,\, x \in \mathbb{R}^d.,$$

for any h in $B(\mathbb{R}^d)$ (real-valued Borel and bounded functions on $\mathbb{R}^d$).

The formal connection between the transition function $P(s,x,t,\mathrm{d}y)$ and the integro-partial differential equation is the Chapman-Kolmogorov identities, namely, the *backward* equation

$$\begin{aligned} &\partial_s P(s,x,t,\mathrm{d}y) = -A(s)P(s,x,t,\mathrm{d}y), \quad \forall s \in [0,t),\, x \in \mathbb{R}^d,\, \mathrm{d}y \\ &\lim_{s \uparrow t} P(s,x,t,\mathrm{d}y) = \delta_x(\mathrm{d}y), \quad \forall t > 0,\, x \in \mathbb{R}^d,\, \mathrm{d}y \end{aligned}$$

and the *forward* equation

$$\begin{aligned} &\partial_t P(s,x,t,\mathrm{d}y) = A^*(t)P(s,x,t,\mathrm{d}y), \quad \forall t > s \geq 0,\, x \in \mathbb{R}^d,\, \mathrm{d}y, \\ &\lim_{t \downarrow s} P(s,x,t,\mathrm{d}y) = \delta_x(\mathrm{d}y), \quad \forall s \geq 0,\, x \in \mathbb{R}^d,\, \mathrm{d}y, \end{aligned}$$

where $\delta_x(\cdot)$ is the Dirac delta measure with unit mass at x, and the operator $A(s)$ is acting on the variable x as usually, but $A^*(t)$ is the adjoint (or dual) operator acting on the y variable, i.e.

$$\int_{\mathbb{R}^d} \varphi(y) \big(A^*(t)P(s,x,t,\cdot)\big)(\mathrm{d}y) := \int_{\mathbb{R}^d} \big(A(t)\varphi\big)(y)\, P(s,x,t,\mathrm{d}y),$$

for every test function φ in $C_0^\infty(\mathbb{R}^d)$. Clearly, the most interesting case is when $P(s,x,t,\mathrm{d}y)$ has a density with respect to the Lebesgue measure, i.e.,

$$P(s,x,t,\mathrm{d}y) := p(s,x,t,y)\,\mathrm{d}y,$$

and $p(s,x,t,y)$ satisfies the backward and forward equations in the usual sense. Note that the backward equation has a terminal condition, whilst the forward equation has a initial condition (in the simplest case $A(s) = A^*(t) = \nabla^2$ is the Laplacian).

As mentioned above, the backward equation follows from Dynkin or Itô formula, and the forward equation can be derived by means of an integration by parts, but it requires much more regularity on the coefficients of $A(t)$. The transition density function $p(s,x,t,y)$ is essentially the main fundamental solution or Green function associated with the parabolic integro-differential operator $\partial_t + A(t)$ with a terminal condition (or $\partial_t - A(t)$ with an initial condition). A comprehensive study on the transition density function $p(s,x,t,y)$ (or Green function) corresponding to non-degenerated diffusion with jumps can be found in Garroni and Menaldi [93, 94], where full details are given only for bounded domains with oblique boundary conditions. However, most of the estimates remain true for the whole space, where the technique is neater and easier. The purely diffusion case is much more well known, and the construction and estimates of the Green function involves the heat kernel, e.g., see the books Eidelman [72], Ladyzhenskaya et al. [147], Lieberman [153], Gilbarg and Trudinger [100], and the papers Ivasišen [114], Solonnikov [232, 233], among others.

On the other hand, if the diffusion term $a(t,x)$ is degenerate (i.e., the condition (5.198) above is not satisfied) then a transition density function is not guarantee. However, the differentiability of the strong solution with respect to the initial data, see Proposition 5.22, ensures that the function

$$u(s,x) := \int_{\mathbb{R}^d} h(y)\, P(s,x,t,\mathrm{d}y) + \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} f(r,y)\, P(s,x,r,\mathrm{d}y) = \\ = \mathbb{E}_{s,x}\{h(x(t))\} + \int_s^t \mathbb{E}_{s,x}\{f(r,x(r))\}\mathrm{d}r$$

is twice-continuously differentiable if the coefficients, f and h are so. Hence

$$\partial_s u(s,x) + A(s)u(s,x) = f(s,x), \quad \forall s \in [0,t),\ x \in \mathbb{R}^d,$$
$$u(t,x) = h(x), \quad \forall x \in \mathbb{R}^d,$$

which is the actual meaning of the backward equation.

This connection with the partial differential equations is of fundamental importance, and a carefully study is needed. For instance, we refer to Bensoussan and Lions [16, 17], Cerrai [36], Doob [60], Engel and Nagel [74], Freidlin [87], Friedman [90], Gihman and Skorohod [97], Taira [243], among others.

To be specific, let us rewrite the integro-differential operator $A(s)$ given by (5.188) as $A(s) = I(s) - L(s)$, where the proper integral part $I(s)$ and the proper differential part $L(s)$ are given by

$$\begin{cases} L(s)\varphi(x) = -\displaystyle\sum_{i,j=1}^d a_{ij}(s,x)\partial_{ij}\varphi(x) - \sum_{i=1}^d a_i(s,x)\partial_i\varphi(x), \\ I(s)\varphi(x) = \displaystyle\int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,s,x), \end{cases} \tag{5.199}$$

where $a_{ij} = a_{ji}$, all coefficients a_{ij} and a_i are measurable, with a structural condition on the Lévy kernel $\mathtt{M}(\mathrm{d}z, s, x)$ of the form

$$\mathtt{M}(B, s, x) = \int_{\{\zeta\,:\,\mathtt{j}(\zeta,s,x)\in B\}} \mathtt{m}(\zeta, s, x)\pi(\mathrm{d}\zeta), \tag{5.200}$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$, where $\mathtt{j}(\cdot)$ and $\mathtt{m}(\cdot)$ are measurable functions taking values in $\mathbb{R}^d$ and $[0,1]$, respectively, and π is a Radon measure on $\mathbb{R}^m_*$ such that for some measurable real-valued function $\bar{\jmath}(\zeta)$ we have

$$\begin{cases} |\mathtt{j}(\zeta, s, x)|\, \mathtt{m}(\zeta, s, x) \le (1 + |x|)\, \bar{\jmath}(\zeta), \\ \displaystyle\int_{\mathbb{R}^m_*} \bar{\jmath}^2(\zeta)\pi(\mathrm{d}\zeta) \le C, \end{cases} \tag{5.201}$$

for every ζ, s, x and for some positive constant C. We may add a zero-order term to the expression of $L(s)$, i.e., to replace $L(s)\varphi(x)$ by $L(s)\varphi(x) + a_0(s,x)\varphi(x)$, which corresponds to kill the initial Markov process with an exponential probability density of parameter $a_0(s,x) \ge 0$.

As discussed in Chapter 2, classic results (e.g. see the books by Friedman [89] or Ladyzhenskaya et al. [147]) ensure (under some assumptions) the existence, uniqueness and suitable estimates (in term of the heat kernel) of the principal fundamental solution (or Green function) $p_L(s, x, t, y)$ corresponding to the parabolic second order partial differential operator $\partial_s - L(s)$ with a terminal condition. For instance, besides uniform ellipticity, typical assumptions are bounded and Hölder continuous coefficients $a_{ij}(s,x)$, $a_i(s,x)$ for (s,x) in $[0,T]\times\mathbb{R}^d$, for every $T > 0$, which yield a Hölder continuous Green function in all its variables (i.e., x, y in $\mathbb{R}^d$, $0 \le s < t \le T$), with Hölder continuous second derivatives in x and first derivative in s, bounded relative to the heat kernel. If the lower order coefficients $a_i(s,x)$ are not longer Hölder continuous then the Green function $p_L(s,x,t,y)$ is not necessarily Hölder continuous in (t,y) and the second derivative in x and first derivative in s are not longer Hölder continuous, however, bounds in term of the heat kernel remain true. Clearly, this involves the so called strong (and weak) solution in Sobolev spaces, which can be pushed even to the case of only continuous higher order coefficients $a_{ij}(s,x)$ (even measurable in (s,x), continuous in x and locally bounded in s) and measurable and bounded (even a linear growth in x is allowed) lower order coefficients $a_i(s,x)$. There is a large variety of results and huge list of references regarding parabolic second order differential equations. From the probabilistic point of view, the reader may consult the statements quoted in the book Stroock and Varadhan [241, Chapter 3 and Appenix], where a L^p-type estimate (involving the heat kernel) is proved for the particular case of coefficients independent of x but not enough in s, see also the book Eidelman [72].

As stated in Section 2.12, the addition of the integral part $I(s)$ destroys the explicit presence of the heat kernel in the estimates of the Green function, so that only a set of key properties are preserved, see Garroni and Menaldi [93, 94].

If p_L denotes the Green function associated with the differential operator $\partial_s + L(s)$ then Green function p associated with the integro-differential operator

$\partial_s + L(s) - I(s)$, we solve a Volterra equation

$$\begin{cases} \text{either} & \text{find } Q_I \text{ such that} & Q_I = Q_L + Q_L \bullet Q_I\,, \\ \text{or} & \text{find } p \text{ such that} & p = p_L + p_L \bullet I(\cdot)p\,, \end{cases}$$

with the relations $Q_L = I(\cdot)p_L$ and $p = p_L + p_L \bullet Q_I$. Recall that the bullet $\bullet$ means the kernel-convolution, i.e., for any $\varphi(x,t,y,s)$ and $\psi(x,t,y,s)$

$$(\varphi \bullet \psi)(x,t,y,s) := \int_0^T \mathrm{d}\tau \int_{\mathbb{R}^d} \varphi(x,t,z,\tau)\psi(z,\tau,y,s)\mathrm{d}z,$$

for every $0 \le s < t \le T$, x and y in $\mathbb{R}^d$. Actually, we express Q_I as the following series

$$Q_I = \sum_{n=1}^{\infty} Q_n\,, \qquad Q_0 = Q_L, \quad Q_n = Q_L \bullet Q_{n-1}\,,\ n \ge 1,$$

where the convergence is in the sense of following Green spaces. Notice that in Section 2.12 we used G and G_L instead of p and p_L, as well as initial condition instead of initial condition.

If p_λ denotes the Green function associated with $\partial_s + L(s) + \lambda(s)$, where $\lambda(s) \ge 0$ is a measurable bounded coefficient of zero-order, then a simple change of variable proves that

$$p_\lambda(s,x,t,y) = \mathrm{e}^{-\int_s^t \lambda(r)\mathrm{d}r} p_L(s,x,t,y),$$

for every x, y in $\mathbb{R}^d$ and $0 \le s < t$. Moreover, if the integral part $I(t)$ has the form

$$I(s)\varphi(x) = \int_{\mathbb{R}^m_*} [\varphi(x + \mathtt{k}(\zeta,s,x)) - \varphi(x)]\mathtt{m}(\mathrm{d}z,s),$$

with $\mathtt{m}(\mathbb{R}^m_*, s) \le C < \infty$, for every s in $[0,T]$, then the Green function p associated with $\partial_s + L(s) - I(s)$ satisfies the equation

$$p(s,x,t,y) = \mathrm{e}^{-\int_s^t \mathtt{m}(\mathbb{R}^m_*,r)\mathrm{d}r} p_L(s,x,t,y) + \int_s^t \mathrm{e}^{-\int_r^t \mathtt{m}(\mathbb{R}^m_*,u)\mathrm{d}u}\mathrm{d}r \times$$
$$\times \int_{\mathbb{R}^d} p(s,x,t,z)\mathrm{d}z \int_{\mathbb{R}^m_*} p_L(r, z + \mathtt{k}(\zeta,r,z), t, y)\mathtt{m}(\mathrm{d}z,s),$$

for every x, y in $\mathbb{R}^d$ and $0 \le s < t$. The reader is referred to the book Skorokhod [230, Section I.2, pp. 23–41] for more probabilistic arguments in this direction.

5.9 Symmetric Markov Jumps

A Markov jump process is characterized by its (proper) integro-differential operator

$$\text{either} \qquad I(t)\varphi(x) = \int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x)]\mathtt{M}(\mathrm{d}z,t,x),$$
$$\text{or} \qquad I(t)\varphi(x) = \int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,t,x),$$

depending on whether the Levy kernel $\mathtt{M}(\mathrm{d}z,t,x)$ integrates the function $z\mapsto|z|$ (jumps of bounded variation) or $z\mapsto|z|^2\wedge|z|$, both with finite first moment. In any case, the bounded kernel $\mathtt{M}_\varepsilon(\mathrm{d}z,t,x)=\mathtt{M}(\mathrm{d}z,t,x)\mathbb{1}_{\{|z|\geq\varepsilon\}}$ produces workable approximations, for most of the key arguments. The dependency of the kernel $\mathtt{M}(\mathrm{d}z,t,x)$ on the variables (t,x) is hard to track in general and structural assumptions are made, such as

$$\begin{cases} I(t)\varphi(x)=\displaystyle\int_{\mathbb{R}^m_*}[\varphi(x+\mathtt{j}(\zeta,t,x))-\varphi(x)- \\ \qquad\qquad -\mathtt{j}(\zeta,t,x)\cdot\nabla\varphi(x)]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{cases} \tag{5.202}$$

or, after using Taylor formula with $\mathtt{j}(\zeta):=\mathtt{j}(\zeta,t,x)$,

$$\begin{aligned} I(t)\varphi(x) &= \int_0^1\mathrm{d}\theta\int_{\mathbb{R}^m_*}\mathtt{j}(\zeta)\cdot[\nabla\varphi(x+\theta\mathtt{j}(\zeta))-\nabla\varphi(x)]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta)= \\ &= \int_0^1(1-\theta)\mathrm{d}\theta\int_{\mathbb{R}^m_*}\mathtt{j}(\zeta)\cdot\big[\nabla^2\varphi\big(x+\theta\mathtt{j}(\zeta)\big)\mathtt{j}(\zeta)\big]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{aligned}$$

as long as φ is smooth.

Contrary to the Wiener process, a space-time homogeneous Markov jump process is not necessarily symmetric, i.e., for

$$\begin{cases} \text{either} & I(t)\varphi(x)=\displaystyle\int_{\mathbb{R}^d_*}[\varphi(x+z)-\varphi(x)]\mathtt{M}(\mathrm{d}z), \\ \text{or} & I(t)\varphi(x)=\displaystyle\int_{\mathbb{R}^d_*}[\varphi(x+z)-\varphi(x)-z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z), \end{cases} \tag{5.203}$$

to be symmetric we need to require

$$\int_{\mathbb{R}^d_*}(|z|^2\wedge|z|)\mathtt{M}(\mathrm{d}z)<\infty,\qquad \mathtt{M}(B)=\mathtt{M}(-B),\quad\forall B\in\mathcal{B}(\mathbb{R}^d_*), \tag{5.204}$$

which effectively reduces the *either-or* above expressions to only one, namely,

$$I\varphi(x)=\int_{\mathbb{R}^d_*}[\varphi(x+z)-\varphi(x)]\mathtt{M}(\mathrm{d}z), \tag{5.205}$$

since any odd function, such as $z\mapsto z$, has mean-zero with respect to $\mathtt{M}(\mathrm{d}z)$. Perhaps, a typical example is the *symmetric Poisson process* where $\mathtt{M}=(\delta_{+1}+\delta_{-1})/2$, i.e.,

$$I\varphi(x)=\frac{1}{2}[\varphi(x+z)+\varphi(x-z)-2\varphi(x)],$$

for any measurable function φ.

Let us consider the space $\mathcal{S}(\mathbb{R}^d)$ of rapidly vanishing at infinity and its dual space $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions. After identifying the Hilbert space $L^2(\mathbb{R}^d)$ with its dual space, we have the inclusions

$$\mathcal{S}(\mathbb{R}^d)\subset L^2(\mathbb{R}^d)\subset\mathcal{S}'(\mathbb{R}^d),$$

and we denote by $(\cdot,\cdot)$ the inner product in $L^2(\mathbb{R}^d)$ and by $\langle\cdot,\cdot\rangle$ the duality pairing.

Let us consider the integro-differential operator I given by (5.205), under the assumptions (5.204), as an unbounded linear operator densely defined on $L^2(\mathbb{R}^d)$, i.e., $I : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ with

$$\langle I\varphi,\psi\rangle = (I\varphi,\psi) = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d_*} [\varphi(x+z)-\varphi(x)]\,\mathtt{M}(\mathrm{d}z)\Big)\psi(x)\,\mathrm{d}x.$$

Approximating $\mathtt{M}(\mathrm{d}z)$ with $\mathtt{M}_\varepsilon(\mathrm{d}z)$, exchanging the order of the integrals and performing the change of variable $x \mapsto y = x+z$ in the integration, we obtain

$$\langle I\varphi,\psi\rangle = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}x \int_{\mathbb{R}^d_*} [\psi(x-z)-\psi(x)]\,\mathtt{M}(\mathrm{d}z),$$

which yields $\langle I\varphi,\psi\rangle = \langle\varphi, I\psi\rangle$, after using the symmetry of the kernel $\mathtt{M}(\mathrm{d}z) = \mathtt{M}(-\mathrm{d}z)$. In turn, the equality

$$\begin{aligned}[\varphi(x+z)-\varphi(x)][\psi(x+z)-\psi(x)] &= \\ &= [\varphi(x+z)\psi(x+z)+\varphi(x)\psi(x)] - [\varphi(x+z)\psi(x)+\varphi(x)\psi(x+z)]\end{aligned}$$

and the previous change of variables provide the symmetric expression

$$\langle I\varphi,\psi\rangle = -\frac{1}{2}\int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d_*} [\varphi(x+z)-\varphi(x)][\psi(x+z)-\psi(x)]\,\mathtt{M}(\mathrm{d}z), \quad (5.206)$$

which makes sense for every bounded and Lipschitz functions φ and ψ. Thus (ignoring the growth at infinite), the symmetric bilinear form $i(u,v) = \langle Iu, v\rangle$ associated with I requires only differentiable functions, while the domain $\mathcal{D}_I$ of I may need twice-differentiable functions. Hence, the completion of $\mathcal{S}(\mathbb{R}^d)$ under the inner product $i(u,v)$ is a Hilbert space, which plays the roll of H^1, when dealing with the Laplacian operator instead of the integro-differential operator I. Clearly, the theory of Dirichlet forms is very hand at this point.

Therefore, instead of allowing the kernel $\mathtt{M}(\mathrm{d}z)$ in (5.203) to depend on (x,t), for symmetric processes, we may allow the kernel $\mathtt{M}(\mathrm{d}z)$ in (5.206) to depends on (x,t), i.e., $\mathtt{M}(\mathrm{d}z,t,x)$ given rise to a family of symmetric bilinear forms. Note that a mixed terms of the form

$$\int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d_*} [\varphi(x+z)-\varphi(x)][\nabla\psi(x)]\,\mathtt{M}_1(\mathrm{d}z)$$

is not included.

In any case, working with the structural assumption (5.202) we calculate the formal adjoint of I as in Garroni and Menaldi [94, Section 2.4, pp. 64–72]. A key condition in this development is assuming that the function $\mathtt{j}(\zeta,t,x)$ is continuously differentiable in x for any fixed (ζ,t) and that there exists a constant $c_0 > 0$ such that for any t, x, x' we have

$$c_0|x-x'| \le |(x-x') + [\mathtt{j}(\zeta,t,x) - \mathtt{j}(\zeta,t,x')]| \le c_0^{-1}|x-x'|. \quad (5.207)$$

This implies that the change of variables $\xi = x + \mathtt{j}(\zeta, t, x)$ is a diffeomorphism of class C^1 in $\mathbb{R}^d$, for any t in $[0, T]$, and ζ in $\mathbb{R}^m_*$. Moreover, the Jacobian of the transformation satisfies

$$c_1^{-1} \le \det[I_d + \nabla \mathtt{j}(\zeta, t, x)] \le C_1, \tag{5.208}$$

for any ζ, t, x and some constants $C_1, c_1 \ge 1$, where I_d is the identity matrix in $\mathbb{R}^d$, $\nabla \mathtt{j}(\zeta, t, x)$ is the matrix of the first partial derivatives in x, and $\det[\cdot]$ denotes the determinant of a matrix. Under this assumption, the change of variable $x \mapsto \xi = x + \mathtt{j}(\zeta, t, x)$ is allowed for each ζ and t fixed. Moreover, we can uniquely define two functions

$$\begin{cases} \xi = x + \mathtt{j}(\zeta, t, x) \quad \text{if and only if} \quad \xi - \mathtt{j}^*(\zeta, t, \xi) = x, \\ \mathtt{m}^*(\zeta, t, \xi) |\det[I_d + \nabla \mathtt{j}(\zeta, t, x)]| = \mathtt{m}(\zeta, t, x). \end{cases} \tag{5.209}$$

where $\mathtt{j}^* : \mathbb{R}^m_* \times [0, T] \times \mathbb{R}^d \to \mathbb{R}^d_*$ and $\mathtt{m}^* : \mathbb{R}^m_* \times [0, T] \times \mathbb{R}^d \to [0, \infty)$. Note that $\mathtt{j}(\zeta, t, x) = \mathtt{j}^*(\zeta, t, \xi)$ and for the particular case when $\mathtt{m}(\zeta, t, x) = 1$ we have $\mathtt{m}^*(\zeta, t, \xi) = |\det[I_d + \nabla \mathtt{j}(\zeta, t, x)]|^{-1}$. Among other things, this tells that ODE problems in strong form as in Section 5.1 are not well adapted to handle symmetric processes with variable coefficients. Weak solutions or martingale problems seems more adequate.

5.9.1 Adjoint with Variable Coefficients

Suppose that I is given by

$$I(t)\varphi(x) = \int_{\mathbb{R}^m_*} [\varphi(x + \mathtt{j}(\zeta, t, x)) - \varphi(x)] \mathtt{m}(\zeta, t, x) \pi(\mathrm{d}\zeta). \tag{5.210}$$

and assume that the functions $\mathtt{j}$ and $\mathtt{m}$ are measurable and satisfy, besides (5.207), the condition

$$\begin{cases} |\mathtt{j}(\zeta, t, x)| \le \bar{\jmath}(\zeta), \qquad 0 \le \mathtt{m}(\zeta, t, x) \le C_0, \\ \displaystyle\int_{\{\bar{\jmath} < 1\}} \bar{\jmath}(\zeta) \pi(\mathrm{d}\zeta) + \int_{\{\bar{\jmath} \ge 1\}} \pi(\mathrm{d}\zeta) < \infty, \end{cases} \tag{5.211}$$

where $\pi(\mathrm{d}\zeta)$ is a Radon measure on $\mathbb{R}^m_*$, $\bar{\jmath}(\zeta)$ is a measurable functions from $\mathbb{R}^m_*$ into $(0, \infty)$ and some constant $C_0 > 0$. Then, we define its (proper part) *formal adjoint* integro-differential operator

$$I_1^*(t)\varphi(x) = \int_{\mathbb{R}^m_*} [\varphi(x - \mathtt{j}^*(\zeta, t, x)) - \varphi(x)] \mathtt{m}^*(\zeta, t, x) \pi(\mathrm{d}\zeta), \tag{5.212}$$

which have essentially the same properties as I, i.e., the functions $\mathtt{j}^*$ and $\mathtt{m}^*$ satisfy (5.211), with the same $\bar{\jmath}$ but perhaps a larger C_0.

If π is a bounded measure then the change of variables $x \mapsto \xi = x + \mathtt{j}(\zeta, t, x)$ yields

$$\int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^m_*} \varphi(x + \mathtt{j}(\zeta, t, x)) \psi(x) \mathtt{m}(\zeta, t, x) \pi(\mathrm{d}\zeta) = \\ = \int_{\mathbb{R}^d} \mathrm{d}\xi \int_{\mathbb{R}^m_*} \varphi(\xi) \psi(\xi - \mathtt{j}^*(\zeta, t, \xi)) \mathtt{m}^*(\zeta, t, \xi) \pi(\mathrm{d}\zeta),$$

and therefore

$$\langle I(t)\varphi,\psi\rangle = \langle \varphi, I_1^*(t)\psi\rangle + (\varphi, i_0(t)\psi), \tag{5.213}$$

for every φ and ψ in $\mathcal{S}(\mathbb{R}^d)$, where the function i_0, defined by

$$i_0(t,x) = \int_{\mathbb{R}^m_*} [\mathtt{m}^*(\zeta,t,x) - \mathtt{m}(\zeta,t,x)]\pi(\mathrm{d}\zeta), \quad \forall t,x,$$

is bounded and measurable. If π is necessarily a bounded measure then we need some extra hypotheses on the coefficients,

Lemma 5.58. *Let $I(t)$ be given by (5.210), and besides assumptions (5.207) and (5.211), suppose that there exist a constant M_1 such that*

$$\begin{cases} |\mathtt{m}(\zeta,t,x) - \mathtt{m}(\zeta,t,x+\mathtt{j}(\zeta,t,x))| \le M_1\bar\jmath(\zeta), \\ |\nabla\mathtt{j}(\zeta,t,x)| \le M_1\bar\jmath_1(\zeta), \end{cases} \tag{5.214}$$

for every ζ,t,x, where $\bar\jmath$ is the same function as in (5.211). Then

$$|\mathtt{m}^*(\zeta,t,\xi) - \mathtt{m}(\zeta,t,\xi)| \le M\bar\jmath(\zeta),$$

for some constant M depending only on the various constants intervening in the assumptions, i.e., the function $i_0(t,x)$ is bounded and measurable, and we have the equality (5.213). Moreover, if for every t,x we have

$$\begin{cases} \displaystyle\int_{\bar\jmath(\zeta)\ge\varepsilon} [\varphi(x+\mathtt{j}(\zeta,t,x)) - \varphi(x)]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta) = \\ \qquad \displaystyle= \int_{\bar\jmath(\zeta)\ge\varepsilon} [\varphi(x-\mathtt{j}^*(\zeta,t,x)) - \varphi(x)]\mathtt{m}^*(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{cases} \tag{5.215}$$

for every $\varepsilon>0$ and any φ in $\mathcal{S}(\mathbb{R}^d)$, then $I(t)$ is a symmetric operator.

Proof. Indeed, since $\nabla\mathtt{j}(\zeta,t,x)$ is a matrix, we have $|\det(I_d+A)-1| \le |A|$, for every A and a convenient norm $|\cdot|$ on the matrix A, which we may assume to be the same used in (5.214) for the matrix $\nabla\mathtt{j}(\zeta,t,x)$. Next, denote by $\bar\jmath(\zeta)$ a function satisfying (5.211) and (5.214), and remark that

$$\begin{aligned} |\mathtt{m}^*(\zeta,t,x) - \mathtt{m}(\zeta,t,x)| &\le |\mathtt{m}(\zeta,t,\xi) - \mathtt{m}(\zeta,t,x)| + \\ &+ C_0|1/\det(I_d+\nabla\mathtt{j}(\zeta,t,x)) - 1| \le M_1(1+C_0c_1^{-1})\bar\jmath(\zeta), \end{aligned}$$

for every ζ,t,x and $\xi = x+\mathtt{j}(\zeta,t,x))$.

Finally, with $\pi_\varepsilon(\mathrm{d}\zeta) = \mathbb{1}_{\{|\zeta|\ge\varepsilon\}}\pi(\mathrm{d}\zeta)$, $\varepsilon>0$, and the change of variables $x\mapsto\xi = x+\mathtt{j}(\zeta,t,x)$ we deduce

$$\begin{aligned} \int_{\mathbb{R}^d}\mathrm{d}x \int_{\mathbb{R}^m_*} \varphi(x+\mathtt{j}(\zeta,t,x))\psi(x)\mathtt{m}(\zeta,t,x)\pi_\varepsilon(\mathrm{d}\zeta) = \\ = \int_{\mathbb{R}^d}\mathrm{d}\xi \int_{\mathbb{R}^m_*} \varphi(\xi)\psi(\xi-\mathtt{j}^*(\zeta,t,\xi))\mathtt{m}^*(\zeta,t,\xi)\pi_\varepsilon(\mathrm{d}\zeta), \end{aligned}$$

which yields (5.213), for every φ and ψ in $\mathcal{S}(\mathbb{R}^d)$. □

Moreover, instead of using (5.202) we may suppose that $I(t)$ is given by

$$\begin{cases} I(t)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} [\varphi(x + \mathtt{j}(\zeta,t,x)) - \varphi(x) - \\ \qquad\qquad - \mathbb{1}_{\{\bar\jmath(\zeta)<1\}} \mathtt{j}(\zeta,t,x)\cdot\nabla\varphi(x)]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{cases} \tag{5.216}$$

with

$$\begin{cases} |\mathtt{j}(\zeta,t,x)| \le \bar\jmath(\zeta), \qquad 0 \le \mathtt{m}(\zeta,t,x) \le C_0, \\ \displaystyle\int_{\{\bar\jmath<1\}} (\bar\jmath(\zeta))^2\pi(\mathrm{d}\zeta) + \int_{\{\bar\jmath\ge 1\}} \pi(\mathrm{d}\zeta) < \infty, \end{cases} \tag{5.217}$$

instead of (5.211), then we propose

$$\begin{cases} I^*_2(t)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} [\varphi(x - \mathtt{j}^*(\zeta,t,x)) - \varphi(x) + \\ \qquad\qquad + \mathbb{1}_{\{\bar\jmath(\zeta)<1\}} \mathtt{j}^*(\zeta,t,x)\cdot\nabla\varphi(x)]\mathtt{m}^*(\zeta,t,x)\pi(\mathrm{d}\zeta). \end{cases} \tag{5.218}$$

The addition of the function $\mathbb{1}_{\{\bar\jmath(\zeta)<1\}}$, is to remark the non-dependency on (t,x) of the function separating small-jumps from long-jumps, i.e., to distinguish the behaviors at the origin and at the infinite of the unbounded Levy kernel.

Because of the extra term, to establish a connection between $I(t)$ and $I^*_2(t)$ we need more assumptions, namely, there exist a constant M_2 such that with the same function $\bar\jmath(\zeta)$ as in (5.217) we have

$$\begin{cases} |\mathtt{m}(\zeta,t,x) - \mathtt{m}(\zeta,t,x + \mathtt{j}(\zeta,t,x))| \le M_1\bar\jmath(\zeta), \\ |\nabla\mathtt{j}(\zeta,t,x)| \le M_1\bar\jmath(\zeta), \end{cases} \tag{5.219}$$

and

$$\begin{cases} |\nabla\mathtt{m}(\zeta,t,x)| \le M_2, \qquad |\nabla\mathtt{j}(\zeta,t,x)| \le M_1\bar\jmath(\zeta), \\ |\nabla\cdot\mathtt{j}(\zeta,t,x) - \nabla\cdot\mathtt{j}(\zeta,t,x + j(\zeta,t,x))| \le M_2(\bar\jmath(\zeta))^2, \\ |\mathtt{m}(\zeta,t,x + \mathtt{j}(\zeta,t,x)) - \mathtt{m}(\zeta,t,x) - \mathtt{j}(\zeta,t,x)\cdot\nabla\mathtt{m}(\zeta,t,x)| \le M_2(\bar\jmath(\zeta))^2, \end{cases} \tag{5.220}$$

for every ζ,t,x. Clearly, we may assume that the function $\bar\jmath$ is the same as in (5.217) and (5.214). Note that $\nabla\cdot\mathtt{j}(\zeta,t,x)$ means the divergence of the function $x \mapsto \mathtt{j}(\zeta,t,x)$, for any fixed ζ and t.

Similarly to above, define

$$\begin{cases} i_0(t,x) \quad = \displaystyle\int_{\mathbb{R}^m_*} \big[\mathtt{m}^*(\zeta,t,x) - \mathtt{m}(\zeta,t,x) + \mathbb{1}_{\{\bar\jmath(\zeta)<1\}}\times \\ \qquad\qquad \times\big(\mathtt{m}(\zeta,t,x)\nabla\cdot\mathtt{j}(\zeta,t,x) + \mathtt{j}(\zeta,t,x)\cdot\nabla\mathtt{m}(\zeta,t,x)\big)\big]\pi(\mathrm{d}\zeta), \end{cases} \tag{5.221}$$

and

$$i_1(t,x) = \int_{\{\bar\jmath<1\}} \big[\mathtt{j}(\zeta,t,x)\mathtt{m}(\zeta,t,x) - \mathtt{j}^*(\zeta,t,x)\mathtt{m}^*(\zeta,t,x)\big]\pi(\mathrm{d}\zeta), \tag{5.222}$$

for every t,x, If π is a bounded measure then an integration by part shows the relation

$$\langle I(t)\varphi,\psi\rangle = \langle\varphi, I^*_2(t)\psi\rangle + (\varphi, i_1(t)\cdot\nabla\psi) + (\varphi, i_0(t)\psi), \tag{5.223}$$

for every φ and ψ in $\mathcal{S}(\mathbb{R}^d)$. In general we have

Lemma 5.59. *Let $I(t)$ be given by (5.216), and besides assumptions (5.207) and (5.217). Then the function $i_1(t,x)$, given by (5.222), is bounded and measurable under condition (5.219), and the function $i_0(t,x)$, given by (5.221), is bounded and measurable under condition (5.220). Moreover, the equality (5.223) holds, and if for every t, x we have*

$$\begin{cases} \displaystyle\int_{\{\bar\jmath<1\}} \big[\mathtt{j}(\zeta,t,x)\mathtt{m}(\zeta,t,x) - \mathtt{j}^*(\zeta,t,x)\mathtt{m}^*(\zeta,t,x)\big]\pi(\mathrm{d}\zeta) = 0, \\ \displaystyle\int_{\bar\jmath(\zeta)\ge\varepsilon} [\varphi(x+\mathtt{j}(\zeta,t,x)) - \varphi(x)]\mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta) = \\ \qquad = \displaystyle\int_{\bar\jmath(\zeta)\ge\varepsilon} [\varphi(x-\mathtt{j}^*(\zeta,t,x)) - \varphi(x)]\mathtt{m}^*(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{cases} \tag{5.224}$$

for every $\varepsilon > 0$ and any φ in $\mathcal{S}(\mathbb{R}^d)$, then $I(t)$ is a symmetric operator.

Proof. Indeed, we should only estimate $i_0(t,x)$ and $i_1(t,x)$. To this purpose, under the relation $\xi = x + \mathtt{j}(\zeta,t,x)$, we note that

$$\begin{aligned} &\mathtt{j}(\zeta,t,\xi) = \mathtt{j}^*(\zeta,t,x) = x - \xi, \\ &\mathtt{m}(\zeta,t,\xi) = \mathtt{m}^*(\zeta,t,x)\det\big(I_d + \nabla\mathtt{j}(\zeta,t,\xi)\big), \end{aligned}$$

which yields

$$\mathtt{j}^*(\zeta,t,x)\mathtt{m}^*(\zeta,t,x) = \mathtt{j}(\zeta,t,\xi)\mathtt{m}(\zeta,t,\xi)/\det\big(I_d + \nabla\mathtt{j}(\zeta,t,\xi)\big),$$

and as previously, we have $|\det(I_d + A) - 1| \le |A|$ for a suitable norm $|\cdot|$ used for matrices. Hence, the function $i_1(t,x)$ is bounded by the integral in $\pi(d\zeta)$ over $\{\bar\jmath < 1\}$ of

$$\begin{aligned} &\big|\mathtt{j}(\zeta,t,x)\mathtt{m}(\zeta,t,x) - \mathtt{j}(\zeta,t,\xi)\mathtt{m}(\zeta,t,\xi)/\det[I_d + \nabla\mathtt{j}(\zeta,t,\xi)]\big| \le \\ &\quad \le C_0|\mathtt{j}(\zeta,t,x) - \mathtt{j}(\zeta,t,\xi)| + \bar\jmath(\zeta)|\mathtt{m}(\zeta,t,x) - \mathtt{m}(\zeta,t,\xi)| + \\ &\quad\quad + C_0c_1^{-1}\bar\jmath(\zeta)\big|\det[I_d + \nabla\mathtt{j}(\zeta,t,\xi)] - 1\big| \le \\ &\quad \le M_1(C_0+1)\bar\jmath(\zeta)|x-\xi| + C_0c_1^{-1}(\bar\jmath(\zeta))^2 \le M(\bar\jmath(\zeta))^2, \end{aligned}$$

with $M = M_1(C_0 + 1 + C_0c_1^{-1})$, i.e., $i_1(t,x)$ is a bounded function in view of assumption (5.219).

To estimate $i_0(t,x)$ we used the inequality

$$\Big|\frac{1}{\det(I_d + A)} - 1 + \mathrm{Tr}(A)\Big| \le c_1^{-1}|A|^2,$$

valid for any symmetric matrix A satisfying $c_1\det(I_d + A) \ge 1$ and some suitable norm $|\cdot|$ for matrices. This implies

$$\begin{aligned} &\big|\mathtt{m}(\zeta,t,\xi)/\det[I_d + \nabla\mathtt{j}(\zeta,t,\xi)] - \mathtt{m}(\zeta,t,\xi)[1 - \nabla\cdot\mathtt{j}(\zeta,t,\xi)]\big| \le \\ &\qquad\qquad \le c_1^{-1}|\nabla\mathtt{j}(\zeta,t,\xi)|^2. \end{aligned}$$

Hence, $i_0(t,x)$ is bounded by the integral in $\pi(d\zeta)$ over $\{\bar\jmath<1\}$ of

$$\begin{aligned}&\big|\mathtt{m}(\zeta,t,\xi)/\det[I_d+\nabla\mathtt{j}(\zeta,t,\xi)]-\mathtt{m}(\zeta,t,x)+\mathtt{m}(\zeta,t,x)\nabla\cdot\mathtt{j}(\zeta,t,x)+\\&\quad+\mathtt{j}(\zeta,t,x)\cdot\nabla\mathtt{m}(\zeta,t,x)\big|\le c_1^{-1}|\nabla\mathtt{j}(\zeta,t,\xi)|^2+\\&\quad+\big|[\mathtt{m}(\zeta,t,\xi)-\mathtt{m}(\zeta,t,x)-\mathtt{j}(\zeta,t,x)\cdot\nabla\mathtt{m}(\zeta,t,x)][1-\nabla\cdot\mathtt{j}(\zeta,t,\xi)]\big|+\\&\quad+\big|\mathtt{j}(\zeta,t,x)\cdot\nabla\mathtt{m}(\zeta,t,x)\,\nabla\cdot\mathtt{j}(\zeta,t,\xi)\big|\le C(\bar\jmath(\zeta))^2,\end{aligned}$$

for some constant $C>0$, in view of assumption (5.220). Thus $i_0(t,x)$ is a bounded function. □

5.9.2 Symmetric Bilinear Forms

Suppose that a (time independent to simplify) symmetric bilinear form $i(\cdot,\cdot)$ of the form

$$\left\{\begin{aligned}i(\varphi,\psi)\quad&=\int_{\mathbb{R}^d}\mathrm{d}x\int_{\mathbb{R}^m_*}[\varphi(x+\mathtt{j}(\zeta,x))-\varphi(x)]\times\\&\qquad\times[\psi(x+\mathtt{j}(\zeta,x))-\psi(x)]\mathtt{m}(\zeta,x)\pi(\mathrm{d}\zeta),\end{aligned}\right.\tag{5.225}$$

defined for every φ and ψ in $\mathcal{S}(\mathbb{R}^d)$, with

$$\left\{\begin{aligned}&|\mathtt{j}(\zeta,x)|\le\bar\jmath(\zeta),\qquad 0\le\mathtt{m}(\zeta,x)\le C_0,\\&\int_{\{\bar\jmath<1\}}(\bar\jmath(\zeta))^2\pi(\mathrm{d}\zeta)+\int_{\{\bar\jmath\ge1\}}\pi(\mathrm{d}\zeta)<\infty.\end{aligned}\right.\tag{5.226}$$

Now, under condition (5.207) and assuming that π is a bounded measure we deduce

$$i(\varphi,\psi)=-\langle I\varphi,\psi\rangle,\quad\forall\varphi,\psi\in\mathcal{S}(\mathbb{R}^d),\tag{5.227}$$

where

$$\left\{\begin{aligned}I\varphi(x)=\int_{\mathbb{R}^m_*}&\big[\big(\varphi(x+\mathtt{j}(\zeta,x))-\varphi(x)\big)\mathtt{m}(\zeta,x)+\\&+\big(\varphi(x-\mathtt{j}^*(\zeta,x))-\varphi(x)\big)\mathtt{m}^*(\zeta,x)\big]\pi(\mathrm{d}\zeta).\end{aligned}\right.\tag{5.228}$$

This integro-differential operator can also be written as

$$\begin{aligned}I\varphi(x)=&\int_{\mathbb{R}^m_*}\big[\varphi(x+\mathtt{j}(\zeta,x))-\varphi(x)-\mathtt{j}(\zeta,x))\cdot\nabla\varphi(x)\big]\mathtt{m}(\zeta,x)\pi(\mathrm{d}\zeta)+\\&+\int_{\mathbb{R}^m_*}\big[\big(\mathtt{m}(\zeta,x)\mathtt{j}(\zeta,x))-\mathtt{m}^*(\zeta,x)\mathtt{j}^*(\zeta,x)\big)\big)\cdot\nabla\varphi(x)\big]\pi(\mathrm{d}\zeta)+\\&+\int_{\mathbb{R}^m_*}\big[\varphi(x-\mathtt{j}^*(\zeta,x))-\varphi(x)+\mathtt{j}^*(\zeta,x)\cdot\nabla\varphi(x)\big]\mathtt{m}^*(\zeta,x))\big]\pi(\mathrm{d}\zeta),\end{aligned}$$

and we have

Lemma 5.60. *Let $i(\cdot,\cdot)$ be the symmetric bilinear form given by (5.225) with assumptions (5.226). Suppose that there exists a constant $M_1 > 0$ such that for the same function $\bar\jmath(\zeta)$ as in (5.226) we have*

$$\begin{cases} |\mathtt{m}(\zeta,x) - \mathtt{m}(\zeta, x+\mathtt{j}(\zeta,x))| \le M_1\bar\jmath(\zeta), \\ |\nabla \mathtt{j}(\zeta,x)| \le M_1\bar\jmath(\zeta), \end{cases} \tag{5.229}$$

for every ζ, x. Then the integro-differential operator I given by (5.228) is defined for every φ in $\mathcal{S}(\mathbb{R}^d)$, and the equality (5.227) holds. □

Sometimes, the symmetric bilinear form is presented as follows:

$$i(\varphi,\psi) \quad = \int_{\mathbb{R}^d\times\mathbb{R}^d} [\varphi(x)-\varphi(y)][\psi(x)-\psi(y)]\mu(\mathrm{d}x,\mathrm{d}y),$$

for any smooth functions φ, ψ, with an assumption like

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \big(|x-y|^2 \wedge 1\big)\mu(\mathrm{d}x,\mathrm{d}y) < \infty.$$

for a symmetric measure $\mu(\mathrm{d}x,\mathrm{d}y) = \mu(\mathrm{d}y,\mathrm{d}s)$ on $\mathbb{R}^d\times\mathbb{R}^d$. The interested reader may check Dirichlet forms, e.g., Fukushima et al. [92], Ma and Röckner [161], among others.

Part III

Reflected SDE with Jumps

Chapter 6

Stochastic Differential Equations II

Let $\mathcal{O}$ be a non-empty proper d-dimensional domain, i.e., an connected open subset of $\mathbb{R}^d$ such that $\dot{\overline{\mathcal{O}}}$, the interior of the closure of $\mathcal{O}$, coincides with $\mathcal{O}$. Following the discussion in the previous chapter, now we focus on stochastic processes with values in either $\mathcal{O}$ or $\overline{\mathcal{O}}$ which satisfy some conditions on the boundary $\partial\mathcal{O}$ of the domain $\mathcal{O}$, e.g., controlled diffusions processes with jumps and with boundary conditions. For instance, consider a stochastic ordinary differential equation of the form

$$\left\{\begin{array}{l} x(t) = x^0 + \displaystyle\int_{t_0}^{t} g(s, x(s), v(s))\mathrm{d}s + \int_{t_0}^{t} \sigma(s, x(s), v(s))\mathrm{d}w(s) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_* \times]t_0, t]} \gamma(\zeta, s, x(s), v(s))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \quad \forall\, t \geq t_0, \end{array}\right.$$

where the drift g, the diffusion σ and the jumps γ coefficients satisfy

(1) $\sigma(t, x, v) = 0$ for any $t \geq t_0$, $v \in \mathbb{R}^q$ and any x in $\mathcal{O}$ near the boundary $\partial\mathcal{O}$,

(2) $x + \gamma(\zeta, t, x, v)$ belongs to $\overline{\mathcal{O}}$ for every ζ in the support of the measure $\pi(\cdot)$ relative to $\mathbb{R}^m$, $t \geq t_0$, $x \in \overline{\mathcal{O}}$ and $v \in \mathbb{R}^q$,

(3) $g(t, x, v) \cdot n(x) \leq 0$ for every $t \geq t_0$, x on the boundary $\partial\mathcal{O}$ and $v \in \mathbb{R}^q$,

1mm] where the exterior normal unit vector $n(x)$ with respect to $\mathcal{O}$ is defined at any boundary point x, and all is refereed to a (standard) $n \times m$ Wiener-Poisson space

$$(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0),$$

with Lévy measure $\pi(\cdot)$. It is clear that because condition (2) there is not jumps outside the (closed) domain $\overline{\mathcal{O}}$ (within itself), because condition (1) there is not diffusion near the boundary $\partial\mathcal{O}$, and condition (3) implies that the *velocity* $g(s, x(s), v(s))$ is not driving the system outside the (closed) domain $\overline{\mathcal{O}}$. Thus, we have a simple case where the above stochastic ordinary differential equation

is restrained to the (closed) domain $\overline{\mathcal{O}}$. Clearly, for condition (1) we only need $\sigma = 0$ on the boundary $\partial\mathcal{O}$, and condition (2) can be generalized to impose that x plus the support of the measure $z \mapsto M(\mathrm{d}z, t, x)$ be included in $\overline{\mathcal{O}}$, for every $t \geq t_0$ and $x \in \overline{\mathcal{O}}$, where $M(\mathrm{d}z, t, x)$ denotes the Lévy kernel.

Another simple situation is when the evolution is stopped at the boundary, i.e., by means to the functional

$$\tau(x^0) := \inf\big\{s \geq t_0 : x(s) \in \mathbb{R}^d \setminus \overline{\mathcal{O}}\big\},$$

we only consider a new process $(s, \omega) \mapsto x(s \wedge \tau(x^0, \omega), \omega)$ which stays inside the (closed) domain $\overline{\mathcal{O}}$. Similarly, in the definition of the functional τ we may use $x(s-)$ and/or $\mathbb{R}^d \setminus \mathcal{O}$ to study other models. Eventually, we could *patch* the stopped process $x(s \wedge \tau(x^0))$ for $t > \tau(x^0)$, with some other process on the boundary $\partial\mathcal{O}$.

On the other hand, diffusion process with jumps are well understood via stochastic differential equations in the whole space $\mathbb{R}^d$. This corresponds to a second-order integro-differential operator of the form

$$\left\{\begin{array}{l} A(t)\varphi(x) = \displaystyle\sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij}\varphi(x) + \sum_{i=1}^{d} a_i(t,x)\partial_i\varphi(x) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z, t, x), \end{array}\right. \tag{6.1}$$

where $(a_{ij}(t,x))$ is a non-negative definite matrix-valued function, $(a_i(t,x))$ is a vector-valued function, the Lévy kernel $\mathtt{M}(\mathrm{d}z, t, x)$ integrate the function $z \mapsto |z|^2 \wedge |z|$, all coefficients are at least measurable and locally bounded, so that $A(t)\varphi(x)$ is defined for any smooth functions in the whole space $\mathbb{R}^d$. The first part is a second-order differential operator, which admits natural (i.e., well known) boundary conditions, but the non-local part needs further study.

If we vanish the jump part then the discussion is more familiar, analytically the boundary $\partial\mathcal{O}$ of an elliptic-parabolic (or degenerate elliptic) problem is decomposed in three pieces, namely

$$\begin{aligned} \partial_1\mathcal{O} &:= \big\{x \in \partial\mathcal{O} \,:\, a_{\partial\mathcal{O}}(x) \neq 0\big\}, \\ \partial_2\mathcal{O} &:= \big\{x \in \partial\mathcal{O} \setminus \partial_1\mathcal{O} \,:\, b_{\partial\mathcal{O}}(x) < 0\big\}, \\ \partial_3\mathcal{O} &:= \big\{x \in \partial\mathcal{O} \setminus \partial_1\mathcal{O} \,:\, b_{\partial\mathcal{O}}(x) \geq 0\big\}, \end{aligned}$$

where

$$\begin{aligned} a_{\partial\mathcal{O}}(x) &:= \sum_{ij} a_{ij}(x) n_i(x) n_j(x), \\ b_{\partial\mathcal{O}}(x) &:= \sum_{ij} [a_i(x) + \partial_j a_{ij}(x)] n_i(x), \end{aligned}$$

$n(x) = (n_i(x), \ldots, n_j(x))$ is the exterior normal unit vector and, for simplicity, the variables t and v have been dropped. The set $\partial_1\mathcal{O}$ is the non-characteristic

portion of the boundary, and the discriminator function $b_{\partial\mathcal{O}}(x)$ is used to delineate on which portions of the boundary data may ($\partial_2\mathcal{O}$) or may not ($\partial_3\mathcal{O}$) be assigned. The well known case is when the matrix (a_{ij}) does not degenerate and therefore the whole boundary is non-characteristic, i.e., $\partial\mathcal{O} = \partial_1\mathcal{O}$. As seen later, the trajectories $x(t)$ of the diffusion process will leave (almost surely) the closed domain $\overline{\mathcal{O}}$ through the part $\partial_1\mathcal{O} \cup \partial_2\mathcal{O}$. The relation between a Markov-Feller semigroup in a closed bounded domain $\overline{\mathcal{O}}$ and an integro-differential of the above form is well know, e.g., results concerning the expression of the infinitesimal generator, the boundary conditions and the construction of the semigroup can be found in Taira [243, Chapters 9 and 10, pp. 321–424], and recently in Skubachevskii [231] and references therein.

The passage from $\mathbb{R}^d$ to manifolds may be delicate and certainly, more notation is necessary, for instance, we refer the reader to the book Ikeda and Watanabe [110, Chapter V, pp. 247–435] for a in deep study. Simple cases can be easily considered, e.g., a hyper cube with periodic boundary conditions can be studied, after the corresponding construction of Wiener processes and Poisson measures in a hyper cube with periodic conditions.

An important role is played by the so-called *local time* associated with a semi-martingale in the following way. First, for any convex function $\varphi(x)$ in $\mathbb{R}^d$ and any d-dimensional semi-martingale $(M(t) : t \geq 0)$ one can review Itô formula to deduce that the process

$$\frac{1}{2}\ell_\varphi(t) := \varphi(M(t)) - \varphi(M(0)) - \sum_{i=1}^{d} \int_{]0,t]} \partial_i^- \varphi(M(s-))\mathrm{d}M_i(s) -$$
$$- \sum_{0<s\leq t} \big[\varphi(M(s)) - \varphi(M(s-)) - \sum_{i=1}^{d} (M_i(s) - M_i(s-))\partial_i^- \varphi(M(s-))\big]$$

is nondecreasing and continuous, where $\partial_i^-\varphi$ means the derivative from the left with respect to the variable x_i. Moreover, if the function φ is continuously differentiable having locally Lipschitz first derivative, then the process $(\ell_\varphi(t) : t \geq 0)$ has local bounded variation. Certainly, if φ is twice-continuously differentiable then

$$\ell_\varphi(t) = \sum_{i,j=1}^{d} \int_{]0,t]} \partial_{ij}\varphi(M(s-))\mathrm{d}[M_i, M_j]^c(s),$$

where $[M_i, M_j]^c$ is the continuous part of the optional quadratic cross-variation of M_i and M_j, which is equal to $\langle M_i^c, M_j^c\rangle$, the predictable quadratic cross-variation of the continuous parts M_i^c and M_j^c of the semi-martingales. In the case of a quasi-left continuous special semi-martingale, the jumps can be re-

written in term of the jump compensator ν^p_M associated with M,

$$\sum_{0<s\le t}\big[\varphi(M(s))-\varphi(M(s-))-\partial_i^-\varphi(M(s-))(M(s)-M(s-))\big]=$$
$$=\int_{\mathbb{R}^d_*\times]0,t]}\big[\varphi(M(s-)+z)-$$
$$-\varphi(M(s-))-\sum_{i=1}^d z_i\partial_i^-\varphi(M(s-))\big]\nu^p_M(\mathrm{d}z,\mathrm{d}s),$$

which is a continuous process, locally of bounded variation, and nondecreasing whenever φ is convex. For instance, the reader may check the books Bertoin [21, Chapters IV, V, pp. 103–154], Dellacherie and Meyer [58, Sections VIII.2.25–29, pp. 348–353], Karatzas and Shreve [124, Section 3.6.D, pp. 212–215].

Applying this formula for the functions $(x-a)^+$, $(x-a)^-$, $|x-a|$, with a fixed a in $\mathbb{R}$ we obtain the expression of the local times (in dimension one), known as Tanaka's formula, i.e., for x^+ and M

$$M^+(t)=M^+(0)+\int_{]0,t]}\mathbb{1}_{\{M(s-)>0\}}\mathrm{d}M(s)+\frac{1}{2}\ell^+(t)+$$
$$+\sum_{0<s\le t}\big[\mathbb{1}_{\{M(s-)\le 0\}}M^+(s)+\mathbb{1}_{\{M(s-)>0\}}M^-(s)\big],$$

while for x^+ and $-M$,

$$M^-(t)=M^-(0)+\int_{]0,t]}\mathbb{1}_{\{M(s-)<0\}}\mathrm{d}M(s)+\frac{1}{2}\ell^-(t)+$$
$$+\sum_{0<s\le t}\big[\mathbb{1}_{\{M(s-)<0\}}M^+(s)+\mathbb{1}_{\{M(s-)\le 0\}}M^-(s)\big].$$

Since $x^-=x^+-x$, we deduce

$$\int_{]0,t]}\mathbb{1}_{\{M(s-)=0\}}\mathrm{d}M(s)=\sum_{0<s\le t}\mathbb{1}_{\{M(s-)=0\}}M(s)+$$
$$+\frac{1}{2}[\ell^+(t)-\ell^-(t)],$$

and the processes ℓ^+ and ℓ^- increase only when $M(s-)=0$.

...***...

Sometimes it is necessary to make a *time changes* in Stieltjes integrals. For a given nondecreasing cad-lag process ℓ with values in $(0,\infty]$ consider

$$\ell^{-1}(s):=\inf\{t\ge 0:\ell(t)>s\},\quad \forall s\ge 0,$$

and $\ell^{-1}(s):=+\infty$ if $\ell(t)\le s$ for all $t\ge 0$. This define a cad-lag process $(\ell^{-1}(s):s\ge 0)$ with the properties

$$\ell^{-1}(s-)=\inf\{t\ge 0:\ell(t)\ge s\},\quad \text{and}\quad \ell[\ell^{-1}(s)]\ge s\quad \forall s\ge 0,$$

$$\ell(t)=\inf\{s\ge 0:\ell^{-1}(s)>t\},\quad \forall t\ge 0.$$

If ℓ is continuous then ℓ^{-1} may not be continuous (when ℓ is not strictly increasing). The following change of variables formula can be obtained. For any nonnegative Borel measurable process f on $[0,\infty)$ we have

$$\int_{[0,\infty)} f(t)\mathrm{d}\ell(t) = \int_0^\infty f(\ell^{-1}(s))\,\mathbb{1}_{\ell^{-1}(s)<\infty}\mathrm{d}s,$$
$$\int_{]a,b]} f(t)\mathrm{d}\ell(t) = \int_{\ell(a)}^{\ell(b)} f(\ell^{-1}(s))\mathrm{d}s,$$
$$\int_{[k(a),k(b)]} f(t)\mathrm{d}\ell(t) = \int_{[a,b]} f(k(t))\mathrm{d}\ell(k(t)),$$

for any continuous non-decreasing process k on the bounded interval $[a,b]$.

If, for a given filtration $(\mathcal{F}(t):t\geq 0)$, the random variables $\omega\mapsto\ell(t,\omega)$ is a stopping for every fixed $t\geq 0$, then we can define the filtration $(\mathcal{F}^\ell(t):t\geq 0)$, where $\mathcal{F}^\ell(t)=\mathcal{F}(\ell(t))$ is the σ-algebra generated by the stopping time $\ell(t)$. Notice that because ℓ is continuous from the right, we have $\mathcal{F}^\ell(t+)=\mathcal{F}^\ell(t)$ for every $t\geq 0$. If the process ℓ is continuous and the initial filtration $(\mathcal{F}(t):t\geq 0)$ is quasi-left continuous, i.e., $\mathcal{F}(\tau)=\mathcal{F}(\tau-)$ for every predictable stopping time, then it is clear that the transformed filtration $(\mathcal{F}^\ell(t):t\geq 0)$ is also quasi-left continuous. Now, if M is a (continuous, local) martingale with respect to $(\mathcal{F}(t):t\geq 0)$ then M^ℓ, where $M^\ell(t):=M(\ell(t))$ is a (continuous, local) martingale with respect to $(\mathcal{F}^\ell(t):t\geq 0)$. Moreover, if $\langle M\rangle(t)$ is the predictable quadratic co-variation of M then $\langle M\rangle(\ell(t))$ is the predictable quadratic co-variation of M^ℓ. In particular, if M is a d-dimensional standard Wiener process then M^ℓ is a continuous square-integrable martingale with predictable quadratic co-variation of $\langle M_i^\ell, M_i^\ell\rangle=\ell$ and $\langle M_i^\ell, M_j^\ell\rangle=0$ for any $i\neq j$. Similarly, if $\nu(\mathrm{d}z,\mathrm{d}t)$ is an integer random measure with jump compensator $\nu^p(\mathrm{d}z,\mathrm{d}t)$ relative to the initial filtration $(\mathcal{F}(t):t\geq 0)$, then $\nu_\ell(\mathrm{d}z,\mathrm{d}t)$ defined by

$$\nu_\ell(B,]a,b]) := \nu(B,]\ell(a),\ell(b)]),\quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\geq 0,$$

is an integer random measure with jump compensator $\nu^p_\ell(\mathrm{d}z,\mathrm{d}t)$ relative to the transformed filtration $(\mathcal{F}^\ell(t):t\geq 0)$, where

$$\nu^p_\ell(B,]a,b]) := \nu^p(B,]\ell(a),\ell(b)]),\quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\geq 0.$$

In particular, if the jump compensator has the form

$$\nu^p(B,]a,b]) = \int_a^b \dot\nu^p(B,s)\mathrm{d}s,\quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\geq 0,$$

for some kernel $\dot\nu^p(B,s)$ such that (1) the process $t\mapsto\dot\nu^p(K,t)$ is predictable and locally bounded for every compact subset K of $\mathbb{R}^d_*$ and (2) $B\mapsto\dot\nu^p(B,t)$ is a measure for every $t\geq 0$, then

$$\nu^p_\ell(B,]a,b]) = \int_{\ell(a)}^{\ell(b)} \dot\nu^p(B,s)\mathrm{d}\ell(s),\quad \forall B\in\mathcal{B}(\mathbb{R}^d_*),\ b>a\geq 0,$$

The read interested in local times, may want to take a look at the book by Marcus and Rosen [166, Chapter 3, pp. 62-120].

...***...

6.1 Simple Boundary Conditions

Here we discuss some simple boundary conditions with a very natural interpretation, particularly, absorbing and sticking boundaries. As a by product, we adapt and extended results (of great interest in some cases) from the previous chapter.

...***...

As it is well established, there is a general probabilistic representation of the solution of a parabolic (and also elliptic) equation through the Itô's stochastic calculus.

Adapting the previous notation, let us consider the operator

$$\begin{cases} A(t)\varphi(x) := A_0(t)\varphi(x) + \sum_{i=1}^{d} g_i(t,x)\partial_i\varphi(x) - a_0(t,x)\varphi(x), \\ A_0(t)\varphi(x) := \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sum_{k=1}^{n}\sigma_{ik}(t,x)\sigma_{jk}(t,x)\Big)\partial_{ij}\varphi(x), \end{cases} \tag{6.2}$$

where the coefficients may depends on the variable (t,x), $a_0 \geq 0$, the (vector) drift $g := (g_1,\dots,g_d)$ replaces the notation $\underline{a}$ and the matrix $\sigma = (\sigma_{ij})$ is a square root of the symmetric nonnegative matrix (a_{ij}) used in the previously used. The drift g and the diffusion σ are assumed regular enough to have a unique solution of the following stochastic differential equation

$$\begin{cases} dX(s) = g(s,X(s))\mathrm{d}s + \sigma(s,X(s))\mathrm{d}W(s), \quad s \geq t, \\ X(t) = x \end{cases} \tag{6.3}$$

where $\{W(t) : t \geq 0\}$ is a standard n-dimensional Wiener process and $X = (X_1,\dots,X_d)$. Also we will use

$$X_0(s) := \exp\Big(-\int_t^s a_0(r,X(r))\mathrm{d}r\Big), \quad \forall s \geq t \geq 0. \tag{6.4}$$

when necessary.

For a given open connected region $\mathcal{O}$ in $\mathbb{R}^d$, with boundary $\partial\mathcal{O}$, we define the first exit time from the closure $\bar{\mathcal{O}}$ as follows

$$\tau(t,x) = \inf\big\{s > t : X(s) \notin \bar{\mathcal{O}}\big\},$$

and $\tau(t,x) = \infty$ if $X(s)$ belongs to $\bar{\mathcal{O}}$ for every $s \geq t$. Clearly, by continuity we have $X(\tau(t,x))$ in $\partial\mathcal{O}$ if $\tau(t,x)$ is finite.

6.1.1 Dirichlet Boundary Conditions

One reverses time for the representation formula, so that instead of having *initial* condition one uses *terminal* condition, i.e., the backward equation

$$\begin{cases} \partial_t u + A(t)u = f(t,x), \quad \forall (t,x) \in [0,T) \times \mathcal{O}, \\ u(T,x) = u_0(x), \quad \forall x \in \mathcal{O}, \\ u(t,x) = \psi(t,x), \quad \forall (t,x) \in [0,T) \times \partial\mathcal{O}, \end{cases} \tag{6.5}$$

where $T > 0$ is given.

By means of Itô formula for $u(t, X(t))$ in t, and with the notation (6.4), we obtain

$$u(t,x) = \mathbb{E}\Big\{ X_0\big(T \wedge \tau(t,x)\big) u\big(T \wedge \tau(t,x), X(T \wedge \tau(t,x))\big) \Big\} + \\ + \mathbb{E}\Big\{ \int_t^{T \wedge \tau(t,x)} X_0(s) f(s, X(s)) \mathrm{d}s,$$

i.e.,

$$u(t,x) = \int_{\tau(t,x)<T} \big[X_0(\tau(t,x)) \psi\big(\tau(t,x), X(\tau(t,x))\big) \big] \mathrm{d}P + \\ + \int_{\tau(t,x) \geq T} \big[X_0(T) u_0(X(T)) \big] \mathrm{d}P + \mathbb{E}\Big\{ \int_t^{T \wedge \tau(t,x)} X_0(s) f(s, X(s)) \mathrm{d}s \Big\}.$$

Clearly, the case $f = 0$, $\psi = 0$ corresponds to the Cauchy problem with (homogeneous) Dirichlet boundary conditions, while the case $f = 0$, $u_0 = 0$ corresponds to the so called Poisson problem.

Next, if $a_0 = 0$ then we set

$$P_T(t,x,\mathrm{d}s,\mathrm{d}y) := P\{\tau(t,x) < T,\ \tau(t,x) \in \mathrm{d}s,\ X(\tau(t,x)) \in \mathrm{d}y\}, \\ G_T(t,x,\mathrm{d}s,\mathrm{d}y) := P\{s < \tau(t,x),\ \tau(t,x) \in \mathrm{d}s,\ X(T \wedge s) \in \mathrm{d}y\},$$

while in general, for $a_0 \geq 0$,

$$P_T(t,x,I,B) := \mathbb{E}\big\{ \mathbb{1}_{\tau(t,x)<T}\ \mathbb{1}_{\{\tau(t,x) \in I\}}\ X_0(\tau(t,x)) \mathbb{1}_{\{X(\tau(t,x)) \in B\}} \big\}, \\ G_T(t,x,I,B) := \mathbb{E}\Big\{ \int_t^{T \wedge \tau(t,x)} \mathbb{1}_{\{\tau(t,x) \in I\}}\ X_0(s) \mathbb{1}_{\{X(s) \in B\}} \mathrm{d}s \Big\}, \\ G_T(t,x,\{T\},B) := \mathbb{E}\big\{ \mathbb{1}_{\tau(t,x) \geq T}\ X_0(T) \mathbb{1}_{\{X(T) \in B\}} \big\},$$

for every $t \geq 0$, x in $\bar{\mathcal{O}}$, for any $I \subset [0,T]$ and $B \subset \mathbb{R}^d$ belonging to a class of sets generating the Borel σ-algebra in $[0,T]$ and $\mathbb{R}^d$, respectively (e.g., semi-closed intervals). It is clear that $P_T(t,x,\cdot,\cdot)$ is supported in $(t,T) \times \partial\mathcal{O}$ while $G_T(t,x,\cdot,\cdot)$ is supported in $(t,T] \times \mathcal{O}$. Since $\tau(t,x) \geq T$ is possible, the measure $\mathrm{d}s \mapsto G_T(t,x,\mathrm{d}s,\mathrm{d}y)$ has an atom at $\{T\}$. Moreover, we may have $P\{\tau(t,x) = \infty\} > 0$ in the degenerate case. Actually, for the degenerate case, the boundary condition is not given in the whole boundary, but only on points belonging to the

support of $P_T(t,x,\cdot,\cdot)$. For instance, if $P\{\tau(t,x)=\infty\}=1$ then $P_T(t,x,\cdot,\cdot)=0$ and no condition on the boundary can be given, the process never reaches the boundary $\partial\mathcal{O}$. Sometimes, it may be convenient to define

$$G_T(t,x,s,B) := \mathbb{E}\Big\{\int_s^{\tau(t,x)} X_0(T\wedge s)\mathbb{1}_{\{X(T\wedge s)\in B\}}\mathrm{d}s\Big\},$$

and understand $G_T(t,x,\mathrm{d}s,B)$ as the (unique) measure induced by the non-increasing function $s\mapsto G_T(t,x,s,B)$ satisfying the equality $G_T(t,x,[s,T],B)=G_T(t,x,s,B)$ for every s in $[t,T]$.

Then we can rewrite

$$\begin{aligned}u(t,x) = &\int_{(t,T)\times\partial\mathcal{O}}\psi(s,y)P_T(t,x,\mathrm{d}s,\mathrm{d}y)+\\ &+\int_{\mathcal{O}}u_0(y)G_T(t,x,\{T\},\mathrm{d}y)+\int_{(t,T)\times\mathcal{O}}f(s,y)G_T(t,x,\mathrm{d}s,\mathrm{d}y),\end{aligned}$$

for any (t,x) in $[0,T]\times\bar{\mathcal{O}}$. Therefore, after reversing the time again (i.e., changing s into $T-t$), the functions P_T and G_T are called the Poisson and Green functions associated with the second-order differential operator A and Dirichlet boundary conditions in $\mathcal{O}$.

For instance, if the coefficients are independent of t and $v(t,x)$ solves the forward equation

$$\begin{cases}\partial_t v = A(t)v+f(t,x), & \forall(t,x)\in(0,T]\times\mathcal{O},\\ v(0,x)=v_0(x), & \forall x\in\mathcal{O},\\ v(t,x)=\psi(t,x), & \forall(t,x)\in(0,T]\times\partial\mathcal{O},\end{cases}\tag{6.6}$$

then we have

$$\begin{aligned}v(t,x) = &\int_{(0,t)\times\partial\mathcal{O}}\psi(t-s,y)P_T(x,\mathrm{d}s,\mathrm{d}y)+\\ &+\int_{\mathcal{O}}v_0(y)G_T(x,t,\mathrm{d}y)+\int_{(0,t)\times\mathcal{O}}f(t-s,y)G_T(x,\mathrm{d}s,\mathrm{d}y),\end{aligned}$$

for any (t,x) in $[0,T]\times\bar{\mathcal{O}}$, where now

$$\begin{aligned}P_T(x,I,B) &:= \mathbb{E}\big\{\mathbb{1}_{\tau(0,x)<T}\,\mathbb{1}_{\{\tau(0,x)\in I\}}\,X_0(\tau(0,x))\mathbb{1}_{\{X(\tau(0,x))\in B\}}\big\},\\ G_T(x,I,B) &:= \mathbb{E}\Big\{\int_0^{T\wedge\tau(0,x)}\mathbb{1}_{\{\tau(0,x)\in I\}}\,X_0(s)\mathbb{1}_{\{X(s)\in B\}}\mathrm{d}s\Big\},\\ G_T(x,\{t\},B) &:= \mathbb{E}\big\{\mathbb{1}_{T\wedge\tau(0,x)\geq t}\,X_0(t)\mathbb{1}_{\{X(t)\in B\}}\big\},\end{aligned}$$

for every $t>0$, x in $\bar{\mathcal{O}}$, for any $I\subset(0,T]$ and $B\subset\mathbb{R}^d$ belonging to a class of sets generating the Borel σ-algebra in $(0,T]$ and $\mathbb{R}^d$, respectively (e.g., semi-closed intervals).

It is probably wise not to write (or really get) the expression with initial condition and time-variable coefficients, may be too messy?

...***...

Before giving details, recall the sample space $D([t_0,\infty),\mathbb{R}^d)$ of cad-lag functions (which is a complete metric space with the Skorokhod topology) and define two maps $\tau(\cdot,\mathcal{O})$ and $\tilde\tau(\cdot,\mathcal{O})$ from $D([t_0,\infty),\mathbb{R}^d)$ into $[0,\infty]$

$$\begin{cases} \tau(x(\cdot),\mathcal{O}) := \inf\{t\geq t_0 : x(t)\notin\mathcal{O}\}, \\ \tilde\tau(x(\cdot),\mathcal{O}) := \inf\{t\geq t_0 : x(t) \text{ or } x(t-)\notin\mathcal{O}\}, \end{cases} \tag{6.7}$$

where $\mathcal{O}$ is a given Borel subset of $\mathbb{R}^d$, and $\tau(x(\cdot),\mathcal{O})=\infty$ if $x(t)$ belongs to $\mathcal{O}$ for every $t\geq t_0$, i.e., when the set to take infimum is empty. Similarly for $\tilde\tau(x(\cdot),\mathcal{O})$, and by convention $x(t_0-)=x(t_0)$. Clearly, these extended functionals are called the *first exit time* from the subset $\mathcal{O}$, and certainly $\tau(x(\cdot),\mathcal{O})=\tilde\tau(x(\cdot),\mathcal{O})$ for every $x(\cdot)$ in the sample space $C([t_0,\infty),\mathbb{R}^d)$ of continuous functions (which is a complete metric space with the locally uniform converge topology). Depending on what is to be emphasized, we use the notation $\tau(x(\cdot),\mathcal{O})$ or simply $\tau(x(\cdot))$ or $\tau(\mathcal{O})$, and similarly for $\tilde\tau(x(\cdot),\mathcal{O})$.

Proposition 6.1. *Let $\mathcal{O}$ open subset of $\mathbb{R}^d$, and $\{x,x_1,x_2,\dots\}$ be functions in $D([t_0,\infty),\mathbb{R}^d)$.*

(a) *First $\tau(x,\overline{\mathcal{O}})=\tilde\tau(x,\overline{\mathcal{O}})$.*

(b) *If $x_n(t)$ converges to $x(t)$ for each $t\geq t_0$ then*

$$\limsup_{n\to\infty}\tau(x_n,\overline{\mathcal{O}})\leq\tau(x,\overline{\mathcal{O}}),$$

i.e., $\tau(\cdot,\overline{\mathcal{O}})$ is upper semi-continuous in $D([t_0,\infty),\mathbb{R}^d)$ for the topology induced by the pointwise convergence.

(c) *If $x_n(s)$ converges to $x(s)$, uniformly on $[t_0,t]$ for each $t>t_0$ then*

$$\liminf_{n\to\infty}\tilde\tau(x_n,\mathcal{O})\geq\tilde\tau(x,\mathcal{O}),$$

i.e., $\tilde\tau(\cdot,\mathcal{O})$ is lower semi-continuous in $D([t_0,\infty),\mathbb{R}^d)$ for the topology induced by the locally uniform converge.

Proof. To show (a) we argue as follows. First $\tau(x,\overline{\mathcal{O}})\geq\tilde\tau(x,\overline{\mathcal{O}})$, for any x. Next, if $t>\tilde\tau(x,\overline{\mathcal{O}})$ then there exists $t'\leq t$ such that $x(t')$ or $x(t'-)$ is not in $\overline{\mathcal{O}}$. If $x(t')$ is not in $\overline{\mathcal{O}}$ then $\tau(x,\overline{\mathcal{O}})\leq t'\leq t$. Alternatively, if only $x(t'-)$ is not in $\overline{\mathcal{O}}$ then for some $s<t'$, the point $x(s)$ does not belong to $\overline{\mathcal{O}}$, i.e, $\tau(x,\overline{\mathcal{O}})\leq s<t'\leq t$. Since t can be taken arbitrary close to $\tilde\tau(x,\overline{\mathcal{O}})$ we deduce $\tau(x,\overline{\mathcal{O}})\leq\tilde\tau(x,\overline{\mathcal{O}})$.

To establish (b), only the case $\tau(x,\overline{\mathcal{O}})<\infty$ need to be considered. So, for any $t>\tau(x,\overline{\mathcal{O}})$ there exists a $t'\leq t$ such that $x(t')$ belongs to $\mathbb{R}^d\setminus\overline{\mathcal{O}}$. Because $\overline{\mathcal{O}}$ is closed and $x_n(t')$ converges to $x(t')$ we deduce that $x_n(t')$ belongs to $\mathbb{R}^d\setminus\overline{\mathcal{O}}$ for n sufficiently large, i.e., $\tau(x_n,\overline{\mathcal{O}})\leq t'\leq t$. Hence the desired result.

Finally to prove (c) for any $t<\tilde\tau(x,\mathcal{O})$ the set $\{x(s),\,x(s-):s\in[t_0,t]\}$ is compact and included in the open set $\mathcal{O}$. Therefore, the locally uniform convergence implies that for n sufficiently large, the set $\{x_n(s),\,x(s-):s\in[t_0,t]\}$ is

included in the open set $\mathcal{O}$, so that $\tilde{\tau}(x_n, \mathcal{O}) > t$. Since t can be taken arbitrary close to $\tilde{\tau}(x, \mathcal{O})$ we conclude. □

Given a cad-lag strong Markov process $(x(t) : t \geq t_0)$ with values in $\mathbb{R}^d$, the expression (6.7) defines a strong Markov functional, i.e., $x(\tilde{\tau} \wedge t)$ and $x(\tau \wedge t)$ are also cad-lag strong Markov process.

6.1.2 Absorbing Boundary

This and the next section may need to be renamed?
Absorbing
Killing
Sticking
Dirichlet
...***...

The absorbing boundary condition is a very natural way of converting a Markov-Feller process acting in the whole space, say $\mathbb{R}^d$, into another Markov (not necessarily Feller) process living in only a domain $\mathcal{O}$ of $\mathbb{R}^d$. The technique uses the first exit time $\tilde{\tau}(x(\cdot), \mathcal{O})$, as defined by (6.7). Since $x(\cdot)$ is a Markov-Feller process, all (cad-lag) paths are determined by the initial condition $x(t_0) = x^0$, so that we can write $\tilde{\tau}(x(\cdot)) = \tilde{\tau}(x(t_0), \omega)$. This random variable $\tilde{\tau}(x(t_0), \omega)$ is a *stopping time* relative to the canonical filtration associated with $x(\cdot)$ for any Borel domain, in particular for any open $\mathcal{O}$ or closed set $\overline{\mathcal{O}}$, or even $\mathcal{O}$ union a (relative open or closed) part of the boundary $\partial\mathcal{O}$.

Hence, if $x(\cdot)$ takes valued in $\mathbb{R}^d$ then the new Markov (not necessarily Feller) process

$$x^\tau(t) := \begin{cases} x(t), & \text{if } t < \tilde{\tau}(x(t_0, \omega) \\ \infty, & \text{otherwise.} \end{cases} \tag{6.8}$$

The symbol ∞ is used to express the fact that $(x^{\tilde{\tau}}(t) : t \geq t_0)$ takes values into $\mathcal{O} \cup \{\infty\}$, which is view as a *one-point compactification* of $\mathcal{O}$. Assuming the initial Markov-Feller process is time-homogeneous, we can take always $t_0 = 0$. The semigroup associated with $(x^{\tilde{\tau}}(t) : t \geq t_0)$ is given by

$$S_{\mathcal{O}}(t)f(x) := \mathbb{E}\{f(x(t \wedge \tilde{\tau}_{\mathcal{O}}))\}, \quad t \geq 0, \tag{6.9}$$

and defined for every real-valued function f in $C_*(\mathcal{O})$, continuous functions defined on vanishing at infinity, i.e., for every $\varepsilon > 0$ there exists a compact subset K of $\mathcal{O}$ such that $|f(x)| \leq \varepsilon$ for any x in $\mathcal{O} \setminus K$. Clearly, any function in $C_*(\mathcal{O})$ is extended to $\mathcal{O} \cup \{\infty\}$, by setting $f(\infty) = 0$. In this construction, the boundary $\partial\mathcal{O}$ is ignored or unseen by the *absorbed process* $x_{\mathcal{O}}(\cdot) = x^{\tilde{\tau}}(\cdot)$. The *particle* (represented by the path) dies (or disappears) as soon as it exits the domain $\mathcal{O}$, and it is sent to the *coffin* state denoted by the symbol ∞.

If the $(S_{\mathcal{O}}(t) : t \geq 0)$, given by (6.9) is a Feller semigroup, and $A_{\mathcal{O}}$ denotes its infinitesimal generator, then the potential function

$$u(x) := \int_0^\infty \mathrm{e}^{-\alpha t} S_{\mathcal{O}}(t)f(x)\mathrm{d}t$$

belongs to the domain $\mathcal{D}(A_{\mathcal{O}})$ of the infinitesimal generator $A_{\mathcal{O}}$ and it is a solution of the following boundary value problem

$$\begin{cases} -Au + \alpha u = f & \text{in} \quad \mathcal{O}, \\ u = 0 & \text{on} \quad \partial\mathcal{O}, \end{cases}$$

for every $\alpha > 0$ and f in $C_*(\mathcal{O})$, where the initial infinitesimal generator A coincides with $A_{\mathcal{O}}$ for functions in $C_*(\mathcal{O})$, and the so-called *Dirichlet boundary condition* is added. Note that when $\mathcal{O}$ is unbounded, we are also imposing a *vanishing* condition at infinity. Recall that, $u = R_\alpha f$, where

$$R(\alpha, A_{\mathcal{O}}) := (\alpha I - A_{\mathcal{O}})^{-1}, \quad \alpha > 0,$$

is the resolvent operator.

The converse (of this argument) is the construction of a Markov process starting from a sub-Markov process, i.e., when the semigroup $S(t)$ satisfies $S(t)1 \leq 1$ instead of $S(t)1 = 1$. In this case, a new state is added where all paths (with initial condition x_0) are sent after a time τ with an exponential distribution to complete the deficit $1 - S(t)1(x^0)$.

This technique can be applied to a stochastic ordinary differential equation in $\mathbb{R}^d$ and a delicate point is the dependency of the first exit time τ and $\tilde{\tau}$, given by (6.7), with respect to the initial data. As seen in Chapter 5, under natural conditions we have (local uniform) continuity of the path with respect to the initial data, and in view of Proposition 6.1, we deduce that $x^0 = x(t_0) \mapsto \tilde{\tau}(x, \mathcal{O})$ and $x^0 = x(t_0) \mapsto \tau(x, \overline{\mathcal{O}})$ are lower and upper semi-continuous, respectively. Hence, if

$$P\{x(t_0) = x^0,\ \tilde{\tau}(x, \mathcal{O}) = \tau(x, \overline{\mathcal{O}})\} = 1,$$

for every x^0 in the open domain $\mathcal{O}$ then $x^0 = x(t_0) \mapsto \tau(x, \overline{\mathcal{O}}) = \tilde{\tau}(x, \mathcal{O})$ is continuous from $\mathcal{O}$ into $[0, \infty]$. Since $\tilde{\tau}(x, \mathcal{O}) = 0$ whenever $x(t_0) = x^0$ belongs to $\partial\mathcal{O}$, to discuss continuity on the boundary we need to assume that the set of all initial (boundary) points which paths exit (the closure $\overline{\mathcal{O}}$) with full probability, i.e.,

$$\partial_0\mathcal{O} := \big\{x^0 \in \partial\mathcal{O} : x(t_0) = x^0,\ P\{\tau(x, \overline{\mathcal{O}}) > 0\} = 0\big\}. \tag{6.10}$$

is the whole boundary $\partial\mathcal{O}$. In view of the strong Markov property, if almost every path $x(\cdot)$ is continuous then the condition $\partial\mathcal{O}_0 = \partial\mathcal{O}$ suffices to ensure that $\tau(x, \overline{\mathcal{O}}) = \tilde{\tau}(x, \mathcal{O})$, for every $x(t_0) = x^0$ in $\overline{\mathcal{O}}$, which yields the continuity of the first exit time with respect to the initial data in the closure $\overline{\mathcal{O}}$. The case of only cad-lag paths need further consideration. For instance, we may have a time t, where $x(t)$ and $x(t-)$ belong to the interior $\mathcal{O}$ and to the boundary $\partial\mathcal{O}$, respectively, which yields the inequality $\tau(x, \overline{\mathcal{O}}) > t \geq \tilde{\tau}(x, \mathcal{O})$. On the other hand, if $\tau(x, \overline{\mathcal{O}})$ is continuous then $\partial\mathcal{O}_0$ is a closed set. In the case of non degenerated diffusion processes the converse also holds.

The main technique is based the existence of *barrier* functions, which should be constructed for each point in $\mathcal{O}_0$, the set of points where some path exits the

closed domain $\overline{\mathcal{O}}$, i.e., the minimal set satisfying

$$P\{x(\tau(x,\overline{\mathcal{O}})) \in \mathcal{O}_0,\ \tau(x,\overline{\mathcal{O}}) < \infty\} = 1, \tag{6.11}$$

for every initial condition $x(t_0) = x^0$ in the closure $\overline{\mathcal{O}}$ (which includes the set of all possible jumps from $\overline{\mathcal{O}}$). If almost every paths are continuous then the strong Markov property shows that $\mathcal{O}_0$ is the set of all initial points (on the boundary $\partial\mathcal{O}$) having paths which exit with non-zero probability, i.e.,

$$\mathcal{O}_0 = \big\{x^0 \in \partial\mathcal{O} : x(t_0) = x^0,\ P\{\tau(x,\overline{\mathcal{O}}) = 0\} > 0\big\}.$$

Moreover, one may hope that $P\{\tau(x,\overline{\mathcal{O}}) = 0\} > 0$ implies $P\{\tau(x,\overline{\mathcal{O}}) = 0\} = 1$, so that $\mathcal{O}_0 = \partial\mathcal{O}_0$, given by (6.11) and (6.10). In general, if paths are cad-lag, we only have $\mathcal{O}_0 \subset \mathbb{R}^d \setminus \mathcal{O}$. By definition of the subset $\mathcal{O}_0$, so-called *regular points*, we have

$$\tau(x(\cdot),\overline{\mathcal{O}}) = \inf\{t \geq t_0 : x(t) \in \mathcal{O}_0\},$$

where $\tau(x(\cdot),\overline{\mathcal{O}})$ is the first exit time from the closure as given by (6.7). Hence, the arguments of Proposition 6.1 show that $\tau(x(\cdot),\overline{\mathcal{O}})$ is lower semi-continuous if the set of regular points $\mathcal{O}_e$ is closed, in which case

$$\tau(x(\cdot),\overline{\mathcal{O}}) = \inf\{t \geq t_0 : x(t) \text{ or } x(t-) \in \mathcal{O}_0\},$$

and the function $x \mapsto \tau(x(\cdot),\overline{\mathcal{O}})$ is continuous with respect to the locally uniform converge.

A typical assumption is to suppose that for each point ξ in $\mathcal{O}_0$ there exists a real-valued function $x \mapsto \phi(\xi,x)$, defined in $\mathbb{R}^d$, smooth in $\overline{\mathcal{O}}$ (i.e., Itô formula can be applied up to the first exit time of the closure $\overline{\mathcal{O}}$) and such that $\phi(\xi,\xi) = 0$,

$$\begin{cases} -A\phi(\xi,\cdot) \geq 1, \quad \text{in} \quad \mathcal{O}, \\ \phi(\xi,\cdot) \geq 0 \quad \text{in} \quad \mathcal{O} \cup \mathcal{O}_0, \\ \phi_0(x) := \inf\limits_{\xi\in\mathcal{O}_0} \phi(\xi,x), \end{cases} \tag{6.12}$$

where $A = A(t)$ is the integro-differential operator (6.1), with coefficients independent of the variable t. The function ϕ_0 satisfies $\phi_0(\xi) = 0$ for every ξ in $\mathcal{O}_0$ and it is upper semi-continuous and non necessarily smooth. More details are needed to check that under (6.12), the set of regular points $\mathcal{O}_0$ is a closed. Now, any regularity imposed on the inf-barrier function ϕ_0 (also called sub-solution) is passed to the first exit time $\tau(x,\overline{\mathcal{O}})$.

Proposition 6.2. *Let $(x(t) : t \geq t_0)$ and $(\bar{x}(t) : t \geq \bar{t}_0)$ be a diffusion process with jumps corresponding the integro-differential operator (6.1), with initial condition $x(t_0) = x^0$ and $\bar{x}(\bar{t}_0) = \bar{x}^0$. Assume that condition (6.12) is satisfied with a finite* inf*-barrier function ϕ_0. Then*

$$\mathbb{E}\{\tau(x,\overline{\mathcal{O}})\} \leq t_0 + \phi_0(x^0),$$

so that $\tau = \tau(x, \overline{\mathcal{O}})$ is finite with full probability. Moreover

$$\mathbb{E}\big\{|\mathrm{e}^{-\alpha\tau} - \mathrm{e}^{-\alpha\bar\tau}|\big\} \le \alpha\mathbb{E}\big\{\mathrm{e}^{-\alpha(\tau\wedge\bar\tau)}|\phi_0(x(\tau\wedge\bar\tau) - \phi_0(\bar x(\tau\wedge\bar\tau)|\big\},$$

for every given $\alpha \ge 0$.

Proof. First, apply Itô formula to the function $x \mapsto \phi(\xi, x)$ and the diffusion process with jumps $x(t)$, between t_0 and $t \wedge \tau$, $\tau = \tau(x, \overline{\mathcal{O}})$, to deduce that

$$\mathbb{E}\{\phi(\xi, x(t\wedge\tau)) + t\wedge\tau\} \le t_0 + \phi(\xi, x(t_0)),$$

after using condition (6.12). This implies $\mathbb{E}\{\tau\} < \infty$, so $\tau < \infty$ almost surely.

Next, for any given constant $\alpha > 0$, apply Itô formula to the function $(t, x) \mapsto \mathrm{e}^{-\alpha t}\,\phi(\xi, x)$, between any stopping time $t_0 \le \theta \le \tau$ and τ, to deduce that the process

$$X(\xi, t) := \phi(\xi, x(t\wedge\tau))\,\mathrm{e}^{-\alpha(t\wedge\tau)} - \\ - \int_{t\wedge\theta}^{t\wedge\tau} [A\phi(\xi, x(s)) - \alpha\phi(\xi, x(s))]\mathrm{e}^{-\alpha(s-\theta)}\mathrm{d}s,$$

is a martingale after θ, for each ξ. Hence

$$X(\xi, \theta) = \mathbb{E}\{X(\xi, t) \mid \mathcal{F}(\theta)\} \ge \mathbb{E}\{\inf_{\xi\in\mathcal{O}_0} X(\xi, t) \mid \mathcal{F}(\theta)\},$$

which implies that

$$\mathbb{E}\Big\{\mathrm{e}^{-\alpha\tau}\phi_0(x(\tau)) + \frac{\mathrm{e}^{-\alpha\theta} - \mathrm{e}^{-\alpha\tau}}{\alpha} \;\Big|\; \mathcal{F}(\theta)\Big\} \le \mathrm{e}^{-\alpha\theta}\phi_0(x(\theta)),$$

as t goes to ∞. Now, take $\theta := \tau\wedge\bar\tau$, where $\bar\tau$ is the first exit time corresponding to the path $\bar x(\cdot)$ with initial condition $\bar x(\bar t_0) = \bar x^0$, to get

$$\mathbb{E}\{\mathrm{e}^{-\alpha(\tau\wedge\bar\tau)} - \mathrm{e}^{-\alpha\tau}\} \le \alpha\mathbb{E}\{\mathbb{1}_{\tau>\bar\tau}[\mathrm{e}^{-\alpha(\tau\wedge\bar\tau)}\phi_0(x(\tau\wedge\bar\tau) - \mathrm{e}^{-\alpha\tau}\phi_0(x(\tau))]\}.$$

Since $\phi_0(x(\tau)) = 0 = \phi_0(\bar x(\bar\tau))$ we deduce

$$\mathbb{E}\{\mathrm{e}^{-\alpha(\tau\wedge\bar\tau)} - \mathrm{e}^{-\alpha\tau}\} \le \alpha\mathbb{E}\{\mathbb{1}_{\tau>\bar\tau}\mathrm{e}^{-\alpha\bar\tau}|\phi_0(x(\bar\tau)) - \phi_0(\bar x(\bar\tau))|\},$$

and the proof is completed by symmetry. □

Actually, the existence of a (nonnegative) barrier function for just one point ξ in $\mathcal{O}_0$ implies that the first exit time from the closure $\overline{\mathcal{O}}$ has a finite moment, so it is finite almost surely. On the other hand, if $\phi_0(x^0) = 0$ then $\tau(x, \overline{\mathcal{O}}) = 0$, $x(t_0) = x^0$. Hence, if ϕ_0 is continuous then $\mathcal{O}_0$ is closed and $\tau(x, \overline{\mathcal{O}})$ is continuous. Moreover, the arguments of the previous Proposition 6.2 extend to any strong Markov process $x(\cdot)$ instead of a diffusion process with jumps.

Considering the strong solution of a stochastic ordinary differential equation, we can use Theorem 5.11 a priori estimates to show that if the inf-barrier function ϕ_0 of condition (6.12) is continuous or (locally) Lipschitz continuous, so is the function $x^0 \mapsto \mathbb{E}\{\mathrm{e}^{-\alpha\tau(x^0)}\}$, with $\tau(x^0) = \tau(x, \overline{\mathcal{O}})$, $x(t_0) = x^0$, $\alpha > 0$.

Moreover, an estimate with the modulus of continuity is obtained from Proposition 6.2, i.e., estimates uniformly with respect to the control parameter. Furthermore, continuity with respect to the initial time t_0 can be established, however, to get (locally) Lipschitz continuity we need to consider the time t as one more state variable, i.e., re-written the time non-homogeneous Markov process in $\mathbb{R}^d$ as an homogeneous Markov process in $\mathbb{R}^{d+1}$.

Note that one of the arguments used in Proposition 6.2 is the fact that the sup (inf) of sub-(super)martingales is also a sub-(super)martingale. Moreover, we can say

$$\begin{cases} -A\phi_0(\cdot) \geq 1, & \text{in martingale sense on} \quad \overline{\mathcal{O}} \setminus \mathcal{O}_0, \\ \phi_0(\xi) = 0, & \forall \xi \in \mathcal{O}_0. \end{cases} \tag{6.13}$$

In general, given three functions α, u and v defined in $\overline{\mathcal{O}}$, we say that $-Au+\alpha u \geq v$ in martingale sense on $\overline{\mathcal{O}} \setminus \mathcal{O}_0$ if the process

$$X(t) = u(x(t \wedge \tau)) \exp\Big[\int_{t_0}^{t\wedge\tau} \alpha(x(r))\mathrm{d}r\Big] + \\ + \int_{t_0}^{t\wedge\tau} v(x(s)) \exp\Big[\int_{t_0}^{s\wedge\tau} \alpha(x(r))\mathrm{d}r\Big]\mathrm{d}s, \quad \forall t \geq t_0$$

is a strong super-martingale, i.e., a (separable) super-martingale where the Doob's sampling theorem holds. For this to make sense, we have to assume that $\alpha \geq 0$ is bounded and Borel measurable, and u and v are bounded (or with polynomial growth) and Borel measurable, and u is upper semi-continuous over the paths, i.e., the function $t \mapsto u(x(t \wedge \tau))$ is (almost surely) upper semi-continuous from the right. Similarly, we say that $-Au + \alpha u \leq v$ in martingale sense on $\overline{\mathcal{O}} \setminus \mathcal{O}_0$ if the process $X(\cdot)$ is a strong sub-martingale, here u must be lower semi-continuous over the paths. Certainly, we do have the property $-Au + \alpha u \leq v$ if and only if $-Au \leq v - \alpha u$

If all jumps are inside the closure $\overline{\mathcal{O}}$, i.e., the jumps coefficient satisfies

$$x \in \overline{\mathcal{O}} \Rightarrow x + \gamma(\zeta, t, x, v) \in \overline{\mathcal{O}}, \quad \forall \zeta, t, v,$$

then $\mathcal{O}_0 \subset \partial\mathcal{O}$, moreover, $\mathcal{O}_0 = \partial_0\mathcal{O}$, i.e., the set of regular points as defined by (6.10). The Dirichlet problem takes the form

$$\begin{cases} Au + \alpha u = f & \text{in} \quad \mathcal{O} \setminus \partial_0\mathcal{O}, \\ u = 0 & \text{on} \quad \partial_0\mathcal{O}. \end{cases}$$

However, if jumps are allowed outside of the closed domain $\overline{\mathcal{O}}$ then $\mathcal{O}_0$ may contain points outside the closure $\overline{\mathcal{O}}$, so that the natural Dirichlet boundary condition $u = 0$ on the $\partial\mathcal{O}$ is incomplete, instead the subset $\mathcal{O}_0$ should be used. Thus, the whole exterior $\mathbb{R}^d \setminus \mathcal{O}$ may be involved in the *boundary* condition. In this case, a condition like *only jumps of order* 1 is necessary to obtain smooth solutions.

If the transition density function in the whole one-dimensional space $\mathbb{R}$ of the Wiener-Poisson process with parameter (γ_i, c_i) is denoted by

$$\Gamma_1(t, x_i, \gamma_i, c_i) := (2\pi t)^{-1/2} \mathrm{e}^{-c_i t} \sum_{k=0}^{\infty} (c_i t)^k \exp\big(-\frac{[x_i + k\gamma_i]^2}{2t}\big),$$

for every $t > 0$ and x_i in $\mathbb{R}$ (γ_i in $\mathbb{R}$ and $c_i > 0$), then the transition density function of the absorbed Wiener-Poisson process in the d-dimensional semi-space $\mathcal{O} = \mathbb{R}^d_+$ is given by

$$\Gamma(t, \tilde{x}, x_d, \xi_d) := \Big[\prod_{i=1}^{d-1} \Gamma_1(t, x_i, \gamma_i, c_i)\Big] \times \\ \times \big[\Gamma_1(t, x_d - \xi_d, \gamma_d, c_d) - \Gamma_1(t, x_d + \xi_d, \gamma_d, c_d)\big],$$

for every $t > 0$, $x = (x_1, \ldots, x_d)$ in $\mathbb{R}^d_+$ and $\xi_d \geq 0$, with $x = (\tilde{x}, x_d)$ and the condition $\gamma_d \geq 0$, so that no jumps are outside the closed semi-space $\bar{\mathbb{R}}^d_+ := \{(\tilde{x}, x_d) : \tilde{x} \in \mathbb{R}^{d-1},\ x_d \geq 0\}$.

A comprehensive study for degenerated diffusion processes can be found the paper Stroock and Varadhan [240], while the analytic counterpart, degenerate Dirichlet problems, is carefully considered in the book Oleinik and Radkevic [192]. Applications to optimal control are developed in a series of articles [170, 169] and Lions and Menaldi [156], among others. Diffusion processes with jumps, even in the non-degenerate case, present another challenge due to the nonlocal character of their (integro-differential elliptic operators) infinitesimal generators. The reader is referred to the book Garroni and Menaldi [94, Chapter 2, pp. 48–81] and references therein.

...***...

Include differentiability of the first exit time of a closed (smooth) subset of $\mathbb{R}^d$.

...***...

6.1.3 Sticking Boundary

The sticking barrier is another very natural boundary condition, which also uses the first exit time (6.7). Intuitively, the *particle* (represented by the path) does not die (nor disappear) when exiting the domain $\overline{\mathcal{O}}$, the path is stopped and instead of being sent to the *coffin* state, it remains there alive. Thus, given an initial Markov-Feller process $x(\cdot)$ acting in the whole space, say $\mathbb{R}^d$, the stopped Markov process $x(t \wedge \tau)$ takes values in $\mathcal{O}_e = \overline{\mathcal{O}} \cup \mathcal{O}_0$, where $\mathcal{O}_0$ is the minimal set satisfying

$$P\{x(\tau) \in \mathcal{O}_0,\ x(\tau-) \in \mathcal{O}_0,\ \tau < \infty\} = 1,$$

which includes the set of all possible jumps from $\overline{\mathcal{O}}$. In particular, if all paths $(x(t) : t \geq 0)$ are continuous then $x(t \wedge \tau)$ takes values in the closure $\overline{\mathcal{O}}$, moreover, for a time-homogeneous case, the strong Markov property shows that

$$\mathcal{O}_0 = \partial_0 \mathcal{O},$$

where

$$\partial_0\mathcal{O} := \big\{x^0 \in \partial\mathcal{O} : P\{\tau(x^0) = 0\} > 0\big\},$$
$$\tau(x^0) := \inf\{t \geq 0 : x(t) \notin \overline{\mathcal{O}},\ x(0) = x^0\},$$

and $x(0) = x^0$ is the initial condition. Note that $\tau = \tau(x^0, \omega)$ is the first exit time from the closure $\overline{\mathcal{O}}$.

The expression (6.9) of the stopped semigroup $(S_{\mathcal{O}}(t) : t \geq 0)$ is also valid for the sticking boundary, but now, we want to consider a space of function larger than $C_*(\mathcal{O})$, namely, $C_*(\mathcal{O}_e)$. Depending on the jumps, we may have $\mathcal{O}_e \subset \overline{\mathcal{O}}$ or even $\mathcal{O}_e = \mathbb{R}^d$. Contrary to the absorbing barrier, the process with sticking barrier may involve (a) the boundary $\partial\mathcal{O}$ if the process continuous, and the whole exterior $\mathbb{R}^d \setminus \mathcal{O}$ if the process is only cad-lag. For any x in $\partial_0\mathcal{O}$ or x in $\mathcal{O}_e \setminus \overline{\mathcal{O}}$, the path exits the domain $\overline{\mathcal{O}}$ with a positive probability, and because the evolution is stopped we have (formally) $A_{\mathcal{O}_e}u(x) = 0$, where $A_{\mathcal{O}_e}$ is the infinitesimal generator of stopped semigroup $(S_{\mathcal{O}}(t) : t \geq 0)$ considered in $C_*(\mathcal{O}_e)$. We are then treating the following boundary value problem

$$\begin{cases} -Au + \alpha u = f \quad \text{in} \quad \mathcal{O}, \\ Au = 0 \quad \text{on} \quad \partial_0\mathcal{O} \cup (\mathcal{O}_e \setminus \overline{\mathcal{O}}), \end{cases}$$

plus a *vanishing* condition at infinity if $\mathcal{O}_e$ is unbounded.

Sometimes, it not desirable to stop *completely* the evolution when exiting the domain $\overline{\mathcal{O}}$, what may be required is to restrict the evolution to be boundary. This is better seen in a simple domain, such as $\mathbb{R}^d_+ = \{x : x_d > 0\}$. In this case, the new stopped Markov process is defined by

$$x^\tau(t) := \big(x_1(t), \dots, x_{d-1}(t), x_d(t \wedge \tau)\big), \quad t \geq 0.$$

The boundary value problem associated with this process is given by

$$\begin{cases} -Au + \alpha u = f \quad \text{in} \quad \mathcal{O}, \\ A_T u = f \quad \text{on} \quad \partial_0\mathcal{O} \cup (\mathcal{O}_e \setminus \overline{\mathcal{O}}), \\ A_N u = 0 \quad \text{on} \quad \partial_0\mathcal{O} \cup (\mathcal{O}_e \setminus \overline{\mathcal{O}}), \end{cases}$$

plus a *vanishing* condition at infinity if $\mathcal{O}_e$ is unbounded, where A_T is the tangential part of A, i.e., for the case $\mathcal{O} = \mathbb{R}^d_+$, the operator A_T is A without the variables x_d, e.g., if $A = \sum_{i=1}^d \partial_i^2$ then $A_T = \sum_{i=1}^{d-1} \partial_i^2$. Similarly, A_N is the normal part of A, i.e., $A_N := A - A_T$.

In the semi-space $\mathbb{R}^d_+$ with a non-degenerate homogeneous (in time and space) diffusion with jumps, i.e., corresponding to an integro-differential operator with constant coefficients (and no jumps outside $\mathbb{R}^d_+$ from itself, $\gamma_d \geq 0$), we can have an explicit expression of the transition density function (after some change of variables, if necessary)

$$G(t, \tilde{x}, x_d, \xi_d) = \Gamma(t, \tilde{x}, x_d - \xi_d) - \Gamma(t, \tilde{x}, x_d + \xi_d) + \\ +2\,\delta_0(\xi_d) \int_{y_d \leq x_d} \Gamma(t, \tilde{x}, y_d)\mathrm{d}y_d,$$

where $\Gamma(t, \tilde{x}, x_d)$ is the transition density function in the whole space $\mathbb{R}^d$, $x = (\tilde{x}, x_d)$ and δ_0 is the Dirac measure at $\xi_d = 0$. Clearly, only the x_d is stopped at the first exist time but, with respect to the variable $\tilde{x}$ the diffusion (with jumps) is unchanged. Note that

$$\int_{[0,\infty)} G(t, \tilde{x}, x_d, \xi_d) \mathrm{d}\xi_d = \int_{\mathbb{R}} \Gamma(t, \tilde{x}, x_d - \xi_d) \mathrm{d}\xi_d,$$

for every x in $\bar{\mathbb{R}}^d_+$, $t > 0$. Moreover, if $(y(t) : t \geq 0)$ denotes the diffusion with jumps with sticking boundary, then

$$P\{y(t) \in B \mid y(0) = x\} = \int_B G(t, \tilde{x} - \tilde{\xi}, x_d, \xi_d) \mathrm{d}\xi,$$

for every x in $\mathbb{R}^d_+$ and $t > 0$. Here, the product expression (in terms of functions of one variable x_i) holds.

On the other hand, if all variables $\tilde{x}$ and x_d are stopped at the first exist time from the closed semi-space $\bar{\mathbb{R}}^d_+$, then the transition (density) function takes the form

$$G(t, \tilde{x}, x_d, \xi_d) = \Gamma(t, \tilde{x}, x_d - \xi_d) - \Gamma(t, \tilde{x}, x_d + \xi_d) + \\ +2\,\delta_0(\xi_d)\delta_0(\tilde{x}) \int_{\mathbb{R}^{d-1}} \mathrm{d}\tilde{y} \int_{y_d \leq x_d} \Gamma(t, \tilde{y}, y_d) \mathrm{d}y_d,$$

and clearly, we have

$$\int_{\bar{\mathbb{R}}^d_+} G(t, \tilde{x} - \tilde{\xi}, x_d, \xi_d) \mathrm{d}\xi = \int_{\mathbb{R}^d} \Gamma(t, \tilde{x} - \tilde{\xi}, x_d - \xi_d) \mathrm{d}\xi = 1,$$

for every $t > 0$. The reader is referred to the examples in Section 2.13 of Chapter 2.

6.1.4 Periodic Boundary Conditions

I do not what to put here, but we should cover this point!

...***...

6.1.5 SDE with Martingales

Check the end of Section 5.1.1 Existence and Uniqueness of Chapter 5.

...***...

6.2 SDE in a Semi-Space

Let us look at the semi-space $\mathbb{R}^d_+$, $r \geq 2$, with $x = (x_1, x_2, \ldots, x_d)$, $x_d \geq 0$, and boundary $\partial\mathbb{R}^d_+ = \{x : x_d = 0\}$, which is identified with whole space $\mathbb{R}^{d-1}$.

...***...

6.2.1 Some Wentzell Boundary Conditions

As early, we consider the following second-order differential boundary operator

$$\begin{cases} B\varphi(x) := B_0\varphi(x) + b_d\partial_d\varphi(x) - \rho\frac{1}{2}\Delta\varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \\ B_0\varphi(x) := \frac{1}{2}\sum\limits_{i,j=1}^{d-1} b_{ij}\partial_{ij}\varphi(x) + \sum\limits_{i=1}^{d-1} b_i\partial_i\varphi(x) - b_0\varphi(x), \end{cases}$$

where (b_{ij}) is a symmetric non-negative definite matrix, $b_0, b_d, \rho \geq 0$, and $\rho > 0$ or $b_d > 0$.

Let us recall that for any convex function $f : \mathbb{R} \to \mathbb{R}$ there exists a countable set $N \subset \mathbb{R}$ outside of which f is differentiable, i.e., the right-hand and the left-hand derivatives agree, $f'_+(x) = f'_-(x)$ for every x in $\mathbb{R} \setminus N$. Moreover

$$f(x) - f(0) = \int_0^x f'_-(r)\mathrm{d}r = \int_0^x f'_+(r)\mathrm{d}r, \quad \forall x \in \mathbb{R}.$$

Actually, f is locally Lipschitz, f'_+ and f'_- are non-decreasing and, by exchanging the monotone limits as $\varepsilon \downarrow 0$ with $h \downarrow 0$ in the expressions

$$\frac{f(x+\varepsilon+h) - f(x+\varepsilon)}{h} \quad \text{and} \quad \frac{f(x-\varepsilon-h) - f(x-\varepsilon)}{h},$$

we see that f'_+ is right-continuous and f'_- is left-continuous. Furthermore, the second derivative f'' exists as a measure. Indeed, because at each point x where f'_- or f'_+ is continuous (i.e., $f'_-(x+) = f'_-(x)$ or $f'_+(x-) = f'_+(x)$) we have $f'_-(x) = f'_+(x) = f'(x)$, it is clear that either

$$f''_-\big([a,b[\big) := f'_-(b) - f'_-(a) \quad \text{or} \quad f''_+\big(]a,b]\big) := f'_+(b) - f'_+(a), \quad \forall b > a,$$

induces the same Radon measure, denoted by f''. Hence, as a Riemann-Stieltjes or Lebesgue-Stieltjes integral, for every continuous function g, which is piecewise continuously differentiable having a compact support in $\mathbb{R}$, we have

$$\int_{-\infty}^{+\infty} g(x)\mathrm{d}f'_\pm(x) = \int_{-\infty}^{+\infty} g(x)f''(\mathrm{d}x) = -\int_{-\infty}^{+\infty} g'(x)f'(x)\mathrm{d}x,$$

where $g'(x)$ is defined for every x outside of a finite set, and $f'(x)$ is defined for every x outside of the countable set N.

Conversely, for any given Radon measure μ and any closed intervals $I \subset J$ the functions

$$\frac{1}{2}\int_I |x-y|\mu(\mathrm{d}y) \quad \text{and} \quad \frac{1}{2}\int_J |x-y|\mu(\mathrm{d}y)$$

are convex on the open interval $\dot{I}$ and their difference is an affine function. Thus, there exists a convex function f in the whole $\mathbb{R}$ such that

$$f(x) = \frac{1}{2}\int_I |x-y|\mu(\mathrm{d}y) + a_I x + b_I, \quad \forall x \in \mathbb{R},$$

for some constants a_I and b_I. Therefore

$$f'_{\pm}(x) = \tfrac{1}{2}\int_I \mathrm{sgn}_{\pm}(x-y)\mu(\mathrm{d}y) + a_I, \quad \forall x \in \dot{I},$$

where the sign functions $\mathrm{sgn}_-(x) := 1$ if $x > 0$ and $\mathrm{sgn}_-(x) = -1$ if $x \le 0$, while $\mathrm{sgn}_-(x) := 1$ if $x \ge 0$ and $\mathrm{sgn}_-(x) = -1$ if $x < 0$. Hence, for every smooth function ϕ with a compact support in the open interval $\dot{I}$ we have

$$\int_{\mathbb{R}} \phi(x)\mu(\mathrm{d}x) = -\int_{\mathbb{R}} \phi'(x) f'_{\pm}(x)\mathrm{d}x,$$

i.e., $f'' = \mu$. Moreover, if μ integrate the function $y \mapsto |x-y|$ for every x (e.g., μ has a compact support) then one can write

$$f(x) = \tfrac{1}{2}\int_{\mathbb{R}} |x-y|\mu(\mathrm{d}y) + ax + b, \quad \forall x \in \mathbb{R},$$
$$f'_{\pm}(x) = \tfrac{1}{2}\int_{\mathbb{R}} \mathrm{sgn}_{\pm}(x-a)\mu(\mathrm{d}a) + a, \quad \forall x \in \mathbb{R},$$

for some constants a and b.

Applying Itô formula to the function $f_\varepsilon : \mathbb{R} \to \mathbb{R}$, $f_\varepsilon(x) = \sqrt{\varepsilon^2 + x^2}$, and the Brownian motion $\beta(t)$, with $\beta(0) = 0$, we have

$$f_\varepsilon(x+\beta(t)) = f_\varepsilon(x) + \int_0^t f'_\varepsilon(x+\beta(s))\mathrm{d}\beta(s) + \tfrac{1}{2}\int_0^t f''_\varepsilon(x+\beta(s))\mathrm{d}s.$$

Since $f'_\varepsilon(x) = x\,(\varepsilon^2 + x^2)^{-1/2}$, we get

$$f_\varepsilon(x+\beta(t)) \to |x+\beta(t)| \quad \text{and} \quad f'_\varepsilon(x+\beta(t)) \to \mathrm{sgn}(x+\beta(t)),$$

as $\varepsilon \to 0$, where the sign function is defined as $\mathrm{sgn}(x) = 1$ if $x > 0$, $\mathrm{sgn}(x) = -1$ if $x < 0$ and $\mathrm{sgn}(x) = 0$ if $x = 0$. Hence we have the following convergence in probability

$$\int_0^t f'_\varepsilon(x+\beta(s))\mathrm{d}\beta(s) \to \int_0^t \mathrm{sgn}(x+\beta(s))\mathrm{d}\beta(s),$$

Now, observe that the process

$$\ell_\varepsilon(t,x) = \tfrac{1}{2}\int_0^t f''_\varepsilon(x+\beta(s))\mathrm{d}s$$

is a non negative, non decreasing process, converging in probability to

$$\ell_\varepsilon(t,x) \to \ell(t,x) := |x+\beta(t)| - |x| - \int_0^t \mathrm{sgn}(x+\beta(s))\mathrm{d}\beta(s). \tag{6.14}$$

By means of Burkhölder-Davis-Gundy inequalities applied to the process

$$\bar{\beta}(t,x) = \int_0^t \mathrm{sgn}(x+\beta(s))\,\mathrm{d}\beta(s),$$

with some $p > 2$. i.e., for some positive constant C_p,

$$\mathbb{E}\{|\bar\beta(t,x) - \bar\beta(t',x)|^p\} \le C_p|t-t'|^{p/2}.$$

It is also clear that $\bar\beta(t,x)$ can be taken jointly measurable in (t,x). Moreover, a direct calculation shows (e.g., see Karatzas and Shreve [124, Section 3.6.B, pages 207]) that

$$\mathbb{E}\{|\bar\beta(t,x) - \bar\beta(t,x')|^p\} \le C_p|x-x'|^{p/2},$$

which implies that the process $\ell(t,x)$ can be chosen continuous in (t,x). Clearly, this also shows that process $\bar\beta(t)$ is a Brownian motion. Moreover $\ell(t,x)$ is non negative, non decreasing, and constant whenever $x + \beta(t) \neq 0$. This process $\ell(t,x)$ is called the *local time* of the Brownian motion $\beta(t)$ at level x. We have also

Theorem 6.3 (Itô-Tanaka)**.** *If f is a convex function, then*

$$f(\beta(t)) = f(0) + \int_0^t f'_-(\beta(s))\,\mathrm{d}\beta(s) + \tfrac{1}{2}\int_0^t \ell(t,x) f''(\mathrm{d}x) \tag{6.15}$$

where we recall that the second derivative of a convex function exists as a measure. Moreover, the occupation times formula holds, i.e., there exists a negligible set outside of which

$$\int_0^t \phi(\beta(s))\,\mathrm{d}s = \int_{\mathbb{R}} \phi(x)\,\ell(t,x)\,\mathrm{d}x$$

for every $t \ge 0$ and for any nonnegative Borel function ϕ.

Proof. First, for any nonnegative continuous function ϕ one can find a convex function f such that $f'' = \phi$ and, by comparing the Itô formula and the Itô-Tanaka formula one gets occupation times formula for ϕ. Next, one can extend the result to any nonnegative Borel function by approximations, i.e., as an application of the monotone class theorem.

To prove Itô-Tanaka formula for any convex function f, one sees that by stopping the Wiener process $\beta(t)$ at the first exit time from a compact set, it suffice to consider only convex functions f having f'' with compact support, i.e.,

$$f(x) = \tfrac{1}{2}\int_{\mathbb{R}} |x-y| f''(\mathrm{d}y) + ax + b, \quad \forall x \in \mathbb{R},$$

for some constants a and b. Hence, by means of (6.14) we get

$$\begin{aligned} f(\beta(t)) &= \tfrac{1}{2}\int_{\mathbb{R}} |\beta(t)-y|\mathrm{d}y + a\beta(t) + b = \\ &= a\beta(t) + f(0) + \tfrac{1}{2}\int_{\mathbb{R}} \Big(\int_0^t \operatorname{sgn}\big(y+\beta(s)\big)\mathrm{d}\beta(s) + \ell(t,y)\Big) f''(\mathrm{d}y). \end{aligned}$$

Since $2\,\mathrm{sgn}(\cdot) = \mathrm{sgn}_+(\cdot) + \mathrm{sgn}_-(\cdot)$ we have

$$\int_{\mathbb{R}} \mathrm{sgn}\big(y+\beta(t)\big) f''(\mathrm{d}y) = f'_+(\beta(t)) + f'_-(\beta(t)),$$

which yields the symmetric Itô-Tanaka formula

$$f(\beta(t)) = f(0) + \tfrac{1}{2}\int_0^t [f'_- + f'_+](\beta(s))\,\mathrm{d}\beta(s) + \tfrac{1}{2}\int_0^t \ell(t,x) f''(\mathrm{d}x).$$

Actually, because $P\{y+\beta(s) = 0\} = 0$, the processes $\mathrm{sgn}_\pm(y+\beta(t))$ and $\mathrm{sgn}(y+\beta(t))$ are versions of each other and the stochastic integral coincides, i.e., we can replace $\mathrm{sgn}(\cdot)$ above with either $\mathrm{sgn}_+(\cdot)$ or $\mathrm{sgn}_-(\cdot)$, and deduce Itô-Tanaka formula (6.15). □

The previous results remain valid for (continuous) semi-martingales, and clearly, the proof of Itô-Tanaka formula requires only the joint measurability of the local time $\ell(t,x)$. The (right-) continuity of $\ell(t,x)$ in x can be proved by means of the occupation times formula and a bound on $\mathbb{E}\{[\ell(t,x)]^p\}$ locally uniformly in x, for each $t>0$ and some $p>2$. Depending on how the smooth functions f_ε approximates $x \mapsto |x|$ (e.g., with $f'_\varepsilon \to \mathrm{sgn}_-$ in lieu of sgn), we may define the local time process as

$$\ell_\pm(t,x) := |x+\beta(t)| - |x| - \int_0^t \mathrm{sgn}_\pm(x+\beta(s))\mathrm{d}\beta(s),$$

instead of (6.14). Thus, to have f'_+, f'_- or the symmetric sum $(f'_+ + f'_-)/2$ we may need to use the local times ℓ_+, ℓ_- or ℓ. For instance, full details on the local times can be found in Karatzas and Shreve [124, Section 3.6, pp. 201–226] or Revuz and Yor [212, Chapter VI, pp. 221–277].

Besides the fact that the $(t,x) \mapsto \ell(t,x)$ could be only right-continuous, it seems to me the above arguments are applicable for a Wiener-Poisson process (i.e., quasi-continuous local-martingale?). Taking a look at Bertoin [21] or Sato [220] book, the local time for non-continuous semi-martingale (e.g., a Levy process) is so, so, so complicate.
What do you think?

After all this preliminaries, by making use of the occupation times formula, one can define the local time $\ell(t,x)$ associated with the process $x+\beta(t)$ as a continuous non decreasing process satisfying

$$\ell(t,x) = \lim_{\epsilon\to 0} \tfrac{1}{2\epsilon}\int_0^t \mathbb{1}_{[-\epsilon,\epsilon]}(x+\beta_r)\mathrm{d}r,$$

which is constant when $x+\beta(t) \neq 0$. Actually, there is a simpler formula, namely

$$\tau(x) := \inf\big\{t \geq 0 : x+\beta(t) \geq 0\big\},$$
$$\ell(t,x) = \tfrac{1}{2}\max\big\{x+\beta(s) : \tau(x) \leq s \leq t\big\},$$

valid in this case. Clearly, one may deduce the continuity in (t, x) directly from the above expressions.

With a multidimensional notation, let $w(t) = \big(w_1(t), \dots, w_d(t)\big)$ be a d-dimensional (standard) Wiener process. The following cases can be considered.

Example 6.1 (Dirichlet)**.** This is the case $b_0 = 1$, $\tilde{b} = 0$, $b_d = 0$, $\rho = 0$ and $(b_{ij}) = 0$. The process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by $X_t = x + w(t)$ with a *lifetime* $\zeta = \tau_x = \inf\{t > 0 : w_d(t) = x_d\}$. We can reformulate with the language of local time $\ell(t, x)$ of $x_d + w_d(t)$ at zero, that is $\zeta = \inf\{t > 0 : \ell(t, x) > 0\} = \tau_x$. □

Example 6.2 (Neumann)**.** This is the case $b_d = 1$, $b_0 = 0$, $\tilde{b} = 0$, $\rho = 0$ and $(b_{ij}) = 0$. The process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by $X_t = x + (w_1(t), w_2(t), \dots, w_{d-1}(t), |w_d(t)|)$. □

Example 6.3 (Absorbing)**.** This is the case $b_0 = 0$, $\tilde{b} = 0$, $b_d = 0$, $\rho = 1$ and $(b_{ij}) = 0$, which is also called *sticking* The process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by $X(t) = x + w(t \wedge \tau_x)$ where $\tau_x = \inf\{t > 0 : w_d(t) = x_d\}$. □

Example 6.4 (Elastic)**.** This is the case $b_0 > 0$, $\tilde{b} = 0$, $b_d > 0$, $\rho = 0$ and $(b_{ij}) = 0$. The process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by $X(t) = x + (w_1(t), w_2(t), \dots, w_{d-1}(t), |w_d(t)|)$ with a *lifetime* $\zeta = \inf\{t > 0 : \ell(t, x) > Z\}$, where Z is an independent (of the Brownian motion) random variable exponentially distributed with parameter $\lambda = \frac{b_0}{b_d}$. Observe that the limit case $b_d = 0$ corresponds to the Dirichlet case, and the other limit case $b_0 = 0$ corresponds to Neumann case. □

Example 6.5 (Sticky)**.** This is the case $b_0 = 0$, $\tilde{b} = 0$, $b_d > 0$, $\rho > 0$ and $(b_{ij}) = 0$. The process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by $X(t) = x + (w_1(r(t)), w_2(r(t)), \dots, w_{d-1}(r(t)), |w_d(r(t))|)$, where the random process $r(t)$ is the inverse function of $c(t) = t + \frac{\rho}{b_d}\ell(t, x)$. Observe that in the limit case of $b_d = 0$, one obtains the absorbing case, and as $\rho \to 0$ one obtains again the Neumann case. □

Example 6.6 (Oblique)**.** This is the case $b_0 = 0$, $\tilde{b} \neq 0$, $b_d > 0$, $\rho = 0$ and $(b_{ij}) = 0$. □

More stuff here...

Example 6.7 (Second-Order)**.** This is the case $b_0 = 0$, $\tilde{b} \neq 0$, $b_d > 0$, $\rho = 0$ and $(b_{ij}) \neq 0$. □

More stuff here...

Example 6.8 (In General)**.** Now, it easier to understand the general case with $b_d > 0$, the process associated to the Laplacian in the space $\mathbb{R}^d_+$ is given by

$$X(t) = x + (w_1(r(t)), w_2(r(t)), \dots, w_{d-1}(r(t)), |w_d(r(t))|),$$

where the random process $r(t)$ is the inverse function of $c(t) = t + \frac{\rho}{b_d}\ell(t,x)$, but with a *lifetime* $\zeta = \inf\{t > 0 : \ell(r(t), x) > Z\}$, where Z is an independent (of the Brownian motion) random variable exponentially distributed with parameter $\lambda = \frac{b_0}{b_d}$. □

...***...

6.2.2 Drift and Diffusion Terms

Suppose that on the interior $\dot{\mathbb{R}}^d_+ = \{x : x_d > 0\}$ we are given a a second-order elliptic (possible degenerate) differential operator (without zero-order term)

$$A(t)\varphi(x) := \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij}\varphi(x) + \sum_{i=1}^{d} a_i(t,x)\partial_i\varphi(x), \tag{6.16}$$

for very continuously differentiable functions φ with a compact support in the closed semi-space $\mathbb{R}^d_+$, where $(a_{ij}(t,x))$ is a symmetric non-negative definite matrix for each (t,x), and all coefficients $a_{ij}(t,x)$ and $a_i(t,x)$ are bounded (linear growth suffices) and continuous functions. A boundary operator of *Wentzell type* has the form

$$\begin{cases} B(t)\varphi(x) := B_0(t)\varphi(x) + b_d(t,x)\partial_d\varphi(x) - \rho(t,x)A(t)\varphi(x), \\ B_0(t)\varphi(x) := \sum_{i,j=1}^{d-1} b_{ij}(t,x)\partial_{ij}\varphi(x) + \sum_{i=1}^{d-1} b_i(t,x)\partial_i\varphi(x), \end{cases} \tag{6.17}$$

where $(b_{ij}(t,x))$ is a symmetric non-negative definite matrix, $b_d(t,x) \geq 0$ and $\rho(t,x) \geq 0$, for every (t,x), and all coefficients are bounded and continuous. Notice that ∂_d is the interior normal derivative on the boundary $\partial\mathbb{R}^d_+$.

A diffusion measure $P = P_{t_0,x}$ generated by the pair $A(t)$ and $B(t)$ with initial condition $P\{x(t_0) = x_0\}$, $x(t) := \omega(t)$, is a probability measure on the canonical space $C([t_0,\infty), \mathbb{R}^d_+)$ such that there exists an adapted, non-decreasing and continuous process $\ell(t)$, so-called *local time*, satisfying the conditions $\ell(t_0) = 0$,

$$\begin{aligned} &\int_{t_0}^{t} \mathbb{1}_{\partial\mathbb{R}^d_+}(x(s))\mathrm{d}\ell(s) = \ell(t), \quad \forall t \geq t_0, \\ &\int_{t_0}^{t} \mathbb{1}_{\partial\mathbb{R}^d_+}(x(s))\mathrm{d}s = \int_{t_0}^{t} \rho(s,x(s))\mathrm{d}\ell(s), \quad \forall t \geq t_0, \end{aligned}$$

and such that the process

$$\begin{aligned} M_\varphi(t) := \varphi(x(t)) - \varphi(x(t_0)) - \int_{t_0}^{t} A(s)\varphi(x(s))\mathrm{d}s - \\ - \int_{t_0}^{t} B(s)\varphi(x(s))\mathrm{d}\ell(s), \end{aligned}$$

is a martingale for any smooth function φ.

Notice that the boundary operator $B(t)$ is determined by the diffusion measure up to a multiplicative constant, i.e., if $\tilde{B}(t)\varphi(x) := \alpha(t,x)B(t)\varphi(x)$, with $\alpha(t,x)$ being a positive and continuous functions on the boundary $\mathbb{R}^d_+$, then

$$\int_{t_0}^t B(s)\varphi(x(s))\mathrm{d}\ell(s) = \int_{t_0}^t \tilde{B}(s)\varphi(x(s))\mathrm{d}\tilde{\ell}(s),$$

where

$$\tilde{\ell}(t) := \int_{t_0}^t \alpha(s,x(s))\varphi(x(s))\mathrm{d}\ell(s), \quad \forall t \geq t_0.$$

On the other hand, the martingale process $M_\varphi(t)$ can be re-written as

$$M_\varphi(t) = \varphi(x(t)) - \varphi(x(t_0)) - \int_{t_0}^t \mathbb{1}_{\dot{\mathbb{R}}^d_+}(x(s))\, A(s)\varphi(x(s))\mathrm{d}s - \\ - \int_{t_0}^t \mathbb{1}_{\partial\mathbb{R}^d_+}(x(s))\, B'(s)\varphi(x(s))\mathrm{d}\ell(s), \quad \forall t \geq t_0,$$

where $B' = B_0 + b_d\partial_d$, i.e., $B = B' - \rho A$, and $\dot{\mathbb{R}}^d_+$ and $\partial\mathbb{R}^d_+$ are the interior and the boundary of the half-space $\mathbb{R}^d_+$.

All this can be written in the form of a stochastic differential equation

$$\mathrm{d}x_i(t) = \mathbb{1}_{\dot{\mathbb{R}}^d_+}(x(t))\left[a_i(t,x(t))\mathrm{d}t + \sum_{k=1}^n \sigma_{ik}(t,x(t))\mathrm{d}w_k(t)\right] + \\ + \mathbb{1}_{\partial\mathbb{R}^d_+}(x(t))\left[b_i(t,x(t))\mathrm{d}\ell(t) + \sum_{k=1}^m \varsigma_{ik}(t,x(t))\mathrm{d}w^\ell_k(t)\right],$$

for $i = 1,\dots,d-1$,

$$\mathrm{d}x_d(t) = \mathbb{1}_{\dot{\mathbb{R}}^d_+}(x(t))\left[a_d(t,x(t))\mathrm{d}t + \sum_{k=1}^n \sigma_{dk}(t,x(t))\mathrm{d}w_k(t)\right] + \\ + b_d(t,x(t))\mathrm{d}\ell(t),$$

and

$$\mathbb{1}_{\partial\mathbb{R}^d_+}(x(t))\mathrm{d}t = \rho(t,x(t))\mathrm{d}\ell(t),$$

where $(w_1,\dots,w_n)$ is a (standard) n-dimensional Wiener process, and the process $(w^\ell_1,\dots,w^\ell_m)$ is an independent m-dimensional continuous square-integrable martingale with predictable quadratic co-variance given by

$$\mathrm{d}\langle w^\ell_i, w^\ell_j\rangle = \begin{cases} \mathrm{d}\ell & \text{if } 1 \leq i,j \leq n, \\ 0 & \text{otherwise,} \end{cases}$$

i.e., $w^\ell_i(s) = w^*_i(\ell(s))$ for some independent (standard) m-dimensional Wiener process $(w^*_1,\dots,w^*_m)$, and

$$a_{ij} = \frac{1}{2}\sum_{k=1}^n \sigma_{ik}\sigma_{jk}, \quad 1 \leq i,j \leq d,$$
$$b_{ij} = \frac{1}{2}\sum_{k=1}^m \varsigma_{ik}\varsigma_{jk}, \quad 1 \leq i,j \leq d-1.$$

relates the coefficients.

Roughly speaking, $b_d(t,x) > 0$ produces a normal *reflection* and $\rho(t,x) > 0$ yields a *sojourn* or *viscosity*, both at time t and state x on the boundary $\partial\mathbb{R}^d_+$. So, if $b_d(t,x) = \rho(t,x) = 0$ then the above boundary problem is not allowed, unless the diffusion is degenerate and the characteristics does not exit the domain $\mathbb{R}^d_+$.

Theorem 6.4. *Assume the coefficients $\sigma_{ik}(t,x)$, $\varsigma_{ik}(t,x)$, $a_i(t,x)$, $b_i(t,x)$ are measurable in (t,x), locally uniform Lipschitz continuous in x and at most of linear growth in x. Also suppose that $\rho(t,x)$ is bounded and continuous, that $\varsigma_{ik}(t,x)$ and $b_i(t,x)$ are also continuous, and that there exists a constant $c > 0$ satisfying*

$$b_d(t,x) \geq c, \quad \forall t \geq t_0,\ x \in \partial\mathbb{R}^d_+,$$
$$\sum_{k=1}^{p} \sigma^2_{dk}(t,x) \geq c, \quad \forall t \geq t_0,\ x \in \mathbb{R}^d_+.$$

Then for any initial condition x_0 at t_0 there exists a solution $x(t)$, $\ell(t)$ of the above stochastic differential equation with boundary conditions, which is unique in law. □

This is a variation of Ikeda and Watanabe [110, Theorem IV.7.2, pp. 222–228], where the proof is given under the assumption that all coefficients are time-independent and bounded. The steps in the proof are as follows. First, a simple normal reflection problem is solved, i.e., with $\rho(t,x) = 0$, $a_d(t,x) = 0$, $\sigma_{d1}(t,x) = 1$ and $\sigma_{dk}(t,x) = 0$ for $k = 2,\dots,n$. Indeed, given a (standard) $n+m$-dimensional Wiener process $(\beta_1,\dots,\beta_{n+m})$ define the processes $x_d(t)$ and $\ell(t)$ by

$$\ell(t) := \max_{t_0 \leq s \leq t} \left\{[\beta_1(s) + x_d(t_0)]^-\right\} \quad \forall t \geq t_0,$$
$$x_d(t) := x_d(t_0) + \beta_1(t) + \ell(t), \quad \forall t \geq t_0,$$

where $[\cdot]^-$ denotes the negative part of a real number. Thus, x_d is the reflecting Wiener process in $[0,\infty)$ and ℓ its local time. Now set $\tilde{x} = (x_1,\dots,x_{d-1})$ and $w_k(t) := \beta_k(t)$ if $1 \leq k \leq n$, and $w^\ell_k(t) := \beta_{m+k}(\ell(t))$ if $n+1 \leq k \leq n+m$, to solve the $(d-1)$-dimensional stochastic differential equation

$$\mathrm{d}\tilde{x}_i(t) = \sum_{k=1}^{n} \sigma_{ik}(t,\tilde{x}(t),x_d(t))\mathrm{d}w_k(t) + a_i(t,\tilde{x}(t),x_d(t))\mathrm{d}t + \\ + \sum_{k=1}^{m} \varsigma_{ik}(t,\tilde{x}(t),x_d(t))\mathrm{d}w^\ell_k(t) + b_i(t,\tilde{x}(t),x_d(t))\mathrm{d}\ell(t),$$

as in the previous chapter. Actually, because w^ℓ is a continuous square-integrable martingale (no in general a Wiener process), we need to change the time t into $t + \ell(t)$ to apply the iteration method of Chapter 5. Notice that only the values of $\varsigma(t,x)$ and $b_i(t,x)$ on the boundary $\partial\mathbb{R}^d_+$ are used, i.e., $x_d(t) = 0$ when $\mathrm{d}w^\ell(t) \neq 0$ or $\mathrm{d}\ell(t) \neq 0$. This yields a $\mathbb{R}^d_+$-valued process x which satisfies all

the required conditions. Next, by means of a rotation of the Wiener process, i.e., a transformation

$$\tilde{w}_k(t) := \sum_{h=1}^{n} \int_{t_0}^{t} r_{kh}(s, x(s)) \mathrm{d}w_h(s),$$

where $r = (r_{hk}(t,x))$ is a n-dimensional orthogonal matrix for each (t,x), we can modify the components of $\sigma(t,x)$, i.e., the matrix $(a_{ij}(t,x))$, and reduce to the previous simple reflection case, after modifying the drift (setting $a_d = 0$) by means of a Girsanov transformation. Finally, if $(x(\cdot), w(\cdot), \ell(\cdot), w^{\ell}(\cdot))$ is a solution corresponding to σ_{ik}, ς_{ik}, a_i, b_i and $\rho = 0$, then apply a time change of the type

$$\tau(t) := t + \int_{t_0}^{t} \rho(x(s)) \mathrm{d}\ell(s), \quad \forall t \geq t_0,$$

as follows

$$\begin{aligned}
&x^{\tau}(t) := x(\tau^{-1}(t)), \qquad \ell^{\tau}(t) := \ell(\tau^{-1}(t)),\\
&w_k^{\ell,\tau}(t) := w_k^{\ell}(\tau^{-1}(t)),\\
&w_k^{\tau}(t) := w_k(\tau^{-1}(t)) + \int_{t_0}^{t} \mathbb{1}_{\partial \mathbb{R}^d_+}\big(x(\tau^{-1}(s))\big) \mathrm{d}\beta_k^*(s), \quad 1 \leq k \leq n,
\end{aligned}$$

where $(\beta_1^*, \ldots, \beta_n^*)$ is another (standard) n-dimensional Wiener process, to obtain the solution $(x^{\tau}(\cdot), w^{\tau}(\cdot), \ell^{\tau}(\cdot), w^{\ell,\tau}(\cdot))$, for the case $\rho(x) \geq 0$. Indeed, the relations

$$\begin{aligned}
&\tau^{-1}(t) = \int_{t_0}^{t} \mathbb{1}_{\dot{\mathbb{R}}^d_+}\big(x(\tau^{-1}(s))\big) \mathrm{d}s,\\
&\int_{t_0}^{t} \mathbb{1}_{\partial \mathbb{R}^d_+}\big(x(\tau^{-1}(s))\big) \mathrm{d}s = \int_{t_0}^{t} \rho(x(\tau^{-1}(s)) \mathrm{d}\ell(\tau^{-1}(s))
\end{aligned}$$

yield the desired existence result.

The uniqueness in law is ensured by the (strong) uniqueness of the $(d-1)$-dimensional stochastic differential equation and the invertibility of the above changes of probability and time.

The condition imposed on the coefficients (reflection and non-degeneracy on the boundary) may be too strong in some cases. Reflection on the boundary can be replaced by

$$b_d(t,x) + \rho(t,x) \geq c > 0, \quad \forall t \geq t_0,\ x \in \partial \mathbb{R}^d_+,$$

so that a reflection or a sojourn is allowed on the boundary. Probabilistically, this transversality condition means that every Markov process on $\partial \mathcal{O}$ is the *trace* on the boundary of some Markov process on $\overline{\mathcal{O}}$. On the other hand, the non-degeneracy condition is only used to construct the local time $\ell(t)$ and the reflecting coordinate $x^d(t)$, but this can be accomplished independently, e.g. see the papers Stroock and Varadhan [239, 240].

Is this correct with t? See Ikeda and Watanabe [110, Section IV.7.2, pp. 217–232].

6.2.3 General Boundary Conditions

Essentially results form paper Anulova [3, 4]?

...***...

Skubachevskii [231] (and references therein) for a full treatment of even more general boundary conditions for Markov-Feller processes.

...***...

6.3 Stochastic Variational Inequalities

In this section we discuss the reflected stochastic differential equation as a set of two inequalities with a complementary condition, typically conditions derived form an orthogonal projection. In a deterministic situation this is very well known, e.g., see the classic books Bensoussan and Lions [16], Kinderlehrer and Stampacchia [131] among others. Here we follows the arguments presented in the papers [171], Menaldi and Robin [173], and the references therein, e.g., Tanaka [247]. This is mainly concerning the strong formulation, but it can be easily adapted to obtain existence and uniqueness of weak solutions. Moreover, some results for stochastic partial differential equations are obtained in Bensoussan and Rascanu [19] and Rascanu [209].

Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t, w(t), \tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) : \zeta \in \mathbb{R}^m_*,\ t \geq 0)$, be a (standard) $n \times m$ Wiener-Poisson space with Lévy measure $\pi(\cdot)$, i.e., in a complete filtered probability space $(\Omega, \mathcal{F}, P, F_t : t \geq 0)$, the stochastic process $(w(t) : t \geq 0)$ is a n-dimensional (standard) Wiener space and $(\nu(B,]0, t]) : B \in \mathbb{R}^m_*, t \geq 0)$ is an independent (standard) Poisson measure with (intensity) Lévy measure $\pi(B) := \mathbb{E}\{\nu(B,]0, t])\}/t$, which satisfies

$$\int_{\mathbb{R}^m_*} \frac{|\zeta|^2}{1 + |\zeta|} \pi(\mathrm{d}\zeta) < \infty,$$

with martingale measure $\tilde{\nu}(B,]0, t]) := \nu(B,]0, t]) - t\pi(B)$, as discussed in previous chapters. This Wiener-Poisson space as well as a $\mathbb{R}^q$-valued adapted process (control process) $(v(t) : t \geq 0)$ are fixed throughout this section.

The coefficients $g(t, x, v)$, $\sigma(t, x, v)$ and $\gamma(\zeta, t, x, v)$ are always supposed Borel measurable, and because we are interested in global solutions defined on a prescribed bounded interval, say $[t_0, t_1]$, we impose a linear growth condition, namely, there exist a constant $C > 0$ such that

$$\left\{ \begin{array}{l} |g(t,x,v)|^2 + |\sigma(t,x,v)|^2 + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^2 \pi(\mathrm{d}\zeta) \leq \\ \qquad\qquad \leq C(1 + |x|^2), \end{array} \right. \tag{6.18}$$

for every (t, x, v) in $[t_0, \infty) \times \mathbb{R}^d \times \mathbb{R}^q$. Also we add a uniform locally Lipschitz condition in the variable x, namely, for any $r > 0$ there exists a positive constant

$M = M(r)$ such that

$$\begin{cases} |g(t,x,v) - g(t,x',v)|^2 + |\sigma(t,x,v) - \sigma(t,x',v)|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v) - \gamma(\zeta,t,x',v)|^2 \pi(\mathrm{d}\zeta) \le M|x-x'|^2, \end{cases} \tag{6.19}$$

for every (t,x,v), (t,x',v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$ with $t \le t_0 + r$, $|x| \le r$ and $|x'| \le r$. Recall that we say uniform (global) Lipschitz condition in the variable x if the bound (6.19) holds true for any (t,x,v), (t,x',v) in $[t_0,t_0+r)\times\mathbb{R}^d\times\mathbb{R}^q$.

The associated integro-differential operator

$$\begin{cases} A(t,v)\varphi(x) = \displaystyle\sum_{i,j=1}^d a_{ij}(t,x,v)\partial_{ij}\varphi(x) + \sum_{i=1}^d g_i(t,x,v)\partial_i\varphi(x) + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_*} [\varphi(x+z) - \varphi(x) - z\cdot\nabla\varphi(x)]\mathtt{M}(\mathrm{d}z,t,x,v), \end{cases} \tag{6.20}$$

where ∂_i and ∂_{ij} denote the partial first x_i and second x_i, x_j derivatives and ∇ is the gradient in the variable x,

$$a_{ij}(t,x,v) := \frac{1}{2}\sum_{k=1}^n \sigma_{ik}(t,x,v)\,\sigma_{jk}(t,x,v), \tag{6.21}$$

and

$$\mathtt{M}(B,t,x,v) = \pi(\{\zeta : \gamma(\zeta,t,x,v) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}^d_*) \tag{6.22}$$

is the jumps or Lévy kernel in $\mathbb{R}^d_*$.

First, the case of a convex domain not necessarily smooth is studied. Next, this is extended to a simply connected smooth domain via a diffeomorphism and finally the technique is adapted to a general smooth domain.

6.3.1 Convex Domain

Our objective is to solve the following problem, so-called *stochastic variational inequality*, by means of a drift penalty argument.

Definition 6.5 (SVI)**.** Given a convex domain $\mathcal{O}$ in $\mathbb{R}^d$ and an initial condition x^0 at time $t_0 \ge 0$, a *strong solution* of a *stochastic variational inequality* is a pair of adapted cad-lag processes $(x(t),\,\eta(t) : t \ge t_0)$ such that:

(1) $x(\cdot)$ takes values in the closure $\overline{\mathcal{O}}$ and $\eta(\cdot)$ is continuous and has locally bounded variation,

(2) the initial conditions $x(t_0) = x^0$ and $\eta(t_0) = 0$ are satisfied,

(3) the following equation is satisfied

$$\begin{cases} \mathrm{d}x(t) + \mathrm{d}\eta(t) = g(t,x(t),v(t))\mathrm{d}t + \sigma(t,x(t),v(t))\mathrm{d}w(t) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} \gamma(\zeta,t,x(t),v(t))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t), \end{cases} \tag{6.23}$$

(4) for every adapted cad-lag process $(y(t) : t \geq t_0)$ with valued in the closure $\overline{\mathcal{O}}$ we have

$$\int_{t_0}^{t_1} [x(s) - y(s)]\mathrm{d}\eta(s) \geq 0, \tag{6.24}$$

for every $t_1 > t_0$. Note that t_0 may be a stopping time and x^0 a $\mathcal{F}_{t_0}$-measurable random variable. □

Usually, $(x(t) : t \geq t_0)$ is called (normal) *reflected diffusion process with jumps* and $(\eta(t) : t \geq t_0)$ is its associated d-dimensional *local time* process. Since $\eta(\cdot)$ is continuous, the last inequality (6.24) is equivalent to

$$\int_{t_0}^{t_1} [x(t-) - y(t-)]\mathrm{d}\eta(t) \geq 0, \tag{6.25}$$

for every $t_1 > t_0$. Thus, no jumps should be outside the closure $\overline{\mathcal{O}}$, i.e., we impose the condition

$$x \in \overline{\mathcal{O}} \quad \Rightarrow \quad x + \gamma(\zeta, t, x, v) \in \overline{\mathcal{O}}, \tag{6.26}$$

for every (ζ, t, v) in $\mathbb{R}^m_* \times [t_0, \infty) \times \mathbb{R}^q$. To eliminate this *only interior jumps* condition, we need to set-up the stochastic variational inequality with the condition (6.25) instead of (6.24) and to take a decision regarding the jumps outside the closed region $\overline{\mathcal{O}}$, e.g., have the jumps also normally reflected. Anyway, in this section we always assume the condition (6.26), perhaps after re-defining $\gamma(\zeta, t, x, v) = 0$ when ζ does not belong to the support of $\pi(\mathrm{d}\zeta)$.

This stochastic variational inequality can be approximated by means of a classic penalty argument applied to a stochastic differential equation in the whole space $\mathbb{R}^d$. Without any loss of generality, one can assume that the coefficients are defined to the whole space $\mathbb{R}^d$, even if they need to be defined only on the closure $\overline{\mathcal{O}}$. Moreover, the assumption (6.26) on the interior jumps is complemented with the condition

$$\gamma(\zeta, t, x, v) = \gamma(\zeta, t, P_{\mathcal{O}}(x), v), \quad \forall \zeta, t, x, v, \tag{6.27}$$

where $P_{\mathcal{O}}(x)$ denotes the orthogonal projection on $\overline{\mathcal{O}}$ of the point x in $\mathbb{R}^d$. Note that $P_{\mathcal{O}}(x) = x$ for every x in the closure $\overline{\mathcal{O}}$.

Define the penalty function $\beta(x) := x - P_{\mathcal{O}}(x)$ (as a column vector), i.e.,

$$\beta(x) = \frac{1}{2}\nabla_x \min_{y \in \overline{\mathcal{O}}}\{|x - y|^2\}, \quad \forall x \in \mathbb{R}^d,$$

and, for every $\varepsilon > 0$, consider the stochastic ordinary differential equation

$$\left\{ \begin{array}{l} \mathrm{d}x^\varepsilon(t) = g(t, x^\varepsilon(t), v(t))\mathrm{d}t + \sigma(t, x^\varepsilon(t), v(t))\mathrm{d}w(t) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} \gamma(\zeta, t, x^\varepsilon(t), v(t))\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}t) - \frac{1}{\varepsilon}\beta(x^\varepsilon(t))\mathrm{d}t, \end{array} \right. \tag{6.28}$$

with the initial condition $x^\varepsilon(t_0) = x^0$. Since β is Lipschitz continuous, under assumptions (6.18) and (6.19) there exists one and only one solution $x^\varepsilon(\cdot)$ of

(6.28) satisfying the given initial conditions. Note that $\beta(x) = 0$ for every x in the closure $\overline{\mathcal{O}}$.

Due to the *large jumps*, we add to assumption (6.18) the condition: there exist constant $\bar{p} > 2$ and $C > 0$ such that

$$\int_{\mathbb{R}^m_*} |\gamma(\zeta, t, x, v)|^{\bar{p}} \pi(\mathrm{d}\zeta) \le C(1 + |x|^{\bar{p}}), \quad \forall t,\, x,\, v. \tag{6.29}$$

When the convex domain $\mathcal{O}$ is unbounded, we need to require a technical assumption, namely, there exists a point a in $\mathbb{R}^d$ and a positive constant $c > 0$ such that

$$(x - a)\beta(x) \ge c|\beta(x)|, \quad \forall x \in \mathbb{R}^d, \tag{6.30}$$

where $|\cdot|$ denotes the Euclidean norm. If $\mathcal{O}$ is bounded then for any a in the interior $\mathcal{O}$ one can find constant $c > 0$, proportional to the distance from a to the boundary $\partial\mathcal{O}$, such that (6.30) is satisfied. For instance, in the one dimensional case for $\overline{\mathcal{O}} := [0, \infty)$, the orthogonal projection is the positive part $P_{\mathcal{O}}(x) = x^+$ and the penalty function is $\beta(x) = x - x^+ = -x^-$. Hence $(x - a)\beta(x) = (a - x)x^- \ge ax^-$ for every x in $\mathbb{R}$ and $a > 0$. Therefore, condition (6.30) is trivially satisfied for the semi-space $\overline{\mathcal{O}} := \mathbb{R}^d_+$.

Since $\overline{\mathcal{O}}$ is convex, the function β is monotone, i.e.,

$$(x' - x)[\beta(x') - \beta(x)] \ge 0, \quad \forall x, x' \in \mathbb{R}^d. \tag{6.31}$$

In particular,

$$(\tilde{x} - x)\beta(x) \le 0, \quad \forall x \in \mathbb{R}^d,\ \tilde{x} \in \overline{\mathcal{O}},$$

and if $\tilde{x} = P_{\mathcal{O}}(x') = x' - \beta(x')$ then

$$(x' - x)\beta(x) \le \beta(x') \cdot \beta(x), \quad \forall x, x' \in \mathbb{R}^d. \tag{6.32}$$

First, a couple of a priori estimates on the stochastic ordinary differential equation (6.28) are established.

Lemma 6.6. *Suppose that the growth conditions (6.18) and (6.29) on the coefficients g, σ and γ are satisfied, as well as the assumptions (6.26), (6.27), (6.30) relative to a convex domain $\mathcal{O}$ in $\mathbb{R}^d$. Then the following a priori estimates on a solution $(x^\varepsilon(t) : t \ge t_0)$ of (6.28) with the initial condition $x^\varepsilon(t_0) = x^0$ holds: for every p in $[1, \bar{p}]$, with $\bar{p}$ as in (6.29), and $t_1 > t_0$ there exists a constant $C > 0$, independent of ε, such that*

$$\mathbb{E}\Big\{ \Big(\frac{1}{\varepsilon} \int_{t_0}^{t_1} |\beta(x^\varepsilon(t))| \mathrm{d}t \Big)^p + \sup_{t_0 \le t \le t_1} |x^\varepsilon(t)|^p \Big\} \le C, \quad \forall \varepsilon > 0, \tag{6.33}$$

and, for every p in $[2, \bar{p}]$, $t_1 > t_0$ and x^0 in $\overline{\mathcal{O}}$ we also have

$$\mathbb{E}\Big\{ \frac{1}{\varepsilon} \int_{t_0}^{t_1} |\beta(x^\varepsilon(t))|^p \mathrm{d}t + \sup_{t_0 \le t \le t_1} |\beta(x^\varepsilon(t))|^p \Big\} \le C\, \varepsilon^{p/2-1}, \tag{6.34}$$

for another constant $C > 0$, independent of ε in $(0,1]$. Moreover, for every $1 \le p \le \bar p$, $0 < q \le \bar p/2 - p/\bar p$, $q < p/2$, $t_1 > t_0$ and x^0 in $\overline{\mathcal O}$, there exists a constant $C > 0$ such that

$$\mathbb{E}\Big\{\Big(\frac{1}{\varepsilon'}\int_{t_0}^{t_1} |\beta(x^{\varepsilon'}(t))|\,|\beta(x^{\varepsilon}(t))|\mathrm{d}t\Big)^p\Big\} \le C\,\varepsilon^q, \tag{6.35}$$

for every ε, $\varepsilon' > 0$. Furthermore, all constants in the above estimates can be taken independent of the initial condition x^0 and t_0, as long as they remain in a bounded region.

Proof. First, we proceed as in Section 5.1.4 of the previous chapter. For a fixed p in $[1,\bar p]$ and a in $\mathbb{R}^d$, we define

$$\alpha(p,a) := p\,\alpha_g(a) + p\,\alpha_\sigma(p,a) + \alpha_\gamma(p,a), \tag{6.36}$$

where

$$\alpha_g(a) := \sup\Big\{\sum_i \frac{(x_i - a_i)\,g_i(t,x,v)}{1+|x-a|^2}\Big\},$$

$$\alpha_\sigma(p,a) := \sup\Big\{\sum_{i,k}\frac{\sigma_{ik}^2(t,x,v)}{1+|x-a|^2} + \\ +(p-2)\sum_{i,j,k}\frac{(x_i-a_i)\,\sigma_{ik}(t,x,v)\,\sigma_{jk}(t,x,v)\,(x_j-a_j)}{(1+|x-a|^2)^2}\Big\},$$

and

$$\alpha_\gamma(p,a) := \sup\Big\{\int_{\mathbb{R}^m_*}\Big[(1+|x-a+\gamma(\zeta,t,x,v)|^2)^{p/2} - \\ -(1+|x-a|^2)^{p/2} - \sum_i p\,(x_i-a_i)\,\gamma_i(\zeta,t,x,v)\times \\ \times(1+|x-a|^2)^{p/2-1}\Big](1+|x-a|^2)^{-p/2}\pi(\mathrm{d}\zeta)\Big\},$$

and the suprema are taken for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, The growth conditions (6.18) and (6.29) on the coefficients g, σ and γ ensure that $\alpha(p,a)$ is finite. Now apply Itô formula to the function

$$(t,x)\mapsto \mathrm{e}^{-\alpha(t-t_0)}\big(1+|x-a|^2\big)^{p/2},\quad \alpha \ge \alpha(p,a)$$

with the stochastic process $x^\varepsilon(t)$ solution of (6.28) to get

$$\mathrm{e}^{-\alpha(t-t_0)}\big(1+|x^\varepsilon(t)-a|^2\big)^{p/2} \le \big(1+|x-a|^2\big)^{p/2} + I(t) - \\ -\frac{p}{\varepsilon}\int_{t_0}^t \big(x^\varepsilon(s)-a\big)\beta(x^\varepsilon(s))\,\mathrm{e}^{-\alpha(s-t_0)}\big(1+|x^\varepsilon(s)-a|^2\big)^{p/2}\mathrm{d}s,$$

where $I(t)$ is a stochastic integral. This implies

$$\mathbb{E}\big\{|x^\varepsilon(t)|^p\big\} \le C,\quad \forall t\in[t_0,t_1],\ \varepsilon>0,$$

for some constant $C > 0$ independent of ε. Moreover, the Davis-Burkhölder-Gundy inequality (i.e., the stochastic integral inequalities (5.46) and (5.47), with $p = 1$, of the previous Chapter), applied to the exact expression of the stochastic integral $I(t)$ yields one part of estimate (6.33), namely

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |x^\varepsilon(t)|^p \Big\} \le C, \quad \forall \varepsilon > 0, \tag{6.37}$$

for some constant C independent of ε.

Now, going back to Itô formula, taking $p = 2$ and using the assumption (6.30), we obtain

$$\begin{aligned} \frac{2c}{\varepsilon} \int_{t_0}^t \mathrm{e}^{-\alpha(s-t_0)} |\beta(x^\varepsilon(s))| \mathrm{d}s \le \big(1 + |x-a|^2\big)^{1/2} + \\ + \int_{t_0}^t \mathrm{e}^{-\alpha(s-t_0)} (x^\varepsilon(s) - a)\sigma(s, x^\varepsilon(s), v(s)) \mathrm{d}w(s) + \\ + \int_{\mathbb{R}^m_* \times]t_0,t]} \mathrm{e}^{-\alpha(s-t_0)} (x^\varepsilon(s) - a)\sigma(s, x^\varepsilon(s), v(s)) \tilde{\nu}(\mathrm{d}z, \mathrm{d}s), \end{aligned}$$

and we get the second part of estimate (6.33).

Recall the expression (6.20) of the integro-differential operator $A(t, v) = A_{g,\sigma}(t, v) + I(t, v)$, where $A_{g,\sigma}(t, v)$ is the local part and $I(t, v)$ the proper integro-differential part. Now, consider the penalty function $\varphi(x) := |\beta(x)|^p$, with p in $[2, \bar{p}]$. Since the gradient satisfies $\nabla |\beta(x)|^2 = 2\beta(x)$, we deduce that

$$\nabla \varphi(x) = p\, \beta(x)\, |\beta(x)|^{p-2}, \quad \forall x \in \mathbb{R}^d,$$

which yields

$$\begin{aligned} &|\partial_i \varphi(x)| \le C |\beta(x)|^{p-1}, \quad \forall x \in \mathbb{R}^d, \\ &|\partial_{ij} \varphi(x)| \le C |\beta(x)|^{p-2}, \quad \text{a.e. } x \in \mathbb{R}^d. \end{aligned}$$

Thus, in view of the growth condition (6.18) on the coefficients g and σ, we proves that for some constant $C > 0$ such that

$$|A_{g,\sigma}(t, v)\varphi(x)| \le C \big[(1 + |x|^2)\, |\beta(x)|^{p-2} + (1 + |x|^2)^{1/2}\, |\beta(x)|^{p-1} \big]$$

for every t, v and almost every x in $\mathbb{R}^d$. On the other hand, the proper integro-differential part $I(t, v)$ satisfies

$$I(t, v)\varphi(x) = \sum_{ij} \int_0^1 (1 - \theta) \mathrm{d}\theta \int_{\mathbb{R}^m_*} \gamma_i \gamma_j \partial_{ij}\, \varphi(x + \theta\gamma) \pi(\mathrm{d}\zeta),$$

with $\gamma = \gamma(\zeta, t, x, v)$, and

$$|\partial_{ij} \varphi(x + \theta\gamma(\zeta, t, x, v))| \le C |\beta(x + \theta\gamma(\zeta, t, x, v))|^{p-2}.$$

In view of condition (6.27) and the property

$$|\beta(x + \theta\gamma(\zeta, t, x, v))| \le |x + \theta\gamma(\zeta, t, x, v) - y|, \quad \forall y \in \overline{\mathcal{O}},$$

we may choose $y = P_{\mathcal{O}}(x) + \theta\gamma(\zeta, t, P_{\mathcal{O}}(x), v)$ to deduce from the growth condition (6.18) that there exists a constant $C > 0$ such that

$$|I(t,v)\varphi(x)| \le C_1\,(1+|x|^2)\,|\beta(x)|^{p-2},$$

for every t, v and almost every x in $\mathbb{R}^d$. This establishes the estimate

$$|A(t,v)\varphi(x)| \le C_0\big[(1+|x|^2)\,|\beta(x)|^{p-2} + (1+|x|^2)^{1/2}\,|\beta(x)|^{p-1}\big],$$

for some constant $C_0 > 0$ and for every t, v and almost every x in $\mathbb{R}^d$.

By means of a mollification of the penalty function $\varphi(x)$, namely,

$$\varphi_n(x) := \Big(\int_{\mathbb{R}^d} |\beta(x-y)|^2 \rho_n(y)\mathrm{d}y\Big)^{p/2},$$

where $(\rho_n : n \ge 1)$ is a sequence of smooth positive kernels with compact support as usual, we can approximate φ by smooth functions to apply Itô formula and to justify the inequality

$$|\beta(x^\varepsilon(t))|^p + \frac{p}{\varepsilon}\int_{t_0}^t |\beta(x^\varepsilon(s))|^p \mathrm{d}s \le I(t) + \\ + C_0 \int_{t_0}^t \big[(1+|x^\varepsilon(s)|^2)\,|\beta(x^\varepsilon(s))|^{p-2} + (1+|x^\varepsilon(s)|^2)^{1/2}\,|\beta(x^\varepsilon(s))|^{p-1}\big]\mathrm{d}s$$

where $I(t)$ is a stochastic integral, namely

$$I(t) := p\int_{t_0}^t \beta(x^\varepsilon(s))\cdot\sigma(s,x^\varepsilon(s))|\beta(x^\varepsilon(s))|^{p-2}\mathrm{d}w(s) + \\ + p\int_{t_0}^t\int_{\mathbb{R}^m_*} \beta(x^\varepsilon(s))\cdot\gamma(\zeta,s,x^\varepsilon(s))|\beta(x^\varepsilon(s))|^{p-2}\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s).$$

Using the Davis-Burkhölder-Gundy inequality, the growth condition (6.18) on the coefficients σ and γ and the inequality

$$(a\,b)^{1/2} \le 2(a/2)^{1/2}\,(b/2)^{1/2} \le \frac{a}{2} + b, \quad \forall a,\, b > 0,$$

we verify that there exists a constant $C > 0$ such that

$$\mathbb{E}\Big\{\sup_{t_0\le t\le t_1}|I(t)|\Big\} \le C\,\mathbb{E}\Big\{\Big(\int_{t_0}^{t_1}(1+|x^\varepsilon(s)|^2)\,|\beta(x^\varepsilon(s))|^{2p-2}\mathrm{d}s\Big)^{1/2}\Big\} \le \\ \le C\,\mathbb{E}\Big\{\Big(\sup_{t_0\le s\le t_1}|\beta(x^\varepsilon(s))|^p\int_{t_0}^{t_1}(1+|x^\varepsilon(s)|^2)\,|\beta(x^\varepsilon(s))|^{p-2}\mathrm{d}s\Big)^{1/2}\Big\} \le \\ \le \frac{1}{2}\,\mathbb{E}\Big\{\sup_{t_0\le s\le t_1}|\beta(x^\varepsilon(s))|^p\Big\} + C\,\mathbb{E}\Big\{\int_{t_0}^{t_1}(1+|x^\varepsilon(s)|^2)\,|\beta(x^\varepsilon(s))|^{p-2}\mathrm{d}s\Big\}.$$

Next, by means of the inequality

$$A\,B \le \frac{A^q}{q} + \frac{B^{q'}}{q'}, \qquad \frac{1}{q}+\frac{1}{q'} = 1,$$

for appropriate non-negative factors A and B, first for $q' := p/(p-1)$ and then for $q' := p/(p-2)$, we deduce estimate (6.34).

To show estimate (6.35) we start with

$$\frac{1}{\varepsilon'}\int_{t_0}^{t_1} |\beta(x^{\varepsilon'}(t))|\,|\beta(x^{\varepsilon}(t))|\mathrm{d}t \le A\,B,$$

with

$$A := \frac{1}{\varepsilon'}\int_{t_0}^{t_1} |\beta(x^{\varepsilon'}(t))|\,\mathrm{d}t,$$
$$B := \sup_{t_0\le t\le t_1} |\beta(x^{\varepsilon}(t))|\,.$$

Since, for $r > 2$ and $r' = r/(r-1)$

$$\mathbb{E}\big\{(A\,B)^p\big\} \le \big(\mathbb{E}\{A^{pr}\}\big)^{1/r}\,\big(\mathbb{E}\{B^{pr'}\}\big)^{1/r'},$$

we deduce form the previous estimates (6.33) and (6.34) that

$$\mathbb{E}\big\{(A\,B)^p\big\} \le C\,\varepsilon^{p/2-1/r}$$

for some constant $C > 0$, which implies estimate (6.35) if the constant r is chosen as large as possible. □

We are ready to present the main result of this section

Theorem 6.7 (SVI)**.** *Let the growth conditions* (6.18), (6.29) *and the uniform locally Lipschitz condition* (6.19) *on the coefficients* g, σ *and* γ *be satisfied. Suppose also assumptions* (6.26), (6.27), (6.30) *relative to a convex domain* $\mathcal{O}$ *in* $\mathbb{R}^d$*. Then there is one and only one solution* $(x(t), \eta(t) : t \ge t_0)$ *of the stochastic variational inequality as described by* Definition 6.5*. Moreover for every* $t_1 > t_0$*, the following convergence holds in probability*

$$\sup_{t_0\le t\le t_1} \big\{|x^{\varepsilon}(t) - x(t)| + |\eta^{\varepsilon}(t) - \eta(t)|\big\} \to 0 \quad \textit{as} \quad \varepsilon \to 0, \tag{6.38}$$

where $(x^{\varepsilon}(t) : t \ge t_1)$ *is the solution of the stochastic ordinary differential equation* (6.28) *with the initial condition* $x^{\varepsilon}(t_0) = x^0$*, for some given* x^0 *in* $\overline{\mathcal{O}}$*, and*

$$\eta^{\varepsilon}(t) := \frac{1}{\varepsilon}\int_{t_0}^{t} \beta(x^{\varepsilon}(s))\mathrm{d}s,$$

with β *being the penalty function. Furthermore, if the coefficients are uniform Lipschitz then we also have*

$$\mathbb{E}\big\{\sup_{t_0\le t\le t_1} \{|x^{\varepsilon}(t) - x(t)|^p + |\eta^{\varepsilon}(t) - \eta(t)|^p\}\big\} \to 0 \quad \textit{as} \quad \varepsilon \to 0,$$

for every $1 \le p \le \bar{p}$ *and locally uniform with respect to the initial condition* x^0

Proof. If $(x(t), \eta(t) : t \geq 0)$ and $(\bar{x}(t), \bar{\eta}(t) : t \geq 0)$ are two solutions of the stochastic variational inequality (6.23), (6.24) then, Itô formula applied to the quadratic function $\xi \mapsto |\xi|^2$ and the process $\xi(t) := x(t) - \bar{x}(t)$ yields

$$\mathbb{E}\{|x(t) - \bar{x}(t)|^2\} \leq C\,\mathbb{E}\Big\{ \int_{t_0}^{t} |x(s) - \bar{x}(s)|^2 \mathrm{d}s \Big\} + \\ + 2\,\mathbb{E}\Big\{ \int_{t_0}^{t} [x(s) - \bar{x}(s)] \cdot [\mathrm{d}\bar{\eta}(s) - \mathrm{d}\eta(s)] \Big\},$$

assuming that the coefficients g, σ and γ are uniformly Lipschitz and the initial data x^0 has a finite second moment. Hence, based on the monotony (6.31) of the penalty function and Gronwall inequality we deduce first $x(\cdot) = \bar{x}(\cdot)$ and next $\eta(\cdot) = \bar{\eta}(\cdot)$. To complete this argument about the uniqueness, similarly to Theorem 5.3 in the previous chapter, for any $r > 0$, define the adapted process

$$\chi_r(t) := \begin{cases} 0 & \text{if } |x(s)| > r \text{ or } |\bar{x}(s)| > r \text{ for some } s > t, \\ 1 & \text{otherwise,} \end{cases}$$

i.e., if $\tau_r(x(\cdot))$ denotes the first exit time from the closed ball of radius r for the process $x(\cdot)$ then

$$\chi_r(t) = 1 \quad \text{if and only if} \quad t < \tau_r,$$

where $\tau_r := \tau_r(x(\cdot)) \wedge \tau_r(\bar{x}(\cdot))$. Since $\chi_r(t) = \chi_r(t)\,\chi_r(s)$ for every $s \leq t$, we have

$$[x(t) - \bar{x}(t)]\chi_r(t) = \int_{t_0}^{t} \chi_r(s)\big[g(s, x(s), v(s)) - g(s, \bar{x}(s), v(s))\big]\mathrm{d}s \\ + \int_{t_0}^{t} \chi_r(s)\big[\sigma(s, x(s), v(s)) - \sigma(s, \bar{x}(s), v(s))\big]\mathrm{d}w(s) + \\ + \int_{\mathbb{R}^m_* \times]t_0, t]} \chi_r(s)\big[\gamma(\zeta, s, x(s), v(s)) - \gamma(\zeta, s, \bar{x}(s), v(s))\big]\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \int_{t_0}^{t} \chi_r(s)[\mathrm{d}\bar{\eta}(s) - \mathrm{d}\eta(s)], \qquad t_0 \leq t \leq \tau_r.$$

Thus, apply Itô formula on the stochastic as above, but on the interval $[t_0, \tau_r)$, to deduce that

$$\mathbb{E}\{|x(t) - \bar{x}(t)|^2 \chi_r(t)\} \leq K\,\mathbb{E}\Big\{ \int_{t_0}^{t\wedge\tau_r} |x(s) - \bar{x}(s)|^2 \chi_r(s) \mathrm{d}s \Big\} \leq \\ \leq K \int_{t_0}^{t} \mathbb{E}\{|x(s) - \bar{x}(s)|^2 \chi_r(s)\}\mathrm{d}s,$$

for every t in $[t_0, t_1]$, and for some constant K depends only on $t_1 - t_0$, r and the constant $M = M(r)$ of hypothesis (6.19). Using Gronwall inequality, we

deduce that for any t in $[t_0, t_1]$ and any $r > 0$ we have $|x(t) - \bar{x}(t)|\chi_r(t) = 0$ with probability one. Since x and $\bar{x}$ are cad-lag, and

$$P\big\{\chi_r(t) = 0 \text{ in } [t_0, t_1]\big\} \le P\Big\{ \sup_{t_0 \le t \le t_1} |x(t)| > r\Big\} + P\Big\{ \sup_{t_0 \le t \le t_1} |\bar{x}(t)| > r\Big\},$$

the a priori estimate (6.33) of Lemma 6.6 yields the existence of a set with probability one where $x(t) = \bar{x}(t)$ for every t in $[t_0, t_1]$.

Now, assume a uniform (global) Lipschitz condition (6.29) on the coefficients g, σ and γ. We claim that for every $1 \le p \le \bar{p}$, $0 < q \le \bar{p}/2 - p/\bar{p}$, $q < p/2$, $t_1 > t_0$ and x^0 in $\overline{\mathcal{O}}$, there exists a constant $C > 0$ such that

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} \{|x^\varepsilon(t) - x^{\varepsilon'}(t)|^p + |\eta^\varepsilon(t) - \eta^{\varepsilon'}(t)|^p\}\Big\} \le C\,(\varepsilon + \varepsilon')^q \tag{6.39}$$

as ε, ε' vanish. To proof this claim, we apply Itô formula to the quadratic function $\xi \mapsto |\xi|^2$ and the process $\xi(t) := x^\varepsilon(t) - x^{\varepsilon'}(t)$ to get

$$\begin{aligned} |x^\varepsilon(t) - x^{\varepsilon'}(t)|^2 \le I(t) + C \int_{t_0}^t |x^\varepsilon(s) - x^{\varepsilon'}(s)|^2 \mathrm{d}s + \\ +2 \int_{t_0}^t [x^\varepsilon(s) - x^{\varepsilon'}(s)] \cdot [\mathrm{d}\eta^{\varepsilon'}(s) - \mathrm{d}\eta^\varepsilon(s)], \end{aligned}$$

where C is a constant depending on the uniform (global) Lipschitz condition (6.29), and $I(t)$ is the stochastic integral

$$\begin{aligned} I(t) := \int_{t_0}^t [x^\varepsilon(s) - x^{\varepsilon'}(s)]\sigma(s)\mathrm{d}w(s) + \\ + \int_{\mathbb{R}^m_* \times]t_0, t]} [x^\varepsilon(s) - x^{\varepsilon'}(s)]\gamma(\zeta, s)\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s), \end{aligned}$$

with

$$\begin{aligned} \sigma(s) &:= \sigma(s, x^\varepsilon(s), v(s)) - \sigma(s, x^{\varepsilon'}(s), v(s)), \\ \gamma(s) &:= \gamma(\zeta, s, x^\varepsilon(s), v(s)) - \gamma(\zeta, s, x^{\varepsilon'}(s), v(s)). \end{aligned}$$

Again, the Davis-Burkhölder-Gundy inequality (i.e., the stochastic integral inequalities (5.46) and (5.47), with $p = 1$, of the previous Chapter), the uniform (global) Lipschitz condition (6.29) and the simple inequality

$$\begin{aligned} \Big(\int_{t_0}^{t_1} |f(t)|^2 \mathrm{d}t\Big)^{1/2} \le \sup_{t_0 \le t \le t_1} |f(t)|^{1/2} \Big(\int_{t_0}^{t_1} |f(t)|\,\mathrm{d}t\Big)^{1/2} \le \\ \le \frac{1}{2} \sup_{t_0 \le t \le t_1} |f(t)| + \frac{1}{2} \int_{t_0}^{t_1} |f(t)|\,\mathrm{d}t, \end{aligned}$$

applied to the expression of $I(t)$ yield

$$\begin{aligned} \mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |I(t)|^p\Big\} \le \frac{1}{2}\,\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |x^\varepsilon(t) - x^{\varepsilon'}(t)|^p\Big\} + \\ + C\,\mathbb{E}\Big\{ \int_{t_0}^{t_1} |x^\varepsilon(s) - x^{\varepsilon'}(s)|^p \mathrm{d}s\Big\}, \end{aligned}$$

for some constant $C > 0$. On the other hand, the monotone property (6.32) on the penalty function β implies

$$\int_{t_0}^{t} [x^{\varepsilon}(s) - x^{\varepsilon'}(s)] \cdot [\mathrm{d}\eta^{\varepsilon'}(s) - \mathrm{d}\eta^{\varepsilon}(s)] \leq$$
$$\leq \left(\frac{1}{\varepsilon} + \frac{1}{\varepsilon'}\right) \int_{t_0}^{t} |\beta(x^{\varepsilon}(s))|\, |\beta(x^{\varepsilon'}(s))|\, \mathrm{d}s.$$

for every positive ε and ε'.

Hence, collecting all pieces and by means of estimate (6.35) in Lemma 6.6, we deduce

$$\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq t_1} |x^{\varepsilon}(t) - x^{\varepsilon'}(t)|^p \Big\} \leq$$
$$\leq C\, \mathbb{E}\Big\{ \int_{t_0}^{t_1} |x^{\varepsilon}(s) - x^{\varepsilon'}(s)|^p \mathrm{d}s \Big\} + C\,(\varepsilon + \varepsilon')^q.$$

which implies the claim (6.39), after using Gronwall inequality and the stochastic equation relating x^{ε} and η^{ε}.

Based on (6.39), we can define the stochastic processes

$$x(t) := \lim_{\varepsilon \to 0} x^{\varepsilon}(t), \qquad \eta(t) := \lim_{\varepsilon \to 0} \eta^{\varepsilon}(t),$$

which clearly satisfy the properties (1), (2) and (3) of Definition 6.5. Also, the same monotone property (6.32) proves that

$$\int_{t_0}^{t} [x^{\varepsilon}(s) - y(s)] \cdot \mathrm{d}\eta^{\varepsilon}(s) \geq 0,$$

for every $t > t_0$, $\varepsilon > 0$ and any adapted cad-lag process $(y(t) : t \geq t_0)$ with values in the closure $\overline{\mathcal{O}}$. Passing to the limit, we check condition (6.24).

To extend these results to the uniform locally Lipschitz case, we remark that under this weaker assumption, all a priori bounds of Lemma 6.6 hold true. Thus, we can review the arguments used in estimate (6.39) to obtain

$$\mathbb{E}\Big\{ \sup_{t_0 \leq t \leq \tau_{\varepsilon,\varepsilon',r}} \{|x^{\varepsilon}(t) - x^{\varepsilon'}(t)|^p + |\eta^{\varepsilon}(t) - \eta^{\varepsilon'}(t)|^p\} \Big\} \leq C\,(\varepsilon + \varepsilon')^q$$

where $\tau_{\varepsilon,\varepsilon',r}$ is the first exit time from the ball of radius r, i.e.,

$$\tau_{\varepsilon,\varepsilon',r} := \inf\big\{ t \geq t_0 : |x^{\varepsilon}(t)| + |x^{\varepsilon'}(t)| > r \big\}.$$

In view of estimate (6.33) of Lemma 6.6, we have

$$\lim_{r \to 0} P\big\{ \tau_{\varepsilon,\varepsilon',r} \leq t_1 \big\} = 0, \qquad \forall t_1 > 0,$$

uniformly in ε and ε'. Hence

$$\sup_{t_0 \leq t \leq t_1} \{|x^{\varepsilon}(t) - x^{\varepsilon'}(t)| + |\eta^{\varepsilon}(t) - \eta^{\varepsilon'}(t)|\} \to 0, \tag{6.40}$$

in probability, as $\varepsilon + \varepsilon'$ vanish. The proof is then completed. □

• *Remark* 6.8. If the initial condition is stochastic i.e., x^0 is a $\mathcal{F}_{t_0}$-measurable $\overline{\mathcal{O}}$-valued random variable, then the convergence in Theorem 6.7 is uniform in x^0, as long as the p-moments of x^0 are uniformly bounded. On the other hand, approximating the coefficient γ by a sequence of functions satisfying condition (6.29), we deduce that the convergence in probability (6.38) of Theorem 6.7 remains true without the extra assumption (6.29) on the integrability of large jumps. It is also clear that if the initial condition x^0 does not belongs to the closure $\overline{\mathcal{O}}$ then the approximation x^ε converges to the solution of the stochastic variational inequality with the initial condition $x(0) = P_{\mathcal{O}}(x^0)$. □

• *Remark* 6.9. The reflected diffusion process with jumps is Lipschitz continuous with respect to the initial condition. Indeed, denote by $(x(t), \eta(t) : t \geq t_0)$ and $(\bar{x}(t), \bar{\eta}(t) : t \geq t_0)$ the solution to stochastic variational inequality (6.23), (6.24) with initial conditions $x(t_0) = x^0$ and $\bar{x}(t_0) = \bar{x}^0$, respectively. Proceeding as in Theorem 5.11 of the previous Chapter, suppose $\alpha \geq \beta(p)$, $0 < p \leq 2$, as defined by the extreme (5.49) in the previous Chapter, corresponding to the coefficients g, σ and γ. Because of the monotone property (6.32) of the penalty function β, then under the uniform (global) Lipschitz condition (6.19) we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{[\alpha - \beta(p)] \displaystyle\int_{t_0}^{t} |x(s) - \bar{x}(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \mathrm{d}s + \\ \qquad + |x(t) - \bar{x}(s)|^p \, \mathrm{e}^{-\alpha(t-t_0)} \Big\} \leq \mathbb{E}\big\{|x(t_0) - \bar{x}(t_0)|^p\big\}, \end{array} \right. \tag{6.41}$$

for every $t \geq t_0$. Furthermore, we have

$$\left\{ \begin{array}{l} \mathbb{E}\Big\{ \displaystyle\sup_{t_0 \leq s \leq t} |x(s) - \bar{x}(s)|^p \, \mathrm{e}^{-\alpha(s-t_0)} \Big\} \leq \\ \qquad \leq M \left[1 + \dfrac{1}{\alpha - \beta(p, \lambda)}\right] \mathbb{E}\big\{(|x(t_0) - \bar{x}(t_0)|^p\big\}, \end{array} \right. \tag{6.42}$$

for every $t \geq t_0$ and for some constant M depending only on p and the bounds of σ and γ through conditions (6.19), which is condition (5.50) in Section 5.1.4. To use $p > 2$ we need to add assumption in Section 5.1.4, relative to the jump coefficient γ. Clearly, this a priori estimates (6.41) and (6.42) are established for the approximation solution uniformly in the parameter ε. □

Let us look at the case of a smooth and convex domain $\mathcal{O}$, i.e., the distance-to-the-boundary function $x \mapsto \mathrm{dist}(x, \partial\mathcal{O})$ can be extended to a twice-differentiable function ρ from $\mathbb{R}^d$ into $\mathbb{R}$, which implies

$$\begin{aligned} &\mathcal{O} = \{x \in \mathbb{R}^d : \rho(x) < 0\}, \\ &\partial\mathcal{O} = \{x \in \mathbb{R}^d : \rho(x) = 0\}, \\ &|\rho(x)| = \mathrm{dist}(x, \partial\mathcal{O}), \quad \forall x \in \mathbb{R}^d \text{ satisfying } \rho(x) > -\delta, \end{aligned}$$

for some constant $\delta > 0$. Since $\nabla\rho(x) = n(x)$ is the exterior unit normal vector at a point x on the boundary $\partial\mathcal{O}$ relative to $\mathcal{O}$, we have $\nabla\rho \cdot \beta = |\beta|$. Thus, by applying Itô formula to the function $\rho(x)$ and the process $x^\varepsilon(s)$, solution of the

stochastic variational inequality, we have

$$\frac{1}{\varepsilon}\int_{t_0}^{t}|\beta(x^\varepsilon(s)|\,\mathrm{d}s=\rho(x)-\rho(x^\varepsilon(t))+I_\rho^\varepsilon(t)+\int_{t_0}^{t}L_\rho^\varepsilon(s)\mathrm{d}s,$$

where $I_\rho^\varepsilon(t)$ is a stochastic integral and $L_\rho^\varepsilon(t)$ is an integro-differential process $A(s,v(s))\rho(x^\varepsilon(s))$. Hence, estimate (6.39) or (6.40) yields

$$\sup_{t_0\le t\le t_1}\Big|\frac{1}{\varepsilon}\int_{t_0}^{t}|\beta(x^\varepsilon(s)|\,\mathrm{d}s-\ell(t)\Big|^p\to 0$$

in L^p or in probability. The limiting process $(\ell(t):t\ge t_0)$ is nonnegative, nondecreasing, continuous, adapted and satisfies

$$\eta(t)=\int_{t_0}^{t}n(x(s))\,\mathrm{d}\ell(s),\quad \forall t\ge 0$$

and

$$\ell(t)=\int_{t_0}^{t}\mathbb{1}_{\{x(s)\in\partial\mathcal{O}\}}\,\mathrm{d}\ell(s),\quad \forall t\ge 0.$$

As mentioned in the previous section, the stochastic process $(\ell(t):t\ge t_0)$ is called the *local time* associated with the reflected diffusion process with jumps $(x(t):t\ge t_0)$.

6.3.2 Oblique Reflection

Following the case of a smooth convex domain, we discuss non-convex cases. Let $\mathcal{O}$ be a smooth domain in $\mathbb{R}^d$, in the sense that its exterior normal unit vector $n(x)$ is uniquely defined at any point x on the boundary $\partial\mathcal{O}$. Besides the coefficients g, σ and γ, suppose given an oblique interior direction $b(t,x)=(b_i:i=1,\dots d)$ as a continuous vector field defined in $[t_0,\infty)\times\partial\mathcal{O}$ which satisfies

$$-b(t,x)\cdot n(x)\ge c>0,\quad \forall t\ge t_0,\ x\in\partial\mathcal{O}. \tag{6.43}$$

In the semi-space $\mathbb{R}^d_+=\{x:x_d>0\}$, this means $b_d(t,x)\ge c>0$, for every $x=(\bar{x},x_d)$, $x_d=0$ as used in Theorem 6.4. For convenience, the vector field b is extended to the whole domain $\overline{\mathcal{O}}$, but it is only used on the boundary $\partial\mathcal{O}$. Now,

Definition 6.10 (OR-SODE)**.** Given an initial condition x^0 in $\overline{\mathcal{O}}$ at time $t_0\ge 0$, a *strong solution* of a stochastic ordinary differential equation with *instantaneous reflection* according to the oblique interior direction b is a pair of adapted cad-lag processes $(x(t),\,\ell(t):t\ge t_0)$ such that:

(1) $x(\cdot)$ takes values in the closure $\overline{\mathcal{O}}$ and $\ell(\cdot)$ is nondecreasing and continuous,

(2) the initial conditions $x(t_0)=x^0$ and $\ell(t_0)=0$ are satisfied,

(3) the following equation is satisfied

$$\begin{cases} \mathrm{d}x(t) = g(t,x(t),v(t))\mathrm{d}t + \sigma(t,x(t),v(t))\mathrm{d}w(t) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_*} \gamma(\zeta,t,x(t),v(t))\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}t) + b(t,x(t))\mathrm{d}\ell(t), \end{cases} \tag{6.44}$$

(4) we have

$$\mathrm{d}\ell(t) = \mathbb{1}_{\{x(t)\in\partial\mathcal{O}\}}\mathrm{d}\ell(t), \tag{6.45}$$

for every $t > t_0$. Note that t_0 may be a stopping time and x^0 a t_0-measurable random variable. □

Usually, $(x(t) : t \geq t_0)$ is called *reflected diffusion process with jumps*, with drift g, diffusion σ, jumps γ, and oblique interior reflecting direction b (or oblique exterior reflecting direction $-b$). Also $(\ell(t) : t \geq t_0)$ is its associated *increasing* process or *local time* process. Notice that depending on the context, sometimes is preferable to use an oblique exterior direction instead of an oblique interior direction, so that a change of sign is found in the literature, i.e., changing b into $-b$. Similarly, the coefficients a_{ij} and a_i (defining the differential operator) may have a negative sign.

If $\mathcal{O}$ is a bounded simply connected subset of $\mathbb{R}^d$ with a smooth connected and orientate boundary $\partial\mathcal{O}$ then the following condition holds: there exists a diffeomorphism Ψ of class C^3 between the closure $\overline{\mathcal{O}}$ and the closed unit ball $\overline{B}$, i.e., a one-to-one map Ψ from a neighborhood of $\overline{\mathcal{O}}$ onto a neighborhood of closed unit ball $\overline{B}$, such that Ψ and its inverse Ψ^{-1} are three times continuously differentiable, and

$$\Psi(\mathcal{O}) = B, \qquad \Psi(\partial\mathcal{O}) = \partial B. \tag{6.46}$$

Such a diffeomorphism Ψ of class C^3 may exist under other smoothness assumptions.

Theorem 6.11. *Suppose $\mathcal{O}$ be a smooth domain satisfying* (6.46) *and $b(t,x)$ be an oblique interior direction* (6.43)*, which is continuously differentiable in t and twice continuously differentiable in x. Then, under the growth condition* (6.18)*, the uniform locally Lipschitz condition* (6.19) *on the coefficients g, σ and γ, there exists a unique solution of the reflected stochastic differential equation of* Definition 6.10*.*

Proof. The point is to build a diffeomorphism between the closure $\overline{\mathcal{O}}$ and the closed unit ball B such that the oblique exterior direction $-b$ is transformed into the outward normal direction. Hence, by means of Itô formula, we consider a stochastic variational inequality on a closed ball $\overline{B}$, which can be solved accordingly to Theorem 6.7 and Remark 6.8. Thus, going back to the initial domain $\overline{\mathcal{O}}$ via the diffeomorphism, the existence and uniqueness of a solution is actually established.

If Ψ denotes the diffeomorphism satisfying (6.46) then for some open subsets $\mathcal{O}_1 \supset \overline{\mathcal{O}}$ and $B_1 \supset B$ such that $\Psi : \mathcal{O}_1 \to B_1$ and its inverse Ψ^{-1} are three

times continuously differentiable. The condition (6.46) implies that for any two curves C_1 and C_2 in $\mathcal{O}_1$ with nonzero angle their images $\Psi(C_1)$ and $\Psi(C_2)$ are also two curves in B_1 with nonzero angle. This fact, which is true for any smooth diffeomorphism, allows us to verify that assumption (6.43) on the oblique interior direction is preserved by Ψ, i.e., for another constant $c > 0$ we have

$$-b^1(t,y)\cdot n^1(y), \quad \forall t \geq t_0,\ y \in \partial B,$$

where $n^1(y)$ is the exterior (or outward) unit normal to the unit ball B at the point x on the boundary ∂B, i.e., $n^1(y) = y/|y|$, and $b^1(t,y) := b(t,x)$, with $\Psi(x) = y$. Thus $b^1 = B\nabla\Psi$, where $\nabla\Psi$ stands for the matrix of the first derivative of Ψ.

Now, our problem has been reduced to the case $\mathcal{O} = B$ (the unit ball) with an oblique internal direction $b^1(t,y)$. To transform b^1 into the interior (inward) unit direction, we consider a continuous vector field $f^t(y)$, twice continuously differentiable in y on the open set $0 < |y| < 2$ and satisfies

$$\begin{aligned} &f^t(y) = -b^1(t,y), \quad \text{if} \quad |y| = 1,\\ &-2\, f^t(y)\cdot y \geq c|y|, \quad \forall y,\\ &f^t(y) = -\frac{y}{|y|}, \quad \text{if} \quad |y| \leq \frac{2}{3}. \end{aligned}$$

Because of the assumptions made on the direction b and the diffeomorphism Ψ, the oblique internal direction b^1 is continuous in (t,y) and twice differentiable in x. Thus, the vector field f^t can be constructed. Next, define a map $\Psi_1 : (t,z) \mapsto y$, by means of an ordinary differential equation as follows: First, for any given a with $|a| = 1$, denote by $y^t(s,a)$ the solution of the initial valued problem

$$\dot{y}^t(s,a) = f^t(y^t(s,a)), \quad y(1/2) = a.$$

for $s \geq 1/2$, where t is a parameter. Set

$$\Psi_1(t,z) = \begin{cases} z, & \text{if } |z| \leq 1/2\\ y^t(|z|, z/|z|), & \text{otherwise} \end{cases}$$

This means that a curve of the form $\{sa : 0 \leq s \leq 2\}$ in z-coordinate becomes the curve

$$\{sa : 0 \leq s \leq 1/2\} \cup \{y^t(s,a) : 1/2 \leq s \leq 2\}$$

in the y-coordinate. This proves that $-b^1(t,y)$, with $\Psi_1(t,z) = y$ is the transformed of the exterior normal direction z in ∂B, not necessarily of unit length. Clearly, the classic theory of ordinary differential equations guarantees that $z \mapsto y = \Psi_1(t,z)$ is a diffeomorphism of class C^2, which is continuously differentiable in t. Therefore,

$$b^1(t,y) = b^2(t,z) = -\lambda(t,z)\, z,$$

where the scalar $\lambda(t,z) \geq c > 0$, for some constant c, and for any $t \geq t_0$, and $y = \Psi_1(t,z)$, y in ∂B.

For a smooth convex domain, such as the unit ball B, there is no difference between a normal reflection and a reflection following an interior direction of the form $-\lambda(t,x)\, n(x)$ Indeed, the process $\eta(\cdot)$ is uniquely determined, so that if $(x(t), \ell(t) : t \geq t_0)$ and $(\bar{x}(t), \bar{\ell}(t) : t \geq t_0)$ are solutions relative to $-\lambda(t,x)\, n(x)$ and $-\bar{\lambda}(t,x)\, n(x)$ with $\lambda(t,x), \bar{\lambda}(t,x) \geq c > 0$, then we have

$$
\begin{aligned}
&x(t) = \bar{x}(t), \quad \forall t \geq t_0,\\
&\lambda(t,x(t))\mathrm{d}\ell(t) = \bar{\lambda}(t,x(t))\mathrm{d}\bar{\ell}(t), \quad \forall t \geq t_0.
\end{aligned}
$$

Finally, if $(z(t), \eta(t))$ is the solution of a normal reflected stochastic ordinary differential equation in the unit ball B with appropriated coefficients, we deduce that

$$
y(t) := \Psi(t, z(t)), \qquad \ell(t) := -\frac{\nabla\Psi(t,z(t))\mathrm{d}\eta(t)}{\lambda(t,z(t))}
$$

satisfies the desired equation, This completes the proof. □

• *Remark* 6.12. The reflecting direction $b(t,x)$ is assumed continuously differentiable in t so that Itô formula can be properly used. Thus, to accommodate a control v on the reflection direction, i.e., $b_0(t,x,v)$, this approach presents some difficulties, since $v = v(t)$ should be a cad-lag adapted process. □

• *Remark* 6.13. Assumption (6.46) is not really needed. It suffices to know that the domain $\mathcal{O}$ can be transformed into a convex domain, via a diffeomorphism of class C^2 with the property of mapping the oblique interior direction b into the interior normal direction, even if the boundary is only piecewise smooth. □

Now we reconsider the penalty argument for a smooth domain, not necessarily convex. Assume that the distance to the boundary $\partial\mathcal{O}$ is a smooth function, i.e.

$$
\begin{cases} x \mapsto \mathrm{dist}(x, \partial\mathcal{O}) \text{ is a twice continuously differentiable} \\ \text{function with bounded first and second derivatives,} \end{cases} \tag{6.47}
$$

for x sufficiently near to the boundary $\partial\mathcal{O}$. This implies that there is a twice continuously differentiable function $\rho : \mathbb{R}^d \to \mathbb{R}$, which first and second derivatives are bounded, such that

$$
\begin{aligned}
&\mathcal{O} = \{x \in \mathbb{R}^d : \rho(x) < 0\},\\
&\partial\mathcal{O} = \{x \in \mathbb{R}^d : \rho(x) = 0\},\\
&|\nabla\rho(x)| \geq 1 \quad \forall x \in \partial\mathcal{O},
\end{aligned}
$$

and $|\rho(x)| = \mathrm{dist}(x, \partial\mathcal{O})$ if $\mathrm{dist}(x, \partial\mathcal{O}) < \delta_1$, for some constant $\delta_1 > 0$, $|\nabla\rho(x)| \geq 1$ for every x belongs to $\mathbb{R}^d \setminus \mathcal{O}$, and $\rho(x)$ is bounded for x in $\overline{\mathcal{O}}$. Note that the domain $\mathcal{O}$ may be unbounded.

The positive part $\rho^+(x)$ behaves (locally) like $\mathrm{dist}(x,\mathcal{O})$, the distance to the set $\mathcal{O}$ (or equivalently, distance to the closure $\overline{\mathcal{O}}$) from the point x, i.e., for every $r>0$ there exist two constants $C_r \geq 1 \geq c_r > 0$ such that

$$c_r \,\mathrm{dist}(x,\mathcal{O}) \leq \rho^+(x) \leq C_r \,\mathrm{dist}(x,\mathcal{O}),. \tag{6.48}$$

for every x in $\mathbb{R}^d$ with $|x| \leq r$. Essentially, this follows from the equality

$$\rho^+(x) = \int_0^1 (x-a)\cdot\nabla\rho\big(a+\theta(x-a)\big)\mathbb{1}_{\{a+\theta(x-a)\notin\mathcal{O}\}}\mathrm{d}\theta,$$

for every x in $\mathbb{R}^d$ and a in $\overline{\mathcal{O}}$.

The exterior unit normal at the boundary point x is given by

$$n(x) = |\nabla\rho(x)|^{-1}\nabla\rho(x), \quad \forall x\in\partial\mathcal{O}.$$

Since the domain is not longer convex, the key inequality $(x-x')\cdot n(x)\geq 0$, for every x on the boundary $\partial\mathcal{O}$ and x' in the closure $\overline{\mathcal{O}}$, is not satisfied anymore. However, the identity

$$\begin{aligned}\rho(x')-\rho(x) &= (x'-x)\cdot\nabla\rho(x) + \\ &+\int_0^1 (x'-x)\cdot\big[\nabla\rho\big(x+\theta(x'-x)\big)-\nabla\rho(x)\big]\mathrm{d}\theta,\end{aligned}$$

and the Lipschitz continuity of ρ yield

$$\begin{cases} \rho^+(x)(x'-x)\cdot\nabla\rho(x) \leq \\ \qquad\qquad \leq \rho^+(x)\rho^+(x') + C\rho^+(x)|x-x'|^2, \quad \forall x,\, x', \end{cases} \tag{6.49}$$

for some constant $C>0$. Properties (6.48) and (6.49) are the equivalent to the monotone properties (6.31) and (6.32) in the convex case.

Even if the domain $\mathcal{O}$ is unbounded, under this regularity on the boundary $\partial\mathcal{O}$, there exists a Lipschitz continuous function $P_{\mathcal{O}}:\mathbb{R}^d\to\mathbb{R}^d$ such that

$$|x-P_{\mathcal{O}}(x)| \leq C_1\,\mathrm{dist}(x,\mathcal{O}) \quad \text{and} \quad P_{\mathcal{O}}(x)\in\overline{\mathcal{O}},$$

for every x satisfying $\mathrm{dist}(x,\mathcal{O})<c_1$, for a sufficiently small constant C_1, $c_1>0$. Notice that $P_{\mathcal{O}}(x)=x$ for every x in $\overline{\mathcal{O}}$.

A priori, all coefficients g, σ and γ are initially defined on the closure $\overline{\mathcal{O}}$. Instead of condition (6.26) on the jumps, we assume

$$\begin{cases} x+\theta\gamma(\zeta,t,x,v)\in\overline{\mathcal{O}}, \quad \forall\theta\in[0,1], \\ \forall(\zeta,t,x,v)\in\mathbb{R}^m_*\times[0,\infty)\times\overline{\mathcal{O}}\times\mathbb{R}^q. \end{cases} \tag{6.50}$$

Then, by means of the function $P_{\mathcal{O}}$, we can extend their definition to the whole space $\mathbb{R}^d$ preserving the assumptions (6.18), (6.19) and (6.29). In particular, taking a smooth function χ_1 from $\mathbb{R}^d$ into $[0,1]$ satisfying $\chi(x)=0$ for any x with $\mathrm{dist}(x,\mathcal{O})>c_1$ (the same constant c_1 as above), and $\chi_1(x)=1$ for every x with $\mathrm{dist}(x,\mathcal{O})<c_1/2$, we may extend the jump coefficient as follows

$$\gamma(\zeta,t,x,v) = \chi_1(x)\gamma(\zeta,t,P_{\mathcal{O}}(x),v), \quad \forall\zeta,t,x,v,$$

so that condition (6.50) yields

$$\begin{cases} \operatorname{dist}(x+\theta\gamma(\zeta,t,x,v),\mathcal{O}) \le C_1 \operatorname{dist}(x,\mathcal{O}), \quad \forall\theta\in[0,1], \\ \forall(\zeta,t,x,v)\in\mathbb{R}^m_*\times[0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q, \end{cases} \tag{6.51}$$

where C_1 is the same constant as above.

We assume that the oblique interior direction $b(t,x)$ has the form

$$b(t,x) = -M(t,x)n(x), \quad \forall t, x, \tag{6.52}$$

where $n(x)$ is the exterior normal at the boundary point x and $M(t,x)$ is a symmetric $d\times d$ square matrix defined and continuous for $t\ge 0$ and x in a neighborhood of the boundary $\partial\mathcal{O}$, which is continuously differentiable in t, twice continuously differentiable in x, and satisfies

$$c|\eta|^2 \le \eta\cdot M(t,x)\eta \le C|\eta|^2, \quad \forall\eta\in\mathbb{R}^d, \tag{6.53}$$

for every t, x and some constants $C\ge c>0$. Certainly, only the lower bound in c is relevant when the domain $\mathcal{O}$ is bounded. Moreover, by constructing a smooth function χ_2 from $\mathbb{R}^d$ into $[0,1]$ such that $\chi_2(x)=1$ in a neighborhood of the boundary $\partial\mathcal{O}$ and $\chi_2(x)=0$ whenever $M(x,\cdot)$ is not defined, we may extend the definition of the matrix $M(t,x) := \chi(x)M(t,x) + \big(1-\chi(x)\big)I_d$, where I_d is the identity $d\times d$ matrix. The property (6.53) on the (symmetric nonnegative) matrix is preserved, i.e., (6.53) holds for every t and x.

Now, we are ready to define the penalty function $\beta(t,x)$, differentiable in t and uniformly Lipschitz continuous in x, by the expression

$$\beta(t,x) := \rho^+(x)M(t,x)\nabla\rho(x), \tag{6.54}$$

where ρ^+ denotes the positive part of the function ρ defining the domain $\mathcal{O}$ under the assumption (6.47). Notice that, in view of the upper bound in (6.53), we have

$$|\beta(t,x)| \le C\rho^+(x)|\nabla\rho(x)|, \quad \forall t,\, x,$$

for some constant $C\ge 1$, while the lower bound in (6.53) yields

$$\nabla\rho(x)\cdot\beta(t,x) \ge c\rho^+(x)|\nabla\rho(x)|^2, \quad \forall t,\, x, \tag{6.55}$$

for some constant $0<c\le 1$.

Given $\varepsilon>0$, consider the stochastic ordinary differential equation

$$\begin{cases} \mathrm{d}x^\varepsilon(t) = g(t,x^\varepsilon(t),v(t))\mathrm{d}t + \sigma(t,x^\varepsilon(t),v(t))\mathrm{d}w(t) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_*}\gamma(\zeta,t,x^\varepsilon(t),v(t))\tilde\nu(\mathrm{d}\zeta,\mathrm{d}t) - \frac{1}{\varepsilon}\beta(t,x^\varepsilon(t))\mathrm{d}t, \end{cases} \tag{6.56}$$

with the initial condition $x^\varepsilon(t_0)=x^0$. Since β is Lipschitz continuous in x, under assumptions (6.18) and (6.19) there exists only and only one solution $x^\varepsilon(\cdot)$ of (6.56) satisfying the given initial conditions. At this point, we have

almost the same properties of used for the stochastic variational inequality used in the convex case. The ε-local time process is defined by

$$\ell^\varepsilon(t) := \frac{1}{\varepsilon}\int_{t_0}^{t} \rho^+(x^\varepsilon(s))|\nabla\rho(x^\varepsilon(s))|\,\mathrm{d}s, \tag{6.57}$$

for every $t \geq t_0$.

Theorem 6.14 (OR-SODE)**.** *Let the growth condition* (6.18) *and the uniform locally Lipschitz condition* (6.19) *on the coefficients g, σ and γ be satisfied. Suppose also assumptions* (6.47), (6.51), (6.52) *and* (6.53) *relative to a smooth domain $\mathcal{O}$ in $\mathbb{R}^d$ and a oblique interior direction b. Then there is one and only one solution $(x(t), \ell(t) : t \geq t_0)$ of the oblique reflected stochastic ordinary differential equation* (6.44), (6.45)*, as described by* Definition 6.10*. Moreover for every $t_1 > t_0$, the following convergence holds in probability*

$$\sup_{t_0\leq t\leq t_1} \big\{|x^\varepsilon(t) - x(t)| + |\ell^\varepsilon(t) - \ell(t)|\big\} \to 0 \quad \textit{as} \quad \varepsilon \to 0, \tag{6.58}$$

where $(x^\varepsilon(t) : t \geq t_1)$ is the solution of the penalized stochastic ordinary differential equation (6.56) *with the initial condition $x^\varepsilon(t_0) = x^0$, for some given x^0 in $\overline{\mathcal{O}}$. Furthermore, if the coefficients are uniform Lipschitz and the growth condition for every $\alpha > 0$,*

$$\begin{cases} |g(t,x,v)|^2 + |\sigma(t,x,v)|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*} |\gamma(\zeta,t,x,v)|^2 \exp\big(\alpha|\gamma(\zeta,t,x,v)|\big)\pi(\mathrm{d}\zeta) \leq C_\alpha, \end{cases} \tag{6.59}$$

for some constant $C_\alpha > 0$ and for every (t,x,v) in $[t_0,\infty)\times\mathbb{R}^d\times\mathbb{R}^q$, is satisfied, then we also have

$$\mathbb{E}\Big\{\sup_{t_0\leq t\leq t_1} \{|x^\varepsilon(t) - x(t)|^p + |\ell^\varepsilon(t) - \ell(t)|^p\}\Big\} \to 0 \quad \textit{as} \quad \varepsilon \to 0, \tag{6.60}$$

for every $p \geq 1$ and locally uniform with respect to the initial condition x^0

Proof. For a given constant $\alpha > 0$, consider the function ϕ_α

$$(t,\xi,x,y,\ell) \mapsto \big(\xi\cdot[M^{-1}(t,x) + M^{-1}(t,y)]\xi\big)\exp[-\alpha(t+\ell)], \tag{6.61}$$

Since $M(t,x)$ is a nonnegative symmetric matrix, the matrix $M^{-1}(t,x)$ is continuously differentiable in t, twice continuously differentiable in x and satisfies

$$c|\eta|^2 \leq \eta\cdot M^{-1}(t,x)\eta \leq C|\eta|^2, \quad \forall\eta\in\mathbb{R}^d,$$

for some constants $C \geq 1 \geq c > 0$. Thus, for some constant $C > 0$, we obtain the bound

$$\begin{aligned} &|\partial_t\phi_\alpha(t,\xi,x,y,\ell)| + |\partial_x^k\phi_\alpha(t,\xi,x,y,\ell)| + \\ &\qquad + |\partial_y^k\phi_\alpha(t,\xi,x,y,\ell)| \leq C|\xi|^2\exp\big(-\alpha(t+\ell)\big), \end{aligned}$$

for every (t, ξ, x, y, ℓ), $k = 1, 2$, where ∂_t, ∂_x^k and ∂_y^k denote the partial derivatives with respect to the variables t, x and y. Also, by means of the property (6.49) we deduce

$$|\partial_\xi \phi_\alpha(t, \xi, x, y, \ell) \cdot M(t, x)\nabla\rho(x) - 2\xi \cdot \nabla\rho(x) \exp\big(-\alpha(t+\ell)\big)| \leq \\ \leq C|\xi|\,|x-y|\,\exp\big(-\alpha(t+\ell)\big),$$

for every (t, ξ, x, y, ℓ) and some constant $C > 0$. This inequality and the Lipschitz continuous property of $\nabla\rho$ yield

$$\partial_\xi \phi_\alpha(t, x-y, x, y, \ell) \cdot M(t, x)\nabla\rho(x) \leq \\ \leq \big[2(y-x)\cdot\nabla\rho(x) + C|x-y|^2\big]\,\exp\big(-\alpha(t+\ell)\big) \leq \\ \leq \big[2\big(\rho(y)-\rho(x)\big) + C|x-y|^2\big]\,\exp\big(-\alpha(t+\ell)\big),$$

for every (t, ξ, x, y, ℓ) and some constant $C > 0$. Hence,

$$\partial_\xi \phi_\alpha(t, x-y, x, y, \ell) \cdot M(t, x)\nabla\rho(x) \leq \\ \leq C|x-y|^2\,\exp\big(-\alpha(t+\ell)\big),$$

for every y in $\overline{\mathcal{O}}$, x on the boundary $\partial\mathcal{O}$, any t and ℓ.

To show the uniqueness, let $(x(t), \ell(t) : t \geq t_0)$ and $(\bar{x}(t), \bar{\ell}(t) : t \geq t_0)$ be two solutions of the oblique reflected stochastic ordinary differential equation (6.44), (6.45). As in Theorem 6.7, for any $r > 0$, define the adapted process

$$\chi_r(t) := \begin{cases} 0 & \text{if } |x(s)| > r \text{ or } |\bar{x}(s)| > r \text{ for some } s > t, \\ 1 & \text{otherwise,} \end{cases}$$

and its associate stopping time τ_r with $\chi_r(t) = 1$ if and only if $t < \tau_r$, to check that

$$[x(t) - \bar{x}(t)]\chi_r(t) = \int_{t_0}^t \chi_r(s)\big[g(s, x(s), v(s)) - g(s, \bar{x}(s), v(s))\big]\mathrm{d}s \\ + \int_{t_0}^t \chi_r(s)\big[\sigma(s, x(s), v(s)) - \sigma(s, \bar{x}(s), v(s))\big]\mathrm{d}w(s) + \\ + \int_{\mathbb{R}^m_* \times]t_0, t]} \chi_r(s)\big[\gamma(\zeta, s, x(s), v(s)) - \gamma(\zeta, s, \bar{x}(s), v(s))\big]\tilde{\nu}(\mathrm{d}\zeta, \mathrm{d}s) + \\ + \int_{t_0}^t \chi_r(s)[\mathbb{1}_{\{x(t)\in\partial\mathcal{O}\}} b(t, x(t))\mathrm{d}\ell(s) - \mathbb{1}_{\{\bar{x}(t)\in\partial\mathcal{O}\}} b(t, \bar{x}(t))\mathrm{d}\bar{\ell}(s)],$$

for every $t_0 \leq t$, in particular for $t \leq \tau_r$. Now, apply Itô formula to the function

$$(t, x, y, \ell) \mapsto \phi_\alpha(t - t_0, x - y, x, y, \ell)$$

with the processes

$$x_t := x(t), \quad y_t := \bar{x}(t), \quad \ell_t := \ell(t) + \bar{\ell}(t)$$

to get

$$\mathbb{E}\big\{|x(t)-\bar{x}(t)|^2\chi_r(t)\exp\big(-\alpha(t-t_0+\ell(t)+\bar{\ell}(t))\big)\big\}\le$$
$$\le (C_r-\alpha c)\,\mathbb{E}\Big\{\int_{t_0}^{t}|x(s)-\bar{x}(s)|^2\chi_r(s)\times$$
$$\times\exp\big(-\alpha(s-t_0+\ell(s)+\bar{\ell}(s))\big)[\mathrm{d}s+\mathrm{d}\ell(s)+\mathrm{d}\bar{\ell}(s)]\Big\}$$

for some constant $C_r>1>c>0$, which may depend on r but not on α. Thus, for α sufficiently large, we get $x(t)=\bar{x}(t)$ for every $t_0\le t\le\tau_r$. Next, as r become large, we deduce $x(\cdot)=\bar{x}(\cdot)$, i.e., the uniqueness holds.

For given constants $\lambda\ge\lambda_0\ge 1$ sufficiently large so that $\rho(x)+\lambda_0\ge 1$, for every x, consider the function

$$\psi_\lambda:(t,x,\ell)\mapsto(\lambda\,\lambda_0+|x|^2)^{1/2}+\lambda\rho(x), \tag{6.62}$$

which is twice continuously differentiable and satisfies

$$\nabla\psi_\lambda\cdot M(t,x)\nabla\rho(x)\ge C\,x\cdot\nabla\rho(x)(\lambda\,\lambda_0+|x|^2)^{-1/2}+c\,|\nabla\rho(x)|^2\ge$$
$$\ge\big[-C+c\,\lambda\,|\nabla\rho(x)|\big]\,|\nabla\rho(x)|,$$

for every x, in view of the assumption (6.53) on the matrix $M(t,x)$. Since $|\nabla\rho|\ge 1$ in $\mathbb{R}^d\setminus\mathcal{O}$, we have

$$\nabla\psi_\lambda\cdot M(t,x)\nabla\rho(x)\,\rho^+(x)\ge[c\,\lambda-C]\,|\nabla\rho(x)|\,\rho^+(x),$$

for every x in $\mathbb{R}^d$. Thus we take λ sufficiently large so that $c\,\lambda-C\ge 1$.

Now, similarly to Lemma 6.6 and under the extra assumption (6.29), apply Itô formula to the function

$$(t,x)\mapsto[\psi_\lambda(x)]^p\,\exp\,[-\alpha(t-t_0)]$$

with the process $x^\varepsilon(t)$ to obtain first a bound on $\sup_t|x^\varepsilon(t)|^p$ and then a bound on the ε-local time process $\ell^\varepsilon(t)$ defined by (6.57), i.e.,

$$\mathbb{E}\Big\{\big(\ell^\varepsilon(t_1)\big)^p+\sup_{t_0\le t\le t_1}|x^\varepsilon(t)|^p\Big\}\le C,\quad\forall\varepsilon>0, \tag{6.63}$$

and, for every p in $[2,\bar{p}]$, $t_1>t_0$ and x^0 in $\overline{\mathcal{O}}$.

Again, similarly to the Lemma 6.6, apply Itô formula to the function $x\mapsto[\rho^+(x)]^p$ with the process $x^\varepsilon(t)$ to obtain, for every p in $[2,\bar{p}]$, $t_1>t_0$ and x^0 in $\overline{\mathcal{O}}$ the estimate

$$\mathbb{E}\Big\{\frac{1}{\varepsilon}\int_{t_0}^{t_1}[\rho^+(x^\varepsilon(t))]^p\mathrm{d}t+\sup_{t_0\le t\le t_1}[\rho^+(x^\varepsilon(t))]^p\Big\}\le C\,\varepsilon^{p/2-1}, \tag{6.64}$$

for another constant $C>0$, independent of ε in $(0,1]$. Moreover, for every $1\le p\le\bar{p}$, $0<q\le\bar{p}/2-p/\bar{p}$, $q<p/2$, $t_1>t_0$ and x^0 in $\overline{\mathcal{O}}$, there exists a constant $C>0$ such that

$$\mathbb{E}\Big\{\Big(\frac{1}{\varepsilon'}\int_{t_0}^{t_1}\rho^+(x^{\varepsilon'}(t))\,\rho^+(x^\varepsilon(t))\mathrm{d}t\Big)^p\Big\}\le C\,\varepsilon^q, \tag{6.65}$$

for every ε, $\varepsilon' > 0$. Furthermore, in the above estimates (6.63), (6.64) and (6.65), all constants can be taken independent of the initial condition x^0 and t_0, as long as they remain in a bounded region and assumption (6.29) holds.

At this point, we have extended all a priori bounds of Lemma 6.6 to a smooth (not necessarily convex) domain $\mathcal{O}$. Assume the growth condition (6.59) and uniform Lipschitz continuity on the coefficients g, σ and γ. Applying Itô formula to the function $\rho(x)$ with the process $x^\varepsilon(t)$ we deduce an exponential bound for the approximate local time process $\ell^\varepsilon(t)$, namely, for every $\alpha > 0$ there exists a $C_\alpha > 0$ such that

$$\mathbb{E}\Big\{ \sup_{t_0 \le s \le t} \exp[\alpha \ell^\varepsilon(s)] \Big\} < C_\alpha, \quad \forall t > t_0, \tag{6.66}$$

for every $\varepsilon > 0$. Note that condition (6.59) implies the extra assumption (6.29) for every $\bar{p} > 2$. The general case is obtained by approximation.

Now, with ϕ_α given by (6.61), apply Itô formula to the function (the same used to show the uniqueness)

$$(t, x, y, \ell) \mapsto \phi_\alpha(t - t_0, x - y, x, y, \ell)$$

with the processes

$$x_t := x^\varepsilon(t), \quad y_t := x^{\varepsilon'}(t), \quad \ell_t := \ell^\varepsilon(t) + \ell^{\varepsilon'}(t)$$

to get

$$\begin{aligned} |x^\varepsilon(t) - x^{\varepsilon'}(t)|^2 \exp\big(-\alpha(t - t_0 + \ell^\varepsilon(t) + \ell^{\varepsilon'}(t))\big) &\le I(t) + Q(t) + \\ &+ (C - \alpha c) \int_{t_0}^t |x^\varepsilon(s) - x^{\varepsilon'}(s)|^2 \times \\ &\times \exp\big(-\alpha[s - t_0 + \ell^\varepsilon(s) + \ell^{\varepsilon'}(s)]\big)[\mathrm{d}s + \mathrm{d}\ell^\varepsilon(s) + \mathrm{d}\ell^{\varepsilon'}(s)], \end{aligned}$$

where the stochastic integral $I(t)$ and the pathwise integral $Q(t)$ satisfy

$$\begin{aligned} \mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |I(t)| \Big\} \le C\, \mathbb{E}\Big\{ \Big[\int_{t_0}^{t_1} |x^\varepsilon(s) - x^{\varepsilon'}(s)|^2 \times \\ \times \exp\big(-2\alpha[s - t_0 + \ell^\varepsilon(s) + \ell^{\varepsilon'}(s)]\big) \mathrm{d}s \Big]^{1/2} \Big\}, \end{aligned}$$

and

$$Q(t) \le C\Big(\frac{1}{\varepsilon} + \frac{1}{\varepsilon'}\Big) \int_{t_0}^t \rho^+(x^\varepsilon(s))\, \rho^+(x^{\varepsilon'}(s))\, \mathrm{d}s$$

for some constants $C > 1 > c > 0$, independent of α, ε. Hence, by means of (6.65) as in Theorem 6.7, we can estimate

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |x^\varepsilon(t) - x^{\varepsilon'}(t)|^{2p} \exp\big(-p\alpha(t - t_0 + \ell^\varepsilon(t) + \ell^{\varepsilon'}(t))\big) \Big\},$$

for every $p \ge 1$, which together with (6.66) yield

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |x^\varepsilon(t) - x^{\varepsilon'}(t)|^p \Big\} \le C_{pq}\, (\varepsilon + \varepsilon')^q,$$

for some constant $C_{pq} > 0$ and any ε, $\varepsilon' > 0$ and any $p > 2q > 2$. Then, from the stochastic ordinary differential equation (6.56) follows

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |\eta^{\varepsilon}(t) - \eta^{\varepsilon'}(t)|^p \Big\} \le C_{pq} (\varepsilon + \varepsilon')^q,$$

where the process $\eta(\cdot)$ is defined by

$$\eta^{\varepsilon}(t) := \frac{1}{\varepsilon} \int_{t_0}^{t} \beta(s, x^{\varepsilon}(s)) \mathrm{d}s,$$

for every $t \ge t_0$.

Now we need to establish that for every $t_1 > t_0$ and $p > 2$ we have

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |\ell^{\varepsilon}(t) - \ell^{\varepsilon'}(t)|^p \Big\} \to 0, \tag{6.67}$$

as $\varepsilon + \varepsilon'$ vanishes. Indeed, by definition

$$\mathrm{d}\eta^{\varepsilon}(t) = \frac{1}{\varepsilon} \rho^{+}(x^{\varepsilon}(t)) M(t, x^{\varepsilon}(t)) \nabla\rho(x^{\varepsilon}(t)) \mathrm{d}s = a_{\varepsilon}(t) \mathrm{d}\ell^{\varepsilon}(t),$$

where

$$a_{\varepsilon}(t) := M(t, x^{\varepsilon}(t)) \nabla\rho(x^{\varepsilon}(t)) \, |\nabla\rho(x^{\varepsilon}(t))|^{-1}.$$

Thus, defining

$$b_{\varepsilon}(t) := M^{-1}(t, x^{\varepsilon}(t)) \nabla\rho(x^{\varepsilon}(t)) \, |\nabla\rho(x^{\varepsilon}(t))|^{-1}.$$

we get

$$\mathbb{E}\Big\{ \sup_{t_0 \le t \le t_1} |b_{\varepsilon}(t) - b_{\varepsilon'}(t)|^p \Big\} \le C_{pq} (\varepsilon + \varepsilon')^q,$$

for a constant $C_{pq} > 0$ as above, and

$$a_{\varepsilon}(t) \cdot b_{\varepsilon}(t) = 1,$$
$$|a_{\varepsilon}(t)| + |b_{\varepsilon}(t)| \le C,$$

for some constant $C > 0$ and every $t \ge t_0$, and $\varepsilon > 0$. Hence,

$$b_{\varepsilon}(t) \cdot \mathrm{d}\eta^{\varepsilon}(t) = b_{\varepsilon}(t) \cdot a_{\varepsilon}(t) \mathrm{d}\ell^{\varepsilon}(t) = \mathrm{d}\ell^{\varepsilon}(t),$$

which yields

$$\ell^{\varepsilon}(t) - \ell^{\varepsilon'}(t) = \int_{t_0}^{t} [b_{\varepsilon}(s) - b_{\varepsilon'}(s)] \cdot \mathrm{d}\eta^{\varepsilon'}(s) + \\ + \int_{t_0}^{t} b_{\varepsilon}(s) \cdot [\mathrm{d}\eta^{\varepsilon}(s) - \mathrm{d}\eta^{\varepsilon'}(s)].$$

The convergence of the first integral (with respect to $\mathrm{d}\eta^{\varepsilon}$) is easily seen, and to check the convergence of the second (pathwise) integral we may proceed as follows. We approximate either (a) the integral with a finite sum or (b) the

integrand with smooth processes. For instance, because the function $(t,x) \mapsto M^{-1}(t,x)\nabla\rho(x)\,|\nabla\rho(x)|^{-1}$ is continuously differentiable, we can define a process

$$b_{\varepsilon,\delta}(t) = h_\delta(t, x^\varepsilon(t)),$$

for some smooth function $h_\delta(t,x)$, which is continuously differentiable in t and twice-continuously differentiable in x, such that for some constant $C > 0$ we have

$$|b_\varepsilon(t) - b_{\varepsilon,\delta}(t)| \le C\,\delta,$$

for every $t \ge t_0$ and $\varepsilon,\, \delta > 0$. Thus,

$$\Big|\int_{t_0}^t b_\varepsilon(s)\cdot[\mathrm{d}\eta^\varepsilon(s) - \mathrm{d}\eta^{\varepsilon'}(s)]\Big| \le C\,\delta\big(\ell^\varepsilon(t) + \ell^{\varepsilon'}(t)\big) + \\ + \int_{t_0}^t b_{\varepsilon,\delta}(s)\cdot[\mathrm{d}\eta^\varepsilon(s) - \mathrm{d}\eta^{\varepsilon'}(s)],$$

by means of Itô formula (integration by parts), the convergence of the above (pathwise) integral in $\mathrm{d}\eta^\varepsilon - \mathrm{d}\eta^{\varepsilon'}$ can be controlled by the stochastic integral

$$\int_{t_0}^t [\eta^\varepsilon(s) - \eta^{\varepsilon'}(s)]\cdot \mathrm{d}b_{\varepsilon,\delta}(s).$$

Collection all, we deduce the convergence (6.67).

Next, from the above properties we verify that the processes

$$x(t) := \lim_{\varepsilon\to 0} x^\varepsilon(t), \qquad \ell(t) := \lim_{\varepsilon\to 0}\ell^\varepsilon(t),$$

are well defined, solve the oblique reflected stochastic ordinary differential equation (6.44), (6.45), and satisfy (6.60).

Finally, as in Theorem 6.7, we can ignore the growth condition (6.59) and uniform Lipschitz continuity on the coefficients g, σ and γ to obtain only the convergence in probability (6.58). □

• *Remark* 6.15. Note that estimates (6.63), (6.64) and (6.65) hold under the extra assumption (6.29), and the smooth boundary condition (6.47), but without assuming the smoothness on the oblique interior direction, only the bounds (6.53) are necessary. The growth condition (6.59) can be substituted by a (6.29) as long as the exponential of the local time is not involved, e.g., a condition to ensure that

$$|\partial_x^k\phi_\alpha(t,\xi,x,y,\ell)| \le C|\xi|^2(1+|x|)^{-k}\exp\big(-\alpha(t+\ell)\big),$$

for every (t,ξ,x,y,ℓ), $k = 1,2$ (similarly for $\partial_y^k\phi_\alpha$). Anyway, there are other possible choices for the functions ϕ_α, ψ_λ. Certainly, Lipschitz continuity with respect to the initial data, i.e., a priori estimates of the type (6.41) and (6.42) hold true. □

Notice that the oblique interior direction $b(t,x)$ is always assumed continuously differentiable in t, so that a controlled reflection of the type $b(t,x,v)$ is not included in this section.

Is there a good technique to construct the reflected diffusion by local coordinates?

6.4 SDE with Boundary Conditions

Case with control on the drift, diffusion and jump coefficients. Again, focus attention to estimates. ?CHECK the validity of estimates with measurable coefficients Section 4.6 for this case.

Check also Dupuis and Ishii [64] and Dupuis and Ramanan [65]. Include the case of a domain given by the intersections of a finite number of semi-spaces (either bounded or unbounded) and a reflection other than the normal.

RECALL: the Skorokhood map is Lipschitz continuous from $D([0,\infty),\mathbb{R}^d)$ into itself with either the Skorokhood topology or the locally uniform topology if the domain has a "piecewise-linear" boundary. If the boundary is just "enough" then this fact seems not to be true...

Check the recent book by Rong [215] and Rong [216].

6.4.1 Strong Formulation

FROM papers Chaleyat-Maurel, El Karoui and Marchal [37] and Lions and Sznitman [157]

CHECK paper Anulova [3, 4] and Anderson [2]

...***...

6.4.2 Weak Formulation

FROM papers Chaleyat-Maurel, El Karoui and Marchal [37] and Lions and Sznitman [157]

...***...

6.5 Girsanov's Change of Measure

Something close to the section 5.4 Measure and Time Changes in Chapter 5, with the reflection.

Control via Girsanov transformation and setting of the dynamical of a control problem (not optimal control!), with out the cost or value function.

Maybe this should go as a Subsection between "Strong Formulation" and "Weak Formulation" of section "SDE with Boundary Conditions".

Je croix que tu peux faire ce partie mieux que moi...

... ...

6.6 Reflected Diffusion with Jumps

Oblique reflection. Some discussion (refer to Chapter 2 and the books Garroni and Menaldi [93, 94]) on the transition function. Explicitly mention Kolmogorov's Forward and Backward equations, relation with the Green Function (and Fundamental Solution) in the stationary and time-dependent cases

Add Feynman-Kac type formula. Mention Itô formula for smooth flows.

... .../**** ***

Chapter 6. This will be your first serious work on the book! 1.- Finish the section on Backward Equation, incorporating the first subsection form Chapter 4.

2.- Section 6.4 (SDE with Boundary Conditions), we think I know from where to take them (see note in the text)

3.- Section 6.5 (Reflected Diffusion w/Jumps), this should be analogous to Section 5.7, as mentioned early.

4.- Section 6.6 (Girsanov's Change of Measure) , this should be analogous to Section 5.3.

Chapter 7

Stochastic Differential Equations III

MAINLY A WORKING CHAPTER ...***...

In the first three section we give details on some analytic techniques used to show the existence and regularity of the density (with respect to the Lebesgue measure) of the jump diffusion with or without boundary condition. We choose the half-space as a model case, in general regions, most of the arguments are based on local coordinates and sharp estimates proved in a half-space.

...***...

7.1 Constant Coefficient in Half-Space

BASED ON PAPER Menaldi and Tubaro [177]

Suppose that on the interior $\dot{\mathbb{R}}^d_+ = \{x : x_d > 0\}$ we are given a second-order constant elliptic differential operator

$$\begin{cases} A\varphi(x) := A_0\varphi(x) + \sum_{i=1}^d a_i \partial_i \varphi(x) - a_0\varphi(x), \\ A_0\varphi(x) := \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_{ij}\varphi(x), \end{cases} \tag{7.1}$$

for every continuously differentiable functions φ with a compact support in the closed half-space $\mathbb{R}^d_+$, where (a_{ij}) is a symmetric positive definite matrix and $a_0 \geq 0$. If necessary, $\underline{a}$ denotes the d-dimensional square matrix (a_{ij}) and also, a (or $\underline{a}$) denotes the d-dimensional vector $(\tilde{a}, a_d)$, where $\tilde{a}$ (or $\underline{\tilde{a}}$) is equal to $(a_1, \ldots, a_{d-1})$. Also, assume that B is a (constant) *Wentzell type* boundary

differential operator of the form

$$\begin{cases} B\varphi(x) := B_0\varphi(x) + b_d\partial_d\varphi(x) - \rho A\varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \\ B_0\varphi(x) := \frac{1}{2}\sum_{i,j=1}^{d-1} b_{ij}\partial_{ij}\varphi(x) + \sum_{i=1}^{d-1} b_i\partial_i\varphi(x) - b_0\varphi(x), \end{cases} \tag{7.2}$$

where (b_{ij}) is a symmetric non-negative definite matrix, $b_0, b_d, \rho \geq 0$, and $\rho > 0$ or $b_d > 0$. If necessary, $\underline{\underline{\tilde{b}}}$ denotes the $(d-1)$-dimensional square matrix (b_{ij}) and also, b (or $\underline{b}$) denotes the d-dimensional vector $(\tilde{b}, b_d)$, where $\tilde{b}$ (or $\underline{\tilde{b}}$) is equal to $(b_1, \dots, b_{d-1})$. Most of the time, the tilde ˜ sign is used to emphasize the $(d-1)$ dimension of the given element, while the underline _ and the double underline $\underline{\underline{\ }}$ refer to a vector and a matrix, respectively. It may be convenient to write $\overline{\overline{b}}_{ij} = \sum_k \varsigma_{ik}\varsigma_{jk}$, where the matrix $\tilde{\varsigma}$ is the product of an orthogonal matrix $\tilde{\varrho}$ and a diagonal matrix $(\tilde{\beta})^{1/2}$, i.e., $\underline{\underline{\tilde{b}}} = \tilde{\varrho}\tilde{\beta}\tilde{\varrho}^*$ and $\varsigma_{ij} = \varrho_{ij}\sqrt{\beta_j}$, where $\beta_1, \dots, \beta_{d-1}$ are the eigenvalues of the matrix (b_{ij}).

Consider the boundary value problem

$$\begin{cases} \partial_t u(t,x) = Au(t,x), \quad \forall t > 0,\ x \in \dot{\mathbb{R}}^d_+, \\ Bu(t,x) + \psi(t,x) = 0, \quad \forall t > 0,\ x \in \partial\mathbb{R}^d_+, \\ u(0,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.3}$$

and the representation formula

$$\begin{cases} u(t,x) = \displaystyle\int_{\mathbb{R}^d_+} G_{A,B}(t, \tilde{x}, x_d, \tilde{y}, y_d)\varphi(y)\mathrm{d}y + \\ \qquad\qquad + \displaystyle\int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t-s, \tilde{x}, x_d, \tilde{y})\psi(s,y)\mathrm{d}\tilde{y}, \end{cases} \tag{7.4}$$

where $G_{A,B}$ and $P_{A,B}$ are the Green and the Poisson functions. Actually, both $G_{A,B}$ and $P_{A,B}$ have the form $G_{A,B}(t, \tilde{x} - \tilde{y}, x_d, y_d)$ and $P_{A,B}(t, \tilde{x} - \tilde{y}, x_d)$.

HAVE TO CHECK THE ADJOINT PROBLEM?

Remark that the adjoint problem to (7.3) is given by

$$\begin{cases} \partial_t u(t,x) = A^* u(t,x), \quad \forall t > 0,\ x \in \dot{\mathbb{R}}^d_+, \\ B^* u(t,x) + \psi(t,x) = 0, \quad \forall t > 0,\ x \in \partial\mathbb{R}^d_+, \\ u(0,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.5}$$

where the adjoint operators

$$A^*\varphi(x) := A_0\varphi(x) - \sum_{i=1}^d a_i\partial_i\varphi(x) - a_0\varphi(x), \tag{7.6}$$

and

$$\begin{cases} B^*_{\rm A}\varphi(x) := B^*_{\rm A0}\varphi(x) + b_d\partial_d\varphi(x) - \rho A^*\varphi(x), \\ B^*_{\rm A0}\varphi(x) := \frac{1}{2}\sum_{i,j=1}^{d-1} b_{ij}\partial_{ij}\varphi(x) - \sum_{i=1}^{d-1} b_i\partial_i\varphi(x) - \Big(b_0 + \frac{a_{dd}a_d}{b_d}\Big)\varphi(x), \end{cases} \tag{7.7}$$

i.e., the coefficient a_i changes into $a^*_i = -a_i$, for $i = 1, \dots, d$ and b_i changes into $b^*_i = -b_i$, for $i = 1, \dots, d-1$, and $b^*_0 = b_0 + a_{dd}a_d/b_d$. Clearly, we have the relation

$$P_{A^*,B^*_{\rm A}}(t, \tilde{x}, x_d, \tilde{y}) = P_{A,B}(t, \tilde{y}, x_d, \tilde{x}), \quad G_{A^*,B^*_{\rm A}}(t, x, y) = G_{A,B}(t, y, x),$$

for the Poisson and Green functions.

Our purpose is to discuss the construction of these Green and Poisson functions as well as their heat-kernel type estimates. Particularly, we denote by D^k_B any derivative of order k with respect to some coefficients of the operator B, i.e., with respect to any ς_{ij} (or ϱ_{ij} or β_i)[1], any $b_0, b_1, \dots, b_d$ and ρ. Also, we denote by D^n_{tx} any partial derivative in the variable t and $x = (x_1, \dots, x_d)$ of parabolic order n, i.e., $n = 2n_0 + n_1 + \cdots + n_d$ with $D^n_{tx} = \partial^{n_0}_t \partial^{n_1}_1 \dots \partial^{n_d}_d$. Therefore, for any n and k there exit positive constants $C_0 = C_{nk}$ and $c_0 = c_{nk}$ such that

$$|D^n_{tx} D^k_B Q_B(t, x)| \le C_0 t^{-(d+n)/2} \exp\Big(-c_0 \frac{|x|^2}{2t}\Big), \tag{7.8}$$

for every $t > 0$ and x in $\mathbb{R}^d_+$.

7.1.1 Heat-Kernel Notation

Denote by Γ_d the heat-kernel (or Gaussian kernel) of dimension d, i.e.,

$$\Gamma_d(t, x) = \Gamma_d(t, \tilde{x}, x_d) := (2\pi t)^{-d/2} \exp\Big(-\frac{1}{2t}\sum_{i=1}^{d} x_i^2\Big), \tag{7.9}$$

with x in $\mathbb{R}^d$, which may be written as $(\tilde{x}, x_d)$ or $(x_1, \dots, x_{d-1}, x_d)$. Note that $\Gamma_d(rt, r^2x) = r^{-d/2}\Gamma_d(t, x)$, for every $r > 0$. As long as confusion does not arrive, it may be convenient to use the notation

$$\Gamma_d(q, x) := (2\pi)^{-d/2} \det(q)^{-1/2} \exp\Big(-\frac{1}{2}\sum_{i,j=1}^{d} x_i q^{ij} x_j\Big), \tag{7.10}$$

for any symmetric (strictly) positive matrix $q = (q_{ij})$ with inverse matrix (q^{ij}). Thus, if $\mathbb{1}$ is the identity matrix then $\Gamma_d(t\mathbb{1}, x) = \Gamma_d(t, \tilde{x}, x_d)$. Clearly, $\Gamma_d(q, x) = \det(q)^{-1/2}\Gamma(\mathbb{1}, q^{-1/2}x)$ and $x \mapsto \Gamma_d(q, x - m)$ is the normal or Gaussian distribution in $\mathbb{R}^d$, with mean vector m and co-variance matrix q. Since q is

[1] in the degenerate case, the inverse mapping $(b_{ij}) \mapsto (\varsigma_{ij})$ is not differentiable

a symmetric positive matrix, it can be written as $q = \rho\lambda\rho^*$, where λ a diagonal matrix (of eigenvalues) and ρ a orthogonal matrix, and then

$$\Gamma_d(q,x) = \det(\lambda)^{-1/2}\Gamma_d(\mathbb{1}, \lambda^{-1/2}\rho^* x) = \prod_{i=1}^{d} \Gamma_1\big(\lambda_i, (\rho^* x)_i\big),$$

where $\lambda_i > 0$ are the eigenvalues of the matrix q.

Note that

$$-\partial_d\Gamma_d(t,x) = \frac{x_d}{t}\Gamma_d(t,x) = (2\pi t)^{-d/2}\,\frac{x_d}{t}\exp\big(-\frac{|x|^2}{2t}\big),$$

for any $t > 0$ and x in $\mathbb{R}^d$, and besides the properties

$$\Gamma_d(t,x) = \prod_{i=1}^{d}\Gamma_1(t,x_i) \quad \text{and} \quad \int_{\mathbb{R}^d}\Gamma_d(t,x)\mathrm{d}x = 1, \quad \forall t > 0,$$

we also have

$$\int_0^\infty \mathrm{d}t \int_{\mathbb{R}^{d-1}} \partial_d\Gamma_d(t,\tilde{x},x_d)\mathrm{d}\tilde{x} = 1, \quad \forall x_d > 0,$$

after using the changes of variables $\frac{\tilde{x}}{\sqrt{2t}} = \tilde{z}$ and $\frac{x_d}{\sqrt{2t}} = s$. However,

$$\int_0^\infty \Gamma_d(t,x)\mathrm{d}t = c_d|x|^{2-d}, \quad \forall x \in \mathbb{R}^d,$$

where c_d[2] is finite only for $d \geq 3$.

The Fourier and the Laplace transforms of the heat-kernel can be calculated, namely

$$\mathcal{F}\big\{\Gamma_d(q,x)\big\}(y) = \int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot y}\Gamma_d(q,x)\mathrm{d}x = \exp\Big(-\frac{1}{2}\sum_{i,j=1}^{d} y_i q_{ij} y_j\Big)$$

and

$$\mathcal{L}\big\{\Gamma_{2k+1}(t,x)\big\}(s) = \int_0^\infty e^{-st}\Gamma_{2k+1}(t,x)\mathrm{d}s = (-1)^k\partial_s^k\frac{1}{\sqrt{2s}}e^{-|x|\sqrt{2s}}.$$

Note that for even dimensions, the Laplace transform involves the modified Bessel functions of the second kind (also called Kelvin or MacDonald functions).

Besides the well known semigroup property

$$\int_{\mathbb{R}^d}\Gamma_d(t,y)\Gamma_d(s,x-y)\mathrm{d}y = \Gamma_d(t+s,x), \quad \forall t,s > 0,\ x \in \mathbb{R}^d,$$

[2]Note that $2c_d = \pi^{-d/2}\Gamma(d/2-1)$, with $\Gamma(\cdot)$ denoting the Gamma function.

based on the Fourier transform, one can check that

$$\int_{\mathbb{R}^d} \Gamma_d(q,y)\Gamma_d(p,x-ry)\mathrm{d}y = \Gamma_d(p+rqr^*,x), \quad \forall x \in \mathbb{R}^d,$$

for any symmetric positive definite matrices p, q and any matrix r. Also, essentially based on the Laplace transform, one can verify that

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \partial_d\Gamma_d(t-s,\tilde{x}-\tilde{y},x_d)\partial_d\Gamma_d(s,\tilde{y},y_d)\mathrm{d}\tilde{y} = -\partial_d\Gamma_d(t,\tilde{x},x_d+y_d),$$

and

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \partial_d\Gamma_d\big(t-s,\tilde{x}+\tilde{b}(t-s)-\tilde{y},x_d+b_d(t-s)\big) \times \\ \times \partial_d\Gamma_d(s,\tilde{y}+\tilde{b}s,y_d+b_ds)\mathrm{d}\tilde{y} = -\partial_d\Gamma_d(t,\tilde{x}+\tilde{b}t,x_d+b_dt+y_d),$$

for every $t, x_d, y_d > 0$, $\tilde{x}$ in $\mathbb{R}^{d-1}$ and $b = (\tilde{b}, b_d)$ in $\mathbb{R}^d$.

Certainly, the heat-kernel satisfies the equation

$$\partial_t\Gamma_d(t,x) = \tfrac{1}{2}\Delta\Gamma_d(t,x), \quad \forall t > 0,\, x \in \mathbb{R}^d,$$

the initial condition

$$\lim_{t\to 0}\int_{\mathbb{R}^d} \Gamma_d(t,x-y)\varphi(y)\mathrm{d}y = \varphi(x), \quad \forall x \in \mathbb{R}^d,$$

and the boundary condition

$$\lim_{x_d\to 0,\, x_d>0}\int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \partial_d\Gamma_d(t-s,\tilde{x}-\tilde{y},x_d)\psi(s,\tilde{y})\mathrm{d}\tilde{y} = -\psi(t,\tilde{x}),$$

for every $t > 0$, $\tilde{x}$ in $\mathbb{R}^{d-1}$, and for any smooth functions with compact support φ and ψ.

This last *jump* condition gives rise the so-called *single-layer* and *double-layer* potentials. Indeed, for any symmetric (strictly) positive matrix $q = (q_{ij})$ with inverse matrix (q^{ij}), the heat-kernel satisfies the equation

$$\partial_t\Gamma_d(tq,x) = \frac{1}{2}\sum_{i,j=1}^d q_{ij}\partial_{ij}\Gamma_d(tq,x), \quad \forall t > 0,\, x \in \mathbb{R}^d,$$

the initial condition remains the same, but for the boundary condition one need to replace the normal derivative ∂_d with the co-normal derivative $\sum_{i=1}^d q_{di}\partial_i$, i.e.,

$$\lim_{x_d\to 0,\, x_d>0}\sum_{i=1}^d \int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} q_{di}\partial_i\Gamma_d\big((t-s)q,\tilde{x}-\tilde{y},x_d\big)\psi(s,\tilde{y})\mathrm{d}\tilde{y} = -\psi(t,\tilde{x}),$$

for every $t > 0$, $\tilde{x}$ in $\mathbb{R}^{d-1}$ and any smooth function with compact support ψ. Note that the kernel $(t, \tilde{x}) \mapsto \partial_i \Gamma_d(tq, \tilde{x}, 0)$ is not integrable for any $i \neq d$, but the cancellation property

$$\int_{\mathbb{R}^{d-1}} \partial_i \Gamma_d(tq, \tilde{x}, x_d) \mathrm{d}\tilde{x} = 0, \quad \forall t > 0,\, x_d \geq 0,\, i \neq d,$$

and the normalization property

$$\int_0^\infty \mathrm{d}t \int_{\mathbb{R}^d_+} q_{dd} \partial_d \Gamma_d(tq, \tilde{x}, x_d) \mathrm{d}\tilde{x} = 1, \quad \forall x_d > 0,$$

hold true. Moreover, one can replace the constant matrix q with a *parametrix* one, i.e., with a bounded Hölder continuous matrix-valued function $q = q(s, \tilde{y})$, and the above boundary-layer (or initial condition) limit remains true. Furthermore, a domain with a smooth boundary (e.g., having a Hölder continuous normal direction) $\mathcal{O}$ can be used instead of the simple half-space $\mathbb{R}^d_+$, where the boundary is flat with a constant normal direction.

7.1.2 Degenerate Equations

To simplify the notation for the heat-kernel (7.9), we use $\tilde{\Gamma}_0(t, \tilde{x}) = \Gamma_{d-1}(t, \tilde{x})$ and $\Gamma_0(t, x) = \Gamma_0(t, \tilde{x}, x_d) = \Gamma_d(t, x)$, for any $t > 0$ and $x = (\tilde{x}, x_d)$ in $\mathbb{R}^d$.

For a *degenerate* second-order differential operator B_0 given by (7.2), we are interested in the following two problems in the (open) half-space $\dot{\mathbb{R}}^d_+ = \mathbb{R}^{d-1} \times (0, \infty)$, without any boundary condition at $\partial \mathbb{R}^d_+$, both with a representation formula, and where the maximum principle ensures the uniqueness, at least for smooth data rapidly decreasing as $|x| \to \infty$. For $\rho > 0$, a parabolic problem in $[0, \infty) \times \dot{\mathbb{R}}^d_+$ with initial condition

$$\begin{cases} B_0 u(t, x) + b_d \partial_d u(t, x) - \rho \partial_t u(t, x) + v(t, x) = 0, \quad \forall t > 0,\, x \in \dot{\mathbb{R}}^d_+, \\ u(0, x) = u_0(x), \quad \forall x \in \mathbb{R}^d_+, \end{cases} \tag{7.11}$$

with

$$\begin{cases} u(t, x) = e^{-b_0(t/\rho)} \displaystyle\int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(\frac{t}{\rho}, \tilde{z}) u_0(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}\frac{t}{\rho}, x_d + b_d \frac{t}{\rho}) \mathrm{d}\tilde{z} + \\ \quad + \displaystyle\int_0^{t/\rho} e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(t - \rho r, \tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z}. \end{cases} \tag{7.12}$$

For $\rho = 0$ and $b_d > 0$, a parabolic problem in $\mathbb{R}^{d-1} \times [0, \infty)$ with terminal condition

$$\begin{cases} B_0 u(\tilde{x}, x_d) + b_d \partial_d u(\tilde{x}, x_d) + v(\tilde{x}, x_d) = 0, \quad \forall x_d > 0,\, \tilde{x} \in \mathbb{R}^{d-1}, \\ \lim_{x_d \to \infty} u(\tilde{x}, x_d) = 0, \quad \forall \tilde{x} \in \mathbb{R}^{d-1}, \end{cases} \tag{7.13}$$

with

$$u(\tilde{x}, x_d) = \int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z}. \tag{7.14}$$

Clearly, if $\tilde{\varsigma}$ vanishes then the above representations are simplified, without the used of the fundamental solution $\tilde{\Gamma}_0 = \Gamma_{d-1}$ as in (7.9).

To verify the representation or inversion formula (7.14), we check

$$\sum_{i=1}^{d} b_i \partial_i \Big[\int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z} \Big] =$$
$$= \int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) \partial_r v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z}$$

and

$$\tfrac{1}{2} \sum_{i,j=1}^{d-1} b_{ij} \partial_{ij} \Big[\int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z} \Big] =$$
$$= \int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) \tfrac{1}{2} \Delta_{\tilde{z}} v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z} =$$
$$= \lim_{\varepsilon \downarrow 0} \int_\varepsilon^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \big[\partial_r \tilde{\Gamma}_0(r, \tilde{z}) \big] v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z}$$

This yields

$$B_0 \Big[\int_0^\infty e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z} \Big] =$$
$$= \lim_{\varepsilon \downarrow 0} \int_\varepsilon^\infty \partial_r \Big[e^{-b_0 r} \mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r, \tilde{z}) v(\tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r) \mathrm{d}\tilde{z} \Big],$$

which prove (7.13). Similarly, we show the validity of the other representation or inversion formula (7.12).

7.1.3 Poisson and Green Functions

Let us assume $A = \frac{1}{2}\Delta$ and introduce the (parabolic) Green function and the Poisson for the Dirichlet problem in the half-space $\dot{\mathbb{R}}^d_+$,

$$\begin{cases} G_D(t, x, y) = \Gamma_d(t, \tilde{x} - \tilde{y}, x_d - y_d) - \Gamma_d(t, \tilde{x} - \tilde{y}, x_d + y_d) = \\ \qquad\qquad = \Gamma_{d-1}(t, \tilde{x} - \tilde{y}) \big[\Gamma_1(t, x_d - y_d) - \Gamma_1(t, x_d + y_d) \big], \end{cases} \tag{7.15}$$

and

$$P_D(t, x, \tilde{y}) = -\partial_d \Gamma_d(t, \tilde{x} - \tilde{y}, x_d), \tag{7.16}$$

where $\Gamma_d(t, \tilde{x}, x_d) = \Gamma_d(t, x)$ for any $x = (\tilde{x}, x_d)$ is given by (7.9). Recall that we write $\Gamma_0(t, x) = \Gamma_0(t, \tilde{x}, x_d) = \Gamma_d(t, \tilde{x}, x_d)$.

Without giving all details, let us mention that the solution of the heat-equation in the half-space with a Dirichlet boundary condition

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\Delta u(t,x) + f(t,x), \quad \forall t>0,\, x\in \dot{\mathbb{R}}^d_+, \\ u(t,x) = g(t,x), \quad \forall t>0,\, x\in \partial\dot{\mathbb{R}}^d_+, \\ u(0,x) = u_0(x), \quad \forall x\in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.17}$$

is given by the expression

$$\begin{cases} u(t,x) = \displaystyle\int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_D(t-s,x,\tilde{y})g(s,\tilde{y})\mathrm{d}\tilde{y} + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_+} G_D(t,x,y)u_0(y)\mathrm{d}y + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d_+} G_D(t-s,x,y)f(s,y)\mathrm{d}y, \end{cases}$$

for any sufficiently smooth data f, g and u_0. Here, we identify the boundary $\partial\dot{\mathbb{R}}^d_+$ with the $(d-1)$-dimensional space $\mathbb{R}^{d-1}$, so that $g(t,x)$ with x in $\partial\dot{\mathbb{R}}^d_+$ can be written as $g(t,\tilde{x})$ with $\tilde{x}$ in $\mathbb{R}^{d-1}$.

The Green function G_D, considered as a distribution in $(0,\infty)\times\dot{\mathbb{R}}^d_+$, satisfies (7.17) with $f(t,x)=\delta(t,x)$, $g=0$ and $u_0=0$, while as a distribution in $\dot{\mathbb{R}}^d_+$, it satisfies (7.17) with $f(t,x)=0$, $g=0$ and $u_0(x)=\delta(x)$. On the other hand, the Poisson function P_D satisfies (7.17) with $f(t,x)=0$, $g(t,\tilde{x})=\delta(t,\tilde{x})$ and $u_0=0$, considered as a distribution in $\mathbb{R}^d_+$.

• *Remark* 7.1. Another typical case is the Green function G_N and the Poisson function P_N with Neumann boundary conditions, i.e.,

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\Delta u(t,x) + f(t,x), \quad \forall t>0,\, x\in \dot{\mathbb{R}}^d_+, \\ \partial_d u(t,x) + g(t,x) = 0, \quad \forall t>0,\, x\in \partial\dot{\mathbb{R}}^d_+, \\ u(0,x) = u_0(x), \quad \forall x\in \dot{\mathbb{R}}^d_+. \end{cases}$$

It is known that,

$$\begin{aligned} G_N(t,x,y) &= \Gamma_d(t,\tilde{x}-\tilde{y},x_d-y_d) + \Gamma_d(t,\tilde{x}-\tilde{y},x_d+y_d), \\ P_N(t,x,\tilde{y}) &= \Gamma_d(t,\tilde{x}-\tilde{y},x_d), \end{aligned}$$

This is discussed as a particular case of what follows. Note the relations $P_D = -\partial_d P_N$ and

$$\begin{aligned} P_D(t,\tilde{x}-\tilde{y},y_d) &= \partial^x_d G_D(t,\tilde{x},0,\tilde{y},y_d), \\ P_D(t,\tilde{x}-\tilde{y},x_d) &= -\partial^y_d G_D(t,x,\tilde{y},0), \end{aligned}$$

where ∂^x_d or ∂^y_d means partial derivatives with respect to the variable x_d or y_d, respectively. □

To solve the heat-equation in the half-space with a Wentzell type boundary condition, i.e.,

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\Delta u(t,x) + f(t,x), \quad \forall t>0,\, x\in \dot{\mathbb{R}}^d_+, \\ Bu(t,x) + g(t,x) = 0, \quad \forall t>0,\, x\in \partial\dot{\mathbb{R}}^d_+, \\ u(0,x) = u_0(x), \quad \forall x\in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.18}$$

with B given by (7.2), we may proceed as follows. If $f = 0$ then $\partial_t u = \frac{1}{2}\Delta u$ and the boundary condition $Bu + g = 0$ is equivalent to the degenerate parabolic equations discussed in the previous subsection, (7.11) and (7.13). Thus, the corresponding Poisson function P_B is obtained by using the representations or inversion formulae (7.12) and (7.14) with P_D, i.e., $P_B(t, x, \tilde{y}) = Q_B(t, \tilde{x} - \tilde{y}, x_d)$, where Q_B is given as follows, for $\rho > 0$,

$$\left\{ \begin{array}{l} Q_B(t,x) = -\partial_d \displaystyle\int_0^{t/\rho} e^{-b_0 r}\mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\times \\ \qquad\qquad \times\Gamma_0(t-\rho r, \tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r)\mathrm{d}\tilde{z}, \end{array} \right. \tag{7.19}$$

and for $\rho = 0$ and $b_d > 0$,

$$\left\{ \begin{array}{l} Q_B(t,x) = -\partial_d \displaystyle\int_0^{\infty} e^{-b_0 r}\mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\times \\ \qquad\qquad \times\Gamma_0(t, \tilde{x} - \tilde{\varsigma}\tilde{z} + \tilde{b}r, x_d + b_d r)\mathrm{d}\tilde{z}. \end{array} \right. \tag{7.20}$$

Note that the variable t is a parameter in the expressions (7.20), and that the partial derivative ∂_d can be calculated inside or outside the integral signs.

• *Remark* 7.2. It is clear that the above integrals defining Q_B are non-singular for $t > 0$ and $x_d > 0$, and that upper estimates of the heat-kernel type (7.8) are necessary to make the above formula workable. Later, we are going to verify these estimates only in particular cases, with explicit calculations. □

To find the expression of the Green function G_B, first we remark that if u is a solution of the Wentzell type boundary condition problem (7.18) with $f = 0$, $g = 0$ and u_0, then Bu is a solution of the Dirichlet problem (7.17) with $f = 0$, $g = 0$ and u_0 replaced by Bu_0. Also we note that

$$\left\{ \begin{array}{l} \displaystyle\int_{\mathbb{R}^d_+} G_D(t,x,y) B_0^y u_0(y)\mathrm{d}y = \int_{\mathbb{R}^d_+} B_0^x G_D(t,x,y) u_0(y)\mathrm{d}y + \\ \qquad\qquad +2b_d \displaystyle\int_{\mathbb{R}^d_+} \partial_d \Gamma_0(t, \tilde{x} - \tilde{y}, x_d + y_d) u_0(y)\mathrm{d}y. \end{array} \right.$$

and

$$\left\{ \begin{array}{l} \displaystyle\int_{\mathbb{R}^d_+} G_D(t,x,y) \Delta^y u_0(y)\mathrm{d}y = \int_{\mathbb{R}^d_+} \Delta^x G_D(t,x,y) u_0(y)\mathrm{d}y - \\ \qquad\qquad -2 \displaystyle\int_{\mathbb{R}^d_+} \partial_d^x \Gamma_0(t, \tilde{x} - \tilde{y}, x_d) u_0(\tilde{y}, 0)\mathrm{d}\tilde{y}, \end{array} \right.$$

i.e., we have

$$\left\{ \begin{array}{l} \displaystyle\int_{\mathbb{R}^d_+} G_D(t,x,y) B^y u_0(y)\mathrm{d}y = 2b_d \int_{\mathbb{R}^d_+} \partial_d \Gamma_d(t, \tilde{x} - \tilde{y}, x_d + y_d) u_0(y)\mathrm{d}y + \\ \quad +\rho \displaystyle\int_{\mathbb{R}^d_+} \partial_d \Gamma_d(t, \tilde{x} - \tilde{y}, x_d) u_0(\tilde{y}, 0)\mathrm{d}\tilde{y} - \int_{\mathbb{R}^d_+} B^x G_D(t,x,y) u_0(y)\mathrm{d}y. \end{array} \right.$$

Now the inversion formulae and the uniqueness (e.g., for $v = B^x G_D$ we get $u = G_D$ from the representation) yield

$$\begin{cases} G_B(t,x,y) = G_D(t,x,y) + 2b_d Q_B(t,\tilde{x}-\tilde{y},x_d+y_d) + \\ \qquad\qquad +\rho\delta_0(y_d)Q_B(t,\tilde{x}-\tilde{y},x_d), \end{cases}$$

where Q_B is given by (7.19) or (7.20) according to the various cases, and δ_0 is the delta measure in the variable y_d.

The expression for the kernel Q_B can be simplify. Indeed, first recall that $b_{ij} = \sum_{k=1}^{d-1} \varrho_{ik}\beta_k\varrho_{jk}$ and $\varsigma_{ij} = \varrho_{ij}\sqrt{\beta_j}$, i.e., $\tilde{\varsigma} = \tilde{\varrho}(\tilde{\beta})^{1/2}$, the diagonal matrix $\tilde{\beta}$, with entries $\beta_i \geq 0$, $i = 1,\dots,d-1$ (in the diagonal) are the eigenvalues of the matrix (b_{ij}), and the orthogonal matrix $\tilde{\varrho}$ satisfied $|\det(\tilde{\varrho})| = 1$ and $|\tilde{\varrho}\tilde{x}| = |\tilde{x}|$, for every $\tilde{x}$ in $\mathbb{R}^{d-1}$. Since $\tilde{\Gamma}_0(r,\tilde{\varrho}\tilde{z}) = \tilde{\Gamma}_0(r,\tilde{z})$, we deduce

$$\int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\tilde{\Gamma}_0(t,\tilde{x}-\tilde{\varsigma}\tilde{z})\mathrm{d}\tilde{z} = \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\tilde{\Gamma}_0(t,\tilde{\varrho}^*\tilde{x}-\tilde{\beta}^{1/2}\tilde{z})\mathrm{d}\tilde{z}.$$

Next, after the individual change of variables $y_i = z_i\sqrt{\beta_i}$ only if $\beta_i > 0$, remarking that $\Gamma_1(r, y_i\sqrt{\beta_i}) = \sqrt{\beta_i}\Gamma_1(\beta_i r, y_i)$ and

$$\int_{\mathbb{R}^d} \Gamma_d(s,y)\Gamma_d(t,x-y)\mathrm{d}y = \Gamma_d(s+t,x), \quad \forall t,\ s>0,\ x\in\mathbb{R}^d,\ d=1,2\dots,$$

we get

$$\begin{aligned} \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\tilde{\Gamma}_0(t,\tilde{x}-\tilde{\varsigma}\tilde{z})\mathrm{d}\tilde{z} &= \int_{\mathbb{R}^{d-1}} \Big(\prod_{i=1}^{d-1}\Gamma_1(\beta_i r,y_i)\Big)\tilde{\Gamma}_0(t,\tilde{\varrho}^*\tilde{x}-\tilde{y})\mathrm{d}\tilde{y} = \\ &= \prod_{i=1}^{d-1}\Gamma_1(t+\beta_i r,(\tilde{\varrho}^*x)_i) = \det(t\tilde{\mathbb{1}}+r\tilde{\beta})^{-1/2}\tilde{\Gamma}_0\big(1,(t\tilde{\mathbb{1}}+r\tilde{\beta})^{-1/2}\tilde{\varrho}^*x\big), \end{aligned}$$

where $\tilde{\mathbb{1}}$ denotes the identity matrix of dimension $d-1$. Alternatively, by means of the Fourier transform we can check that the convolution is indeed a centered normal distribution with (invertible) covariance matrix $(t\mathbb{1}+r\tilde{\varsigma}\tilde{\varsigma}^*)$, i.e.,

$$\begin{aligned} \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\tilde{\Gamma}_0(t,\tilde{x}-\tilde{\varsigma}\tilde{z})\mathrm{d}\tilde{z} &= \Gamma_{d-1}(t\tilde{\mathbb{1}}+r\tilde{\varsigma}\tilde{\varsigma}^*,x) = \\ &= (2\pi)^{-d/2}\big[\det(t\tilde{\mathbb{1}}+r\tilde{\varsigma}\tilde{\varsigma}^*)\big]^{-1/2}\exp\big[-|(t\tilde{\mathbb{1}}+r\tilde{\varsigma}\tilde{\varsigma}^*)^{-1/2}x|^2\big], \end{aligned}$$

with the notation (7.10), which agrees with the previous expression.

This is to say that the formulae for Q_B can be reduced, without the integral in $\mathbb{R}^{d-1}$, i.e, for $\rho > 0$,

$$Q_B(t,x) = -\partial_d \int_0^{t/\rho} e^{-b_0 r}\Gamma_d\big((t-\rho r)\mathbb{1}+r\underline{\underline{b}},x+br\big)\mathrm{d}r, \tag{7.21}$$

and for $\rho = 0$ and $b_d > 0$,

$$Q_B(t,x) = -\partial_d \int_0^{\infty} e^{-b_0 r}\Gamma_d\big(t\mathbb{1}+r\underline{\underline{b}},x+br\big)\mathrm{d}r, \tag{7.22}$$

where $\underline{\underline{b}}$ is the matrix (b_{ij}) enlarged by zeros to be a square d-dimensional matrix, and the notation (7.10) is used. Clearly, the expression (7.21) becomes (7.22) as ρ approaches zero. Thus, formula (7.21) represents all cases, with the convention $t/\rho = \infty$ if $\rho = 0$.

• *Remark* 7.3. To summarize, we have shown that

$$\begin{cases} P_B(t,x,\tilde{y}) = Q_B(t,\tilde{x}-\tilde{y},x_d), \\ G_B(t,x,y) = \big[\Gamma_d(t,\tilde{x}-\tilde{y},x_d-y_d) - \Gamma_d(t,\tilde{x}-\tilde{y},x_d+y_d)\big] + \\ \qquad\qquad + 2b_d Q_B(t,\tilde{x}-\tilde{y},x_d+y_d) + \rho\delta_0(y_d)Q_B(t,\tilde{x}-\tilde{y},x_d), \end{cases}$$

are the Poisson and Green functions corresponding to the heat-equation with a Wentzell type boundary condition (7.18), where the kernel Q_B is given by the formula (7.21) and satisfies

$$BQ_B(t,x) = \partial_d \Gamma_d(t,x), \quad \forall t>0,\, x\in\mathbb{R}^d_+.$$

Clearly, these equalities prove the heat-kernel type estimates for P_B and G_B, knowing the validity of (7.8) for the kernel Q_B. Moreover, even if B may contain second-order derivative the expression BQ_B satisfies heat-kernel type estimates as $\partial_d\Gamma_d$, i.e., a singularity comparable to first-order derivatives. □

Elastic Case

In the elastic case, i.e., for $\rho = 0$, $\varsigma = 0$ and $b_d > 0$ we can compute the above integral, by first calculating

$$\int_0^\infty e^{-b_0 r}\Gamma_0(t,x+b\,r)\,\mathrm{d}r = (2\pi t)^{-d/2}\int_0^\infty e^{-b_0 r}e^{-\frac{|x+br|^2}{2t}}\,\mathrm{d}r.$$

Indeed, by observing that

$$|x+br|^2 = |x|^2 + 2(x\cdot b)r + |b|^2 r^2 = \Big[|b|r + \frac{(x\cdot b)}{|b|}\Big]^2 + |x|^2 - \frac{(x\cdot b)^2}{|b|^2}$$

we have

$$\int_0^\infty e^{-b_0 r}\Gamma_0(t,x+br)\,\mathrm{d}r = \\ = (2\pi t)^{-d/2} e^{\frac{(x\cdot b+b_0 t)^2 - |b|^2|x|^2}{2|b|^2 t}} \int_0^\infty e^{-\frac{1}{2t}\left(|b|r+\frac{x\cdot b+b_0 t}{|b|}\right)^2}\,\mathrm{d}r.$$

Hence, the following natural change of variables

$$\rho = \frac{1}{\sqrt{2t}}\Big(|b|r + \frac{x\cdot b + b_0 t}{|b|}\Big),$$

yields

$$\int_0^\infty e^{-b_0 r}\Gamma_0(t,x+br)\,\mathrm{d}r = \\ = (2\pi t)^{-d/2} e^{\frac{[x\cdot b+b_0 t]^2 - |b|^2|x|^2}{2|b|^2 t}} \frac{\sqrt{2t}}{|b|}\int_{\frac{x\cdot b+b_0 t}{\sqrt{2t}|b|}}^\infty e^{-\rho^2}\,\mathrm{d}\rho.$$

Now, take the derivative of this expression with respect to x_d to get

$$Q_e(t,x) = \frac{1}{|b|^2}\,\Gamma_0(t,x)\,\Big\{b_d + \\ +\sqrt{2}\,\frac{|b|^2 x_d - b_d[b\cdot x + b_0 t]}{|b|\sqrt{t}}\,e^{\frac{[b\cdot x+b_0 t]^2}{2t|b|^2}}\int_{\frac{b\cdot x+b_0 t}{\sqrt{2t}|b|}}^{\infty} e^{-z^2}\,\mathrm{d}z\Big\},$$

or

$$\begin{cases} Q_e(t,x) = |b|^{-2}\,\Gamma_0(t,x)\,\Big\{b_d + \sqrt{\pi}\,\dfrac{|b|^2 x_d - b_d[b\cdot x + b_0 t]}{|b|\sqrt{2t}}\,\times \\ \qquad\qquad \times \exp\Big(\dfrac{[b\cdot x + b_0 t]^2}{2t|b|^2}\Big)\mathrm{Erfc}\Big(\dfrac{b\cdot x + b_0 t}{\sqrt{2t}|b|}\Big)\Big\}. \end{cases} \tag{7.23}$$

using the complementary error function $\mathrm{Erfc}(\cdot)$. In particular, for the Neumann problem, i.e., $b_0 = 0$, $b_d = 1$ and $\tilde{b} = 0$, we found $Q_N = \Gamma_0$ which yields the well know formulae

$$P_N(t,x,\tilde{y}) = \Gamma_d(t,\tilde{x}-\tilde{y},x_d),$$
$$G_N(t,x,y) = \Gamma_d(t,\tilde{x}-\tilde{y},x_d-y_d) + \Gamma_d(t,\tilde{x}-\tilde{y},x_d+y_d),$$

as expected.

The above explicit formula allow a simple verification of the lower and upper heat-kernel estimates for the elastic case, i.e., by means of the bounds

$$\begin{cases} \dfrac{2}{r+\sqrt{r^2+2}} \le \sqrt{\pi}\,e^{r^2}\,\mathrm{Erfc}(r) \le \dfrac{2}{r+\sqrt{r^2+1}}, \quad \forall r \ge 0, \\ 2 = \mathrm{Erfc}(-\infty) < \mathrm{Erfc}(r) \le \mathrm{Erfc}(0) = 1, \quad \forall r \le 0, \end{cases}$$

we can estimate the expression

$$R_e(t,x) = b_d + \sqrt{\pi}\,\frac{|b|^2 x_d - b_d[b\cdot x + b_0 t]}{|b|\sqrt{2t}}\,e^{\frac{[b\cdot x+b_0 t]^2}{2t|b|^2}}\mathrm{Erfc}\Big(\frac{b\cdot x + b_0 t}{\sqrt{2t}|b|}\Big)$$

appearing in the definition (7.23) of the kernel Q_e. Indeed, since

$$|b|^2 x_d - b_d[b\cdot x + b_0 t] = |\tilde{b}|^2 x_d - b_d(\tilde{b}\cdot\tilde{x} + b_0 t) \le |b|\,|\tilde{b}|\,|x|, \quad \forall t \ge 0,\ x \in \mathbb{R}^d_+,$$

if $b\cdot x + b_0 t \le 0$ we have $\tilde{b}\cdot\tilde{x} \le -b_d x_d - b_0 t$,

$$[b\cdot x + b_0 t]^2 = (\tilde{b}\cdot\tilde{x})^2 + 2(\tilde{b}\cdot\tilde{x})(b_d x_d + b_0 t) + (b_d x_d + b_0 t)^2 \le |\tilde{b}|^2|\tilde{x}|^2 - (b_d x_d + b_0 t)^2$$

and

$$\frac{[b\cdot x + b_0 t]^2}{2|b|^2 t} \le \frac{|\tilde{b}|^2|\tilde{x}|^2 - (b_d x_d)^2}{2|b|^2 t} = \frac{|x|^2}{2t} - \frac{b_d^2|x|^2}{2|b|^2 t}$$

then

$$b_d \le R_e(t,x) \le b_d + \sqrt{2\pi}\,|\tilde{b}|\,\frac{|x|}{\sqrt{t}}\,\exp\Big(\frac{|x|^2}{2t} - \frac{b_d^2|x|^2}{2|b|^2 t}\Big), \tag{7.24}$$

while if $b \cdot x + b_0 t \geq 0$ then the calculations are longer. Begin with

$$R_e(t,x) \leq \frac{2|b|^2 x_d + b_d\big[\sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t} - (b\cdot x + b_0 t)\big]}{(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t}},$$

and use

$$\begin{aligned}
&\big[\sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t} - (b\cdot x + b_0 t)\big] = \\
&\quad = \frac{\big[(b\cdot x + b_0 t)^2 + |b|^2 4t - (b\cdot x + ct)^2\big]}{\big[\sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t} + (b\cdot x + b_0 t)\big]} = \\
&\quad = \frac{|b|^2 4t}{\sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t} + (b\cdot x + b_0 t)},
\end{aligned}$$

to get

$$R_e(t,x) \leq \frac{2|b|^2 x_d\big[(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t}\big] + b_d |b|^2 4t}{\big[(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t}\big]^2}.$$

Similarly, we obtain

$$R_e(t,x) \geq \frac{2|b|^2 x_d\big[(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 2t}\big] + b_d |b|^2 2t}{\big[(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 2t}\big]^2}.$$

This shows that for

$$r = \frac{(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 4t}}{\sqrt{t}} \quad \text{and} \quad \rho = \frac{x_d}{\sqrt{t}}$$

we have

$$\begin{aligned}
&R_e(t,x) \leq \max\big\{b_d,\, 2|b|^2(\rho r^{-1} + 2b_d r^{-2})\big\}, \\
&2|b| \leq r \leq 2|b| + 2\frac{|b|^2}{b_d}\rho, \quad \rho \geq 0,
\end{aligned}$$

i.e., if $b \cdot x + b_0 t \geq 0$ then

$$R_e(t,x) \leq b_d + \frac{x_d}{\sqrt{t}}, \quad \forall t > 0,\, x \in \mathbb{R}^d_+. \tag{7.25}$$

A lower bound is found similarly, namely, for

$$r = \frac{(b\cdot x + b_0 t) + \sqrt{(b\cdot x + b_0 t)^2 + |b|^2 2t}}{\sqrt{t}} \quad \text{and} \quad \rho = \frac{x_d}{\sqrt{t}}$$

we have

$$\begin{aligned}
&R_e(t,x) \geq \min\big\{b_d,\, 2|b|^2(\rho r^{-1} + b_d r^{-2})\big\}, \\
&2|b| \leq r, \quad r^2 \leq 4\Big(\frac{[b\cdot x + b_0 t]^2}{t} + |b|^2\Big), \quad \rho \geq 0,
\end{aligned}$$

i.e., if $b \cdot x + b_0 t \geq 0$ then

$$R_e(t,x) \geq \frac{b_d}{2}\Big(\frac{[b\cdot x + b_0 t]^2}{|b|^2 t} + 1\Big)^{-1}, \quad \forall t>0,\, x\in\mathbb{R}^d_+. \tag{7.26}$$

These estimates (7.24), (7.25) and (7.26) yield upper and lower estimates on Q_e, i.e., for every $c_0 > 0$ such that $b_d^2 c_0 < |b|$ there exists a constant $C_0 > 0$ (depending only on c_0, $|b|$ and $b_d > 0$) satisfying

$$Q_e(t,x) \leq C_0 t^{-d/2} \exp\Big(-c_0\frac{|x|^2}{2t}\Big), \quad \forall t>0,\, x\in\mathbb{R}^d_+, \tag{7.27}$$

and for any $c_1 > 1$ there exists a constant $C_1 > 0$ (depending only on c_1, $|b|$ and $b_d > 0$) satisfying

$$Q_e(t,x) \geq C_1(1+b_0 t)^{-1} t^{-d/2} \exp\Big(-c_1\frac{|x|^2}{2t}\Big), \quad \forall t>0,\, x\in\mathbb{R}^d_+. \tag{7.28}$$

Moreover, upper bound estimate can also be found for all derivatives of Q_e as in (7.8).

The elastic case for $b_b \to 0$ and $\tilde\varsigma = 0$ yields

$$Q_e(t,x) = \Gamma_0(t,x)\sqrt{\pi}\,\frac{x_d}{\sqrt{2t|\tilde b|}}\,\exp\Big(\frac{[\tilde b\cdot\tilde x + b_0 t]^2}{2t|\tilde b|^2}\Big)\mathrm{Erfc}\Big(\frac{\tilde b\cdot\tilde x + b_0 t}{\sqrt{2t|\tilde b|}}\Big).$$

However, we see that as $\tilde b\cdot\tilde x \to -\infty$, the heat-kernel type estimate is lost.

Sticky Case

In the sticky case, i.e., if $\rho > 0$, but $b_0 = 0$ and $\varsigma = 0$ then

$$\begin{aligned} Q_s(t,x) &= -\partial_d \int_0^{t/\rho} \Gamma_0\big(t-\rho r, x+br\big)\mathrm{d}r = \\ &= -(2\pi)^{-d/2}\,\partial_d \int_0^{t/\rho} (t-\rho r)^{-d/2} \mathrm{e}^{-\frac{|x-br|^2}{2(t-\rho r)}}\,\mathrm{d}r. \end{aligned}$$

This integral can be computed using the complementary error function $\mathrm{Erfc}(\cdot)$ for any odd dimension d. Otherwise, the function $\Phi_\nu(t,x)$ defined by the integral

$$\Phi_\nu(t,x) = \frac{1}{2^\nu}\int_0^t \frac{1}{s^{\nu+1}}\,\mathrm{e}^{-(s+\frac{x^2}{4s})}\,\mathrm{d}s, \qquad \forall t,x,\nu>0$$

should be used. Observe that

$$\Phi_{\nu+1}(t,x) = -\frac{1}{x}\frac{\partial}{\partial x}\Phi_\nu(t,x), \qquad \forall t,x,\nu>0.$$

and that for $t\to\infty$ the function reduces to the so-called modified Bessel functions of second kind defined by

$$K_\nu(x) = \frac{1}{2}\Big(\frac{x}{2}\Big)^\nu \int_0^\infty \frac{1}{s^{\nu+1}}\,\mathrm{e}^{-(s+\frac{x^2}{4s})}\,\mathrm{d}s, \qquad x,\nu>0,$$

which are also called Kelvin or MacDonald functions.

We have

$$
\begin{aligned}
Q_s(t,x) = -\frac{1}{\rho}\frac{1}{(2\pi)^{d/2}}\,\exp\Big[(x+b\tfrac{t}{\rho})\cdot(b\frac{1}{\rho})\Big]\times \\
\times\Big[b_d\,|b|^{d-2}\frac{1}{\rho^{d-1}}\,\Phi_{\frac{d}{2}-1}\Big(\frac{|b|^2t}{2\rho^2},\frac{1}{\rho}|b||x+b\,\frac{t}{\rho}|\Big)- \\
-(x_d+b_d\frac{t}{\rho})\,|b|^d\,\frac{1}{\rho^d}\,\Phi_{\frac{d}{2}}\Big(\frac{|b|^2t}{2\rho^2},\frac{1}{\rho}|b||x+b\,\frac{t}{\rho}|\Big)\Big],
\end{aligned}
\tag{7.29}
$$

where we remark the homogeneity in b/ρ as expected.

Clearly, dimension $d=1$ corresponds to $\nu=-1/2$. In this case, we can calculate

$$
\begin{cases}
\displaystyle\int_0^t \mathrm{e}^{-(s+\frac{x^2}{4s})}\,\frac{\mathrm{d}s}{\sqrt{s}} = \mathrm{e}^{-x}\int_{\frac{x-2\,t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z - \mathrm{e}^{x}\int_{\frac{x+2\,t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z = \\
\qquad = \dfrac{\sqrt{\pi}}{2}\,\mathrm{e}^{-x}\,\mathrm{Erfc}(\dfrac{x-2\,t}{2\sqrt{t}}) - \dfrac{\sqrt{\pi}}{2}\,\mathrm{e}^{x}\,\mathrm{Erfc}(\dfrac{x+2\,t}{2\sqrt{t}}).
\end{cases}
\tag{7.30}
$$

In particular, for $t\to\infty$, one gets

$$
\int_0^{\infty}\mathrm{e}^{-(s+\frac{x^2}{4s})}\,\frac{\mathrm{d}s}{\sqrt{s}} = \mathrm{e}^{-x}\,\sqrt{\pi}.
$$

Indeed, performing the substitution $s=r^2$ one get

$$
\int_0^{t}\mathrm{e}^{-(s+\frac{x^2}{4s})}\,\frac{\mathrm{d}s}{\sqrt{s}} = 2\int_0^{\sqrt{t}}\mathrm{e}^{-(r^2+\frac{x^2}{4r^2})}\,\mathrm{d}r.
$$

Now observe that

$$
r^2+\frac{x^2}{4r^2} = \Big(r-\frac{x}{2r}\Big)^2 + x,
$$

and that the invertible substitution $\rho = r-\frac{x}{2r}$ yields

$$
r=\tfrac{1}{2}\,(\rho+\sqrt{\rho^2+2x}) \quad\text{ and }\quad 2\mathrm{d}r = \Big(1+\frac{\rho}{\sqrt{\rho^2+2x}}\Big)\,\mathrm{d}\rho.
$$

Hence

$$
\int_0^{t}\mathrm{e}^{-(s+\frac{x^2}{4s})}\,\frac{\mathrm{d}s}{\sqrt{s}} = \mathrm{e}^{-x}\int_{-\infty}^{\frac{2t-x}{2\sqrt{t}}}\mathrm{e}^{-\rho^2}\,\Big(1+\frac{\rho}{\sqrt{\rho^2+2x}}\Big)\,\mathrm{d}\rho,
$$

which can be written as (7.30), after remarking that the integration with respect the measure $\big(\rho/\sqrt{\rho^2+2x}\big)\mathrm{d}\rho$ is zero on symmetric intervals about zero.

Thus, for example we have

$$
\Phi_{-\frac{1}{2}}(t,x) = \sqrt{2}\,\Big(\mathrm{e}^{-x}\int_{\frac{x-2t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z - \mathrm{e}^{x}\int_{\frac{x+2t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z\Big)
$$

and

$$\Phi_{\frac12}(t,x) = -\frac{1}{x}\,\frac{\partial}{\partial x}\Phi_{-\frac12}(t,x) = \frac{\sqrt{2}}{x}\,\Big(\mathrm{e}^{-x}\int_{\frac{x-2t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z + \mathrm{e}^{x}\int_{\frac{x+2t}{2\sqrt{t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z\Big),$$

which gives

$$Q_s(t,x) = \tfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\frac{t+2cx}{4c^2}}\,\Big[-\frac{t+2cx}{4c^2}\,\Phi_{\frac12}\big(\tfrac{t}{8c^2},\tfrac{t+2cx}{4c^2}\big) + \Phi_{-\frac12}\big(\tfrac{t}{8c^2},\tfrac{t+2cx}{4c^2}\big)\Big],$$

and using the complementary error function $\mathrm{Erfc}(\cdot)$,

$$Q_s(t,x) = -\tfrac{2}{\sqrt{\pi}}\,\mathrm{e}^{\frac{t+cx}{2c^2}}\int_{\frac{t+cx}{c\sqrt{2t}}}^{\infty}\mathrm{e}^{-z^2}\mathrm{d}z = \exp\Big(\frac{t+cx}{2c^2}\Big)\mathrm{Erfc}\Big(\frac{t+cx}{c\sqrt{2t}}\Big),$$

for $d=1$, $b=b_d=1$ and $\rho/2=c$, see the notation of Example 2.8 in Chapter 2.

7.1.4 General Constant Coefficients

We begin with the particular case

$$A\varphi(x) = \tfrac12\Delta\varphi(x) + a_d\partial_d\varphi(x), \tag{7.31}$$

where the starting point is the following (parabolic) fundamental solution

$$\Gamma_0(t,\tilde{x},x_d-a_dt) = (2\pi t)^{-d/2}\exp\Big(-\frac{|\tilde{x}|^2+|x_d-a_dt|^2}{2t}\Big),$$

for any $t>0$ and $x=(\tilde{x},x_d)$ in $\mathbb{R}^d$. Similarly,

First, note that any solution of an homogeneous Dirichlet problem in the half-space

$$\begin{cases}\partial_t u(t,x) = \frac12\Delta u(t,x) + a_d\partial_d u(t,x) + f(t,x), \quad \forall t>0,\, x\in\dot{\mathbb{R}}^d_+,\\ u(t,\tilde{x},0) = g(t,\tilde{x}), \quad \forall t>0,\, x\in\mathbb{R}^{d+1},\\ u(0,x) = u_0(x), \quad \forall x\in\dot{\mathbb{R}}^d_+,\end{cases}$$

can be extended to a problem in the whole space by setting

$$u(t,\tilde{x},x_d) := g(t,\tilde{x}) - e^{-2a_dx_d}[u(t,\tilde{x},-x_d) - g(t,\tilde{x})] \quad \text{if} \quad x_d<0.$$

Indeed, because of the zero boundary condition, this extension produces a C^1-continuous function across the boundary which is a solution of the Cauchy problem

$$\begin{cases}\partial_t u(t,x) = \frac12\Delta u(t,x) + a_d\partial_d u(t,x) + \hat{f}(t,x), \quad \forall t>0,\, x\in\mathbb{R}^d,\\ u(0,x) = \hat{u}_0(x), \quad \forall x\in\mathbb{R}^d.\end{cases}$$

in the whole space, where $\hat{u}_0$ and $\hat{f}$ are the corresponding extensions.

This yields an explicit representation, namely,

$$u(t,x) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Gamma_0(t-s, \tilde{x}-\tilde{y}, x_d - a_d(t-s) - y_d)\hat{f}(s,y)\mathrm{d}y + \\ + \int_{\mathbb{R}^d} \Gamma_0(t, \tilde{x}-\tilde{y}, x_d - a_d t - y_d))\hat{u}_0(y)\mathrm{d}y,$$

i.e., the expression of the Green function is

$$\begin{cases} G_{a_d,D}(t,x,y) := \Gamma_0(t, \tilde{x}-\tilde{y}, x_d - a_d t - y_d) - \\ \qquad\qquad -\mathrm{e}^{2a_d y_d}\Gamma_0(t, \tilde{x}-\tilde{y}, x_d - a_d t - y_d). \end{cases} \tag{7.32}$$

after a change of variables.

To find the expression of the Poisson function for the Dirichlet problem in the half-space corresponding to the operator (7.31), we begin with the double-layer potential jump relation for Γ_0, i.e., as already mentioned,

$$\lim_{x_d \downarrow 0} \int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \big[-\partial_d \Gamma_0(t-s, \tilde{x}-\tilde{y}, x_d)\big]\psi(s,\tilde{y})\mathrm{d}\tilde{y} = \psi(t,\tilde{x}),$$

for every continuous function with a compact support ψ, for any $t > 0$ and $\tilde{x}$ in $\mathbb{R}^{d-1}$. Since

$$\frac{x_d}{t}\Gamma_0(t-s, \tilde{x}-\tilde{y}, x_d - a_d t) = e^{a_d x_d - a_d^2(t-s)/2}\big[-\partial_d \Gamma_0(t-s, \tilde{x}-\tilde{y}, x_d)\big],$$

we also have

$$\lim_{x_d \downarrow 0} \int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \big[\frac{x_d}{t-s}\Gamma_0(t-s, \tilde{x}-\tilde{y}, x_d - a_d t)\big]\psi(s,\tilde{y})\mathrm{d}\tilde{y} = \psi(t,\tilde{x}),$$

for any continuous function with a compact support ψ. Since the functions $\partial_d \Gamma_0(t, x - a_d t)$ and $a_d \Gamma_0(t, x - a_d t)$ solve the homogenous equation, the same holds true for their sum, i.e., for

$$P_{a_d,D}(t,x,\tilde{y}) := \frac{x_d}{t}\Gamma_0(t, \tilde{x}-\tilde{y}, x_d - a_d t), \tag{7.33}$$

which is then the Poisson function corresponding to (7.31).

Once the Poisson function $P_{a_d,D}$ is known, an alternative way to find the expression (7.32) of the Green function $G_{a_d,D}$ is to propose

$$G_{a_d,D}(t,x,y) = \Gamma_0(t, \tilde{x}-\tilde{y}, x_d - a_d t - y_d) - V(t, \tilde{x}-\tilde{y}, x_d, y_d)$$

and to solve the non-homogeneous Dirichlet problem in the half-space satisfied by V. This yields

$$V(t, \tilde{x}, x_d, y_d) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{a_d,D}(t-s, x, \tilde{y})\Gamma_0(s, \tilde{y}, y_d - a_d s)\mathrm{d}\tilde{y},$$

which can be explicitly computed to reproduce (7.32).

Going back to the general case (7.1), first, let us check that we can choose a change of variables $z = cx$, i.e., $z_i = \sum_j c_{ij}x_j$, such that the purely second order differential operator $\sum_{ij} a_{ij}\partial_{ij}$ in the variable x belonging to $\mathbb{R}^d_+$ becomes the Δ in the variable z also belonging to $\mathbb{R}^d_+$. Indeed, we find an orthogonal (rotation) matrix σ transforming $\sum_{ij} a_{ij}\partial^x_{ij}$ into $\sum_i \alpha_i \partial^y_{ii}$ for $y = \sigma x$, where α_i are the eigenvalues of the matrix (a_{ij}). Next, a diagonal matrix α, $\xi_i = (\alpha_i)^{-1/2} y_i$, transforms it into Δ^ξ. However, the semi-space $x_d \geq 0$ has became $\sum_{i=1}^d \sigma_{id}\sqrt{\alpha_i}\xi_i \geq 0$. Then, another orthogonal matrix γ will take back this semi-space into the half-space $z_d \geq 0$. Only the d-column of γ is determined by the condition

$$\gamma_{id} = \sigma_{id}\sqrt{\alpha_i}\Big(\sum_{j=1}^d \alpha_j \sigma^2_{jd}\Big)^{-1/2}, \quad \forall i = 1, \dots, d,$$

and the other columns are arbitrary chosen keeping γ an orthogonal matrix. Thus $c = \gamma\alpha^{-1/2}\sigma$, with

$$c_{ij} = \sum_{k=1}^d \gamma_{ik}(\alpha_k)^{-1/2}\sigma_{kj}, \quad \forall i, j = 1, \dots, d.$$

Note that

$$a_{ij} = \sum_{k=1}^d \sigma_{ki}\alpha_k\sigma_{kj}, \quad \text{and} \quad \sum_{k,\ell=1}^d c_{ik}a_{k\ell}c_{j\ell} = \delta_{ij}, \quad \forall i, j = 1, \dots, d,$$

where $\delta_{k\ell} = 1$ only when $k = \ell$ (and 0 otherwise), while the inverse matrix (a^{ij}) of (a_{ij}) is given by

$$a^{ij} = \sum_{k=1}^d \sigma_{ki}(\alpha_k)^{-1}\sigma_{kj} = \sum_{k=1}^d c_{ki}c_{kj}, \quad \forall i, j = 1, \dots, d.$$

Because $\mathbb{R}^d_+$ is invariant,

$$c = \begin{pmatrix} \tilde{c} & 0 \\ * & c_{dd} \end{pmatrix}, \quad c_{dd} = \frac{1}{\sqrt{a_{dd}}},$$

where the last row $(*, c_{dd})$ of the matrix c is $(c_{d1}, \dots, c_{dd})$. Also, one has the estimate

$$\min_i\{\lambda_i\} \leq \det(\tilde{c})\sqrt{\det(a)} \leq \max_i\{\lambda_i\},$$

with $a = (a_{ij})$ and λ_i its eigenvalues. In short, we have $cc^* = a^{-1}$ and $cac^* = 1$, and therefore $|cx|^2 = x \cdot a^{-1}x$, where $\cdot$ denotes the scalar or dot product and a^{-1} is the inverse of the matrix a.

At the same time the coefficients of the boundary differential operator of Wenzell type B, denoted by $\bar B$, become

$$\bar b_{ij} = \sum_{k,\ell=1}^{d-1} c_{ik} b_{k\ell} c_{j\ell} \quad \text{and} \quad \bar b_i = \sum_{k=1}^{d} c_{ik} b_k, \quad \forall i,j = 1,\dots,d-1,$$

and $\bar b_d = c_{dd} b_d = b_d (a_{dd})^{-1/2}$. Thus, except for the lower terms coefficients in A given by (7.1), i.e., for A_0 instead of A, we are reduced to the case studied in the precedent subsections.

Next, to incorporate the lower order coefficients a_i, for $i = 1,\dots,d-1$, we make the change of variables $x_i = z_i - a_i t$, i.e., $\partial_i^x = \partial_i^z$ and $\partial_t = \partial_t - \sum_{i=1}^{d-1} a_i \partial_i^x$. Clearly, to add the zero-order term a_0 we use the factor $e^{-a_0 t}$ in all the expression. Thus, the expressions

$$\begin{cases} G_{A,D}(t,x,y) := e^{-a_0 t}\big[\Gamma_d\big(t, c(\tilde x - \tilde a t - \tilde y, x_d - a_d t - y_d)\big) - \\ \qquad\qquad - \mathrm{e}^{2c_{dd}a_d y_d}\Gamma_d\big(t, c(\tilde x - \tilde a t - \tilde y, x_d - a_d t + y_d)\big)\big], \\ P_{A,D}(t,x,\tilde y) := e^{-a_0 t}\dfrac{c_{dd}x_d}{t}\big[\Gamma_d\big(t, c(\tilde x - \tilde a t - \tilde y, x_d - a_d t)\big)\big], \end{cases}$$

provide the Green and Poisson functions in the half-space $\mathbb{R}^d_+$ for the operator A (7.1) with Dirichlet boundary conditions. Note that $\tilde a = (a_1,\dots,a_{d-1})$ and that $\Gamma_d = \Gamma_0$ if the d-dimensional Gaussian kernel (7.9). Clearly, by using the matrix $\underline{\underline{a}} = (a_{ij})$ and the normal distribution notation (7.10), we can write

$$\begin{cases} G_{A,D}(t,x,y) := e^{-ta_0}\big[\Gamma_d\big(t\underline{\underline{a}}, \tilde x - t\underline{\tilde a} - \tilde y, x_d - ta_d - y_d\big) - \\ \qquad\qquad - \mathrm{e}^{2a_d y_d/\sqrt{a_{dd}}}\Gamma_d\big(t\underline{\underline{a}}, \tilde x - t\underline{\tilde a} - \tilde y, x_d - ta_d + y_d\big)\big], \\ P_{A,D}(t,x,\tilde y) := e^{-ta_0}\dfrac{x_d}{t\sqrt{a_{dd}}}\big[\Gamma_d\big(t\underline{\underline{a}}, \tilde x - t\underline{\tilde a} - \tilde y, x_d - ta_d\big)\big], \end{cases}$$

where the vector single-underline notation is used, $\underline{\tilde a} = (a_1,\dots,a_{d-1})$, to emphasize the difference with the matrix double-underline notation. Note that

$$F_A(t,x) = \Gamma_d\big(t\underline{\underline{a}}, \tilde x - t\underline{\tilde a}, x_d - ta_d\big)$$

is the fundamental solution corresponding to the operator A.

Finally, we may repeat the previous subsection, with $G_{A,D}$ and $P_{A,D}$ instead of G_D and P_D, or alternatively, we can make the change of variables

$$u(t,\tilde x, x_d) \mapsto v(t,\tilde\xi,\xi_d), \qquad u(t,\tilde x, x_d) := e^{-a_0 t} v\big(t, c(\tilde x - \tilde a t, x_d)\big)$$

and $G_{a_d,D}$ and $P_{a_d,D}$ (associated with $\bar B$) instead of G_D and P_D, to obtain an expression (all in term of a kernel $Q_{A,B}$ similar to Q_B) of Green and Poisson functions for corresponding to the operator (7.1) with Wentzell type boundary conditions, i.e., with the convention $t/\rho = \infty$ when $\rho = 0$, we have

$$\begin{cases} Q_{A,B}(t,x) = e^{-a_0 t}\displaystyle\int_0^{t/\rho} e^{-b_0 r}\Big(\frac{x_d + b_d r}{a_{dd}(t-\rho r)}\Big)\times \\ \qquad\qquad \times\Big(\displaystyle\prod_{i=1}^{d}\Gamma_1(\alpha_i t + \beta_i r - \alpha_i\rho r, x_i + b_i r - a_i t)\Big)\mathrm{d}r, \end{cases} \tag{7.34}$$

where $\alpha_d = a_{dd}$, $\beta_d = 0$, α_i, β_i, $i = 1, \dots, d-1$, are the eigenvalues of (a_{ij}), (b_{ij}), and Γ_1 if the Gaussian kernel (7.9) with $d = 1$. Note that the last factor in the above product is $\Gamma_1\big(a_{dd}(t-\rho r), x_d + b_d r - a_d t\big)$, and by means of the normal distribution (7.10), we have

$$\prod_{i=1}^{d-1} \Gamma_1(\alpha_i t + \beta_i r - \alpha_i \rho r, x_i + b_i r - a_i t) = \\ = \Gamma_{d-1}\big((t-\rho r)\underline{\underline{\tilde{a}}} + r\underline{\underline{\tilde{q}}}, x + r\underline{\tilde{b}} - t\underline{\tilde{a}}\big),$$

where $\underline{\tilde{a}} = (a_1, \dots, a_{d-1})$, $\underline{\tilde{b}} = (b_1, \dots, b_{d-1})$, and the $(d-1)$-dimensional square matrices $\underline{\underline{\tilde{a}}} = (a_{ij})$, $\underline{\tilde{b}} = (b_{ij})$, $\underline{\underline{\tilde{q}}} = \underline{\underline{\tilde{c}}}\,\underline{\tilde{b}}\,\underline{\underline{\tilde{c}}}^*$ or perhaps just $\underline{\underline{\tilde{q}}} = [(\underline{\underline{\tilde{a}}}\underline{\tilde{b}})(\underline{\underline{\tilde{a}}}\underline{\tilde{b}})^*]^{1/2}$, or simply $\underline{\underline{\tilde{q}}} = \underline{\tilde{b}}$, i.e., $q_{ij} = ?$

. . . exp(HELP) . . .
Is all this correct? Is this the right expression of $Q_{A,B}$? How do the coefficients a_{id} and a_{dj} affect the calculation? It is true that if β_i are the eigenvalues of (b_{ij}) then β_i/α_i are the eigenvalues of $(\bar{b}_{ij})$? What do you think?

7.2 Variable Coefficients in Half-Space

Let A be a second-order (uniformly) elliptic differential operator with bounded and Hölder continuous coefficients the open half-space $\dot{\mathbb{R}}^d_+ = \{x : x_d > 0\}$, i.e.,

$$\begin{cases} A(t)\varphi(x) := A_0(t)\varphi(x) + \sum_{i=1}^{d} a_i(t,x)\partial_i\varphi(x) - a_0(t,x)\varphi(x), \\ A_0(t)\varphi(x) := \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij}\varphi(x), \end{cases} \tag{7.35}$$

where $a_0 \geq 0$, $a_{ij} = a_{ji}$ for any i, j, and for some positive constants c_0, C_0 and any $t > 0$, x in $\mathbb{R}^d_+$ we have

$$c_0|\xi|^2 \leq \sum_{i,j=1}^{d} a_{ij}(t,x)\xi_i\xi_j \leq C_0|\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \tag{7.36}$$

and

$$a_{ij},\, a_i,\, a_0 \in C_b^{\alpha}([0,\infty) \times \mathbb{R}^d_+), \quad \forall i,\, j, \tag{7.37}$$

for every continuously differentiable functions φ with a compact support in the closed half-space $\mathbb{R}^d_+$. Also, let B be a (uniform) *Wentzell type* boundary second-

order differential operator with bounded and Hölder continuous coefficients, i.e.,

$$\begin{cases} B(t)\varphi(x) := B_0(t)\varphi(x) + b_d(t,\tilde{x})\partial_d\varphi(x) - \rho(t,\tilde{x})A(t)\varphi(x), \\ B_0(t) := \frac{1}{2}\sum_{i,j=1}^{d-1} b_{ij}(t,\tilde{x})\partial_{ij} + \sum_{i=1}^{d-1} b_i(t,\tilde{x})\partial_i - b_0(t,\tilde{x}), \end{cases} \tag{7.38}$$

where $b_{ij} = b_{ji}$ is a symmetric non-negative definite matrix, $b_0, b_d, \rho \geq 0$, and for some positive constant c_0 and any $t > 0$, $\tilde{x})$ in $\mathbb{R}^{d-1}$ we have

$$\rho(t,\tilde{x}) \geq c_0 \quad \text{or} \quad b_d(t,\tilde{x}) \geq c_0, \tag{7.39}$$

and

$$b_{ij},\, b_i,\, b_0, \rho \in C_b^\alpha([0,\infty)\times\mathbb{R}^{d-1}), \quad \forall i,\, j. \tag{7.40}$$

Note that all coefficients are trivially extended to the whole half-space. Sometimes, we may need to use the notation $A\varphi(x) = A(t)\varphi(x) = A(t,x)\varphi(x)$ and $B\varphi(x) = B(t)\varphi(x) = B(t,\tilde{x})\varphi(x)$ to emphasize the (t,x)-dependency of the coefficients.

Consider the boundary value problem

$$\begin{cases} \partial_t u(t,x) = A(t)u(t,x), \quad \forall t > t_0,\, x \in \dot{\mathbb{R}}^d_+, \\ B(t)u(t,x) + \psi(t,x) = 0, \quad \forall t > t_0,\, x \in \partial\mathbb{R}^d_+, \\ u(t_0,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.41}$$

and the representation formula

$$\begin{cases} u(t,x) = \displaystyle\int_{\mathbb{R}^d_+} G_{A,B}(t,\tilde{x},x_d,t_0,\tilde{y},y_d)\varphi(y)\mathrm{d}y + \\ \qquad\quad + \displaystyle\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t,\tilde{x},x_d,s,\tilde{y})\psi(s,\tilde{y})\mathrm{d}\tilde{y}, \end{cases} \tag{7.42}$$

where $G_{A,B}$ and $P_{A,B}$ are the Green and the Poisson functions.

If the coefficients are smooth, then it is convenient to define the formal adjoint operators

$$\begin{cases} A^*(t)\varphi(x) := A_0^*(t)\varphi(x) + \displaystyle\sum_{i=1}^d \partial_i\big(a_i^*(t,x)\varphi(x)\big) - a_0^*(t,x)\varphi(x), \\ A_0^*(t)\varphi(x) := \frac{1}{2}\displaystyle\sum_{i,j=1}^d \partial_j\big(a_{ij}(t,x)\partial_i\varphi(x)\big), \end{cases} \tag{7.43}$$

and

$$\begin{cases} B^*(t)\varphi(x) := B_0^*(t)\varphi(x) + b_d(t,\tilde{x})\partial_d\varphi(x) - \rho A^*\varphi(x), \\ B_0^*(t) := \frac{1}{2}\displaystyle\sum_{i,j=1}^{d-1} \partial_j\big(b_{ij}(t,\tilde{x})\partial_i\,\cdot\,\big) + \sum_{i=1}^{d-1} \partial_i\big(b_i^*(t,\tilde{x})\,\cdot\,\big) - b_0^*(t,\tilde{x}), \end{cases} \tag{7.44}$$

where the adjoint coefficients may be computed as follows

$$a_{ij}(t,x) = a^*_{ij}(t,x), \qquad a^*_0(t,x) = a_0(t,x) + \sum_{i=1}^{d} \partial_i a_i(t,x),$$
$$a^*_i(t,x) = -a_i(t,x) - \sum_{j=1}^{d} \partial_j a_{ij}(t,x),$$

and

$$b^*_{ij}(t,\tilde{x}) = b_{ij}(t,\tilde{x}), \qquad b^*_0(t,\tilde{x}) = b_0(t,\tilde{x}) + \sum_{i=1}^{d-1} \partial_i b_i(t,\tilde{x}),$$
$$b^*_i(t,\tilde{x}) = -b_i(t,\tilde{x}) - \sum_{j=1}^{d-1} \partial_j b_{ij}(t,\tilde{x}),$$

Remark that in the construction of the Green and Poisson functions we require $a_0(t,x) \geq 0$ and $b_0(t,\tilde{x}) \geq 0$ (among other assumptions) but not necessarily $a^*_0(t,x) \geq 0$ and $b^*_0(t,\tilde{x}) \geq 0$. Thus, the adjoint problem does not always satisfy the conditions for the direct construction, however, the above relation (7.64) shows the existence (with estimates) also for the adjoint problem.

7.2.1 Green and Poisson Representations

In this sub-section we discuss a probabilistic version (which applies to the variable coefficients case) of the arguments used to construct the Green and Poisson functions for constant coefficients. Let us begin with the explicit representation of the solution of a (possible) degenerate differential equation in the half-space associated with the boundary operator B.

The explicit expressions (7.12) and (7.14) can be written in terms of a stochastic equations. Indeed, with $\tilde{X}(t,\tilde{x},t_0,r) = \tilde{X}(r)$ consider

$$\tilde{X}(r) = \tilde{z} + \int_{t_0}^{r} \varsigma(t,\tilde{X}(r))\mathrm{d}\tilde{w}(r) + \int_{t_0}^{r} \tilde{b}(t,\tilde{X}(r))\mathrm{d}r, \tag{7.45}$$

and $X_d(t,\tilde{x},x_d,t_0,r) = X_d(r)$, $X_0(t,\tilde{z},t_0,r) = X_0(r)$ are given by

$$X_d(r) := x_d + \int_{t_0}^{r} b_d(t,\tilde{X}(r))\mathrm{d}r, \qquad X_0(r) := \int_{t_0}^{r} b_0(t,\tilde{X}(r))\mathrm{d}r, \tag{7.46}$$

valid even for variable coefficients.

First, note that by means of Itô formula, for any given u smooth function, we have

$$\mathbb{E}\big\{e^{-X_0(T)}u(t,\tilde{X}(T),X_d(T))\big\} - u(t,\tilde{x},x_d) =$$
$$= \int_{t_0}^{T} \mathbb{E}\big\{e^{-X_0(r)}\big[B_0(t,\tilde{X}(r))u(t,\tilde{X}(r),X_d(r)) +$$
$$+b_d(t,\tilde{X}(r))\partial_d u(t,\tilde{X}(r),X_d(r))\big]\big\}\mathrm{d}r,$$

so that as $T \to \infty$ we have

$$u(t,\tilde{x},x_d) = \int_{t_0}^{\infty} \mathbb{E}\big\{e^{-X_0(t,\tilde{x},t_0,r)} v(t,\tilde{X}(t,\tilde{x},t_0,r),X_d(t,\tilde{x},x_d,t_0,r))\big\}\mathrm{d}r,$$
$$B_0 u(t,\tilde{x},x_d) + b_d(t,\tilde{x})\partial_d u(t,\tilde{x},x_d) + v(t,\tilde{x},x_d) = 0.$$

for any $t > t_0$, x in $\mathbb{R}^d_+$, usually used for $x_d = 0$. Note that t plays the role of a parameter. This reduces to (7.14) when the coefficients are constants.

To check (7.12), we add another equation, namely, with $T(t,\tilde{x},t_0,r) = T(r)$, $\tilde{X}(r) = \tilde{X}(t,\tilde{x},t_0,r)$, $X_d(r) = X_d(t,\tilde{x},x_d,t_0,r)$ and $X_0(r) = X_0(t,\tilde{x},t_0,r)$, for $t \geq t_0$, where

$$\begin{cases} T(r) = t - \displaystyle\int_{t_0}^{r} \rho(T(r),\tilde{X}(r))\mathrm{d}r, \\ \tilde{X}(r) = \tilde{z} + \displaystyle\int_{t_0}^{r} \varsigma(T(r),\tilde{X}(r))\mathrm{d}\tilde{w}(r) + \int_{t_0}^{r} \tilde{b}(T(r),\tilde{X}(r))\mathrm{d}r, \end{cases} \tag{7.47}$$

and

$$\begin{cases} X_d(r) := x_d + \displaystyle\int_{t_0}^{r} b_d(T(r),\tilde{X}(r))\mathrm{d}r, \\ X_0(r) := \displaystyle\int_{t_0}^{r} b_0(T(r),\tilde{X}(r))\mathrm{d}r, \\ \tau(t,\tilde{x},t_0) := \inf\big\{r > t_0 : T(t,\tilde{x},t_0,r) = t_0\big\}. \end{cases} \tag{7.48}$$

Again, Itô formula yields

$$\begin{aligned} \mathbb{E}\big\{e^{-X_0(\tau)} u(T(\tau),\tilde{X}(\tau)),X_d(\tau))\big\} - u(t,\tilde{x},x_d) = \\ = \mathbb{E}\Big\{\int_{t_0}^{\tau} e^{-X_0(r)}\big[B_0(T(r),\tilde{X}(r))u(T(r),\tilde{X}(r),X_d(r)) + \\ + b_d(T(r),\tilde{X}(r))\partial_d u(T(r),\tilde{X}(r),X_d(r)) - \\ - \rho(T(r),\tilde{X}(r))\partial_t u(T(r),\tilde{X}(r),X_d(r))\big]\mathrm{d}r\Big\}, \end{aligned}$$

i.e.,

$$\begin{aligned} u(t,\tilde{x},x_d) = \mathbb{E}\big\{e^{-X_0(\tau)} u(t_0,\tilde{X}(\tau),X_d(\tau)\big\} + \\ + \mathbb{E}\Big\{\int_{t_0}^{\tau} e^{-X_0(r)} v(T(r),\tilde{X}(r),X_d(r))\mathrm{d}r\Big\}, \\ B_0 u(t,\tilde{x},x_d) + b_d(t,\tilde{x})\partial_d u(t,\tilde{x},x_d) - \rho(t,\tilde{x})\partial_t u(t,\tilde{x},x_d) + v(t,\tilde{x},x_d) = 0. \end{aligned}$$

for any $t > t_0$, x in $\mathbb{R}^d_+$. Note the initial condition at t_0 and that, as one may expect, the variable x_d plays an independent role, only X_d depends on x_d.

It would be appropriate to study under which conditions on the coefficients $b_{ij}(t,\tilde{x})$, $b_i(t,\tilde{x})$, $b_0(t,\tilde{x})$ and $\rho(t,\tilde{x})$ the upper bounds estimates of the heat-kernel type are preserved from v into u, i.e., if for some constants $C_0, c_0 > 0$ and some $s > 0$, y in $\mathbb{R}^d_+$,

$$|v(t,x)| \leq C_0(t-s)^{-d/2}\exp\Big(-c_0\frac{|x-y|^2}{t-s}\Big), \quad \forall t > s,\ x \in \mathbb{R}^d_+,$$

then $u(t,x)$ also has a heat-kernel type upper bound, i.e., the same estimate holds true for $u(t,x)$, perhaps with a large constant C_0 and a smaller positive constant c_0. We will need the same type of estimates for d replaced for $(d-1)$, and estimates on the first and second derivative. However, no such a study was found by the authors, even for the constant coefficients case.

Now, following the discussion in Section 6.2.2 of the previous Chapter with $n=d$ and $m=d-1$, let $X(t)=(\tilde{X}(t),X_d(t))$ be the (diffusion process) solution of the martingale problem relative to the operators A_1 and B_1, i.e., there is a local time $\ell(t)$, satisfying the conditions $\ell(0)=0$,

$$\int_{t_0}^t \mathbb{1}_{\{X_d=0\}}\mathrm{d}\ell(s)=\ell(t),\quad \forall t\geq t_0,$$
$$\int_{t_0}^t \mathbb{1}_{\{X_d=0\}}\mathrm{d}s=\int_{t_0}^t \rho(s,X(s))\mathrm{d}\ell(s),\quad \forall t\geq t_0,$$

and such that the process

$$M_\varphi(t):=\varphi(X(t))-\varphi(x_0)-\int_{t_0}^t A_1(s)\varphi(X(s))\mathrm{d}s-$$
$$-\int_{t_0}^t B_1(s)\varphi(X(s))\mathrm{d}\ell(s),$$

is a martingale for any smooth function φ. Here A_1 and B_1 do not have a zero-order term, i.e., $A=A_1-a_0$ and $B=B_1-b_0$. Certainly, some regularity conditions (e.g., locally Lipschitz in x with a linear growth) and $b_d(t,\tilde{x})+\rho(t,\tilde{x})\geq c_0>0$ are required for the existence of such a diffusion process. Thus $X=X(t_0,x_0,t)$ and $\ell=\ell(t_0,x_0,t)$.

Adding another coordinate $X_0(t)=X_0(t_0,x_0,t)$ defined by

$$X_0(t):=\int_{t_0}^t \left[a_0(s,\tilde{X}(s),X_d(s))+\mathbb{1}_{\{X_d(s)=0\}}\,b_0(s,\tilde{X}(s))\right]\mathrm{d}s,\quad \forall t\geq t_0,$$

form Itô formula one deduces that the following expression is a martingale for any smooth function φ,

$$e^{-X_0(t)}\varphi(t,X(t))-\varphi(t_0,x_0)-\int_{t_0}^t e^{-X_0(t)}\left[\partial_s+A(s)\right]\varphi(s,X(s))]\mathrm{d}s-$$
$$-\int_{t_0}^t e^{-X_0(t)}\left[\partial_s+B(s)\right]\varphi(s,X(s))\mathrm{d}\ell(s),$$

which includes the a_0 and b_0 terms.

Now, if $u(t,x)$ is a smooth solution of the PDE problem in $\mathbb{R}^d_+$ with a *terminal condition* relative to the operators A and B, i.e.,

$$\begin{cases}[\partial_t+A(t)]u(t,x)=f(t,x),\quad \forall x\in\dot{\mathbb{R}}^d_+,\,t<T,\\ u(T,x)=u_T(x),\quad \forall x\in\dot{\mathbb{R}}^d_+,\\ B(t)u(t,\tilde{x},0)=\psi(t,\tilde{x}),\forall\tilde{x}\in\mathbb{R}^{d-1},\,t<T,\end{cases}$$

I do not know too well how to write this stuff! Can you help me how to include all cases in the representation formula?
What do you think?
PLEASE CHECK CAREFULLY!

7.2.2 Green Identity in Half-Space

For a second-order differential operator A considered on the half-space or semi-space $\mathbb{R}^d_+$, as (7.35) we define the co-normal differential operators ∂_A and ∂^*_A,

$$\begin{cases} \partial_{\mathrm{A}}\varphi(x) := \frac{1}{2}\sum_{i=1}^{d}\big(a_{id}(t,\tilde{x},0)\partial_i\varphi(x)\big), \\ \partial^*_{\mathrm{A}}\varphi(x) := \frac{1}{2}\sum_{j=1}^{d}\partial_j\big(a_{dj}(t,\tilde{x},0)\varphi(x)\big) - a_d(t,\tilde{x},0)\varphi(x) \end{cases} \tag{7.49}$$

on the boundary $\partial\mathbb{R}^d_+ \simeq \mathbb{R}^{d-1}$. Note that for $A = \frac{1}{2}\Delta$ we have $\partial_{\mathrm{A}} = \partial^*_{\mathrm{A}} = \frac{1}{2}\partial_d$.

First, we study the case with coefficients independent of t, and we integrate by parts two given smooth functions with compact support u and v to get

$$\sum_{i,j=1}^{d}\int_{\mathbb{R}^d_+} a_{ij}(x)\big(\partial_{ij}u(x)\big)v(x)\mathrm{d}x = -\sum_{i,j=1}^{d}\int_{\mathbb{R}^d_+}\big(\partial_i u(x)\big)\partial_j\big(a_{ij}(x)v(x)\big)\mathrm{d}x - \\ -\sum_{i=1}^{d}\int_{\mathbb{R}^{d-1}} a_{id}(\tilde{x},0)\big(\partial_i u(\tilde{x},0)\big)v(\tilde{x},0)\mathrm{d}\tilde{x},$$

$$-\sum_{i,j=1}^{d}\int_{\mathbb{R}^d_+} u(x)\partial_{ij}\big(a_{ij}(x)v(x)\big)\mathrm{d}x = \sum_{i,j=1}^{d}\int_{\mathbb{R}^d_+}\big(\partial_i u(x)\big)\partial_j\big(a_{ij}(x)v(x)\big)\mathrm{d}x + \\ +\sum_{j=1}^{d}\int_{\mathbb{R}^{d-1}} u(\tilde{x},0)\partial_j\big(a_{dj}(\tilde{x},0)v(\tilde{x},0)\big)\mathrm{d}\tilde{x},$$

and

$$\sum_{i=1}^{d}\int_{\mathbb{R}^d_+} a_i(x)\big(\partial_i u(x)\big)v(x)\mathrm{d}x = -\sum_{i=1}^{d}\int_{\mathbb{R}^d_+} u(x)\partial_i\big(a_i(x)v(x)\big)\mathrm{d}x - \\ -\int_{\mathbb{R}^{d-1}} a_d(\tilde{x},0)u(\tilde{x},0)v(\tilde{x},0)\mathrm{d}\tilde{x},$$

to get the following Green identity

$$\begin{cases} \displaystyle\int_{\mathbb{R}^d_+} \big[\big(Au(x)\big)v(x) - u(x)\big(A^*v(x)\big)\big]\mathrm{d}x = \\ \qquad\qquad = -\displaystyle\int_{\mathbb{R}^{d-1}} \big[\big(\partial_{\mathrm{A}}u(\tilde{x},0)\big)v(\tilde{x},0) - u(\tilde{x},0)\big(\partial^*_{\mathrm{A}}v(\tilde{x},0)\big)\big]\mathrm{d}\tilde{x}, \end{cases} \tag{7.50}$$

with the notation (7.49) with coefficients independent of t.

If the coefficients depends also on t, then add the relation

$$\int_{t_1}^{t_2} \big(\partial_t u(t,x)\big)v(t,x)\mathrm{d}t = -\int_{t_1}^{t_2} u(t,x)\big(\partial_t v(t,x)\big)\mathrm{d}t + \\ +\big[u(t_2,x)v(t_2,x) - u(t_1,x)v(t_1,x)\big],$$

and integrate in t the equality (7.50) to obtain the following parabolic Green identity

$$\begin{cases} \displaystyle\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^d_+} \big[\big(\mathcal{A}u(t,x)\big)v(t,x) - u(t,x)\big(\mathcal{A}^*v(t,x)\big)\big]\mathrm{d}x = \\ \quad = \displaystyle\int_{\mathbb{R}^d_+} \big[u(t_2,x)v(t_2,x) - u(t_1,x)v(t_1,x)\big]\mathrm{d}x + \\ \quad\quad + \displaystyle\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}} \big[\big(\partial_{\mathrm{A}}u(t,\tilde{x},0)\big)v(t,\tilde{x},0) - u(t,\tilde{x},0)\big(\partial^*_{\mathrm{A}}v(t,\tilde{x},0)\big)\big]\mathrm{d}\tilde{x}, \end{cases} \tag{7.51}$$

for any $t_2 > t_1 \geq 0$, where $\mathcal{A} = \partial_t - A(t)$ and $\mathcal{A}^* = -\partial_t - A^*(t)$.

If $(s,y) \mapsto G(t,x,s,y)$ satisfies

$$\mathcal{A}^*G(t,x,\cdot,\cdot) = 0, \qquad G(t,x,t,\cdot) = \delta_x, \tag{7.52}$$

for $t > s \geq 0$ and x, y in $\mathbb{R}^d_+$, and $(t,x) \mapsto u(t,x)$ is a solution of the problem

$$\mathcal{A}u = f, \qquad u(t_0,x) = u_0(x), \tag{7.53}$$

then taking $v(s,y) = G(t,x,s,y)$ in (7.51) with $t_1 = t_0 \geq 0$ and $t_2 = t$ we deduce

$$\int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^d_+} f(s,y)G(t,x,s,y)\mathrm{d}y = u(t,x) - \int_{\mathbb{R}^d_+} u_0(y)G(t,x,t_0,y)\mathrm{d}y + \\ + \int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^{d-1}} \big[\big(\partial_{\mathrm{A}}u(s,\tilde{y},0)\big)G(t,x,s,\tilde{y},0) - u(s,\tilde{y},0)\big(\partial^*_{\mathrm{A}}G(t,x,s,\tilde{y},0)\big)\big]\mathrm{d}\tilde{y}.$$

Now, if we add the complementary boundary conditions

$$\partial^*_{\mathrm{A}}G(t,x,\cdot,\cdot,0) = 0, \quad \text{and} \quad \partial_{\mathrm{A}}u(\cdot,\cdot,0) + \psi(\cdot,\tilde{x}) = 0,$$

we have the representation

$$u(t,x) = \int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^d_+} G(t,x,s,y)f(s,y)\mathrm{d}y + \int_{\mathbb{R}^d_+} G(t,x,t_0,y)u_0(y)\mathrm{d}y + \\ + \int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^{d-1}} G(t,x,s,\tilde{y},0)\psi(s,\tilde{y})\mathrm{d}\tilde{y},$$

or the alternative complementary boundary conditions

$$G(t,x,s,\tilde{y},0)=0 \quad \text{and} \quad u(t,\tilde{x},0)=\psi(t,\tilde{x}),$$

we get

$$u(t,x)=\int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^d_+}G(t,x,s,y)f(s,y)\mathrm{d}y+\int_{\mathbb{R}^d_+}G(t,x,t_0,y)u_0(y)\mathrm{d}y+ \\ +\int_{t_0}^{t}\mathrm{d}s\int_{\mathbb{R}^{d-1}}\big(-\partial_{\mathrm{A}}^{*}G(t,x,s,\tilde{y},0)\big)\psi(s,\tilde{y})\mathrm{d}\tilde{y}.$$

Therefore, the Green and Poisson functions satisfy

$$\begin{cases} P_{A,D}(t,x,s,\tilde{y})=-\partial_{\mathrm{A}}^{*}G_{A,D}(t,x,s,\tilde{y},0),\\ G_{A,D}(t,x,s,y)=G^{*}_{A,D}(s,y,t,x),\\ P_{A,N}(t,x,s,\tilde{y})=G_{A,N}(t,x,s,\tilde{y},0),\end{cases} \tag{7.54}$$

for Dirichlet and co-normal (Neumann) boundary conditions. Note that the co-normal first differential operator $\partial_{\mathrm{A}}^{*}$, defined by (7.49), is acting on the variable y. Moreover, the Green function is found as solving the adjoint problem in the variable (s,y), i.e., (7.52) plus a complementary boundary condition.

Let us now consider the case of a boundary differential operator B given by (7.38). Clearly when $\rho=0$, for the tangential derivative we have

$$\int_{\mathbb{R}^{d-1}}\big[\big(B_0u(\tilde{x},0)\big)v(\tilde{x},0)-u(\tilde{x},0)\big(B_0^{*}v(\tilde{x},0)\big)\big]\mathrm{d}\tilde{x}=0,$$
$$\int_{\mathbb{R}^{d-1}}\big(\partial_{\mathrm{A}}u(\tilde{x},0)\big)v(\tilde{x},0)\mathrm{d}\tilde{x}=-\int_{\mathbb{R}^{d-1}}u(\tilde{x},0)\big(\tilde{\partial}_{\mathrm{A}}^{*}v(\tilde{x},0)\big)\mathrm{d}\tilde{x}+ \\ +\int_{\mathbb{R}^{d-1}}\big(a_{dd}(t,\tilde{x},0)\partial_d u(\tilde{x},0)\big)v(\tilde{x},0)\mathrm{d}\tilde{x},$$

where

$$\tilde{\partial}_{\mathrm{A}}^{*}\varphi(x)=\tfrac{1}{2}\sum_{i=1}^{d-1}\partial_i\big(a_{id}(t,\tilde{x},0)\varphi(x)\big).$$

If

$$a_{dd}(t,\tilde{x},0)=2b_d(t,\tilde{x}),\quad \forall\, t\geq 0,\ \tilde{x}\in\mathbb{R}^{d-1}, \tag{7.55}$$

then

$$\int_{\mathbb{R}^{d-1}}\big[\big(Bu(\tilde{x},0)\big)v(\tilde{x},0)-u(\tilde{x},0)\big(B_0^{*}+\tilde{\partial}_{\mathrm{A}}^{*}\big)v(\tilde{x},0)\big]\mathrm{d}\tilde{x}= \\ =\int_{\mathbb{R}^{d-1}}\big(\partial_{\mathrm{A}}u(\tilde{x},0)\big)v(\tilde{x},0)\mathrm{d}\tilde{x},$$

which yields

$$\begin{cases}\displaystyle\int_{\mathbb{R}^d_+}\big[\big(Au(x)\big)v(x)-u(x)\big(A^{*}v(x)\big)\big]\mathrm{d}x=\\ \qquad\displaystyle=-\int_{\mathbb{R}^{d-1}}\big[\big(Bu(\tilde{x},0)\big)v(\tilde{x},0)-u(\tilde{x},0)\big(B_{\mathrm{A}}^{*}v(\tilde{x},0)\big)\big]\mathrm{d}\tilde{x},\end{cases} \tag{7.56}$$

and

$$\left\{\begin{aligned}&\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^d_+}\big[\big(\mathcal{A}u(t,x)\big)v(t,x)-u(t,x)\big(\mathcal{A}^*v(t,x)\big)\big]\mathrm{d}x=\\&\quad=\int_{\mathbb{R}^d_+}\big[u(t_2,x)v(t_2,x)-u(t_1,x)v(t_1,x)\big]\mathrm{d}x+\\&\qquad+\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\big[\big(Bu(t,\tilde x,0)\big)v(t,\tilde x,0)-u(t,\tilde x,0)\big(B^*_{\mathrm{A}}v(t,\tilde x,0)\big)\big]\mathrm{d}\tilde x,\end{aligned}\right.\tag{7.57}$$

where condition (7.55) is enforced,

$$B^*_{\mathrm{A}}:=B^*_0+\tilde\partial^*_A+\partial^*_{\mathrm{A}}=B^*_0+2\tilde\partial^*_A+\partial_d\big(a_{dd}\cdot\,\big)-a_d,$$

and $\mathcal{A}=\partial_t-A(t)$, $\mathcal{A}^*=-\partial_t-A^*(t)$.

Therefore, if the above functions $v(s,y):(s,y)\mapsto G(t,x,s,y)$ and $u(s,y)$ satisfy the boundary condition

$$B^*_{\mathrm{A}}G(t,x,s,\tilde y,0)=0\quad\text{and}\quad Bu(t,\tilde x,0)=\psi(t,\tilde x),$$

then, the Green and Poisson functions, corresponding to the differential operators A and B given by (7.35) and (7.38), satisfy

$$P_{A,B}(t,x,s,\tilde y)=G_{A,B}(t,x,s,\tilde y,0),\tag{7.58}$$

for the case $\rho=0$ and $2b_d=a_{dd}$.

• *Remark* 7.4. Remark that if the normalization condition (7.55) is not a priori satisfied, then the PDE boundary problem (7.41) can be re-stated (equivalently) with a new boundary boundary operator B satisfying (7.55). This is, given a boundary differential operator B as in (7.38) one can define another operator $\hat B$ by the relation $2b_d\hat B=a_{dd}B$. Thus, A and $\hat B$ satisfy the (7.55) and their Green and Poisson functions satisfy (7.58) with $\hat B$ in lieu of B. However the equation yields

$$G_{A,B}=G_{A,\hat B}\quad\text{and}\quad P_{A,B}\,2b_d=P_{A,\hat B}\,a_{dd}.$$

Hence, (7.58) becomes

$$P_{A,B}(t,x,s,\tilde y)\,2b_d(s,\tilde y)=G_{A,B}(t,x,s,\tilde y,0)\,a_{dd}(s,\tilde y,0),$$

for the case $\rho=0$. Clearly, this relation is proved for smooth coefficients (i.e., when the adjoint differential operator A^* and B^* are defined), but it remains true (by passage to the limit) for Hölder continuous operators where these Green and Poisson functions can be constructed. □

Next, to treat the case $2b_d=a_{dd}$ and $\rho>0$, we use the integration by parts

$$\begin{aligned}&\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\rho(t,\tilde x)\big[\big(\partial_tu(t,x)\big)v(t,x)+u(t,x)\big(\partial_tv(t,x)\big)\big]\mathrm{d}\tilde x=\\&\qquad=\int_{\mathbb{R}^{d-1}}\Big[\rho(t,\tilde x)u(t,x)v(t,x)\Big]_{t=t_1}^{t=t_2}\mathrm{d}\tilde x-\\&\qquad\quad-\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\big(\partial_t\rho(t,\tilde x)\big)u(t,x)v(t,x)\mathrm{d}\tilde x,\end{aligned}$$

so that for u satisfying (7.53) and v such that $\partial_t v(t,x) = -A^*(t)v(t,x)$ we deduce

$$
\begin{aligned}
-\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\big[\rho(t,\tilde{x})\big(A(t)u(t,x)\big)v(t,x)-u(t,x)\big(A^*(t)v(t,x)\big)\big]\mathrm{d}\tilde{x} = \\
=\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\rho(t,\tilde{x})f(t,x)v(t,x)\mathrm{d}\tilde{x}- \\
-\int_{\mathbb{R}^{d-1}}\Big[\rho(t,\tilde{x})u(t,x)v(t,x)\Big]_{t=t_1}^{t=t_2}\mathrm{d}\tilde{x}+ \\
+\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\big(\partial_t\rho(t,\tilde{x})\big)u(t,x)v(t,x)\mathrm{d}\tilde{x},
\end{aligned}
$$

Thus, setting $t=s$, $x=y$, assuming $v(t_2,\tilde{y},0)=0$, letting $y_d\to 0$ and using the identity (7.57), we obtain

$$
\left\{
\begin{aligned}
&\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d_+}\big[\big(\mathcal{A}u(s,y)\big)v(s,y)-u(s,y)\big(\mathcal{A}^*v(s,y)\big)\big]\mathrm{d}y = \\
&\quad=-\int_{t_1}^{t_2}\mathrm{d}t\int_{\mathbb{R}^{d-1}}\rho(s,\tilde{y})f(s,\tilde{y},0)v(s,\tilde{y},0)\mathrm{d}\tilde{x}- \\
&\qquad-\int_{\mathbb{R}^{d-1}}\rho(t_1,\tilde{y})u(t_1,\tilde{y},0)v(t_1,\tilde{y},0)\mathrm{d}\tilde{y}+ \\
&\qquad+\int_{\mathbb{R}^d_+}\big[u(t_2,y)v(t_2,y)-u(t_1,y)v(t_1,y)\big]\mathrm{d}y+ \\
&\qquad+\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^{d-1}}\big[\big(Bu(s,\tilde{y},0)\big)v(s,\tilde{y},0)-u(s,\tilde{y},0)\big(B^*_{\mathrm{A}}v(s,\tilde{y},0)\big)\big]\mathrm{d}\tilde{y},
\end{aligned}
\right. \tag{7.59}
$$

where condition (7.55) is enforced and the A-adjoint of the boundary operator B is given by

$$
B^*_{\mathrm{A}} := B^*_0 + 2\tilde{\partial}^*_{\mathrm{A}} + \partial_d\big(a_{dd}\cdot\big) - a_d - \rho A^* + (\partial_t\rho), \tag{7.60}
$$

i.e.,

$$
\begin{aligned}
&B^*_{\mathrm{A}}(t)\varphi(x) := B^*_{A0}(t)\varphi(x) + \partial_d\big(b_d(t,\tilde{x})\varphi(x)\big) - \rho(t,\tilde{x})A^*, \\
&B^*_{A0}(t)\varphi(x) := \tfrac{1}{2}\sum_{i,j=1}^{d-1} b_{ij}(t,\tilde{x})\partial_{ij}\varphi(x)\big) + \sum_{i=1}^{d-1}\bar{b}^*_i(t,\tilde{x})\partial_i\varphi(x) - \bar{b}^*_0(t,\tilde{x})\varphi(x), \\
&b^*_i(t,\tilde{x}) = -b_i(t,\tilde{x}) - \tfrac{1}{2}\sum_{j=1}^{d-1}\partial_j b_{ij}(t,\tilde{x}) + \sum_{i=1}^{d-1} a_{id}(t,\tilde{x},0), \\
&b^*_0(t,\tilde{x}) = b_0(t,\tilde{x}) + \sum_{i=1}^{d-1}\partial_i b_i(t,\tilde{x}) - \partial_t\rho(t,\tilde{x}) + a_d(t,\tilde{x},0) - \sum_{i=1}^{d-1}\partial_i a_{id}(t,\tilde{x},0),
\end{aligned}
$$

after using the symmetry $a_{di}=a_{id}$.

Now, take $t_2=t$, $t_1=t_0$ and $v(s,y)=G(t,x,s,y)$ satisfying (7.52). It is clear that either the heat-kernel type estimates or the terminal condition ensure that

$$
G(t,x,s,\tilde{y},0)\to 0 \quad \text{as} \quad s\to t, \quad \forall t>0,\ x=(\tilde{x},x_d),\ \tilde{x}\in\mathbb{R}^{d-1},\ x_d>0,
$$

i.e., $v(t,\tilde{y},0)=0$. If G and u satisfy the boundary condition

$$B_{\rm A}^* G(t,x,s,\tilde{y},0)=0 \quad \text{and} \quad Bu(t,\tilde{x},0)=\psi(t,\tilde{x}),$$

then we obtain the representation formula

$$\begin{aligned} u(t,x)=&\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^d_+} G(t,x,s,y)f(s,y)\mathrm{d}y+\int_{\mathbb{R}^d_+} G(t,x,t_0,y)u_0(y)\mathrm{d}y+\\ &+\int_{t_0}^t \mathrm{d}s\int_{\mathbb{R}^{d-1}} G(t,x,s,\tilde{y},0)\psi(s,\tilde{y})\mathrm{d}\tilde{y}+\\ &+\int_{t_0}^t \mathrm{d}t\int_{\mathbb{R}^{d-1}} G(t,x,s,\tilde{y},0)\rho(s,\tilde{y})f(s,\tilde{y},0)\mathrm{d}\tilde{y}+\\ &+\int_{\mathbb{R}^{d-1}} G(t,x,t_0,\tilde{y},0)\rho(t_0,\tilde{y})u_0(\tilde{y},0)\mathrm{d}\tilde{y}, \end{aligned}$$

Since the representation formula and the maximum principle yield the uniqueness of the Green and Poisson functions, we deduce that the Green and Poisson functions satisfy

$$\begin{cases} P_{A,B}(t,x,s,\tilde{y})=Q(t,x,s,\tilde{y},0),\\ G_{A,B}(t,x,s,y)=Q(t,x,s,y)+Q(t,x,s,\tilde{y},0)\rho(s,\tilde{y})\delta(y_d), \end{cases} \tag{7.61}$$

for the case $\rho\geq 0$ and $2b_d=a_{dd}$, where the kernel Q satisfies

$$\mathcal{A}^*Q(t,x,s,y)=0, \quad Q(t,x,t,y)=\delta_x(y), \quad B_{\rm A}^*Q(t,x,s,\tilde{y},0)=0, \tag{7.62}$$

for $t>s\geq 0$ and x, $y=(\tilde{y},y_d)$ in $\mathbb{R}^d_+$, with $\mathcal{A}^*=-\partial_t-A^*$, A^* and $B_{\rm A}^*$ given by (7.43) and (7.60), both acting on the variables (s,y). Certainly, the problem (7.62) is really meaningful when some estimates of heat-kernel type are satisfied. Note that we do have

$$G_{A,B}(t,x,s,\tilde{y},y_d)\to P_{A,B}(t,x,s,\tilde{y}) \quad \text{as} \quad y_d\to 0,$$

in a pointwise sense.

• *Remark* 7.5. The kernel Q can be regarded as the Green function corresponding to $\rho>0$ but with data vanishing on the boundary, i.e., $f(t,\tilde{x},0)=0$ and $u_0(\tilde{x},0)=0$, for any $t\geq 0$ and $\tilde{x}$ in $\mathbb{R}^{d-1}$. The argument used in Remark 7.4 can be used with $Q_{A,\hat{B}}$ as early by means of (7.62) with $\hat{B}$ in lieu of B, and $\rho\hat{B}=B$ so that one may express

$$\begin{aligned} &P_{A,B}(t,x,s,\tilde{y})=Q_{A,B}(t,x,s,\tilde{y},0),\\ &G_{A,B}(t,x,s,y)=G_{A,D}(t,x,s,y)+Q_{A,B}(t,x,s,y)\frac{2b_d(s,\tilde{y})}{a_{dd}(s,\tilde{y},0)}+\\ &\qquad\qquad +Q_{A,B}(t,x,s,\tilde{y},0)\rho(s,\tilde{y})\delta(y_d), \end{aligned}$$

where $G_{A,D}$ is the Green functions with Dirichlet boundary conditions and $Q_{A,B}$ is a suitable kernels. Note that $G_{A,D}(t,x,s,y)=0$ for either $x_d=0$ or $y_d=0$. □

To treat the case $b_d = 0$ and $\rho > 0$ we proceed as above, for u satisfying (7.53), an integration by parts (with the tangential derivatives) yields

$$\begin{aligned}\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \big[\big(Bu(s,\tilde y,0)\big)v(s,\tilde y,0) - u(s,\tilde y,0)\big(B_\rho^* v(s,\tilde y,0)\big)\big]\mathrm{d}\tilde y = \\ = -\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \rho(s,\tilde y) f(s,\tilde y,0) v(s,\tilde y,0)\mathrm{d}\tilde y - \\ - \int_{\mathbb{R}^{d-1}} \Big[\rho(s,\tilde y)u(s,\tilde y,0)v(s,\tilde y,0)\Big]_{s=t_0}^{s=t}\mathrm{d}\tilde y,\end{aligned}$$

where $B_\rho^* = B_0^* - \big(\partial_t \rho\,\cdot\,\big)$. Hence, taking $v(s,\tilde y,0) = P_B(t,\tilde x,s,\tilde y)$ satisfying

$$B_\rho^* P_B(t,\tilde x,s,\tilde y) = 0 \quad \text{and} \quad \rho(t,\cdot)P_B(t,\tilde x,t,\cdot) = \delta_{\tilde x},$$

we deduce

$$\begin{aligned}\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \psi(s,\tilde y) P_B(t,\tilde x,s,\tilde y)\mathrm{d}\tilde y = \\ = -\int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \rho(s,\tilde y) f(s,\tilde y,0) P_B(t,\tilde x,s,\tilde y)\mathrm{d}\tilde y + \\ + u(t,\tilde x) - \int_{\mathbb{R}^{d-1}} \rho(t_0,\tilde y) u_0(\tilde y,0) P_B(t,\tilde x,t_0,\tilde y)\mathrm{d}\tilde y.\end{aligned}$$

Now, we use the Green identity (7.51) for Dirichlet boundary conditions, i.e., the representation formula with $G = G_{A,D}$ to get

$$\begin{aligned}u(t,x) = \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^d_+} G(t,x,s,y) f(s,y)\mathrm{d}y + \int_{\mathbb{R}^d_+} G(t,x,t_0,y)u_0(y)\mathrm{d}y + \\ + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} \big(-\partial_{\mathrm{A}}^* G(t,x,s,\tilde y,0)\big)u(s,\tilde y)\mathrm{d}\tilde y.\end{aligned}$$

Hence, if we define

$$P_{A,B}(t,x,s,\tilde y) := \int_s^t \mathrm{d}\tau \int_{\mathbb{R}^{d-1}} \big(-\partial_{\mathrm{A}}^* G_{A,D}(t,x,\tau,\tilde\xi,0)\big) P_B(\tau,\tilde\xi,s,\tilde y)\mathrm{d}\tilde\xi$$

then

$$\begin{aligned}u(t,x) := \int_{\mathbb{R}^d_+} G_{A,D}(t,x,t_0,y)u_0(y)\mathrm{d}y + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^d_+} G_{A,D}(t,x,s,y) f(s,y)\mathrm{d}y + \\ + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t,x,s,\tilde y)\psi(s,\tilde y)\mathrm{d}\tilde y + \\ + \int_{t_0}^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t,x,s,\tilde y)\rho(s,\tilde y) f(s,\tilde y,0)\mathrm{d}\tilde y + \\ + \int_{\mathbb{R}^{d-1}} P_{A,B}(t,x,t_0,\tilde y)\rho(t_0,\tilde y)u_0(\tilde y,0)\mathrm{d}\tilde y,\end{aligned}$$

for $b_d = 0$ and $\rho > 0$. Note that in this case, the conditions on the boundary and on the interior are independent, i.e., first get P_B and $G_{A,B} = G_{A,D}$ independently, and then build $P_{A,B}$ as a singular integral.

Therefore, if the coefficients are smooth, then the adjoint problem (with a terminal condition) to (7.41) is given by

$$\begin{cases} -\partial_t u(t,x) = A^*(t)u(t,x), \quad \forall t < t_0,\ x \in \dot{\mathbb{R}}^d_+, \\ B^*_{\rm A}(t)u(t,x) + \psi(t,x) = 0, \quad \forall t < t_0,\ x \in \partial\mathbb{R}^d_+, \\ u(t_0,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.63}$$

for any given $t_0 \geq 0$, with $B^*_{\rm A}$ given by (7.60), with $2b_d = a_{dd}$. Clearly, if $\rho = 0$ then we have the relation

$$\begin{cases} P_{A^*,B^*_{\rm A}}(t,\tilde{x},x_d,t_0,\tilde{y}) = P_{A,B}(t_0,\tilde{y},x_d,t,\tilde{x}), \\ G_{A^*,B^*_{\rm A}}(t,x,t_0,y) = G_{A,B}(t_0,y,t,x), \end{cases} \tag{7.64}$$

for the Poisson and Green functions. Note that when $\rho > 0$ a δ-measure appears on the boundary and the above relation is modified as described in Remarks 7.4 and 7.5 when $b_d > 0$.

7.2.3 Successive Approximations

Perhaps the first step in the construction of the fundamental functions for variables coefficients is to study a Volterra equation for heat-type kernels $Q(t,x,\tau,\xi)$, namely,

$$\begin{cases} Q(t,x,\tau,\xi) = Q_0(t,x,\tau,\xi) + (Q_0 \star Q)(t,x,\tau,\xi), \\ (Q_0 \star Q)(t,x,\tau,\xi) := \displaystyle\int_\tau^t {\rm d}s \int_{\mathbb{R}^d} Q_0(t,x,s,y)Q(s,y,\tau,\xi){\rm d}y, \end{cases} \tag{7.65}$$

where the given kernel Q_0 satisfies the estimates

$$|Q_0(t,x,\tau,\xi)| \leq C_0(t-\tau)^{-\frac{d+2-\alpha}{2}} \exp\Big(-c_0\frac{|x-\xi|^2}{t-\tau}\Big), \tag{7.66}$$

for any $t > \tau$, and x,ξ in $\mathbb{R}^d$, and some $\alpha > 0$. For a given $c_0 > 0$, it is convenient to denote by $[\![Q_0]\!]_{(\alpha)}$ the smallest constant C_0 for which the bound (7.66) is satisfied.

Essentially based on the Beta-function and the equality

$$\begin{aligned} \int_{\mathbb{R}^d} \exp\Big(-c_0\frac{|x-y|^2}{t-s}\Big)\exp\Big(-c_0\frac{|y-\xi|^2}{s-\tau}\Big){\rm d}y = \\ = \Big[\Big(\frac{\pi}{c_0}\Big)\Big(\frac{(t-s)(s-\tau)}{t-\tau}\Big)\Big]^{\frac{d}{2}} \exp\Big(-c_0\frac{|x-\xi|^2}{t-\tau}\Big), \end{aligned}$$

one can prove that the sequence of kernel Q_n defined by the recurrence as

$$Q_{n+1}(t,x,\tau,\xi) = (Q_0 \star Q_n)(t,x,\tau,\xi)$$

satisfies

$$[\![Q_n]\!]_{(n\alpha)} \leq \frac{q_\alpha}{(n!)^{\frac{\alpha}{2}}}, \quad \forall n = 1,2,\dots,$$

where the constant q_α depends only on c_0, C_0, α and d. Hence, the Volterra (7.65) has a (unique) solution given by the series

$$Q(t,x,\tau,\xi) = \sum_{n=0}^{\infty} Q_n(t,x,\tau,\xi),$$

where the limit is uniformly within compact sets in $\{(t,x,\tau,\xi) : t > \tau,\ x,\xi \in \mathbb{R}^d\}$.

This same argument can be use with the Green function in half-space $\mathbb{R}^d_+$, where the Volterra equation has the form

$$\begin{cases} Q(t,x,\tau,\xi) = Q_0(t,x,\tau,\xi) + (Q_0 \star Q)(t,x,\tau,\xi), \\ (Q_0 \star Q)(t,x,\tau,\xi) := \displaystyle\int_\tau^t \mathrm{d}s \int_{\mathbb{R}^d_+} Q_0(t,x,s,y) Q(s,y,\tau,\xi)\mathrm{d}y, \end{cases} \tag{7.67}$$

with a given kernel Q_0 defined within any $\mathbb{R}^d_+$ instead $\mathbb{R}^d$, and also, with the Poisson function in half-space $\mathbb{R}^d_+$, where now the Volterra equation has the form

$$\begin{cases} R(t,x,\tau,\tilde\xi) = R_0(t,x,\tau,\tilde\xi) + (R_0 \tilde\star R)(t,x,\tau,\tilde\xi), \\ (R_0 \tilde\star R)(t,x,\tau,\tilde\xi) := \displaystyle\int_\tau^t \mathrm{d}s \int_{\mathbb{R}^{d-1}} R_0(t,x,s,\tilde y) R(s,\tilde y,0,\tau,\tilde\xi)\mathrm{d}y, \end{cases} \tag{7.68}$$

with $y = (\tilde y, y_d)$ and $R(s,y,\tau,\tilde\xi) = R(s,\tilde y, y_d, \tau, \tilde\xi)$. The Volterra equation (7.67) works very similar to the initial equation (7.65) in $\mathbb{R}^d$, and its (unique) solution is expressed as (convergent) series of kernels $R_{n+1} = R_0 \tilde\star R_n$. However, to study the Volterra equation (7.68) we need to have a kernel satisfying

$$|R_0(t,x,\tau,\xi)| \le C_0 (t-\tau)^{-\frac{d+1-\alpha}{2}} \exp\Big(- c_0 \frac{|x-\xi|^2}{t-\tau}\Big), \tag{7.69}$$

for any $t > \tau$, and x, ξ in $\mathbb{R}^d$, i.e., the heat-kernel type estimates in $\mathbb{R}^d_+$ like (7.66) with $d-1$ instead of d. If we just keep heat-kernel type estimates like (7.66) in $\mathbb{R}^d_+$, then, because the kernel convolution $\tilde\star$ is only in dimension $(d-1)$ the second integral in s involves a factor of the form $(t-s)^{(\alpha-3)/2}$, which is not integrable if $\alpha \le 1$. Alternatively, we may assume that the kernel R_0 satisfies a variation of (7.66), namely

$$\begin{cases} |R_0(t,x,\tau,\tilde\xi)| \le C_0 \Big(\dfrac{x_d}{\sqrt{t-\tau}} + 1\Big)(t-\tau)^{-\frac{d+1-\alpha}{2}} \times \\ \qquad\qquad \times \exp\Big(- c_0 \dfrac{|x-\xi|^2}{t-\tau}\Big), \end{cases} \tag{7.70}$$

for any $t > \tau$, and x, ξ in $\mathbb{R}^d_+$, and some $\alpha > 0$. Thus, if $\tilde{[\![} R_0 \tilde{]\!]}_{(\alpha)}$ denotes the smallest constant C_0 for which the bound (7.70) is satisfied, then we have

$$\tilde{[\![} R_n \tilde{]\!]}_{(n\alpha)} \le \frac{r_\alpha}{(n!)^{\frac{\alpha}{2}}}, \quad \forall n = 1, 2, \dots,$$

where the constant r_α depends only on c_0, C_0, α and d.

Next, to obtain a Hölder estimate of the k-type, namely,

$$\begin{cases} |Q(t,x,\tau,\xi) - Q(t',x',\tau',\xi')| \le C_0\big[|t-t'|^{\alpha/2} + |x-x'|^\alpha + \\ \qquad +|\tau-\tau'|^{\alpha/2} + |\xi-\xi'|^\alpha\big](t-\tau)^{-\frac{d+2+k}{2}} \exp\Big(-c_0\frac{|x-\xi|^2}{t-\tau}\Big), \end{cases} \tag{7.71}$$

for any $t > \tau$, $t' > \tau'$, and x, ξ, x', ξ' in $\mathbb{R}^d$, with $(t-\tau)|x'-\xi'|^2 \le (t'-\tau')|x-\xi|^2$, for the same $\alpha > 0$ and $c_0 > 0$, is harder. A more complicate argument (essentially based on some cancellation property of Q_0) is used to show the validity of (7.71) for the kernel Q, solution of the Volterra equation (7.65) with $k = 0$. Similarly for the kernel R.

Fundamental Solution

The problem is set in the whole space, and boundary conditions are replaced by growth conditions on the functions and its derivatives. For instances, comprehensive details on this classic case can be found in the books Friedman [88] or Ladyzhenskaya et al. [147].

The fundamental solution $G(t,x,s,y)$ defined for $t > s \ge 0$ and x, y in $\mathbb{R}^d$ is expressed as

$$F(t,x,s,y) = F_0(t-s, x-y; s, y) + F_0 \star Q(t,x,s,y),$$

where $F_0(t,x;s,y)$ is the fundamental solution with *freezed* coefficients and Q is a kernel to be determined, This is usually refer to as *parametix method*. Clearly, constant or parameterized by (s,y) means

$$F_0(t,x;s,y) = e^{-ta_0(s,y)}\Gamma_d\big(t\sqrt{\underline{\underline{a}}}(s,y), x - t\underline{a}(s,y)\big),$$

with the notation (7.10), only the part with the matrix $\underline{\underline{a}}$ are most relevant, the terms with a_0 and the vector $\underline{a}$ may be omitted, i.e., they can be part of the kernel Q.

If $A(s,y)$ denotes the second-order differential operator (7.35) with parameterized coefficients (but acting on the variable x) and set

$$Q_0(t,x,s,y) := [A(s,y) - A(t,x)]F_0(t-s, x-y; s, y),$$

then the kernel Q is found as the solution of the Volterra equation

$$Q = Q_0 + Q_0 \star Q,$$

which can be solved by the method of successive approximations in view of the non-degeneracy and bounded Hölder continuity assumptions (7.36), (7.37) on the coefficients, and the heat-kernel type estimates proved on the explicit expression of F_0.

The next step is to establish the validity heat-kernel estimates for the fundamental solution F, based on the above expression.

Dirichlet Conditions

Essentially, the Green function with Dirichlet boundary conditions is constructed with the same arguments used to build the fundamental solution, but the initial G_0 is the Green function with constant (or parameterized) coefficients corresponding to Dirichlet boundary conditions. Again, G_0 has an explicit expression as discussed in previous sections. In this case, the Volterra equation is solved by the method of successive approximations in the half-space $\mathbb{R}^d_+$ and the Green function G_D is obtained as a series.

However, the arguments to construct the Poisson function are more delicate since the heat-kernel type estimates have a stronger singularity, $(t-s)^{-1/2}$ higher than the Green function. If the coefficient were smooth, then the Poisson function can be calculated as normal derivative of the adjoint Green function with Dirichlet boundary conditions, via Green identity. For bounded Hölder continuous coefficients, the expression

$$P_D := P_0 + G_D \star [A_0 - A]P_0,$$

provided the Poisson function, where $P_0(t-s, x-y; s, y)$ is the Poisson function corresponding to constant (or parameterized) coefficients and

$$[A_0 - A]P_0(t-s, x-y; s, y) := [A(s,y) - A(t,x)]P_0(t-s, x-y; s, y),$$

with both differential operators A_0 and A acting on the variable x. The hard point is to establish the heat-kernel type estimates for $P_D - P_0$ given by the above relation. Essentially, some kind of integration by parts is used to relate the singular integral $G_D \star [A_0 - A]P_0$ with the non-singular (weak or integrable) integral $[A_0 - A]G_D \star P_0$.

Alternatively, one may begin with the fundamental solution for variable coefficients denoted by $F(t, x, s, y)$ and then one solves the Dirichlet problem in the variables t and x,

$$AF_1 = 0 \ \text{ in } \ \mathbb{R}^d_+ \quad \text{ and } \quad F_1 = F \ \text{ on } \ \partial\mathbb{R}^d_+,$$

with vanishing initial condition, and finally setting $G = F - F_1$ as the Green function with Dirichlet boundary conditions. Here, the point is to show the estimates necessary to allow the construction of the kernel F_1. At the same time, this procedure produces estimates for the Poisson kernel P_D.

For instance, details can be found in Ivasišen [114] and Solonnikov [233] for parabolic systems. Also, in the books Eidelman [72] and Friedman [88] the interested reader will find some useful discussion.

Oblique Derivative

This is the case where the assumptions $\rho = 0$, $b_0 \geq 0$, normalization $2b_d = a_{dd}$, no second-order derivatives, non-degeneracy (7.36), (7.39) and bounded Hölder continuous coefficients (7.37), (7.40) are imposed.

The arguments are similar to those of the fundamental solution, but a two-step method is necessary, one step to make variable the coefficients of the interior

differential operator A and another step for the boundary operator B. Indeed, first set

$$G_1 = G_0 + G_0 \star Q,$$

and determine the kernel Q by the Volterra equation

$$Q = Q_0 + Q_0 \star Q, \qquad Q_0 := [A_0 - A]G_0,$$

where $G_0(t-s, x-y; s, y)$ is the Green function corresponding to constant (or parameterized) coefficients, and again both differential operators $A_0 = A(s, y)$ and $A = A(t, x)$ act on the x variable. In view of the heat-kernel type estimates on G_0, this Volterra equation is solved by the method of successive approximations in the half-space $\mathbb{R}^d_+$. The Green function G_1 and the Poisson function, according to Remark 7.4, are related by the equality

$$P_1(t, x, s, \tilde{y}) := G_1(t, x, s, \tilde{y}, 0)\frac{a_{dd}(s, \tilde{y}, 0)}{2b_d(s, \tilde{y})},$$

which corresponds to interior variable coefficients and constant (or parameterized) coefficients on the boundary, i.e., satisfying on the boundary

$$B_0 G_1 = 0 \quad \text{and} \quad B_0 P_1 = \tilde{\delta},$$

where the boundary differential operator $B_0 = B(s, y)$ is acting on the variable x, and $\tilde{\delta}$ is the delta measure on $(t, \tilde{x})$ concentrated at $(s, \tilde{y})$.

The next step is to set

$$P = P_1 + P_1 \tilde{\star} R, \quad \text{and} \quad G = G_1 + P \tilde{\star} [B_0 - B]G_1,$$

and to determine the kernel R by solving

$$R = R_0 + R_0 \tilde{\star} R, \qquad R_0 := [B_0 - B]P_1,$$

where G_1 and P_1 are as above, and both boundary differential operators $B_0 = B(s, y)$ and $B = B(t, x)$ are acting on the variable x. Note that because P_1 is the Poisson function one has

$$BP = \tilde{\delta} + [B - B_0]P_1 + R - R_0 \tilde{\star} R = \tilde{\delta},$$

and

$$BG = \tilde{\delta} + [B - B_0]G_1 + [B_0 - B]G_1 = 0,$$

which reproduces the desired equations.

Since the boundary operator B does not contain second-order derivatives, the kernel R_0 has a weak (integrable) singularity $(t-s)^{(-d-1+\alpha)/2}$, and so that the Volterra equation for R is solvable, and heat-type estimates are possible for the (surface) kernel convolution

Alternatively, first we may set

$$P_1 = P_0 + P_0 \tilde{\star} R,$$

and determining R by solving the Volterra equation

$$R = R_0 + R_0 \tilde{\star} R, \qquad R_0 := [B_0 - B]P_0,$$

where $P_0(t-s, \tilde{x}-\tilde{y}, x_d)$ is the Poisson function corresponding to constant (or parameterized) coefficients. Thus, once P_1 has been found, the expression

$$G_1 = G_0 + P_1 \tilde{\star} [B_0 - B]G_0$$

gives the Green function, i.e.,

$$A_0 G_1 = \delta, \quad \text{and} \quad BG_1 = 0.$$

Next, we have to solve the Volterra equation

$$Q = Q_0 + Q_0 \star Q, \qquad Q_0 := [A_0 - A]G_1$$

and then $G = G_1 + G_1 \star Q$ results the expression of the Green function corresponding to A and B.

Certainly, a great effort is needed to establish the heat-kernel type estimates for the Green functions $G = G_{A,B}$ and its derivatives. Note that in this case, the Poisson function $P_{A,B}$ is equal to the Green function on the boundary, i.e., $P_{A,B}(t,x,s,\tilde{y}) := G_{A,B}(t,x,s,\tilde{y},0)$, provided the normalization conditions $a_{dd}(t,\tilde{x},0) = 2b_d(t,\tilde{x})$ holds. Full details can be found in Garroni and Solonnikov [95], Ivasišen [114] and Solonnikov [233].

Sticky Boundary

This is the case $\rho > 0$, $b_0 \geq 0$ and $b_d > 0$. Hence, we normalize by setting $\rho = 1$, i.e., defining a new boundary operator $\hat{B}$ by the relation $\rho\hat{B} = B$. Then, we proceed as in the previous case of oblique derivative or alternatively, one may begin setting

$$\hat{P}_1 = \hat{P}_0 + \hat{P}_0 \tilde{\star} R,$$

and determining R by solving the Volterra equation

$$R = \hat{R}_0 + \hat{R}_0 \tilde{\star} R, \qquad \hat{R}_0 := [\hat{B}_0 - \hat{B}]P_0,$$

where $\hat{P}_0(t-s, \tilde{x}-\tilde{y}, x_d)$ is the Poisson function corresponding to constant (or parameterized) coefficients. Remark that $[\hat{B}_0 - \hat{B}]$ contains only derivatives up to the first order (even if $\rho > 0$, the term with A_0 is unchanged), and so the previous Volterra equation can be solved. Thus, once $\hat{P}_1$ has been found, the expression

$$\hat{G}_1 = \hat{G}_0 + \hat{P}_1 \tilde{\star} [\hat{B}_0 - \hat{B}]\hat{G}_0$$

gives associated Green function, i.e.,

$$A_0 \hat{G}_1 = \delta, \quad \text{and} \quad \hat{B}G_1 = 0.$$

Note that $\hat{B} - A_0$ is a first order differential boundary operator. Next, we have to solve the Volterra equation

$$\hat{Q} = Q_0 + Q_0 \star \hat{Q}, \qquad Q_0 := [A_0 - A]G_1$$

and then

$$Q_{A,\hat{B}} = \hat{G}_1 + \hat{G}_1 \star \hat{Q},$$
$$P_{A,B}(t,x,s,\tilde{y}) := Q_{A,\hat{B}}(t,x,s,\tilde{y},0)\,\frac{a_{dd}(s,\tilde{y},0)}{2b_d(s,\tilde{y})}$$

is the expression of the Poisson function corresponding to A and B, and

$$G_{A,B}(t,x,s,y) = Q_{A,\hat{B}}(t,x,s,y) + Q_{A,\hat{B}}(t,x,s,\tilde{y})\,\rho(s,\tilde{y})\,\delta(y_d)$$

is the Green function, see Remark 7.5.

Independent Conditions

This is the case $\rho > 0$, $b_0 \geq 0$ and $b_d = 0$. Clearly, this reduces to the Dirichlet boundary condition, and really independent conditions when $b_0 = 0$, see Section 7.2.2.

Second-order derivatives

When the boundary differential operator contains second-order (tangential) derivative in x_i, $i = 1, \dots, d-1$, the calculations are more delicate, but essentially the same arguments are valid. In particular, as in with Dirichlet boundary conditions, one may begin with the fundamental solution for variable coefficients denoted by $F(t,x,s,y)$ and then one solves the boundary value problem in the variables t and x,

$$AF_1 = 0 \ \text{ in } \ \mathbb{R}^d_+ \quad \text{ and } \quad BF_1 = BF \ \text{ on } \ \partial\mathbb{R}^d_+,$$

with vanishing initial condition, and finally setting $G = F - F_1$ as the Green function with the boundary conditions given by the operator B. The point here is that BF is a smooth (Hölder continuous) function for x on the boundary $\partial\mathbb{R}^d_+$, as long as y is in the interior of $\mathbb{R}^d_+$. Thus, general Theorem can be used to find a unique solution F_1, but a lot of effort should be done to produce sharp estimates leading to the mentioned heat-kernel estimates, e.g., see Eidelman [72] and Solonnikov [233].

7.2.4 Heat-Kernel Estimates

As discussed in the previous Subsection 7.2.2, the Green and Poisson functions have various forms, depending on the boundary conditions. The operators A and B are as in Section 7.2, under the assumptions (7.35), (7.36), (7.37) for A in $\mathbb{R}^d_+$, and (7.38), (7.39), (7.40) for B on the boundary $\partial\mathbb{R}^d_+ = \mathbb{R}^{d-1}$.

To establish a clear relation with the probabilistic counterpart, we use *terminal condition* instead of initial condition (as customary used in analysis and described in the previous subsection) to define Green and Poisson function, this involves only a change of the variables t into $T-t$, where T is the terminal time. Since the time derivative is not involved (directly) in the boundary conditions, the part concerning ρA (which could be expressed using $\rho\partial_t$) of the Wentzell boundary operator B does not change sign.

For a fixed terminal time $T > 0$, we consider the boundary value problem

$$\begin{cases} \partial_t u(t,x) + A(t)u(t,x) = f(t,x), \quad \forall (t,x) \in [0,T[\times\dot{\mathbb{R}}^d_+, \\ B(t)u(t,\tilde{x},0) + \psi(t,\tilde{x}) = 0, \quad \forall (t,\tilde{x}) \in [0,T[\times\mathbb{R}^{d-1}, \\ u(T,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.72}$$

with the compatibility condition

$$B(T)\varphi(\tilde{x},0) + \rho(T,\tilde{x})f(T,\tilde{x},0) + \psi(T,\tilde{x}) = 0, \tag{7.73}$$

for any $\tilde{x}$ in $\mathbb{R}^{d-1}$, with the notation (7.38). Clearly, this compatibility condition comes form replacing the term ρA by $-\rho\partial_t$ in the expression of the Wentzell operator B and taking the limit of the boundary condition as t approaches T. As mentioned early, we can distinguish three cases.

Dirichlet Type

The Dirichlet boundary conditions includes $b_d = 0$, $\rho = 0$, $b_0 > 0$, and in this case, the tangential first and second-order coefficients (b_i, $b_{ij} : i,j = 1,\dots,d-1$) are irrelevant, and may be taken all zeros. In this case, the singularity of the Poisson function is higher than the one of the Green function.

The Green and Poisson functions with terminal condition $G_{A,D}(t,x,s,y)$ and $P_{A,D}(t,x,s,\tilde{y})$ are defined for $0 \le t < s \le T$, x, y in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$. For a fixed s, y and $\tilde{y}$ the functions $(t,x) \mapsto G_{A,D}(t,x,s,y)$ and $(t,x) \mapsto P_{A,D}(t,x,s,\tilde{y})$ are twice continuously differentiable in x and once in t and satisfy

$$\begin{cases} [\partial_t + A(t)]G_{A,D}(t,x,s,y) = \delta(t-s,x-y), \quad \forall (t,x) \in [0,s[\times\dot{\mathbb{R}}^d_+, \\ G_{A,D}(t,\tilde{x},0,s,y) = 0, \quad \forall (t,\tilde{x}) \in [0,s[\times\mathbb{R}^{d-1}, \\ \lim_{t\to s} G_{A,D}(t,x,s,y) = \delta(x-y), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.74}$$

and

$$\begin{cases} [\partial_t + A(t)]P_{A,D}(t,x,s,\tilde{y}) = 0, \quad \forall (t,x) \in [0,s[\times\dot{\mathbb{R}}^d_+, \\ P_{A,D}(t,\tilde{x},0,s,\tilde{y}) = \delta(t-s,\tilde{x}-\tilde{y}), \quad \forall (t,\tilde{x}) \in [0,s[\times\mathbb{R}^{d-1}, \\ \lim_{t\to s} P_{A,D}(t,x,s,\tilde{y}) = 0, \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.75}$$

where the meaning of the δ measures is given by the representation formula

$$\left\{\begin{array}{l} u(t,x) = \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^d_+} G_{A,D}(t,x,s,y) f(s,y)\mathrm{d}y + \\ \qquad + \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,D}(t,x,s,\tilde{y})\psi(s,\tilde{y})\mathrm{d}\tilde{y} + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_+} G_{A,D}(t,x,T,y)\varphi(y)\mathrm{d}y, \end{array}\right. \tag{7.76}$$

where $u(t,x)$ is the solution of (7.72) with Dirichlet boundary conditions and smooth (bounded and Hölder continuous) data f, ψ and φ satisfying the compatibility condition $\psi(T,\tilde{x}) + \varphi(\tilde{x},0) = 0$ for any $\tilde{x}$ in $\mathbb{R}^{d-1}$.

Clearly, the construction of $G_{A,D}$ and $P_{A,D}$ involves only the operator A with the assumptions (7.35), (7.36), (7.37) on the coefficients. To state the heat-kernel estimates we denote by ∂^ℓ, $\ell = 0,1,2$ any derivative in x of order ℓ or the derivative in t for $\ell = 2$. Then, there exist two constants $C \geq c > 0$ such that for every $0 \leq t < s \leq T$, x, y in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$, we have

$$\left\{\begin{array}{l} |\partial^\ell G_{A,D}(t,x,s,y)| \leq C|s-t|^{\frac{2-d-\ell}{2}} \exp\Big(-c\dfrac{|x-y|^2}{s-t}\Big), \\ |\partial^\ell P_{A,D}(t,x,s,\tilde{y})| \leq C|s-t|^{\frac{2-d-\ell-1}{2}} \exp\Big(-c\dfrac{|\tilde{x}-\tilde{y}|^2 + x_d^2}{s-t}\Big), \end{array}\right. \tag{7.77}$$

and

$$\left\{\begin{array}{l} |\partial^\ell G_{A,D}(t,x,s,y) - \partial^\ell G_{A,D}(t',x',s',y')| \leq \\ \qquad \leq C\big[|t-t'|^{\frac{\alpha}{2}} + |x-x'|^\alpha + |s-s'|^{\frac{\alpha}{2}} + |y-y'|^\alpha\big] \times \\ \qquad\qquad \times\Big[|s-t|^{\frac{2-d-\ell-\alpha}{2}} \exp\Big(-c\dfrac{|x-y|^2}{s-t}\Big)\Big], \end{array}\right. \tag{7.78}$$

$$\left\{\begin{array}{l} |\partial^\ell P_{A,D}(t,x,s,\tilde{y}) - \partial^\ell P_{A,D}(t',x',s',\tilde{y}')| \leq \\ \qquad \leq C\big[|t-t'|^{\frac{\alpha}{2}} + |x-x'|^\alpha + |s-s'|^{\frac{\alpha}{2}} + |\tilde{y}-\tilde{y}'|^\alpha\big] \times \\ \qquad\qquad \times\Big[|s-t|^{\frac{2-d-\ell-1-\alpha}{2}} \exp\Big(-c\dfrac{|\tilde{x}-\tilde{y}|^2 + x_d^2}{s-t}\Big)\Big], \end{array}\right. \tag{7.79}$$

for $|s-t| \leq |s'-t'|$, $|x-y| \geq |x'-y'|$ and $|\tilde{x}-\tilde{y}|^2 + x_d^2 \geq |\tilde{x}'-\tilde{y}'|^2 + x_d'^2$, where $0 < \alpha < 1$ is the Hölder exponent of the coefficients. In what follows, we will refer these estimates on $\partial^\ell G_{A,D}$ and $\partial^\ell P_{A,D}$ as heat-kernel estimates (sup and Hölder) of order ℓ and $\ell - 1$ respectively.

As already mentioned, the Hölder continuity of the higher-order coefficients $(a_{ij} : i,j = 1,\dots d)$ plays a primary role in the construction and estimation of the Green and Poisson functions. However, if we drop the Hölder continuity of the lower order coefficients $(a_0,\, a_i : i = 1,\dots d)$ then only the estimate relative to the Hölder continuity of $\partial^\ell G_{A,D}$ and $\partial^\ell P_{A,D}$ in the variable s and y is actually lost.

Neumann Type

The Neumann type or oblique derivative includes $b_d > 0$, $\rho = 0$ and $b_0 \geq 0$. Properly speaking, oblique derivative correspond to the case $b_0 = 0$ and zero tangential second-order coefficients $b_{ij} = 0$ for any $i, j = 1, \ldots, d-1$. There is not essential changes-of-behavior in the Green and Poisson functions due to the addition of the non-zero tangential second-order coefficients, only more difficulties in the construction and the proof of suitable estimates. One may introduce some modification in the behavior as $T \to \infty$ of the Green and Poisson functions by adding a zero-order term $b_0 > 0$, as seen in the particular case of elastic Brownian motion.

The Green and Poisson functions $G_{A,B}(t, x, s, y)$ and $P_{A,B}(t, x, s, \tilde{y})$ with terminal condition (and boundary condition given by B) are defined for $0 \leq t < s \leq T$, x, y in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$. As in the Dirichlet case, they are twice continuously differentiable in x and once in t and satisfy

$$\begin{cases} [\partial_t + A(t)]G_{A,B}(t, x, s, y) = \delta(t - s, x - y), \quad \forall (t, x) \in [0, s[\times \dot{\mathbb{R}}^d_+, \\ B(t)G_{A,B}(t, \tilde{x}, 0, s, y) = 0, \quad \forall (t, \tilde{x}) \in [0, s[\times \mathbb{R}^{d-1}, \\ \lim_{t \to s} G_{A,B}(t, x, s, y) = \delta(x - y), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.80}$$

and

$$\begin{cases} [\partial_t + A(t)]P_{A,B}(t, x, s, \tilde{y}) = 0, \quad \forall (t, x) \in [0, s[\times \dot{\mathbb{R}}^d_+, \\ B(t)P_{A,B}(t, \tilde{x}, 0, s, \tilde{y}) = \delta(t - s, \tilde{x} - \tilde{y}), \quad \forall (t, \tilde{x}) \in [0, s[\times \mathbb{R}^{d-1}, \\ \lim_{t \to s} P_{A,B}(t, x, s, \tilde{y}) = 0, \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.81}$$

where the meaning of the δ measures is given by the representation formula

$$\begin{cases} u(t, x) = \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^d_+} G_{A,B}(t, x, s, y) f(s, y) \mathrm{d}y + \\ \qquad + \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t, x, s, \tilde{y}) \psi(s, \tilde{y}) \mathrm{d}\tilde{y} + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_+} G_{A,B}(t, x, T, y) \varphi(y) \mathrm{d}y, \end{cases} \tag{7.82}$$

where $u(t, x)$ is the solution of (7.72) with $\rho = 0$ and smooth (bounded and Hölder continuous) data f, ψ and φ satisfying the compatibility condition

$$B(T)\psi(T, \tilde{x}) + \varphi(\tilde{x}, 0) = 0,$$

for any $\tilde{x}$ in $\mathbb{R}^{d-1}$.

As mentioned early, since $\rho = 0$ we have

$$P_{A,B}(t, x, s, \tilde{y})\, 2b_d(s, \tilde{y}) = G_{A,B}(t, x, s, \tilde{y}, 0)\, a_{dd}(s, \tilde{y}, 0), \tag{7.83}$$

for any $0 \leq t < s \leq T$, x in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$. Contrary to the Dirichlet case, the Green and the Poisson functions have the same degree of singularity.

Namely, both $\partial^\ell G_{A,B}$ and $\partial^\ell P_{A,B}$ satisfy the heat-kernel estimates of order $\ell = 0, 1, 2$. Moreover, also there exist two constants $C \geq c > 0$ such that for every $0 \leq t < s \leq T$, x, y in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$, we have

$$\begin{cases} |B(t)G_{A,B}(t,x,s,y)| \leq C|s-t|^{\frac{1-d}{2}} \exp\Big(-c\dfrac{|x-y|^2}{s-t}\Big), \\ |B(t)P_{A,B}(t,x,s,\tilde{y})| \leq C|s-t|^{\frac{1-d}{2}} \exp\Big(-c\dfrac{|\tilde{x}-\tilde{y}|^2+x_d^2}{s-t}\Big), \end{cases} \tag{7.84}$$

which are rather clear when the boundary operator $B(t)$ does not contain tangential second-order terms, but very essential in the general case. As mentioned early, the constants C and c depend only on the Hölder continuous constant of the higher-order coefficients $(a_{ij} : i, j = 1, \dots, d)$, b_d, and the constants C_0 and c_0 in assumptions (7.36) and (7.39). Clearly, a companion Hölder type estimate to (7.84) also holds, i.e.,

$$\begin{aligned} &|B(t)G_{A,B}(t,x,s,y) - B(t')G_{A,B}(t',x',s',y')| \leq \\ &\qquad \leq C\big[|t-t'|^{\frac{\alpha}{2}} + |x-x'|^\alpha + |s-s'|^{\frac{\alpha}{2}} + |y-y'|^\alpha\big] \times \\ &\qquad\qquad \times\Big[|s-t|^{\frac{2-d-\ell-\alpha}{2}} \exp\Big(-c\frac{|x-y|^2}{s-t}\Big)\Big], \end{aligned} \tag{7.85}$$

for $|s-t| \leq |s'-t'|$, $|x-y| \geq |x'-y'|$ and $|\tilde{x}-\tilde{y}|^2 + x_d^2 \geq |\tilde{x}'-\tilde{y}'|^2 + x_d'^2$, where $0 < \alpha < 1$ is the Hölder exponent of the coefficients.

Wentzell Type

The general Wentzell type allows $\rho > 0$, and depending on whether b_d is positive we see two distinct situations. Again, the tangential first and second-order coefficients have little effect in the qualitative behavior. In this case, the Green function has a delta-measure concentrated on the boundary. Because ρ is strictly positive, it is convenient to factor out ρ, i.e., to write $\rho\hat{B} = B = B_1 - \rho A$ and use the boundary conditions given through $\hat{B} = \hat{B}_1 - A$, where the only non-tangential derivative in B_1 and $\hat{B}_1$ is $b_d\partial_d$.

First one finds the Green and Poisson functions corresponding to the problem (7.72) for the particular case where $\varphi(\tilde{x}, 0) = 0$, $\psi(T, \tilde{x}) = 0$ and $f(T, \tilde{x}, 0) = 0$. These particular Green and Poisson functions $\hat{G}_{A,\hat{B}}(t,x,s,y)$ and $\hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})$ are twice continuously differentiable in x and once in t and satisfy

$$\begin{cases} [\partial_t + A(t)]\hat{G}_{A,\hat{B}}(t,x,s,y) = \delta(t-s,x-y), \quad \forall (t,x) \in [0,s[\times\dot{\mathbb{R}}^d_+, \\ [\partial_t + \hat{B}_1(t)]\hat{G}_{A,\hat{B}}(t,\tilde{x},0,s,y) = 0, \quad \forall (t,\tilde{x}) \in [0,s[\times\mathbb{R}^{d-1}, \\ \lim\limits_{t\to s} \hat{G}_{A,\hat{B}}(t,x,s,y) = \delta(x-y), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.86}$$

and

$$\begin{cases} [\partial_t + A(t)]\hat{P}_{A,\hat{B}}(t,x,s,\tilde{y}) = 0, \quad \forall (t,x) \in [0,s[\times\dot{\mathbb{R}}^d_+, \\ [\partial_t + \hat{B}(t)]\hat{P}_{A,\hat{B}}(t,\tilde{x},0,s,\tilde{y}) = \delta(t-s,\tilde{x}-\tilde{y}), \quad \forall (t,\tilde{x}), \\ \lim\limits_{t\to s} \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y}) = 0, \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.87}$$

where the meaning of the δ measures is given by the representation formula

$$\begin{cases} u(t,x) = \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^d_+} \hat{G}_{A,\hat{B}}(t,x,s,y) f(s,y)\mathrm{d}y + \\ \qquad + \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^{d-1}} \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})\psi(s,\tilde{y})\mathrm{d}\tilde{y} + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_+} \hat{G}_{A,\hat{B}}(t,x,T,y)\varphi(y)\mathrm{d}y, \end{cases} \tag{7.88}$$

where $u(t,x)$ is the solution of (7.72) with $\rho > 0$ and smooth (bounded and Hölder continuous) data f, ψ and φ satisfying the restrictions $\varphi(\tilde{x},0) = 0$, $\psi(T,\tilde{x}) = 0$ and $f(T,\tilde{x},0) = 0$, for any $\tilde{x}$ in $\mathbb{R}^{d-1}$.

For simplicity we have used $\hat{B}$, but clearly, by using $\rho\partial_t + B_1$ instead of $\partial_t + \hat{B}_1$, we get the relation

$$\begin{aligned} &\hat{G}_{A,B}(t,x,s,y) = \hat{G}_{A,\hat{B}}(t,x,s,y), \\ &\hat{P}_{A,B}(t,x,s,\tilde{y})\,\rho(s,\tilde{y}) = \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y}), \quad \forall t,x,s,y,\tilde{y}. \end{aligned}$$

If b_d is strictly positive, i.e., $b_d(t,\tilde{x}) \geq c_0 > 0$ then

$$\hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})\,2b_d(s,\tilde{y}) = \hat{G}_{A,\hat{B}}(t,x,s,\tilde{y},0)\,a_{dd}(s,\tilde{y},0),$$

for any $0 \leq t < s \leq T$, x in $\mathbb{R}^d_+$ and $\tilde{y}$ in $\mathbb{R}^{d-1}$. However, if $b_d = 0$ then both kernels are independent, actually $\hat{G}_{A,\hat{B}} = G_{A,D}$ the Green function with Dirichlet boundary conditions. In any case, the complete Green and Poisson functions have the form

$$\begin{cases} P_{A,B}(t,x,s,\tilde{y}) = \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})\,\dfrac{1}{\rho(s,\tilde{y})}, \\ G_{A,B}(t,x,s,y) = \hat{G}_{A,\hat{B}}(t,x,s,y) + \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})\delta(y_d), \end{cases} \tag{7.89}$$

or equivalently

$$\begin{cases} P_{A,B}(t,x,s,\tilde{y}) = Q_{A,B}(t,x,s,\tilde{y},0), \\ G_{A,B}(t,x,s,y) = G_{A,D}(t,x,s,y) + Q_{A,B}(t,x,s,\tilde{y},0)\dfrac{2b_d(s,\tilde{y})}{a_{dd}(s,\tilde{y},0)} + \\ \qquad\qquad + Q_{A,B}(t,x,s,\tilde{y},0)\rho(s,\tilde{y})\delta(y_d), \end{cases}$$

for some suitable kernel, this means that two extra terms, namely,

$$\begin{cases} \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^{d-1}} \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y}) f(s,\tilde{y},0)\mathrm{d}\tilde{y} + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_+} \hat{P}_{A,\hat{B}}(t,x,T,\tilde{y})\varphi(\tilde{y},0)\mathrm{d}\tilde{y}, \end{cases} \tag{7.90}$$

to the representation formula (7.88).

Finally, both kernels $\partial^\ell \hat{G}_{A,\hat{B}}(t,x,s,y)$ and $\partial^\ell \hat{P}_{A,\hat{B}}(t,x,s,\tilde{y})$ satisfy the heat-kernel estimate of order $\ell = 0,1,2$, as well as the extra boundary estimates like (7.84) and (7.85).

Local Description

The construction and estimation of the Green and Poisson functions for parabolic system can be found in Eidelman [72], Ivasišen [114] and Solonnikov [233], in particular this applies to second-order parabolic equations. The parametrix technique goes as follows. First a model problem is considered in the whole space or the half-space with constant coefficients, where specific expressions are found for the fundamental, Green and Poisson functions. Next one considers parametric coefficients, i.e., the operators A and B with variable coefficients are frozen and, for an equation with constant coefficients, say $A(s,x)$ and $B(s,y)$ the previous expression (valid for constant coefficients) is used. For instance, denote by $(t,x) \mapsto G_0(s-t, x-y; s, y)$ one of those expression, either the fundamental, the Green or the Poisson functions with parametric coefficients in (s,y). Then, the kernel $G_0(t,x;s,y)$ satisfies the heat-estimates of a suitable order, namely, for any $\ell = 0, 1, 2, \dots$ there exist two constants $C_\ell \geq c_\ell > 0$ such that for every $0 < t \leq T$, s in $[0,T]$, x and y in $\mathbb{R}^d_+$ we have

$$|\partial^\ell G_0(t,x;s,y)| \leq C_\ell t^{\frac{2-d-\ell}{2}} \exp\Big(- c_\ell \frac{|x|^2}{t}\Big) \tag{7.91}$$

and

$$\left\{ \begin{aligned} &|\partial^\ell G_0(t,x;s,y) - \partial^\ell G_0(t,x;s',y')| \leq \\ &\qquad \leq C_\ell \big[|s-s'|^{\frac{\alpha}{2}} + |\tilde y - \tilde y'|^\alpha \big] \Big[t^{\frac{2-d-\ell}{2}} \exp\Big(- c_\ell \frac{|x|^2}{t}\Big)\Big], \end{aligned} \right. \tag{7.92}$$

where $0 < \alpha < 1$ is the Hölder exponent of the coefficients. Certainly, some changes are necessary for the Poisson function. Remark that there is an almost explicit expression (either directly of via a Fourier-Laplace transform) for the kernel G_0. Recall that $B(s,\tilde y) G_0(t,x;s,\tilde y,0) = P_{0,D}(t,x;s,\tilde y)$, the (parametrix) Poisson function with Dirichlet boundary conditions, and so the derivative $\partial^\ell B(s,\tilde y) G_0(t,x;s,\tilde y,0)$ satisfied suitable heat-kernel estimates, i.e., as above of order $\ell+1$, even if B includes second-order derivatives. Besides, all tangential derivative can be integrated by parts, which yields a cancellation property on the boundary.

At this point, one propose $G(t,x,s,y) = G_0(s-t, x-y; s, y) + G_1(t,x,s,y)$ to get the fundamental, Green or Poisson function for variable coefficients. It is found that $\partial^\ell G_1$ satisfies some heat-kernel estimates of order $\ell + \alpha$, i.e., an integrable singularity even for $\ell = 2$. Finally, one uses local coordinates to extend the result from the half-space to a smooth domain $\mathcal{O}$ of $\mathbb{R}^d$. Certainly, all this argument need a lot of complicate details, found in the above references.

7.3 Nonlocal Operator in Half-Space

In this section we follow the technique developed in the books Garroni and Menaldi [93, 94] to treat the case of a second-order integro-differential operator with Wentzell boundary conditions. Only a guide is given, full details are out of the scope of this work.

First we take a look at an integro-differential operator acting on the half-space $\mathbb{R}^d_+$,

$$\begin{cases} I(t)\varphi(x) = \displaystyle\int_{\mathbb{R}^m_*} \big[\varphi(x+\mathtt{j}(\zeta,t,x)) - \varphi(x) - \mathtt{j}(\zeta,t,x)\cdot\nabla\varphi(x)\big] \times \\ \qquad\qquad\qquad\qquad \times \mathtt{m}(\zeta,t,x)\pi(\mathrm{d}\zeta), \end{cases} \tag{7.93}$$

where the coefficients satisfy

$$\begin{cases} |\mathtt{j}(\zeta,t,x)|\,\mathbb{1}_{\{\mathtt{m}(\zeta,t,x)>0\}} \le \bar\jmath(\zeta), \qquad 0 \le \mathtt{m}(\zeta,t,x) \le 1, \\ \displaystyle\int_{\{\bar\jmath<1\}} [\bar\jmath(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{\bar\jmath\ge 1\}} \bar\jmath(\zeta)\pi(\mathrm{d}\zeta) \le C_0, \end{cases} \tag{7.94}$$

for every ζ, t, x and for some positive constants C_0, $0 \le \gamma < 2$ and a positive measurable function $\bar\jmath(\cdot)$. We also need Hölder continuity, i.e., there exist a positive measurable function (again denoted by) $\bar\jmath(\cdot)$ and some constant $M_0 > 0$ such that for any t, t' x, x' and ζ we have

$$\begin{cases} |\mathtt{j}(\zeta,t,x) - \mathtt{j}(\zeta,t',x')| \le \bar\jmath(\zeta)[|t-t'|^{\alpha/2} + |x-x'|^\alpha], \\ |\mathtt{m}(\zeta,t,x) - \mathtt{m}(\zeta,t',x')| \le M_0[|t-t'|^{\alpha/2} + |x-x'|^\alpha], \\ \displaystyle\int_{\{\bar\jmath<1\}} [\bar\jmath(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{\bar\jmath\ge 1\}} \pi(\mathrm{d}\zeta) \le M_0. \end{cases} \tag{7.95}$$

We need to assume that $\mathtt{j}(\zeta,t,x)$ is continuously differentiable in x for any fixed ζ, and that there exist a constant $c_0 > 0$ such that

$$c_0|x-x'| \le |(x-x') + \theta[\mathtt{j}(\zeta,t,x) - \mathtt{j}(\zeta,t,x')]| \le c_0^{-1}|x-x'|, \tag{7.96}$$

for any t, x, x' and $0 \le \theta \le 1$.

Now, a (complete) second-order integro-differential is the combination of the $A(t)$ in the previous Section 7.2 with $I(t)$, called again $A(t)$, i.e.,

$$\begin{cases} A(t)\varphi(x) := A_2\varphi(x) + I(t)\varphi(x), \\ A_2(t)\varphi(x) := A_0(t)\varphi(x) + \displaystyle\sum_{i=1}^d a_i(t,x)\partial_i\varphi(x) - a_0(t,x)\varphi(x), \\ A_0(t)\varphi(x) := \frac{1}{2}\displaystyle\sum_{i,j=1}^d a_{ij}(t,x)\partial_{ij}\varphi(x), \end{cases} \tag{7.97}$$

where the coefficients $a_0 \ge 0$, $a_{ij} = a_{ji}$ for any i, j are bounded, Hölder continuous and uniformly elliptic, i.e., (7.36) and (7.37) are satisfied.

To setup the problem in the half-space, we need to localize the integral part of $I(t)$, to make it acting on functions defined only on the half-space $\mathbb{R}^d_+$. In this case, the condition is simple, namely

$$[x_d + \mathtt{j}_d(\zeta,t,x)]\,\mathbb{1}_{\{\mathtt{m}(\zeta,t,x)>0\}} \ge 0, \quad \forall (\zeta,t,x) \in \mathbb{R}^m_* \times [0,T] \times \mathbb{R}^d_+. \tag{7.98}$$

i.e., all jumps $x+\mathtt{j}(\zeta,t,x)$ from x in $\mathbb{R}^d_+$ should belongs to $\mathbb{R}^d_+$. Alternatively, one may use an extension operator $\varphi \mapsto P\varphi$ and to re-define $I(t)\varphi(x) = I(t)(P\varphi)(x)$, where certainly $(P\varphi)(x) = \varphi(x)$ for every x in $\mathbb{R}^d_+$. For instance, for Dirichlet boundary conditions, we may use a zero-extension, but some extra difficulties appear due to the fact that such a zero-extension preserves Lipschitz continuous functions (vanishing on the boundary), but not necessarily C^2 functions, see the discussion in Section 2.11.

Technically, one may add an integro-differential part to the Wentzell boundary operator, but again some delicate issues appear. For instance, the interested reader may check the book by Skubachevskii [231].

Consider the boundary operator $B(t)$ given by (7.38) and satisfying (7.39), (7.40). Under the previous assumptions, our interest is on the boundary value problem with terminal condition,

$$\begin{cases} \partial_t u(t,x) + A(t)u(t,x) = f(t,x), \quad \forall (t,x) \in [0,T[\times \dot{\mathbb{R}}^d_+, \\ B(t)u(t,x) + \psi(t,x) = 0, \quad \forall (t,x) \in [0,T[\times \partial\mathbb{R}^d_+, \\ u(T,x) = \varphi(x), \quad \forall x \in \dot{\mathbb{R}}^d_+, \end{cases} \tag{7.99}$$

and the representation formula

$$\begin{cases} u(t,x) = \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^d_+} G_{A,B}(t,x,s,y) f(s,y)\mathrm{d}y + \\ \qquad + \displaystyle\int_t^T \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{A,B}(t,x,s,\tilde{y})\psi(s,\tilde{y})\mathrm{d}\tilde{y} + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d_+} G_{A,B}(t,x,T,y)\varphi(y)\mathrm{d}y, \end{cases} \tag{7.100}$$

where $G_{A,B}$ and $P_{A,B}$ are the Green and the Poisson functions. As in the purely differential case, if $\rho > 0$ then the Green function $G_{A,B}$ contains a delta-measure on the boundary, and the above formula is suitable modified.

7.3.1 Constant Coefficients

For a general constant coefficients non-local operator I, we can take $\mathtt{j}(\zeta,t,x) = \mathtt{j}(\zeta)$ and $\mathtt{m}(\zeta,t,x) = \mathtt{m}(\zeta)$,

$$I\varphi(x) = \int_{\mathbb{R}^m_*} \big[\varphi(x+\mathtt{j}(\zeta)) - \varphi(x) - \mathtt{j}(\zeta)\cdot\nabla\varphi(x)\big]\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta).$$

Clearly, we can approximate I with

$$I_\varepsilon\varphi(x) = \int_{|\mathtt{j}(\zeta)|\geq\varepsilon} \big[\varphi(x+\mathtt{j}(\zeta)) - \varphi(x) - \mathtt{j}(\zeta)\cdot\nabla\varphi(x)\big]\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta), \tag{7.101}$$

as $\varepsilon \to 0$, and the assumption (7.94) reduces to $0 \leq \mathtt{m}(\zeta) \leq m_0$ and

$$\int_{\{|\mathtt{j}(\zeta)|<1\}} [\mathtt{j}(\zeta)]^\gamma \pi(\mathrm{d}\zeta) + \int_{\{|\mathtt{j}(\zeta)|\geq 1\}} |\mathtt{j}(\zeta)|\pi(\mathrm{d}\zeta) \leq C_0, \tag{7.102}$$

for some constants $C_0 > 0$ and $0 \le \gamma < 2$.

First, if $F_0(t,x)$ is the fundamental function corresponding to the constant coefficients differential operator A_2 then one may propose $F = F_0 + F_0 \star Q$ where $\star$ means the kernel convolution in $[0,\infty) \times \mathbb{R}^d$, i.e.,

$$[\varphi \star \phi](t,x) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \varphi(t-s, x-y)\psi(s,y)\mathrm{d}y, \quad \forall t \ge 0,\, x \in \mathbb{R}^d.$$

This yields a Volterra equation for either F or Q, namely,

$$F = F_0 + F_0 \star IQ, \qquad Q = IF_0 + IF_0 \star Q.$$

If $Q_0 = IF_0$ then the formal series

$$F = \sum_{k=0}^{\infty} F_k, \quad F_k = F_0 \star IF_{k-1} \quad \text{and} \quad Q = \sum_{k=0}^{\infty} Q_k, \quad Q_k = Q_0 \star Q_{k-1},$$

represent the (unique) solutions. Clearly, the presence of the non-local operator I makes disappear the heat-kernel type estimates and the difficulty is the convergence of these series.

For the particular case

$$I\varphi(x) = \mathtt{m}[\varphi(x+\mathtt{j}) - \varphi(x)],$$

one can calculate explicitly the solution. Indeed, by means of the identity

$$F_0(t+s,x) = \int_{\mathbb{R}^d} F_0(t, x-z)\, F_0(s,z)\mathrm{d}z = [F_0(t,\cdot) * F_0(s,\cdot)](x)$$

we get

$$F_k(t,\cdot) = \frac{t^k}{k!} I^k F_0(t,\cdot) \quad \text{and} \quad Q_k(t,\cdot) = \frac{t^k}{k!} I^{k+1} F_0(t,\cdot), \tag{7.103}$$

where

$$I^k\varphi(x) = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} \varphi(x + i\mathtt{j})\mathtt{m}^k.$$

Hence

$$\begin{aligned} F(t,x) \quad &= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} \frac{(\mathtt{m}t)^k}{k!} F_0(t, x+i\mathtt{j}) = \\ &= e^{-\mathtt{m}t} \sum_{k=0}^{\infty} \frac{(\mathtt{m}t)^k}{k!} F_0(t, x+k\mathtt{j}), \quad \forall t > 0,\, x \in \mathbb{R}^d. \end{aligned}$$

A posteriori, we can check the convergence of the series (and all its derivatives), but the heat-kernel estimates are lost, there are many singular points, not just the origin.

Now for a general I, we approximate I with I_ε to establish the relation (7.103), but the expression of the power of I is more complicate, and simple explicit calculations are not longer possible. As studied in Garroni and Menaldi [93], several semi-norm are introduced, but only two are necessary (with a slight change of notation) for the constant coefficient case, namely, for any kernel $\varphi(t,x)$ and k real (usually non-negative) we define

$$\begin{cases} K_0(\varphi,k) = \inf\big\{K_0 \ge 0 \,:\, |\varphi(t,x)| \le K_0 t^{-1+(k-d)/2},\ \forall t,x\big\}, \\ K_n(\varphi,k) = \inf\big\{K_n \ge 0 \,:\, \displaystyle\int_{\mathbb{R}^n} |\varphi(t,\tilde x^{(n)},x^{(n)})|\mathrm{d}\tilde x^{(n)} \le \\ \qquad\qquad \le K_n t^{-1+(k-d+n)/2},\ \forall t, x^{(n)}\big\}, \\ K_d(\varphi,k) = \inf\big\{K_d \ge 0 \,:\, \displaystyle\int_{\mathbb{R}^d} |\varphi(t,x)|\mathrm{d}x \le K_d t^{-1+k/2},\ \forall t\big\}, \\ K(\varphi,k) = \max\big\{K_0(\varphi,k),\dots,K_d(\varphi,k)\big\}, \end{cases} \tag{7.104}$$

where $x = (\tilde x^{(n)}, x^{(n)})$, $\tilde x^{(n)} = (x_1,\dots,x_n)$, $x^{(n)} = (x_{n+1},\dots,x_d)$, $n = 1,\dots,d-1$. Actually, K_0, K_{d-1} and K_d are the most relevant semi-norms. In view of the heat-kernel estimates satisfied by the fundamental function F_0, the semi-norm $K(\partial^\ell F_0, 2-\ell)$ is finite, for any $\ell = 0,1,\dots$, where ∂^ℓ denotes any derivative of order ℓ_x in x and order ℓ_t in t with $\ell = 2\ell_t + \ell_x$.

The constant $0 \le \gamma < 2$ in assumption (7.102) plays an important role. Indeed, for $0 \le \gamma \le 1$, the operator I may have a simpler form, i.e,

$$I\varphi(x) = \int_{\mathbb{R}^m_*} \big[\varphi(x + \mathtt{j}(\zeta)) - \varphi(x)\big]\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta),$$

which yields

$$K(I\varphi, 1+k-\gamma) \le m_0 C_0\,[K(\varphi,1+k) + K(\nabla\varphi,k)].$$

Similarly, for $1 \le \gamma \le 2$, by means of the expression

$$I\varphi(x) = \int_0^1 \mathrm{d}\theta \int_{\mathbb{R}^m_*} \mathtt{j}(\zeta)\cdot\big[\nabla\varphi(x+\theta\mathtt{j}(\zeta)) - \nabla\varphi(x)\big]\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta),$$

we get

$$K(I\varphi, 2+k-\gamma) \le m_0 C_0\,[K(\nabla\varphi,1+k) + K(\nabla^2\varphi,k)].$$

Hence $K(\partial^\ell F_k, 2+k(2-\gamma)-\ell)$ is finite, and moreover, setting

$$K^2(\varphi,k) = \max\big\{K(\varphi,2+k),\, K(\nabla\varphi,1+k),\, K(\nabla^2\varphi,k)\big\},$$

and using the fact that

$$F_0(t+s,\cdot) = F_0(t,\cdot) * F_0(s,\cdot), \quad \text{and} \quad IF_0(t+s,\cdot) = IF_0(t,\cdot) * F_0(s,\cdot),$$

we have the estimate (as in Garroni and Menaldi [93, Chapter 3]) CAN WE DO A BETTER/SIMPLER EXPLANATION THAN IN G-M FOR THIS CASE?

$$K^2(\partial^\ell F_k, k(2-\gamma)-\ell) \le C_1^k (k!)^{-(2-\gamma)}\, K^2(\partial^\ell F_0, -\ell),$$

for any k and ℓ and for some constant C_1, depending only on the dimension d and the constants C_0 and m_0 in (7.102). Clearly, this provides the convergence (and estimates) of the series defining the fundamental solution F corresponding to the second-order integro-differential operator A with constant coefficients.

Now, to construct the Green and Poisson functions for constant coefficients we are in almost the same situation as in Subsection 7.1.3. Clearly, by imposing the following restriction

$$\mathtt{j}_d(\zeta) \geq 0, \quad \forall \zeta \in \mathbb{R}^m_*, \tag{7.105}$$

we are able to localize the operator I to $\mathbb{R}^d_+$.

Dirichlet Conditions

To simplify the presentation, we take $A = \frac{1}{2}\Delta$ as in Subsection 7.1.3 and by means of the procedure discussed in Subsection 7.1.4 we can add a first-order and zero-order constant differential part. Thus, as in the purely differential case, the Poisson function have the form

$$P_{I,D}(t, x, \tilde{y}) = \frac{x_d}{t} F(t, \tilde{x} - \tilde{y}, x_d), \tag{7.106}$$

where $F(t, \tilde{x}, x_d) = F(t, x)$ for any $x = (\tilde{x}, x_d)$ is the fundamental solution constructed above, i.e., the series

$$F = \sum_{k=0}^{\infty} F_k, \qquad F_k(t, \cdot) = \frac{t^k}{k!} I^k F_0(t, \cdot),$$

with F_0 the heat-kernel.

Indeed, we need to check that

$$\int_0^t \frac{x_d}{s} \mathrm{d}s \int_{\mathbb{R}^{d-1}} F_k(s, \tilde{y}, x_d) \psi(t - s, \tilde{x} - \tilde{y}) \mathrm{d}\tilde{y} \to 0 \quad \text{as} \quad x_d \to 0, \tag{7.107}$$

for any $k = 1, 2, \ldots$. By means of the explicit expression of the heat-kernel F_0 and the condition (7.105) we deduce

$$\int_{\mathbb{R}^{d-1}} |I F_0(t, \tilde{y}, x_d)| \mathrm{d}\tilde{y} \leq C_1 t^{-1-\gamma/2} \exp\big(-\frac{x_d^2}{2t}\big),$$

which yields (7.107) for $k = 1$. Similarly, we have

$$\int_{\mathbb{R}^{d-1}} |F_k(t, \tilde{y}, x_d)| \mathrm{d}\tilde{y} \leq C_k t^{-1+k(2-\gamma)/2} \exp\big(-\frac{x_d^2}{2t}\big),$$

for any $k = 2, \ldots,$ and then

$$\int_0^t \frac{x_d}{s} \mathrm{d}s \int_{\mathbb{R}^{d-1}} P_{I,D}(s, \tilde{y}, x_d) \psi(t - s, \tilde{x} - \tilde{y}) \mathrm{d}\tilde{y} \to \psi(t, \tilde{x}),$$

as $x_d \to 0$, for any smooth function ψ.

Next, we have to check that $P_{I,D}$ solves the homogeneous integro-differential equation in $\mathbb{R}^d_+$, i.e.,

$$[\partial_t - \frac{1}{2}\Delta]P_{I,D}(t,x) = 0, \quad \forall t > 0,\, x \in \mathbb{R}^d_+. \tag{7.108}$$

To this purpose, it suffices to show the above equation for I_ε instead of I, see (7.101). Moreover, we can move the differential part to A_2, i.e., add a first order coefficients $\underline{a}$ of the form

$$\underline{a} = \int_{\{|\mathtt{j}(\zeta)|\geq\varepsilon\}} \mathtt{j}(\zeta)\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta),$$

so that we are reduced to the case

$$I\varphi(x) = \int_{\mathbb{R}^m_*} \big[\varphi(x+\mathtt{j}(\zeta)) - \varphi(x)\big]\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta),$$

under the condition $0 \leq \mathtt{m}(\zeta) \leq m_0$ and

$$\int_{\mathbb{R}^m_*} |\mathtt{j}(\zeta)|\pi(\mathrm{d}\zeta) \leq C_0,$$

actually, $\mathtt{j}(\zeta)$ is bounded and $\pi(\mathbb{R}^m_*) < \infty$, and $A_2 = \frac{1}{2}\Delta + \underline{a}\cdot\nabla$.

Thus, define $\varphi_d(x) = x_d$ and

$$I'_d\varphi(x) = \int_{\mathbb{R}^m_*} \varphi(x+\mathtt{j}(\zeta))\mathtt{j}_d(\zeta)\mathtt{m}(\zeta)\pi(\mathrm{d}\zeta)$$

to see that

$$[I(\varphi_d\varphi)](x) = \varphi_d(x)I\varphi(x) + I'_d\varphi(x).$$

Hence

$$\big[a_dF_0(t,x) - \partial_dF_0(t,x)\big] = \varphi_d(x)F_0(t,x)$$

and therefore

$$\big[a_dtIF_0(t,x) - tI\partial_dF_0(t,x)\big] = \varphi_d(x)IF_0(t,x) + I'_dF_0(t,x).$$

Again

$$I\big[a_dtIF_0(t,x) - tI\partial_dF_0(t,x)\big] = \varphi_d(x)I^2F_0(t,x) + II'_dF_0(t,x) + I'_dIF_0(t,x)$$

and because I and I'_d commute, we deduce by induction

$$I^{k-1}\big[a_dtIF_0(t,x) - tI\partial_dF_0(t,x)\big] = \varphi_d(x)I^kF_0(t,x) + kI'_dI^{k-1}F_0(t,x),$$

for any $k \geq 1$. This proves that

$$\big[a_dF_k(t,x) - \partial_dF_k(x)\big] = \frac{x_d}{t}F_k(x) + I'_dF_{k-1}(x), \quad \forall k \geq 1.$$

Next, by means of the series (7.103) we obtain

$$\big[a_d F(t,x) - \partial_d F(t,x)\big] = \frac{x_d}{t} F(t,x) + I'_d F(t,x), \quad \forall t>0,\ x\in\mathbb{R}^d. \tag{7.109}$$

Since F, $\partial_d F$ and $I'_d F$ are solutions of the homogeneous equation then $x_d F(t,x)/t$ is also a solution, i.e., the equation (7.108) is satisfied and $P_{I,D}$ is indeed the Poisson function associated to $\frac{1}{2}\Delta + I$ with Dirichlet boundary conditions.

Now, to find the Green function we need to calculate the integral

$$V(t,\tilde{x},x_d,y_d) = \int_0^t \frac{x_d}{t-s}\mathrm{d}s \int_{\mathbb{R}^{d-1}} F(t-s,\tilde{x}-\tilde{y},x_d)F(s,\tilde{y},y_d)\mathrm{d}\tilde{y},$$

where the most singular term, namely

$$\int_0^t \frac{x_d}{t-s}\mathrm{d}s \int_{\mathbb{R}^{d-1}} F_0(t-s,\tilde{x}-\tilde{y},x_d)F_0(s,\tilde{y},y_d)\mathrm{d}\tilde{y} = F_0(t,\tilde{x},x_d+y_d),$$

is computed exactly, and all other lower order terms can be estimated with the semi-norms

$$\begin{cases} K_0(\varphi,k) = \inf\big\{K_0\geq 0 \,:\, |\varphi(t,\tilde{x},x_d)| \leq \\ \qquad\qquad \leq K_0 t^{-1+\frac{k-d}{2}} \exp\big(-\dfrac{x_d^2}{3t}\big),\ \forall t,\tilde{x},x_d\big\}, \\ K_{d-1}(\varphi,k) = \inf\big\{K_{d-1}\geq 0 \,:\, \displaystyle\int_{\mathbb{R}^{d-1}} |\varphi(t,\tilde{x},x_d)|\mathrm{d}x \leq \\ \qquad\qquad \leq K_{d-1} t^{-1+\frac{k-1}{2}} \exp\big(-\dfrac{x_d^2}{3t}\big),\ \forall t,x_d\big\}, \\ K(\varphi,k) = \max\big\{K_0(\varphi,k), K_{d-1}(\varphi,k)\big\}, \end{cases} \tag{7.110}$$

as in the case of the fundamental solution.

As mentioned early, we can add constant first-order and zero-order terms $(a_i : i=1,\dots,d)$ and a_0 to the operator A. Moreover, we can studied the dependency on the coefficients of the Green and Poisson function to be able to discuss the parametrix method.

An alternative way to construct the Green function corresponding to the second-order integro-differential operator $A = A_2 + I$ with Dirichlet boundary conditions in $\mathbb{R}^d_+$ is to proceed as in the case of the fundamental solution, by solving the following Volterra equation for either G or Q, namely,

$$G = G_0 + G_0 \star IQ, \qquad Q = IG_0 + IG_0 \star Q,$$

where G_0 is the Green function corresponding to the purely differential part A_2, and now the (non-commutative) kernel convolution $\star$ is in $[0,\infty]\times\mathbb{R}^d_+$, i.e.,

$$(\varphi\star\psi)(t,\tilde{x},x_d,y_d) = \int_0^t \mathrm{d}s \int_{\mathbb{R}^d_+} \varphi(t-s,\tilde{x}-\tilde{z},x_d,z_d)\psi(s,\tilde{z},z_d,y_d)\mathrm{d}z,$$

with $z = (\tilde{z}, z_d)$. If $Q_0 = IG_0$ then the formal series

$$G = \sum_{k=0}^{\infty} G_k, \quad G_k = G_0 \star IG_{k-1} \quad \text{and} \quad Q = \sum_{k=0}^{\infty} Q_k, \quad Q_k = Q_0 \star Q_{k-1}.$$

provide the (unique) solutions. Clearly, we do have the property

$$G_0(t+s, \tilde{x}, x_d, y_d) = \int_{\mathbb{R}^d_+} G_0(t, \tilde{x} - \tilde{z}, x_d, z_d) G_0(s, \tilde{z}, z_d, y_d) \mathrm{d}z,$$

but we do not have anymore the equality $I(\varphi \star \psi) = \varphi \star (I\psi)$, which would yield $G_k(t, \cdot) = t^k I^k G_0(t, \cdot)/k!$. On the other hand, we need to modify the definition of the semi-norms, namely, for kernels $\varphi(t, x, y_d)$,

$$\begin{cases} K_0(\varphi, k) = \inf \big\{ K_0 \geq 0 \,:\, |\varphi(t, x, y_d)| \leq K_0 t^{-1+(k-d)/2},\ \forall t, x, y_d \big\}, \\ K_d(\varphi, k) = \inf \big\{ K_d \geq 0 \,:\, \displaystyle\int_{\mathbb{R}^d_+} |\varphi(t, \tilde{z}, z_d, y_d)| \mathrm{d}z + \\ \qquad + \displaystyle\int_{\mathbb{R}^d_+} |\varphi(t, \tilde{z}, x_d, z_d)| \mathrm{d}z \leq K_d t^{-1+k/2},\ \forall t, x_d, y_d \big\}, \\ K(\varphi, k) = \max \big\{ K_0(\varphi, k), K_d(\varphi, k) \big\}, \end{cases} \tag{7.111}$$

where $x = (\tilde{x}, x_d)$. Also, we may use semi-norms of the type $K_n(\cdot, \cdot)$, for $n = 1, \ldots, d-1$. Moreover, estimating the G is simple, but it may be complicate to handle $\partial^\ell G$ for $\ell = 1, 2, \ldots$. However, in view of the condition (7.105), the semi-norm

$$\begin{aligned} K_0(\varphi, k) = \inf \big\{ K_0 \geq 0 \,:\, |\varphi(t, \tilde{x}, x_d, y_d)| \leq \\ \leq K_0 t^{-1+\frac{k-d}{2}} \exp\Big[-\frac{(x_d - y_d)^2}{3t} \Big],\ \forall t, x_d, y_d \big\}, \\ K_{d-1}(\varphi, k) = \inf \big\{ K_{d-1} \geq 0 \,:\, \int_{\mathbb{R}^{d-1}} |\varphi(t, \tilde{x}, x_d, y_d)| \mathrm{d}x \leq \\ \leq K_{d-1} t^{-1+\frac{k-1}{2}} \exp\Big[-\frac{(x_d - y_d)^2}{3t} \Big],\ \forall t, x_d, y_d \big\}, \end{aligned}$$

can be used, and one can essentially redo the calculations of the purely differential case. Thus the Green function $G = G_{A,D}$ has almost the same estimates as the fundamental solution. If the Green function is found first, then one may obtain the Poisson function as

$$P_{I,D}(t, x) = \partial_d^y G_{I,D}(t, x, 0),$$

where ∂_d^y is the derivative with respect to the last variable, i.e., y_d.

Wentzell Conditions

The arguments developed in Subsection 7.1.3 can be used in this context, i.e., we begin with the case $A = \frac{1}{2}\Delta + I$ and then we use the arguments in Subsec-

tion 7.1.4. Thus, for the constant coefficients, the functions

$$\begin{cases} P_{I,B}(t,x,\tilde{y}) = Q_{I,B}(t,\tilde{x}-\tilde{y},x_d), \\ G_{I,B}(t,x,y) = G_{I,D}(t,\tilde{x}-\tilde{y},x_d,y_d)+ \\ \qquad\qquad + 2b_d Q_{I,B}(t,\tilde{x}-\tilde{y},x_d+y_d) + \rho\delta_0(y_d)Q_{I,B}(t,\tilde{x}-\tilde{y},x_d), \end{cases}$$

are the Poisson and Green functions corresponding to the second-order integro-differential operator $A = \frac{1}{2}\Delta + I$ with constant coefficients and Wentzell boundary conditions B in $\mathbb{R}^d_+$, where the kernel $Q_{I,B}$ is given by the formula

$$\begin{cases} Q_{I,B}(t,x) = \displaystyle\int_0^{t/\rho} e^{-b_0 r}\mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\times \\ \qquad\qquad \times \dfrac{x_d + b_d r}{t-\rho r} F(t-\rho r, \tilde{x}-\tilde{\varsigma}\tilde{z}+\tilde{b}r, x_d + b_d r)\mathrm{d}\tilde{z}, \end{cases} \tag{7.112}$$

for $\rho > 0$, and

$$\begin{cases} Q_{I,B}(t,x) = -\partial_d \displaystyle\int_0^{\infty} e^{-b_0 r}\mathrm{d}r \int_{\mathbb{R}^{d-1}} \tilde{\Gamma}_0(r,\tilde{z})\times \\ \qquad\qquad \times \dfrac{x_d + b_d r}{t} F(t, \tilde{x}-\tilde{\varsigma}\tilde{z}+\tilde{b}r, x_d + b_d r)\mathrm{d}\tilde{z}, \end{cases} \tag{7.113}$$

for $\rho = 0$ and $b_d > 0$, where $F = F_I$ is the fundamental solution corresponding to $A = \frac{1}{2}\Delta + I$. Moreover, we have

$$BQ_{I,B}(t,x) = -\frac{x_d}{t}F(t,x) = P_{I,D}(t,x), \quad \forall t > 0,\ x \in \mathbb{R}^d_+,$$

where $P_{I,D}$ denotes the Poisson function with Dirichlet boundary conditions. Clearly, these equalities prove the heat-kernel estimates for the main singular part of $P_{I,B}$ and $G_{I,B}$, and the K semi-norms estimate for the non-local part. Moreover, even if B may contain second-order derivative the expression $BQ_{I,B}$ satisfies heat-kernel type estimates as $P_{I,D}$, i.e., a singularity comparable to first-order derivatives.

On the other hand, we could add a tangential non-local part to the boundary Wentzell operator B and use the arguments of the construction of the fundamental solution F_A in term of a series, i.e., add the integro-differential operator

$$J\varphi(x) = \int_{\mathbb{R}^m_*} \big[\varphi(\tilde{x}+\tilde{\jmath}(\zeta)) - \varphi(\tilde{x}) - \tilde{\jmath}(\zeta)\cdot\tilde{\nabla}\varphi(\tilde{x})\big]\pi(\mathrm{d}\zeta).$$

to B. In this case, we have to modify the expression (7.112) and (7.113) defining the kernel $Q_{I,B}$, which would be given in tern of a series similar to the one used to express the fundamental solution F_A. Again, as mentioned early, we can add constant first-order and zero-order terms $(a_i : i = 1,\dots,d)$ and a_0 to the operator A.

7.3.2 Variable Coefficients

This is more complicate, we are going to give only some indications of how this can be proven using the technique in the books Garroni and Menaldi [93, 94], where the read can find full details for the case of oblique boundary conditions in a smooth region of $\mathbb{R}^d$. Moreover, by means of local coordinates, the case of the half-space $\mathbb{R}^d_+$ is the reference, i.e., to which the structure is locally isomorphic. Note that the Poisson function obtained form the Green function in the various cases. The only case that may require some special attention for the construction of the Poisson function is the Dirichlet boundary conditions where the coefficients are only Hölder continuous, in this case, one has to redo suitable estimates.

The assumptions are as stated in the beginning of this section and we assume $\gamma + \alpha < 2$ to simplify the estimates, in any case we do suppose $0 \le \gamma < 2$. Then, we follows the guidelines of Section 7.2.3 on the successive approximation method, but we need to use various semi-norms to solve the corresponding Volterra equations.

7.3.3 Density and Processes

THIS SHOULD BE A CONTINUATION OF SECTION 7.2.1 ON GREEN AND POISSON REPRESENTATION, WITH THE EQUATION WITH TERMINAL CONDITION (instead of initial).

WHAT DO YOU THINK?

7.4 Forward-Backward Equations

May be interesting? If this section is added, then Section 4.5 (representation of martingales) may be moved as a subsection of this Backward Equations. This will make shorter Chapter 4, which is relatively long.

Check books Mao [165, Chapter 7, pp. 233–268] and Yong and Zhou [261, Chapter 7, pp. 345–400].

7.5 Function-Valued Integrals

Our main interest are the following situations:

Case 1 Let $\{w_k(t) : t \geq 0,\, k = 1, 2, \ldots\}$ be a sequence of independent standard real valued Wiener processes. Now, let $(Z, \mathcal{Z})$ be a measurable space, where the σ-algebra $\mathcal{Z}$ is countably generated with an increasing sequence $\{Z_n\}$ of subsets in $\mathcal{Z}$ such that $Z = \bigcup_n Z_n$. Moreover, let $\{p_k(B,t) : t \geq 0,\, B \in \mathcal{Z},\, k = 1, 2, \ldots\}$ be a sequence of independent real valued standard Poisson measures with (same) Levy measure $\pi(B) = \mathbb{E}\{p_k(B,t)\}/t$ defined on $(Z, \mathcal{Z})$, $\pi(B) < \infty$ for every B in $\mathcal{Z}_n = \{B \in \mathcal{Z} : B \subset Z_n\}$. Denote by $\{q_k(B,t) : t \geq 0,\, B \in \mathcal{Z},\, k = 1, 2, \ldots\}$ the Poisson martingale measure $q_k(B,t) = p_k(B,t) - t\pi(B)$. Furthermore, let $(\Omega, \mathbb{F}, P)$, $\mathbb{F} = \{\mathcal{F}(t) : t \geq 0\}$ be a filtered probability space where $w_k(\cdot)$ and $q_k(B,\cdot)$ are martingales relative to $\mathbb{F}$, for any B in $\mathcal{Z}_n$. Therefore the expression

$$M(t,x) = M_0 + \sum_k \int_0^t \sigma_k(s,x)\mathrm{d}w_k(s) + \sum_k \int_{Z\times]0,t]} \gamma_k(\zeta,s,x) q_k(\mathrm{d}\zeta, \mathrm{d}s),$$

where M_0 is a real-valued random variable $\mathcal{F}_0$ measurable, $\{\sigma_k(t,x) : t \geq 0,\, x \in \mathcal{O},\, k = 1, 2, \ldots\}$ and $\{\gamma_k(t,x) : t \geq 0,\, x \in \mathcal{O},\, k = 1, 2, \ldots\}$ are sequences of adapted and predictable (respectively) random fields such that

$$\sum_k \int_0^t |\sigma_k(s,x)|^2 \mathrm{d}s + \sum_k \int_0^t \mathrm{d}s \int_Z |\gamma_k(\zeta,s,x)|^2 \pi(\mathrm{d}\zeta) < \infty,$$

almost surely, for any $t > 0$ and x in $\mathcal{O}$.

Case 2 Clearly, this can be re-written in term of a separable Hilbert space K, with norm $|\cdot|_K$ and scalar product $(\cdot,\cdot)_K$. Given a real-valued $\mathcal{F}_0$-measurable random variable M_0, $(w(t) : t \geq 0)$ is cylindrical K-valued standard Wiener process, and $q(\zeta, \mathrm{d}s)$ is a K-valued standard Poisson martingale measure,

$$M_t(x) = M_0 + \int_0^t \big(\sigma(s,x), \mathrm{d}w(s)\big)_K + \int_{Z\times]0,t]} \big(\gamma_k(\zeta,s,x), q(\mathrm{d}\zeta, \mathrm{d}s)\big)_K,$$

where the random fields $(\sigma(s,x) : s \geq 0,\, x \in \mathcal{O})$ and $(\gamma(s,x) : s \geq 0,\, x \in \mathcal{O})$ are adapted and predictable, respectively, such that

$$\int_0^t |\sigma(s,x)|_K^2 \mathrm{d}s + \int_0^t \mathrm{d}s \int_Z |\gamma(\zeta,s,x)|_K^2 \pi(\mathrm{d}\zeta) < \infty,$$

almost surely, for any $t > 0$ and x in $\mathcal{O}$.

Case 3 In term of two separable Hilbert spaces H and K, and the space of Hilbert-Schmidt (linear) operators $L_{\text{HS}}(K,H)$, we are given almost the same elements, a H-valued $\mathcal{F}_0$-measurable random variable M_0, $(w(t) : t \geq 0)$ is cylindrical K-valued Wiener process, and $q(\zeta, \mathrm{d}s)$ is a cylindrical K-valued Poisson martingale measure, we have the expression

$$M_t = M_0 + \int_0^t \sigma(s)\mathrm{d}w(s) + \int_{Z\times]0,t]} \gamma(\zeta,s) q(\mathrm{d}\zeta, \mathrm{d}s),$$

where the $L_{\text{HS}}(K,H)$-valued processes $(\sigma(s) : s \geq 0)$ and $(\gamma(s) : s \geq 0)$ are adapted and predictable, respectively, such that

$$\int_0^t \|\sigma(s)\|_{\text{HS}}^2 \mathrm{d}s + \int_0^t \mathrm{d}s \int_Z \|\gamma(\zeta,s)\|_{\text{HS}}^2 \pi(\mathrm{d}\zeta) < \infty, \quad \forall t > 0,$$

almost surely, where $\|\cdot\|_{\text{HS}}$ denotes the Hilbert-Schmidt norm. Note that $\sigma^*(s)\sigma(s)$ and $\gamma^*(\zeta,s)\gamma(\zeta,s)$ are nuclear or trace-class operator on K.

It is clear that *Case 1* is really *Case 2* with $\mathbb{K}=\ell^2(\mathbb{R})$, or in general, *Case 2* becomes *Case 1* after a basis have been chosen. The interest of *Case 1* or *Case 2* is to treat B-valued martingales, where the Banach function space B is of the type $C^{m,\alpha}(\mathcal{O})$, $m\geq 0$, $0<\alpha<1$, or Sobolev spaces $W_0^{m,p}(\mathcal{O})$, with $m\geq 0$, $1<p<\infty$. For *Case 3*, the Hilbert space H is a Sobolev space, e.g., $H_0^m(\mathcal{O}$, with $m\geq 0$.

I need your help here...

7.5.1 Function-valued Martingales

In this section we continue the discussion initiated early on martingales, now considering processes $\{M(t):t\in T\}$ with values in some function spaces. To simplify the notation, we take $T=[0,\infty)$ and sometimes we write $(M_t:t\geq 0)$. Typical (separable) function spaces are Lebesgue spaces $L^p(\mathcal{O})$, $1\leq p<\infty$, Sobolev spaces (which are reflexive) and continuous-type spaces $C^k(\bar{\mathcal{O}})$, $k=0,1,\ldots$, where $\mathcal{O}$ is a domain in $\mathbb{R}^d$ (which are not reflexive). Other typical function spaces are the Hölder spaces $C^\alpha(\bar{\mathcal{O}})$, $0<\alpha<1$, which are nor separable neither reflexive. Note that, given a measure space $(\Omega,\mathcal{F},P)$ and a Banach space B, a Borel measurable function $f\colon\Omega\to B$ is limit a.e. of a sequence of (measurable) simple functions if and only if the image $f(\Omega\setminus N)$ is separable for some set N of P-measure zero, and so, a bounded and Borel measurable function may not be integrable.

Let H be an abstract separable Hilbert space with its inner product $(\cdot,\cdot)_H$ and norm $\|\cdot\|_H$, and B be a separable Banach space of functions (i.e., B is a subset of real or complex valued functions defined in some domain in $\mathbb{R}^d$) with its norm $\|\cdot\|_B$. It is clear that a vector (H or B) valued process $X=(X_t:t\in T)$ is a measurable map from a measurable space $(\Omega,\mathcal{F})$ into the Borel-product spaces $(H^T,\mathcal{B}^T(H))$ or $(B^T,\mathcal{B}^T(B))$. The total order defined in index set T is used to generate an increasing family of σ-algebras $(\mathcal{F}_t:t\in T)$ on $(\Omega,\mathcal{F})$. As soon as a probability is given on $(\Omega,\mathcal{F})$, the process X is identified with its P-equivalent class, and the probability P_X on $(H^T,\mathcal{B}^T(H))$ or $(B^T,\mathcal{B}^T(B))$ is defined as image of P through X. To realize (or to construct) a process X, we want to prescribe the values of P_X on a sufficiently large family of sets in $\mathcal{B}^T(H)$ or $\mathcal{B}^T(B)$ so that a (unique) probability measure P_X can be defined. In this sense, the initial probability measure P and the measurable map X are tied together. Thus, when discussing processes, we refer to either P or X, if the other item X or P is understood by default. However, as mentioned early, due essentially to the fact that the Borel product σ-algebra $\mathcal{B}^T(H)$ or $\mathcal{B}^T(B)$ is too small for the product space H^T or B^T, the above project requires a little correction, i.e., the initial (product) topology in B^T or H^T has to be changed when T is uncountable, i.e., a sub-space (such as continuous or cad-lag functions from T into H or B) of the spaces H^T or B^T is used. Within these sub-spaces, an element (or function defined in T) can be regarded as a unique extension of its restriction to any countable subset of the index set T. This procedure effectively

reconciliate the concepts of version and equality of stochastic processes.

All these points were discussed early in the chapter, however, the new *complication* now is the fact that the spaces H and B have (in general) no a finite dimension with at least two topologies (weak and strong). For instance, the interested reader may consult the book Métivier [178]
(who else here?)
for a carefully analysis.

The *martingale property* for a H or B valued process $(M_t : t \in T)$ involves an increasing family of (completed) σ-algebras $(\mathcal{F}_t : t \in T)$. However, even in the case of a space of functions, any sub σ-algebra of $\mathcal{F}$ is generated by B-valued random variables X, and as one may expect, the real-valued family of random variables $X(x)$, for x in $\mathcal{O}$ is disregarded in this context.

First, in the case of a function Banach space

$$\mathbb{E}\{|M_t(x)|\} < \infty, \ \forall t, \qquad \mathbb{E}\{M_t(x) \mid \mathcal{F}_s\} = M_s(x), \ \text{ a.s. } \ \forall t > s, \, x.$$

Since $M_t(x)$ is a real-valued process for each (or almost every) x in the domain $\mathcal{O}$, this last condition is clearly understood. Note that implicitly, one assume that $\mathbb{E}\{\|M_t\|_B\} < \infty$ implies $\mathbb{E}\{|M_t(x)|\} < \infty$ for every or almost every x, depending on whether B is a Lebesgue or a Hölder (or continuous-type) Banach space.

The conditional expectation can be considered as an orthogonal projection operator in the Hilbert space $L^2(H) = L^2(\Omega, \mathcal{F}, P; H)$, i.e., if $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$ and $L^2(\mathcal{G}, H) = L^2(\Omega, \mathcal{G}, P; H)$ then

$$\mathbb{E}\{ \cdot \mid \mathcal{G}\} \colon L^2(H) \longrightarrow L^2(\mathcal{G}, H),$$

$$\mathbb{E}\{\big(\mathbb{E}\{x \mid \mathcal{G}\}, y\big)_H\} = \mathbb{E}\{(x, y)_H\}, \quad \forall y \in L^2(\mathcal{G}, H).$$

However, the conditional expectation for Banach-valued random variables is defined as follows. For every simple function

$$f = \sum_{i=1}^{n} \mathbb{1}_{F_i} b_i, \quad b_i \in B, \ F_i \in \mathcal{F},$$

one defines

$$\mathbb{E}\{f \mid \mathcal{G}\} = \sum_{i=1}^{n} \mathbb{E}\{\mathbb{1}_{F_i} \mid \mathcal{G}\} b_i.$$

This defines a continuous linear mapping from a dense (by definition of the space L^1) linear subspace of $L^1(\mathcal{F}) = L^1(\Omega, \mathcal{F}, P; B)$ into $L^1(\mathcal{F}) = L^1(\Omega, \mathcal{G}, P; B)$ with norm less than or equal to 1. Clearly, this mapping can be uniquely extended to the whole space $L^1(\mathcal{F})$. Certainly, this definition agrees with the previous one.

On the other hand, as early noted, an important role is played by square-integrable martingales, i.e.,

$$\mathbb{E}\big\{\|M_t\|_H^2\big\} < \infty, \ \forall t, \qquad \mathbb{E}\{M_t \mid \mathcal{F}_s\} = M_s, \ \text{ a.s. } \ \forall t > s,$$

with the previous meaning. Since $t \mapsto \|M_t\|^2$ is a sub-martingale, one can define the quadratic optional (or predictable) variation processes $[M]_t$ (or $\langle M\rangle_t$) as H-valued processes and $[\|M\|]_t$ (or $\langle\|M\|\rangle_t$) as real-valued processes. Certainly, we may have a square integrable B-valued martingale if $\mathbb{E}\{|M_t|_B^2\} < \infty$ and we can localize these definitions by means of stopping times.

It is worth to note that even for the simple case of $H = \mathbb{R}^2$, we may have some difficulties (???)

Definition 7.6 (K-valued). TO BE ADJUSTED...
Let $(\Omega, \mathcal{F}, P)$ be a probability space, K be a separable Hilbert space with scalar product $(\cdot,\cdot)$, and R be a nuclear (trace-class) self-adjoint positive definite operator R on K. A continuous process $(w(t) : t \geq 0)$ with independent increments and values in K satisfying

$$\mathbb{E}\big\{\exp\big(\mathtt{i}(w(t),k)\big)\big\} = \exp\big(-\frac{t}{2}(Rk,k)\big), \quad \forall t>0,\, k\in K,$$

is called a K-valued Wiener process with covariance (operator) R. Similarly, given a measure π on $K_* = K \setminus \{0\}$ satisfying

$$\int_{K_*} (Rk,k)\pi(\mathrm{d}k) < \infty,$$

a cad-lag process $(q(t) : t \geq 0)$ with independent increments and values in K satisfying

$$\mathbb{E}\big\{\exp\big(\mathtt{i}(q(t),k)\big)\big\} = \\ = \exp\Big(t\int_{K_*}\big[\mathrm{e}^{\mathtt{i}(Rh,k)} - 1 - \mathtt{i}(Rh,k)\big]\pi(\mathrm{d}h)\Big), \quad \forall t>0,\, k\in K,$$

is called a K-valued compensated Poisson point (or purely jumps) process with Levy measure π_R, $\pi_R(B) = \pi(R^{-1}B)$, B in $\mathcal{B}(K_*)$. Analogously, let m be a σ-finite measure on K_* and $\mathcal{B}_0(K_*) = \{B \in \mathcal{B}(K_*) : m(B) < \infty\}$. A family $\{w(B,t) : t \geq 0,\, B \in \mathcal{B}_0(K_*)\}$ (or $\{p(B,t) : t \geq 0,\, B \in \mathcal{B}_0(K_*)\}$) of continuous (or cad-lag) processes with independent increments and values in K, indexed by B in $\mathcal{B}_0(K_*)$, is called a K-valued Wiener (or Poisson) measure with characteristic (R, m) if

$$\mathbb{E}\big\{\exp\big(\mathtt{i}(w(B,t),k)\big)\big\} = \exp\Big(\frac{t}{2}m(B)(Rk,k)\Big), \quad \forall t>0,\, k\in K,$$

for every B in $\mathcal{B}_0(K_*)$ (or

$$\mathbb{E}\big\{\exp\big(\mathtt{i}(p(B,t),k)\big)\big\} = \\ = \exp\Big(t\sum_{j=1}^{\infty} m(e_jB)\big[\mathrm{e}^{\mathtt{i}(Re_j,k)} - 1\big]\Big), \quad \forall t>0,\, k\in K,\, B\in\mathcal{B}_0(K_*),$$

where $\{e_j\}$ is an orthonormal basis in K and e_jB is the projection of B in the direction e_j, i.e., $e_jB = \{(b,e_j)e_j : b \in B\}$. Note that $m(B) = \sum_j m(e_jB)$ and

so the above series is convergent). The expression $\tilde{p}(B,t) = p(B,t) - \mathbb{E}\{p(B,t)\}$ (or $w(B,t)$ is called a Poisson (or Wiener) martingale (compensated) measure with values in K. □

Given an orthonormal basis $\{e_j\}$ in K and a measure m on K with $m(\{0\}) = 0$, one has $m(B) = \sum_j m(e_j B)$, where $e_j B$ is the projection of B in the direction e_j, i.e., $e_j B = \{(b, e_j)e_j : b \in B\}$. Moreover, the vector valued measure $\hat{m}(B) = \sum_j m(e_j B)e_j$ satisfies

$$\int_{K_*} \big(\hat{f}(k), \hat{m}(\mathrm{d}k)\big) = \sum_{j=1}^{\infty} \int_{K_*} f(k) m(e_j \mathrm{d}k) = \int_{K_*} f(h) m(\mathrm{d}h)$$

for any integrable function f with $\hat{f} = \sum_j f\big((e_j, k)\big)e_j$.

It is clear that the image space $K_1 = R^{1/2}K$ becomes a Hilbert space with the natural scalar product $(h,k)_1 = (R^{1/2}h, R^{-1/2}k)$, which is dense and continuously imbedded in K. Thus we have the triplet $K_1 \subset K = K' \subset K_1'$, where the prime $'$ denotes the dual space via Riesz' representation theorem. Moreover, as in Section 1.14 in Chapter 1, we may use the Hilbert space $L^2(]0,\infty[, K)$ instead of just K to describe the characteristic function, which would include the fact that the processes have independent increments, e.g.,

$$\mathbb{E}\Big\{\mathrm{e}^{\mathtt{i}\langle \dot{w}, k\rangle}\Big\} = \exp\Big(-\frac{1}{2}\int_0^{\infty} (Rk(t), k(t))\mathrm{d}t\Big),$$

for any k in $L^2(]0,\infty[, K)$. Similarly, for $q(t)$ one has

$$\mathbb{E}\Big\{\mathrm{e}^{\mathtt{i}\langle \dot{q}, k\rangle}\Big\} = \exp\Big(\int_0^{\infty} \mathrm{d}t \int_{K_*} \big[\mathrm{e}^{\mathtt{i}(Rh, k(t))} - 1 - \mathtt{i}(Rh, k(t))\big]\pi(\mathrm{d}h)\Big).$$

Next, after a proper justification (e.g., estimating the moments), one select a continuous or cad-lag version of $\langle \dot{w}, \mathbb{1}_{(0,t)}k\rangle$ or $\langle \dot{q}, \mathbb{1}_{(0,t)}k\rangle$. However, for $w(B,t)$ or $p(B,t)$ one should use $L^2_m(]0,\infty[\times K, K)$ (with $q = p - \mathbb{E}p$),

$$\mathbb{E}\Big\{\mathrm{e}^{\mathtt{i}\langle \dot{w}, \phi\rangle}\Big\} = \exp\Big(-\frac{1}{2}\int_0^{\infty} \mathrm{d}t \int_{K_*} \big(R\phi(t,k), \phi(t,k)\big) m(\mathrm{d}k)\Big),$$

or

$$\mathbb{E}\Big\{\mathrm{e}^{\mathtt{i}\langle \dot{q}, \phi\rangle}\Big\} = \\ = \exp\Big(\sum_{j=1}^{\infty} \int_0^{\infty} \mathrm{d}t \int_{K_*} \big[\mathrm{e}^{\mathtt{i}(Re_j, \phi(t,k))} - 1 - \mathtt{i}(Re_j, \phi(t,k))\big] m(e_j \mathrm{d}k)\Big),$$

where $m(e_j \mathrm{d}k)$ denotes the projection measure on the direction e_j. Then, after a proper justification (e.g., using the separability of the σ-algebras), one select a *measure* version of $\langle \dot{w}, \phi\rangle$ or $\langle \dot{q}, \phi\rangle$ with $\phi = k\mathbb{1}_{(0,t)}\mathbb{1}_B$.

Given K and R as above, by means of Sazonov's Theorem (e.g., Kallianpur and Xiong [123, Theorem 2.3.4, pp. 68–70]) one get a realization of the above Levy processes (Wiener and compensated Poisson point processes) and Poisson

measure. The condition that R is a trace-class operator is necessary and forbid the choice $R = I$, when K is not of finite dimension. However, the transformation $R^{1/2}$ applied to the previous K-valued processes provide the K_1'-valued processes $R^{-1/2}w(t)$ and $R^{-1/2}q(t)$, which are called *cylindrical* Wiener process and *cylindrical* compensated Poisson point process. Therefore, for any complete orthonormal basis $\{e_i : i = 1, 2, \ldots\}$, the projections $(w(t), e_i)$, $(q(t), e_i)$ and $(p(B,t), e_i)$ form independent sequences of real-valued standard Wiener process, (standard) compensated Poisson point (or purely jumps) process and (standard) Poisson measure, and the following series

$$\sum_i (w(t), e_i), \qquad \sum_i (q(t), e_i)$$

are almost surely convergent in K_1', for any $t > 0$ and B in $\mathcal{B}_0(K_*)$. Conversely, if $\{w_i\}$, $\{q_i\}$ and $\{p_i\}$ are sequences of independent real-valued standard Wiener processes, (standard) compensated Poisson point (or purely jumps) processes and (standard) Poisson measures, then the expressions

$$\sum_i w_i(t) e_i, \qquad \sum_i q_i(t) e_i$$

are almost surely convergent in K_1', for any $t > 0$ and B in $\mathcal{B}_0(K_*)$, and provided a realization of the above cylindrical K_1'-valued processes.

It should be clear the equivalence between (standard) compensated Poisson point (or purely jumps) processes and (standard) Poisson measures, as in the finite dimensional case. Indeed, if $p(B,t)$ is first given, then stochastic integral

$$\big(q(t), k\big) = \int_{K_* \times]0,t]} \big(\tilde{p}(\mathrm{d}k, \mathrm{d}s), k\big),$$

where $\tilde{p}(B,t) = p(B,t) - \mathbb{E}\{p(B,t)\}$, for any $t > 0$ and $B \in \mathcal{B}_0(K_*)$, yields the compensated jump process. Conversely, if $q(t)$ is first given, then the jumps $\{j(t) = q(t) - q(t-) : t > 0\}$ define the random measure

$$p(B,t) = \sum_{0<s\le t} j(s) \mathbb{1}_{\{j(s)\in B\}}, \quad t \ge 0,$$

for any B in $\mathcal{B}(K_0)$ separated form the origin. Note that $\mathbb{E}\{\big(p(B,t), k\big)\} = t\,(Rk, k)\,\pi(B)$, for any $t > 0$, k in K and B in $\mathcal{B}_0(K_*)$.

If the measure π on $\mathcal{B}(K_*)$ satisfies only

$$\int_{K_*} \big((Rk, k) \wedge 1\big)\pi(\mathrm{d}k) < \infty, \tag{7.114}$$

then we need to adjust the definition of a compensated Poisson point process, which is a particular case but its first or second moment may be infinite due to the long (or big) jumps. Moreover, we may consider a compound Poisson process with finite first moment, i.e., a Poisson measure with Levy measure π_R on $K_* = K \setminus \{0\}$ satisfying

$$\int_{K_*} \big(1 + \sqrt{(Rk, k)}\big)\pi(\mathrm{d}k) < \infty,$$

A realization of such a process can be simplified, since $p(t) = p(K_*, t)$ defines a K-valued process. Indeed let us consider a sequence of independent identically distributed $\mathbb{R} \times K_*$-valued random variables $\{(\tau_j, x_j) : j = 1, 2, \ldots\}$, where $\{\tau_j\}$ are exponentially distributed with parameter $\pi_R(K_*)$ and $\{x_j\}$ have $\pi_R(\cdot)/\pi_R(K_*)$ as law and is independent of τ_j. Since

$$\pi_R(K_*)\,\mathbb{E}\{|x_j|\} = \int_{K_*} \sqrt{(Rk,k)}\,\pi(\mathrm{d}k) < \infty,$$
$$\sum_j \mathbb{E}\{\mathbb{1}_{\{\tau_1+\cdots+\tau_j \le t\}}\} < \infty,$$

the inequality

$$\mathbb{E}\{|p(t)|\} \le \sum_j \mathbb{E}\{|x_j|\}\,\mathbb{E}\{\mathbb{1}_{\{\tau_1+\cdots+\tau_j \le t\}}\}$$

allows us to define

$$p(t) = \sum_j x_j\,\mathbb{1}_{\{\tau_1+\cdots+\tau_j \le t\}}, \quad \forall t \ge 0,$$

almost surely as a cad-lag (piecewise constant) process with values in K. In general, if only assumption (7.114) is satisfied then one approximate π_R with π_R^ε, $\pi_R^\varepsilon(B) = \pi_R(B \cap \{k : \varepsilon \le |k| \le 1/\varepsilon\})$, so that, first we apply the precedent construction to π_R^ε and next we let ε vanishing. The fact that R is a trace-class operator is used to insure the existence of a sequence of independent identically distributed K_*-valued random variables $\{x_j : j = 1, 2, \ldots\}$ with law $\pi_R(\cdot)/\pi_R(K_*)$. RIGHT?

.../////...

Definition 7.7 (fv-martingale)**.** TO BE ADJUSTED...
A (general) martingale with states in separable Banach space B of real functions (defined in some domain $\mathcal{O} \subset \mathbb{R}^d$) is a (complete) probability measure P on $(\Omega, \mathcal{F})$, together with a measurable mapping M (P-equivalence class) from $(\Omega, \mathcal{F})$ into $(B^T, \mathcal{B}^T(B))$ and an increasing family of completed σ-algebras $(\mathcal{F}_t : t \in T)$ on $(\Omega, \mathcal{F})$ satisfying the martingale property

$$\mathbb{E}\big\{|M_t(x)|\big\} < \infty,\ \forall t, \qquad \mathbb{E}\{M_t(x) \,|\, \mathcal{F}_s\} = M_s(x),\ \text{ a.s. } \forall t > s,$$

for each (or almost every) x in $\mathcal{O}$. If the family of σ-algebras $(\mathcal{F}_t : t \in T)$ is not mentioned, then it is assumed $(\mathcal{F}_t : t \in T)$ is the history $(\mathcal{H}_t : t \in T)$ of the process $(M_t : t \in T)$, i.e., $\mathcal{H}_t$ is generated by the random variables $\{M_s : s \le t\}$ and the null sets. Moreover, we say that the martingale is *cad-lag* if $(\mathcal{F}_t : t \in T)$ is a filtration satisfying the usual conditions and except on a set of P-probability zero, the paths of $(M_t(x) : t \in T)$ are cad-lag for every (or almost every) x. The martingale is *continuous* if their paths are continuous. Furthermore, since $M_t(x)$ takes values in $\mathbb{R}$, we may define also *super-* or *sub-*martingale by replacing the equal sign by either $\le$ or $\ge$ in the above condition. Finally, if the process M is valued in an abstract Hilbert $\mathbb{H}$ then only (general) square-integrable martingales are defined. □

Eventually, one can easily generalize the definition of martingales taking values into a separable dual space of an abstract Banach space and to a countably Hilbertian space.

Now, we need to know that we can always construct a version of M which is cad-lag. Essentially, one uses the ideas of real-valued sub-martingales for each x in a countable dense set of $\mathcal{O}$ and then specific properties of the topology in the Banach space of functions B to conclude. Certainly, this is obvious for square-integrable martingales with values in some abstract Hilbert space H. (Hope this is TRUE?)

RECALL to include the case where a random-martingale can be considered as a Hilbert-valued martingale, see G-K

PERHAPS RECALL to include the case of a countably Hilbert space valued stochastic integral, which includes Rigged Hilbert/Banach Spaces!...

Let $\mathcal{S} = \cap_{n=0}^{\infty} \mathcal{S}_n$ be a nuclear countably Hilbertian space as in Definition 2.5 of Chapter 2, i.e.,

$$\mathcal{S} \subset \cdots \subset \mathcal{S}_n \subset \cdots \subset \mathcal{S}_1 \subset \mathcal{S}_0 = \mathcal{S}_0' \subset \mathcal{S}_{-1} \subset \mathcal{S}_{-n} \subset \cdots \subset \mathcal{S}',$$

where all inclusions are continuous and dense, and the norm and the inner product in the space $\mathcal{S}_n$ and its dual space $\mathcal{S}_{-n}$ are denoted by $\|\cdot\|_n$, $(\cdot,\cdot)_n$, and $\|\cdot\|_{-n}$, $(\cdot,\cdot)_{-n}$, respectively.

7.5.2 Martingale Measures

Usually, (real-valued) martingales measures are obtained from (random) integer measures, as a part of the so-called random measures on Blackwell spaces, which are measurable spaces $(E, \mathcal{E})$ with the disintegration (or regular conditional probability) property, i.e., if X is any $(E, \mathcal{E})$-valued random variable on any probability space $(\Omega, \mathcal{F}, P)$ and if $\mathcal{G}$ is any sub σ-field of $\mathcal{F}$ then X admits a regular conditional probability with respect to $\mathcal{G}$. Integer measures provide a way of study the jumps of cad-lag processes, and the key point is the predictable compensators, which is a random measure. Thus, starting from an integer measure $\{\nu(B,t) : t \geq 0,\, B \in \mathcal{E}\}$ one get its predictable compensator $\{\nu^p(B,t) : t \geq 0,\, B \in \mathcal{E}\}$, and then $\tilde{\nu}(B,t) = \nu(B,t) - \nu^p(B,t)$ is a (real-valued, local) martingale measure. The most important example is the (real-valued) standard Poisson measure, i.e., an integer measure $p(B,t)$ with a deterministic predictable compensator of the form $\mathbb{E}\{p(B,t)\} = t\pi(B)$, where π is a σ-finite measure on $(E, \mathcal{E})$. Actually, if such a π (called intensity measure) is given on as Blackwell space $(E, \mathcal{E})$ then a standard (also called homogeneous) Poisson measure with intensity measure π can be constructed, e.g., see Jacod and Shiryaev [117, Section II.1, pp. 64–74]. A standard Poisson measure with intensity measure π is a family of Poisson process $\{p(B,t) : t \geq 0\}$, indexed by B in $\mathcal{E}_0 = \{B \in \mathcal{E} : \pi(B) < \infty\}$, with mean $\mathbb{E}\{p(B,t)\} = t\pi(B)$, and for any finite number of disjoint subsets $B_1, \ldots, B_n$ in $\mathcal{E}_0$, the processes $p(B_1,t), \ldots, p(B_n,t)$ are independent. Clearly, $q(B,t) = p(B,t) - t\pi(B)$ is a family of martingales indexed by B. Also, it can be defined as a family of cad-lag processes with

characteristic function

$$\mathbb{E}\big\{\exp\big(\mathtt{i}rp(B,t)\big)\big\} = \exp\Big(t\pi(B)\big[\mathrm{e}^{\mathtt{i}r}-1\big]\Big), \quad \forall r\in\mathbb{R},$$

for every $t>0$ and B in $\mathcal{E}_0$. Hence, given a (real-valued) strictly positive square-integrable function γ, the integral (defined ω-per-ω or via a stochastic integral, almost surely)

$$p_\gamma(B,t) = \int_B \gamma(\zeta)p(\mathrm{d}\zeta,t), \quad \forall t>0,\, B\in\mathcal{E}_0,$$

yields a family of cad-lag processes with characteristic function

$$\mathbb{E}\big\{\exp\big(\mathtt{i}rp_\gamma(B,t)\big)\big\} = \exp\Big(t\int_B\big[\mathrm{e}^{\mathtt{i}r\gamma(\zeta)}-1\big]\pi(\mathrm{d}\zeta)\Big),$$

for every $t>0$, r in $\mathbb{R}$ and B in $\mathcal{E}_0$. Thus, $\pi_\gamma=\pi\gamma^{-1}$, i.e., $\pi_\gamma(R)=\pi(\gamma^{-1}(R))$, for any Borel subset R of $\mathbb{R}_*=\mathbb{R}\setminus\{0\}$, is the Levy measure of p_γ.

IS all this business with γ correct? For instance, since

$$\int_E |\gamma(\zeta)|^2\pi(\mathrm{d}\zeta)<\infty$$

we can simply use the stochastic integral

$$\int_B \gamma(\zeta)p(\mathrm{d}\zeta,t) = \int_{B\times(0,t]}\gamma(\zeta)q(\mathrm{d}\zeta,\mathrm{d}s) + t\int_B\gamma(\zeta)\pi(\mathrm{d}\zeta)$$

to define $p_\gamma(B,t)$, for any $t>0$ and B in $\mathcal{E}_0$.

On the other hand, let $(Z,\mathcal{Z})$ be a measurable space, where the σ-algebra $\mathcal{Z}$ is countably generated, e.g., $Z=\mathbb{R}_*^m=\mathbb{R}^m\setminus\{0\}$ with its Borel sets $\mathcal{Z}$. Furthermore we re given an increasing sequence $\{Z_n\}$ of subsets in $\mathcal{Z}$ such that $Z=\bigcup_n Z_n$, and set $\mathcal{Z}_n=\{B\in\mathcal{Z}: B\subset Z_n\}$, e.g., in the case of $Z=\mathbb{R}_*^m$, we choose $Z_n=\{x\in\mathbb{R}_*^m:\frac{1}{n}\le|x|\le n\}$.

Definition 7.8. A family of random variables $q(B,t)$, indexed by (B,t), with $t\in\mathbb{R}_+$ and $B\in\mathcal{Z}$, is a (local) martingale measure if

(1) $q(B,\cdot)$ is a (local) square integrable martingale for any B in $\bigcup_n\mathcal{Z}_n$.

(2) $q(B_1\cup B_2\cup\ldots\cup B_n,t)=q(B_1,t)+q(B_2,t)+\ldots+q(B_n,t)$ a.s., for any t and any pairwise disjoint sets $B_1,B_2,\ldots,B_n\in\mathcal{Z}_n$.

(3) For every (t,ω) there exists a measure ϱ_t on $(Z,\mathcal{Z})$ such that ϱ_t is finite on Z_n, the process $\varrho_\cdot(B)$ is predictable for every $B\in\mathcal{Z}_n$ and

$$\langle q(B,\cdot)\rangle_t = \int_{]0,t]}\varrho_s(B)\,\mathrm{d}V_s<\infty$$

for every t and $B\in\bigcup\mathcal{Z}_n$ and some increasing predictable cad-lag (real valued) processes with $V_0=0$.

(4) $\langle q(B_1,\cdot),q(B_2,\cdot)\rangle_t=\langle q(B_1\cap B_2,\cdot)\rangle_t$ for every t and $B_1,B_2\in\bigcup_n\mathcal{Z}_n$. □

• *Remark* 7.9. The property (4) easily implies the notion of orthogonal martingale measure (see Definition 3.33, i.e. $\langle q(B_1,\cdot), q(B_2,\cdot)\rangle_t = 0$ if $B_1 \cap B_2 = \emptyset$. On the other hand the property of being orthogonal implies (4), namely we can write $B_1 = (B_1 \smallsetminus B_2) \cup (B_1 \cap B_2)$ and $q(B_1,t) = q(B_1 \smallsetminus B_2, t) + q(B_1 \cap B_2, t)$ by using (2). By using the analogous decomposition for $q(B_2,t)$ we can write

$$q(B_1,t)q(B_2,t) = q(B_1 \smallsetminus B_2,t)q(B_2,t) + q(B_1 \cap B_2,t)^2 + \\ + q(B_1 \cap B_2,t)q(B_2 \smallsetminus B_1,t).$$

Now, the orthogonality property implies (4). □

• *Remark* 7.10. Perhaps the two most typical examples are the Gaussian and Poisson orthogonal martingale measure. Given a σ-measure μ on $(Z,\mathcal{Z})$, a Gaussian orthogonal martingale measure $W(B,t)$ is a family of independent Wiener processes indexed by B, $\mu(B) < \infty$) such that $\mathbb{E}\{W(B_1,t)W(B_2,s)\} = (t \wedge s)\mu(B_1 \cap B_2)$. On the other hand, given a σ-measure ν on $(Z,\mathcal{Z})$ and a function h in $L^2(\nu)$, a Poisson orthogonal martingale measure $P(B,t)$ is a family of h-compensated Poisson processes indexed by B, $\nu(B) < \infty$, i.e., $P(B,t) = \int_B h(z)\big(N(\mathrm{d}z,t) - t\nu(\mathrm{d}z)\big)$, where $N(B,t)$ is a family of independent Poisson processes such that $\mathbb{E}\{N(B,t)\} = t\nu(B)$, (i.e., a Poisson random measure). It is clear that $\langle W(B,\cdot)\rangle_t = t\mu(B)$ and $\langle P(B,\cdot)\rangle_t = t\nu(B)$, i.e., condition (3) of Definition 7.8 with $V(t) = t$ and a deterministic processes ϱ_t. □

Let us denote $\Pi(B,t) := \langle q(B,\cdot)\rangle_t$; for every $\mathcal{P}\times\mathcal{Z}$ measurable function φ such that

$$\int_{Z\times]0,t]} \varphi^2(\zeta,s)\,\Pi(\mathrm{d}\zeta,\mathrm{d}s) < \infty \tag{7.115}$$

with probability 1, we can define the stochastic integral

$$I_t(\varphi) = \int_{Z\times]0,t]} \varphi(\zeta,s)\,q(\mathrm{d}\zeta,\mathrm{d}s)$$

in the usual way (by approximating with step predictable processes) and $I_t(\varphi)$ is a (local) martingale.

Let us choose a countable family $\{B_1, B_2, \ldots\}$ that generates $\mathcal{Z}$ and such that $q(B_i,t) < \infty$ for any i, any t and any ω. For every (s,ω), the sequence of functions $\{\mathbb{1}_{B_1}, \mathbb{1}_{B_2}, \ldots\}$ is *total* in every Hilbert space $L^2_{(s,\omega)}(Z, \varrho_s(\omega,\mathrm{d}\zeta))$. For each $L^2_{(s,\omega)}(Z, \varrho_s(\omega,\mathrm{d}\zeta))$, we can construct an orthonormal base

$$g^{(1)}_{(s,\omega)}(z),\ g^{(2)}_{(s,\omega)}(z),\ g^{(3)}_{(s,\omega)}(z), \ldots$$

by applying the Gram-Schmidt ortho-normalization procedure to the previous total sequence, in such a way that any $(s,\omega,\zeta) \to g^{(n)}_{(s,\omega)}(z)$ is $\mathcal{P}\times\mathcal{Z}$-measurable. Let us consider the sequence of real-valued (local) martingales

$$h_n(t) := I_t\big(g^{(n)}_{(s,\omega)}(\cdot)\big) = \int_{Z\times]0,t]} g^{(n)}_{(s,\omega)}(\zeta)\,q(\mathrm{d}\zeta,\mathrm{d}s).$$

We have $\langle h_n(\cdot)\rangle_t = V_t$ and $\langle h_n(\cdot), h_m(\cdot)\rangle_t = 0$ for $n \neq m$. Now define

$$H(t) = \sum_{n=1}^{\infty} \tfrac{1}{n}\, h_n(t)\, e_n, \qquad e_n = (0,0,0,\dots,0,1,0,\dots,0,\dots),$$

where the 1 is the n-coordinate, i.e. the set $\{e_1, e_2, e_3, \dots\}$ is the standard orthonormal base of ℓ^2, the space of all sequences $(a_n)_{n=1}^{\infty}$ such that $\sum_n a_n^2 < \infty$. The stochastic process $H(t)$ is a local square integrable martingale in ℓ^2.

On the other hand we can consider the sequence

$$\varphi_n(s,\omega) = \int_Z \varphi(\zeta, s)\, g^{(n)}_{(s,\omega)}(\zeta)\, \varrho_s(\omega, \mathrm{d}\zeta)$$

of components of $\varphi(s,\cdot)$ in each $L^2_{(s,\omega)}(Z, \varrho_s(\omega,\mathrm{d}\zeta))$. Let us consider the sequence of stochastic integrals

$$j_n(t) := \int_{]0,t]} \varphi_n(s)\, \mathrm{d}h_n(s)$$

and it is easy to see that the series $j(t) := j_1(t) + j_2(t) + \cdots$ converges in the $L^2(\Omega)$ sense. We recall that this can be interpreted as the stochastic integral

$$j(t) = \int_{]0,t]} \langle \phi(s), \mathrm{d}H(s)\rangle$$

where $\phi(t)\colon \{a \in \ell^2 : \sum_n \frac{1}{n^2}|a_n|^2 < \infty\} \to \mathbb{R}$ is the linear map defined by $\langle \phi(t,\omega), e_n\rangle = n\,\varphi_n(t,\omega)$. We have

Theorem 7.11. *The stochastic integral with respect to a (local) martingale measure can be represented as a stochastic integral with respect to an ℓ_2-valued martingale:*

$$\int_{Z\times]0,t]} \varphi(\zeta,s)\, q(\mathrm{d}\zeta, \mathrm{d}s) = \int_{]0,t]} \langle \phi(s), \mathrm{d}H(s)\rangle,$$

with the previous notation.

Proof. Exchanging the series with the integral, it is easy to see that

$$\int_{Z\times]0,t]} \varphi(\zeta,s)\, q(\mathrm{d}\zeta, \mathrm{d}s) = \sum_{n=1}^{\infty} \int_{]0,t]} \varphi_n(s)\, \mathrm{d}h_n(s);$$

the right hand side is just the stochastic integral $j(t)$. □

Now let us consider function-valued integrators. Given a filtered space $(\Omega, \mathcal{F}, P, \mathcal{F}_t : t \geq 0)$ and a measure space $(A, \mathcal{A}, \mu)$, for each point a in A, let us assume that $M_c(t,a)$ is a real-valued continuous (local) square-integrable martingale with predictable quadratic co-variation $\alpha_{a,b}(t) = \langle M_c(\cdot,a), M_c(\cdot,b)\rangle(t)$ and $M_j(t,a)$ is a $\mathbb{R}^d$-valued purely jump (local) square-integrable martingale

with predictable co-compensator $\nu^p_{a,b}$ in $\mathbb{R}^m_* \times (0,\infty)$, $m = 2d$, i.e., the predictable compensator corresponding to the integer random measure $\nu_{a,b}$ associated with the $\mathbb{R}^m$-valued process $t \mapsto \big(M_j(t,a), M_j(t,b)\big)$, which yields the (local) martingale measure $\tilde\nu_{a,b} = \nu_{a,b} - \nu^p_{a,b}$. To simplify notation, we set $\alpha_a(t) := \alpha_{a,a}(t)$, $\tilde\nu_a := \tilde\nu_{a,a}$ and $\nu^p_a := \nu^p_{a,a}$. If that $f(t,a)$ and $g(\zeta,t,a)$ are two real-valued predictable processes belonging to $L^2_{\rm loc}(\alpha_a)$ and $L^2_{\rm loc}(\nu^p_a)$, respectively, then the stochastic integral

$$I(t,a) = \int_0^t f(s,a) M_c(\mathrm{d}s,a) + \int_{\mathbb{R}^m_* \times (0,t]} g(\zeta,s,a)\tilde\nu_a(\mathrm{d}\zeta,\mathrm{d}s)$$

is defined for every a in A. Now, one wants to regard $I(t,a)$ as a process in both variables (t,a). If all processes f, g, M_c and M_j are jointly measurable (which imply that $\alpha_{a,b}(t)$, $\tilde\nu_{a,b}(B,(0,t])$ and $\nu^p_{a,b}(B,(0,t])$ are too), then the stochastic integral process $I(t,a)$ is also jointly measurable.

Indeed, to check this point, it suffices to see that if $\{I_n(t,a) : n \geq 1\}$ is a sequence of jointly measurable predictable processes such that

$$q_{t,a}(I_n, I, \varepsilon) := P\Big\{|I_n(t,a) - I(t,a)| > \varepsilon\Big\} \to 0, \quad \text{as} \quad n \to \infty,$$

for every $\varepsilon > 0$, $t \geq 0$ and a in A, then defining inductively the sequence of indexes $\{n(k,t,a) : k \geq 1\}$ by $n(0,t,a) = 1$ and

$$n(k,t,a) = \inf\Big\{r > n(k-1,t,a) \ : \ \sup_{n,m \geq r} q_{t,a}(I_n, I_m, 2^{-k}) \leq 2^{-k})\Big\},$$

one deduce that the series

$$\sum_{k=1}^{\infty} \Big\{|I_{n(k+1,a)}(t,a) - I_{n(k,a)}(t,a)|\Big\}$$

is almost surely convergent, proving that there is a jointly measurable (also cadlag or continuous in t) process $\bar I(t,a)$ such that $\bar I(\cdot,a) = I(\cdot,a)$, almost surely, for each a in $\mathcal{A}$. Hence, the stochastic integral can be taken jointly measurable.

Note that the predictable quadratic co-variation associated with the continuous part of the stochastic integral $I(t,a)$ is the process

$$\int_0^t f(s,a) f(s,b) \mathrm{d}\alpha_{a,b}(t),$$

while

$$\int_{(0,t]\times\mathbb{R}^m_*} g(\zeta,s,a) g(\zeta,s,b) \nu^{\mathrm{p}}_{a,b}(\mathrm{d}\zeta,\mathrm{d}s)$$

is the predictable co-compensator.

Once the measurability question is resolved, a Fubini's type theorem for stochastic integral follows, i.e., under the above joint measurability conditions

and assuming that $M_c(t,a) = M_c(t)$ and $M_j(t,a) = M_j(t)$ are independent of a, one has

$$\int_A I(t,a)\mu(\mathrm{d}a) = \int_0^t \Big(\int_A f(s,a)\mu(\mathrm{d}a)\Big) M_c(\mathrm{d}s) + \\ + \int_{\mathbb{R}^m_* \times (0,t]} \Big(\int_A g(\zeta,s,a)\mu(\mathrm{d}a)\Big)\tilde{\nu}(\mathrm{d}\zeta,\mathrm{d}s)$$

provided that the processes f and g are (locally) square-integrable with respect to the product measure $\mu(\mathrm{d}a) \times \mathrm{d}\alpha(t) \times \mathrm{d}P$ and $\mu(\mathrm{d}a) \times \nu^p(\mathrm{d}\zeta,\mathrm{d}t) \times \mathrm{d}P$, respectively.

Can we have a more general version of Fubini's theorem? Something like starting from a (local) martingale-measure in $\Upsilon(\mathrm{d}a,t)$ in the product $A \times [0,\infty)$, and the *sections* martingale-measures $M(\mathrm{d}t,a)\mu(\mathrm{d}a)$ and $\mu(\mathrm{d}a,t)M(\mathrm{d}t)$ to make-up Fubini's theorem.
What do you think?

7.6 Function-valued Stochastic Equation?

NOW, THE STUFF FROM Gyongy-Krylov..., UPDATED!

7.6.1 Stochastic Differentials and Flows

Itô and energy formulae

THIS HAS A NON-EMPTY INTERSECTION WITH "FLOWS" Perhaps, we should check book Kunita [143]!

7.6.2 Existence and Uniqueness

NOW, THE STUFF FROM Gyongy-Krylov..., by now, only mainly "Existence and Uniqueness" remain...

THIS IS NOT STOCHASTIC PDE, BUT IT IS RELATED, THE COEFFICIENTS ARE FUNCTION-VALUED LOCALLY LIPSCHITZ WITH LINEAR GROWTH (OR POLYNOMIAL GROWTH WITH A PRIORI BOUNDS), THE MARTINGALE IS ALSO FUNCTION-VALUED, OK?

MAINLY, the coefficients g, σ and γ were (in Chapter 5) functions from $[0,\infty) \times \mathbb{R}^d$, $[0,\infty) \times \mathbb{R}^d$ and $\mathbb{R}^m_* \times [0,\infty) \times \mathbb{R}^d$ into $\mathbb{R}^d$, $\mathbb{R}^d \times \mathbb{R}^n$ and $\mathbb{R}^d$, respectively, and now we will have $\mathbb{R}^n$ and $\mathbb{R}^m$ replaced by a Hilbert space K and $\mathbb{R}^m_*$ replaced by $K_* = K \setminus \{0\}$, and the space where the solution live $\mathbb{R}^d$ replaced by another Hilbert space H. Usually, this will be "applicable when "H will have a finite dimension or the coefficient will be linear bounded operators. The driven processes are K-valued Levy process, i.e., first a Wiener process with a trace-class covariance operator R_1 denoted by $\{w(t) : t \geq 0\}$ and a second a Poisson point process with a trace-class covariance operator R_2 and

Levy measure π, this means that

$$\mathbb{E}\{e^{i(w(t),k)}\} = \exp\big(-t(R_1k,k)/2\big),$$
$$\mathbb{E}\{e^{i(p(t),k)}\} = \exp\Big(t\int_{K_*}\big[e^{i(R_2h,k)} - 1 - i(R_2h,k)\big]\pi(dh)\Big),$$

for every $t \geq 0$ and k in K, where $(\cdot,\cdot) = (\cdot,\cdot)_K$ denotes the inner product in K. Certainly π satisfies

$$\int_{K_*}\big(|k|^2 \wedge |k|\big)\pi(dk) < \infty,$$

where $|\cdot| = |\cdot|_K$ is the norm in K.

7.7 Applications of Backward Integration

This section is taken from [50], where we use the notion of backward integration to treat, in a simpler and unifying way, various results obtained with other technique. This includes Girsanov and Feynman-Kac formulae.

Suppose given functions $g\colon [0,T]\times\mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma\colon [0,T]\times\mathbb{R}^d \mapsto L(\mathbb{R}^d,\mathbb{R}^\ell)$ and $\gamma\colon \mathbb{R}^m_* \times [0,T]\times\mathbb{R}^d \mapsto \mathbb{R}^d$ satisfying the assumptions:

(a) The coefficients $g(t,x)$, $\sigma(t,x)$ and $\gamma(z,t,x)$ are always supposed Borel measurable, and because we are interested in global solutions defined on a prescribed bounded interval, say $[0,T]$, we impose a linear growth condition, namely, there exists a constant $C > 0$ such that

$$|g(t,x)|^2 + |\sigma(t,x)|^2 + \int_{\mathbb{R}^m_*}|\gamma(z,t,x)|^2\pi(dz) \leq C(1+|x|^2), \tag{7.116}$$

for every (t,x) in $[0,T]\times\mathbb{R}^d$. Thus, the initial condition x must be an $\mathcal{F}(t_0)$-measurable random variable (most of the time, a deterministic value).

(b) A clean existence and uniqueness theory is developed adding a uniform locally Lipschitz condition in the variable x, namely, for any $r > 0$ there exists a positive constant $M = M(r)$ such that

$$\begin{cases} |g(t,x) - g(t,x')|^2 + |\sigma(t,x) - \sigma(t,x')|^2 + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*}|\gamma(z,t,x) - \gamma(z,t,x')|^2\pi(dz) \leq M|x-x'|^2, \end{cases} \tag{7.117}$$

for every (t,x), (t,x') in $[0,T]\times\mathbb{R}^d$ with $|x| \leq r$ and $|x'| \leq r$.

(c) The functions $g(t,x)$, $\sigma(t,x)$ and $\gamma(z,t,x)$ are twice continuously differentiable in x and locally bounded, i.e., if $\partial_x^\ell g(t,x)$, $\partial_x^\ell\sigma(t,x)$ and $\partial_x^\ell\gamma(z,t,x)$ denote any of the derivatives up to the order $\ell \leq 2$ then for any $r > 0$, we have

$$|\partial_x^\ell g(t,x)|^2 + |\partial_x^\ell\sigma(t,x)|^2 + \int_{\mathbb{R}^m_*}|\partial_x^\ell\gamma(z,t,x)|^2\pi(dz) \leq K_r^\ell, \tag{7.118}$$

for any $0 \le t \le T$, $|x| \le r$ and some constant K_r^ℓ.

Now, for a given $T > 0$, x in $\mathbb{R}^d$ and s in $[0, T]$, let us consider the stochastic differential equation

$$\begin{cases} X(t) = x + \displaystyle\int_s^t g(r, X(r))\mathrm{d}r + \int_s^t \sigma(r, X(r))\mathrm{d}w(r) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}_*^m \times]s,t]} \gamma(z, r, X(r))\tilde{p}(\mathrm{d}z, \mathrm{d}r), \end{cases} \tag{7.119}$$

with t in $]s, T]$. Note that one may replace $\sigma(r, X(r))$ and $\gamma(z, r, X(r))$ with $\sigma(r, X(r-))$ and $\gamma(z, r, X(r-))$ in both stochastic integrals.

It is well known that, under conditions (7.116), (7.117) and (7.118), the stochastic differential equation (7.119) has a unique solution, that we denote by $X(t, s, x)$, t in $[s, T]$, such that $X(t, s, x)$ is twice differentiable on x. Clearly this is set on a complete filtered probability space $(\Omega, \mathcal{F}, P)$ with a (standard) Wiener process w and a (standard) Poisson measure p with Lévy measure π. We also consider the flow operators $U_{s,t}$ acting on functions $\varphi\colon \mathbb{R}^d \times \Omega \mapsto \mathbb{R}$ defined as

$$U_{s,t}\varphi(x) = \varphi(X(t, s, x)), \qquad 0 \le s \le t \le T.$$

It is easy to check that the backward evolution property:

$$U_{s,t} = U_{s,r}U_{r,t}, \qquad 0 \le s \le r \le t \le T,$$

holds thanks to the semigroup law of $X(t, s, x)$. In particular if φ is deterministic then we are interested in the process (random field) $(s, x) \mapsto \varphi(X(t, s, x)) = U_{s,t}\varphi(x)$ which we denote by $u_\varphi^t(s, x)$.

7.7.1 Definitions

Let us recall briefly the backward integration for a standard Wiener process and a Poisson measure. In a given probability space $(\Omega, \mathcal{F}, P)$ let $(w(t),\ t \ge 0)$ be a $\mathbb{R}^\ell$-valued Wiener process and let $\{p(\cdot, t) : t \ge 0\}$ be a standard Poisson measure with Lévy (characteristic or intensity) measure $\pi(\cdot)$ in $\mathbb{R}_*^m = \mathbb{R}^m \setminus \{0\}$, and (local) martingale measure $\{\tilde{p}(\cdot, t) : t \ge 0\}$, $\tilde{p}(\cdot, t) = p(\cdot, t) - t\pi(\cdot)$. Given a fix $T > 0$, define the standard Wiener process $\hat{w}_T(t) = w(T) - w(T - t)$ and the a standard Poisson measure $\hat{p}_T(\cdot, t) := p(\cdot, T) - p(\cdot, T - t)$, for t in $[0, T)$, with Lévy (characteristic or intensity) measure $\hat{\pi}_T(\cdot) = \pi(T) - \pi(T - t)$, and (local) martingale measure $\tilde{\hat{p}}_T(\cdot, t) = \hat{p}_T(\cdot, t) - t\hat{\pi}_T(\cdot)$. Then the backward integral is the (forward) integral with respect to $\hat{w}_T$ and $\tilde{\hat{p}}_T$.

Define the two-index family of sub σ-algebras $\{\mathcal{F}_a^b : t > s \ge 0\}$ or $\{\mathcal{F}(b, a) : b \ge a \ge 0\}$ generated by the increments $w(t) - w(s)$ and $p(B, t) - p(B, s)$, $b \ge t > s \ge a$, $B \in \mathbb{R}_*^m$. Now, for a given $T > 0$, define $\{\hat{\mathcal{F}}_t^T : t \le T\}$ also denoted by $\{\hat{\mathcal{F}}_T(t) : t \le T\}$, where $\hat{\mathcal{F}}_T(t) = \cap_{\varepsilon > 0}\mathcal{F}(T, t + \varepsilon)$, clearly the single-index family of sub σ-algebras $\{\hat{\mathcal{F}}_T(t) : t \le T\}$ is decreasing and left-continuous, called *backward filtration*. An elementary or simple process (backward-predictable) has the form

(1)-*either* $f(t,\omega) = f_i(\omega)$ if $t_{i-1} \le t < t_i$ with some $i = 1,\dots,n$, where $0 \le t_0 < t_1 < \cdots < t_n = T$ are real numbers and f_i is a $\bar{\mathcal{F}}_T(t_i)$ measurable bounded random variable for any i, and $f(t,\omega) = 0$ otherwise;
(2)-*or* $g(z,t,\omega) = g_{i,j}(\omega)$ if $t_{i-1} \le t < t_i$ and z belongs to K_j with some $i = 1,\dots,n$, and $j = 1,\dots,m$, where $0 \le t_0 < t_1 < \cdots < t_n = T$ are real numbers, K_j are disjoint sets with compact closure in $\mathbb{R}^m_*$ and $g_{i,j}$ is a $\bar{\mathcal{F}}_T(t_i)$ measurable bounded random variable for any i, and $g(z,t,\omega) = 0$ otherwise.

It is clear what the backward integral should be for any backward-predictable processes $f(t)$ and $g(z,t)$, namely

$$\int_0^T f(s)\,\hat{\mathrm{d}}w(s) := \sum_{i=1}^n f_i\,[w(t_i) - w(t_{i-1})],$$
$$\int_a^b f(s)\,\hat{\mathrm{d}}w(s) := \int_0^T f(s)\,\mathbb{1}_{[a,b)}(s)\,\hat{\mathrm{d}}w(s),$$

and

$$\int_{\mathbb{R}^m_*\times[0,T)} g(z,s)\,\tilde{p}(\mathrm{d}z,\hat{\mathrm{d}}s) := \sum_{i=1}^n\sum_{j=1}^m g_{i,j}\,\tilde{p}(K_j\times]t_{i-1},t_i]),$$
$$\int_{\mathbb{R}^m_*\times[a,b)} g(z,s)\,\tilde{p}(\mathrm{d}z,\hat{\mathrm{d}}s) := \int_{\mathbb{R}^m_*\times[0,T)} g(z,s)\,\mathbb{1}_{[a,b)}(s)\,\tilde{p}(\mathrm{d}z,\hat{\mathrm{d}}s),$$

for every $b > a \ge 0$.

Then for any right-continuous backward-adapted process (i.e., adapted to the filtration $\{\bar{\mathcal{F}}_T(t) : t \le T)$ of the form $g(z,t,\omega) = g_j(t,\omega)$ when z belongs to K_j, we can calculate the backward stochastic integral, namely,

$$\int_{\mathbb{R}^m_*\times[t,T)} \sum_{j=1}^m g_j(s)\,\mathbb{1}_{K_j}(z)\,p(\mathrm{d}z,\hat{\mathrm{d}}s) := \sum_{j=1}^m \sum_{k=1}^{p(t,K_j,\omega)} g_j(\theta_k(\omega,K_j),\omega),$$

for any $t \ge 0$, where $\theta_k(\omega,K_j)$ is the time of the k jump of the Poisson process $t \mapsto p(K_j,t)$. In the case of a compound-Poisson process as above, we may forget about the K dependency, and make the previous pathwise definition, both concepts agree, i.e., if the Poisson measure $p(B,t)$ is approximated by the ε-Poisson measure $p(B,t)$ with finite Levy measure $\pi_\varepsilon(B) := \pi(\{z \in B : |z| \ge \varepsilon\})$ then the backward stochastic integral can be written as above.

The Poisson measure $p(\mathrm{d}z,\mathrm{d}s)$ with Lévy measure π satisfies $p(\mathbb{R}^m_*,\{0\}) = 0$ and can be approximated by another Poisson measure $p_\varepsilon(\mathrm{d}z,\mathrm{d}s)$ with Lévy measure $\pi_\varepsilon = \mathbb{1}_{K_\varepsilon}\pi$, where the support $K_\varepsilon = \{0 < \varepsilon \le |z| \le 1/\varepsilon\}$ of π_ε is a compact on $\mathbb{R}^m_*$, i.e., all jumps smaller than ε or larger than $1/\varepsilon$ have been eliminated. The integer measure p_ε is associated with a compound Poisson process and has a finite (random) number of jumps, i.e., for any $T > 0$ there is an integer $N = N(T,\omega)$, points $z_i = z_i(T,\omega)$ in K_ε for $i = 1,\dots,N$ and positive reals $\theta_i = \theta_i(T,\omega)$, $i = 1,\dots,N$ such that $p_\varepsilon(B,]a,b],\omega) = \sum_{n=1}^N \mathbb{1}_{z_i\in B}\,\mathbb{1}_{a<\theta_i\le b}$, for every $B \in \mathcal{B}(\mathbb{R}^m_*)$, $0 \le a < b \le T$. In this case, the forward stochastic integral

can be written as

$$\int_{\mathbb{R}^m_*\times(0,T]} f(z,s)\,\tilde{p}_\varepsilon(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^{N} f(z_i,\theta_i-) - \int_0^T \mathrm{d}s \int_{K_\varepsilon} f(z,s)\pi(\mathrm{d}z),$$

for any adapted (forward, i.e. to $\mathcal{F}_0^s$, $s\geq 0$) cad-lag process $f(z,s)$, continuous in z. On the other hand, the backward stochastic integral is written as

$$\int_{\mathbb{R}^m_*\times[0,T)} g(z,s)\,\tilde{p}_\varepsilon(\mathrm{d}z,\hat{\mathrm{d}}s) = \sum_{i=1}^{N} g(z_i,\theta_i) - \int_0^T \mathrm{d}s \int_{K_\varepsilon} g(z,s)\pi(\mathrm{d}z),$$

for any adapted (backward, i.e. to $\mathcal{F}_s^T$, $s\leq T$) cad-lag process $g(z,s)$, continuous in z. Recall that elementary forward processes are left-hand continuous while elementary backward processes are right-continuous. However, after taking limits for elementary processes, both, cad-lag and cag-lad adapted (either forward or backward) processes are integrable, but one takes the cag-lad version for the forward integral and the cad-lag version for the backward integral.

Finally, the backward stochastic integral is extended to all backward predictable processes, including all cag-lad (i.e., left-hand continuous having right-hand limit) processes, satisfying

$$\int_0^T |f(t)|^2\mathrm{d}t<\infty \quad \text{and} \quad \int_0^T \mathrm{d}t\int_{\mathbb{R}^m_*}|g(z,t)|^2\pi(\mathrm{d}z)<\infty$$

almost surely, similarly to Chapter chap-4.

It is then clear that

$$\int_a^b f(t)\,\hat{\mathrm{d}}w(t) = \int_{T-b}^{T-a} \hat{f}_T(t)\,\mathrm{d}\hat{w}_T(t),$$
$$\int_{\mathbb{R}^m_*\times[a,b)} g(z,t)\,\tilde{p}(\mathrm{d}z,\hat{\mathrm{d}}t) = \int_{\mathbb{R}^m_*\times(T-b,T-a]} \hat{g}_T(z,t)\,\tilde{\hat{p}}_T(\mathrm{d}z,\mathrm{d}t),$$

for any $0\leq a<b\leq T$, where $\hat{f}_T(t):=f(T-t)$ and $\hat{g}_T(z,t):=g(z,T-t)$.

Note that by using backward integration with respect to a Wiener and a Poisson measure we avoid possible difficulties with time reversal.

7.7.2 Backward Itô Formula

For any nonnegative integer k, we denote by $C^k(\mathbb{R}^d)$ (resp. $C_b^k(\mathbb{R}^d)$) the space of all functions from $\mathbb{R}^d$ into $\mathbb{R}$ which are uniformly continuous (resp. uniformly continuous and bounded) together with their derivatives of order less or equal than k.

In what follows, we prove a backward Itô formula for the process $\varphi(X(t,s,x))$, where φ in $C^2(\mathbb{R}^d)$ and $X(t,s,x)$ is the solution to (7.119). The proof is based on Taylor formula and explicitly exploits the fact that $X(t,s,x)$ is the solution to the differential stochastic equation (7.119).

We denote by $\mathcal{L}_s$ and $\mathcal{M}_s$ the linear operators defined for φ in $C_b^2(\mathbb{R}^d)$

$$\mathcal{L}_s\varphi(x) = \mathcal{L}_s^0\varphi(x) + \mathcal{L}_s^\gamma\varphi(x), \tag{7.120}$$

where

$$\mathcal{L}_s^0\varphi(x) = \tfrac{1}{2}\mathrm{Tr}[D^2\varphi(x)\sigma(s,x)\sigma^*(s,x)] + \big(g(s,x), D\varphi(x)\big),$$
$$\mathcal{L}_s^\gamma\varphi(x) = \int_{\mathbb{R}_*^m} [\varphi(x+\gamma(z,s,x)) - \varphi(x) - (\gamma(z,s,x), D\varphi(x))]\pi(\mathrm{d}z),$$

and

$$\begin{cases} \mathcal{M}_s\varphi(x) = \sigma(s,x)^* D\varphi(x), \\ \mathcal{N}_s(z)\varphi(x) = \varphi(x+\gamma(z,s,x)) - \varphi(x). \end{cases} \tag{7.121}$$

To simplify the notation, we may use

$$\big(\mathcal{M}_s\varphi(x), y\big) = \big(\sigma(s,x)y, D\varphi(x)\big), \quad \forall y \in \mathbb{R}^d,$$

where $(\cdot,\cdot)$ is the scalar (or dot) product in $\mathbb{R}^d$. We have

Proposition 7.12. *Assume conditions (7.116), (7.117) and (7.118) and let φ in $C^2(\mathbb{R}^d)$. Then the random field $u_\varphi^t(s,x) = \varphi(X(t,s,x))$ satisfies the stochastic partial differential equation*

$$\begin{cases} u_\varphi^t(s,x) = \varphi + \displaystyle\int_s^t \mathcal{L}_r u_\varphi^t(r,x)\mathrm{d}r + \int_s^t \big(\mathcal{M}_r u_\varphi^t(r,x), \hat{\mathrm{d}}w(r)\big) + \\ \qquad\qquad + \displaystyle\int_{[s,t[\times\mathbb{R}_*^m} \mathcal{N}_r(z) u_\varphi^t(r,x)\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r), \end{cases} \tag{7.122}$$

where we recall that $\hat{\mathrm{d}}w(t)$ and $\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r)$ stand for backward stochastic integration.

Proof. The arguments are similar to those in Ikeda and Watanabe [110, Theorem II.5.1, p. 66]. The first step is to prove the result without the small jumps, e.g., we approximate the function γ in (7.119) and the measure π. Indeed, without any lost of generality we assume $\gamma(z,t,x)$ continuous in z and we set $\pi_n(B) = \pi(\{z \in B \,:\, n|z| \geq 1,\, |z| \leq n\})$, $n = 1, 2, \ldots$ and B any Borel subset of $\mathbb{R}_*^m$, so that if p_n is the corresponding Poisson measure then there exist an increasing sequence of stopping times $\{\tau_1, \tau_2, \ldots\}$ and an adapted sequence of random variables (jumps) $\{z_1, z_2, \ldots\}$ such that

$$p_n(]0,t] \times \mathbb{R}_*^m) = \sum_{m=1}^{\infty} z_i \mathbb{1}_{\{\tau_i \leq t\}}.$$

Clearly, $0 < \tau_i < \tau_{i+1}$ if $\tau_i < \infty$ and τ_i is the time of the i jump given by z_i.

Consider the process $X_n(t) = X_n(t,s,x)$ solution to the following stochastic differential equation

$$\begin{cases} X_n(t) = x + \displaystyle\int_s^t g(r, X_n(r))\mathrm{d}r + \int_s^t \sigma(r, X_n(r))\mathrm{d}w(r) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}_*^m\times]s,t]} \gamma(z,r,X_n(r))\tilde{p}_n(\mathrm{d}z,\mathrm{d}r). \end{cases} \tag{7.123}$$

Such solution $X_n(t)$ has properties similar to those of equation (7.119), and if $u^t_{\varphi,n}(s,x) = \varphi(X_n(t,s,x))$ then converges (in probability) to $u^t_\varphi(s,x)$, together with their first and second derivatives in x, uniformly for $0 \le s \le t \le T$ and $|x| \le r$, for any fixed $r > 0$. Moreover, we can write the stochastic differential equation (7.123) as

$$X_n(t) = x + \int_s^t g(r, X_n(r))\mathrm{d}r + \int_s^t \sigma(r, X_n(r))\mathrm{d}w(r) + \\ + \sum_i \gamma(z_i, \tau_i, X_n(\tau_i))\mathbb{1}_{\{s<\tau_i\le t\}} - \int_s^t \mathrm{d}r \int_{\mathbb{R}^m_*} \gamma(z, r, X_n(r))\pi_n(\mathrm{d}z),$$

for any n.

Hence we have

$$\begin{cases} \varphi(X_n(t,s,x)) - \varphi(x) = \\ \quad = \sum_i \{\varphi(X_n(t, s\vee\tau_{i-1}, x)) - \varphi(X_n(t, s\vee\tau_i-, x))\} + \\ \qquad + \sum_i \{\varphi(X_n(t, s\vee\tau_i-, x)) - \varphi(X_n(t, s\vee\tau_i, x))\}. \end{cases} \tag{7.124}$$

and we can deal with first sum as a continuous process. Therefore, we assume temporarily that

$$X_n(t) = x + \int_s^t \tilde g_n(r, X_n(r))\mathrm{d}r + \int_s^t \sigma(r, X_n(r))\mathrm{d}w(r), \tag{7.125}$$

where

$$\tilde g_n(t,x) = g(t,x) - \int_{\mathbb{R}^m_*} \gamma(z,t,x)\pi_n(\mathrm{d}z),$$

i.e., without changing notation we are working between two consecutive jumps, namely, in $]\tau_i, \tau_{i+1}[$ or alternatively, we do consider the jumps.

Now, let Σ_{st} be the set of all decompositions (or partitions) $\sigma = \{s = s_0 < s_1 < \ldots < s_N = t\}$ of the interval $[s,t]$, partially ordered in the usual way. For any σ in Σ_{st} we set

$$|\sigma| = \max\big\{s_k - s_{k-1} \,:\, k = 1, \ldots, N\big\}.$$

and we have

$$\varphi(X_n(t,s,x)) - \varphi(x) = \sum_{k=1}^N \big[\varphi(X_n(t, s_{k-1}, x)) - \varphi(X_n(t, s_k, x))\big] = \\ = \sum_{k=1}^N \big[\varphi(X_n(t, s_k, X_n(s_k, s_{k-1}, x))) - \varphi(X_n(t, s_k, x))\big].$$

To simplify the notation, we set

$$\begin{aligned}
&\varphi_n(t,s,x) = \varphi\big(X_n(t,s,x)\big) = u^n_{t,\varphi}(s,x),\\
&\varphi'_n(t,s,x) = D_x\big[\varphi\big(X_n(t,s,\cdot)\big)\big](x),\\
&\varphi''_n(t,s,x) = D^2_x\big[\varphi\big(X_n(t,s,\cdot)\big)\big](x),
\end{aligned}$$

where D_x means derivative in the variable x. It follows

$$\begin{aligned}
\varphi(X_n(t,s,x)) - \varphi(x) &= \sum_{k=1}^N \big(\varphi_n(t,s_k,x), X_n(s_k,s_{k-1},x) - x\big) + \\
&= \tfrac{1}{2}\sum_{k=1}^N \big(\varphi''_n(t,s_k,x)(X_n(s_k,s_{k-1},x)-x), X_n(s_k,s_{k-1},x) - x\big) + \\
&\qquad\qquad +R_1(|\sigma|),
\end{aligned}$$

Then

$$\begin{aligned}
&\sum_{k=1}^N \big(\varphi'_n(t,s_k,x), X_n(s_k,s_{k-1},x) - x\big) = \\
&\qquad = \sum_{k=1}^N \big(\varphi'_n(t,s_k,x), \tilde g(s_k,x)\big)(s_k - s_{k-1}) + \\
&\qquad\quad + \sum_{k=1}^N \big(\varphi'_n(t,s_k,x), \sigma(s_k,x)(w(s_k)-w(s_{k-1}))\big) + R_2(|\sigma|),
\end{aligned}$$

and

$$\begin{aligned}
&\tfrac{1}{2}\sum_{k=1}^N \big(\varphi''_n(t,s_k,x)(X_n(s_k,s_{k-1},x)-x), X_n(s_k,s_{k-1},x) - x\big) = \\
&\qquad = \tfrac{1}{2}\sum_{k=1}^N \big(\varphi''_n(t,s_k,x)\sigma(s_k,x), \sigma(s_k,x)\big)(s_k - s_{k-1}) + R_3(|\sigma|).
\end{aligned}$$

Gathering all, we find

$$\left\{\begin{aligned}
&\varphi(X_n(t,s,x)) - \varphi(x) = \sum_{k=1}^N \big(\varphi'_n(t,s_k,x), \tilde g(s_k,x)\big)(s_k - s_{k-1}) + \\
&\qquad + \sum_{k=1}^N \big(\varphi'_n(t,s_k,x), \sigma(s_k,x)(w(s_k)-w(s_{k-1}))\big) + \\
&\qquad + \tfrac{1}{2}\sum_{k=1}^N \big(\varphi''_n(t,s_k,x)\sigma(s_k,x), \sigma(s_k,x)\big)(s_k - s_{k-1}) + R(|\sigma|),
\end{aligned}\right. \tag{7.126}$$

where

$$R(|\sigma|) = R_1(|\sigma|) + R_2(|\sigma|) + R_3(|\sigma|).$$

Now, by proceeding as in the usual proof of Itô formula, taking into account that we have

$$X_n(s_k, s_{k-1}, x) - x = \int_{s_{k-1}}^{s_k} \tilde{g}(r, X_n(r, s_{k-1}, x))\mathrm{d}r + \\ + \int_{s_{k-1}}^{s_k} \sigma(r, X_n(r, s_{k-1}, x))\mathrm{d}w(r),$$

we deduce $R(|\sigma|) \to 0$ in probability. So now, by setting $\varphi_n(s) = \varphi(t, s, x) = \varphi(X_n(t, s, x))$ with $X_n(t, s, x)$ given by (7.125), and letting $|\sigma|$ tend to 0 in (7.126), we obtain

$$\varphi_n(s) = \varphi + \int_s^t \mathcal{L}_r^0 \varphi_n(r)\mathrm{d}r - \int_s^t \int_{\mathbb{R}_*^m} \big(\gamma(z, r, x), D\varphi_n(r)\big)\pi_n(\mathrm{d}z)\mathrm{d}r + \\ + \int_s^t \big(\mathcal{M}_r \varphi_n(r), \hat{\mathrm{d}}w(r)\big),$$

This means that for the first sum in (7.124) yields the contribution

$$\varphi(x) + \int_s^t \mathcal{L}_r^0 u_{\varphi,n}^t(r, x)\mathrm{d}r - \int_s^t \int_{\mathbb{R}_*^m} \big(\gamma(z, r, x), Du_{\varphi,n}^t(r, x)\big)\pi_n(\mathrm{d}z)\mathrm{d}r + \\ + \int_s^t \big(\mathcal{M}_r u_{\varphi,n}^t(r), \hat{\mathrm{d}}w(r)\big),$$

where $u_{\varphi,n}^t(s, x) = \varphi(X_n(t, s, x))$ with $X_n(t, s, x)$ given by (7.123).

For the second sum in (7.124), i.e., the sum of jumps, we set $\gamma_n(z, s) = \gamma(z, s, X_n(t, s-, x))$ and we have

$$\begin{aligned}
&\sum_i \{\varphi(X_n(t, s \vee \tau_i -, x)) - \varphi(X_n(t, s \vee \tau_i, x))\} = \\
&= \sum_i \{\varphi(X_n(t, \tau_i -, x)) - \varphi(X_n(t, \tau_i, x))\}\mathbb{1}_{\{s < \tau_i \le t\}} = \\
&= \int_{[s,t)\times\mathbb{R}_*^m} \big[\varphi(X_n(t, r-, x) + \gamma(z, r, x)) - \varphi(X_n(t, r-, x))\big] p_n(\mathrm{d}z, \mathrm{d}r) = \\
&= \int_{[s,t)\times\mathbb{R}_*^m} \big[\varphi(X_n(t, r-, x) + \gamma(z, r, x)) - \varphi(X_n(t, r-, x))\big] \tilde{p}_n(\mathrm{d}z, \mathrm{d}r) + \\
&\quad + \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big[\varphi(X_n(t, r-, x) + \gamma(z, r, x)) - \varphi(X_n(t, r-, x))\big] \pi_n(\mathrm{d}z),
\end{aligned}$$

thus, establishing (7.122), replacing $\tilde{p}$ with $\tilde{p}_n$, for $u_{\varphi,n}^t(s, x)$.

Now, we can pass to the limit for $n \to \infty$. One have $X_n(t, s, x) \to X(t, s, x)$, $D_x X_n(t, s, x) \to D_x X(t, s, x)$ and $D_x^2 X_n(t, s, x) \to D_x^2 X(t, s, x)$, a.s. (also uniformly in finite intervals with respect to s and x). Consequently, $u_{\varphi,n}^t(s, x) \to u_\varphi^t(s, x)$ and $\mathcal{L}_r^0 u_{\varphi,n}^t(s, x) \to \mathcal{L}_r^0 u_\varphi^t(s, x)$ a.s. (also uniformly in finite intervals with respect to s and x).

Also, it is easy to see, by the dominated convergence theorem, that

$$\int_0^t \mathcal{L}_r^0 u_{\varphi,n}^t(r,x)\mathrm{d}r \to \int_0^t \mathcal{L}_r^0 u_\varphi^t(r,x)\mathrm{d}r, \quad \text{a.s.},$$

$$\int_0^t \big(\mathcal{M}_r u_{\varphi,n}^t(r), \hat{\mathrm{d}}w(r) \to \int_0^t \big(\mathcal{M}_r u_\varphi^t(r), \hat{\mathrm{d}}w(r) \ \text{ in } \ L^2([0,T]\times\Omega),$$

$$\begin{aligned}&\int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big[\varphi(X_n(t,r-,x)+\gamma(z,r,x)) - \varphi(X_n(t,r-,x))\big]\pi_n(\mathrm{d}z)\\ &\to \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big[\varphi(X(t,r-,x)+\gamma(z,r,x)) - \varphi(X(t,r-,x))\big]\pi(\mathrm{d}z),\end{aligned}$$

a.s., and

$$\begin{aligned}&\int_{[s,t)\times\mathbb{R}_*^m} \big[\varphi(X_n(t,r-,x)+\gamma(z,r,x)) - \varphi(X_n(t,r-,x))\big]\tilde{p}_n(\mathrm{d}z,\mathrm{d}r)\\ &\to \int_{[s,t)\times\mathbb{R}_*^m} \big[\varphi(X(t,r-,x)+\gamma(z,r,x)) - \varphi(X(t,r-,x))\big]\tilde{p}(\mathrm{d}z,\mathrm{d}r),\end{aligned}$$

where the convergence is meant in $L^2([0,T]\times\Omega)$. Thus the proof is now complete. □

• *Remark* 7.13. Using Stratonovich integral, instead of backward Itô integral, the second order term in $\mathcal{L}_t$ disappears, consequently equation (7.122) reduces to the first order stochastic partial differential equation studied by Kunita [143]. In other words equation (7.122) is essentially a *first order equation*, see also Krylov and Rozovskii [218]. □

• *Remark* 7.14. If φ in $C_b^2(\mathbb{R}^d)$, then by taking expectation in (7.122) we readily see that function $v(s,x)$

$$v(s,x) = \mathbb{E}\big\{\varphi(X(t,s,x))\big\}, \quad 0 \le s \le t \le T,$$

gives a solution of backward Kolmogorov equation

$$\begin{cases} \partial_s v(s,x) + \mathcal{L}_s v(s,x) = 0, & \forall s \in]0,t], \\ v(t,x) = \varphi(x), & \forall x \in \mathbb{R}^d, \end{cases}$$

in the whole space $\mathbb{R}^d$. □

• *Remark* 7.15. If the function φ also depends on the backward time s, i.e., $u_{t,\varphi}(s,x) = \varphi(s, X(t,s,x))$ then the backward Itô formula is modified by adding a term $\partial_s\varphi(s, X(t,s,x)))$ into the operator $\mathcal{L}$. □

Let us introduce three mappings $F\colon [0,T]\times\mathbb{R}^d\mapsto\mathbb{R}$, $G\colon [0,T]\times\mathbb{R}^d\mapsto\mathbb{R}^\ell$ and $H\colon \mathbb{R}^m_*\times[0,T]\times\mathbb{R}^d\mapsto(-1,\infty)$, and let us assume that F and G are continuous on $[0,T]\times\mathbb{R}^d$ and

$$\begin{cases} \mathbb{E}\Big\{\displaystyle\int_s^T \mathrm{d}r \int_{\mathbb{R}^d_*} |H(z,r,X(r,s,x))|^2\pi(\mathrm{d}z)\Big\}<\infty, \\ \mathbb{E}\Big\{\displaystyle\int_s^T \mathrm{d}r \int_{\mathbb{R}^d_*} \big[H(z,r,X(r,s,x)) - \\ \qquad -\ln\big(1+H(z,r,X(r,s,x))\big)\big]\pi(\mathrm{d}z)\Big\}<\infty, \end{cases} \tag{7.127}$$

for any s in $[0,T]$. Then we consider the following linear stochastic differential equation, for any t in $]s,T]$,

$$\begin{cases} \eta_s(t) = 1 + \displaystyle\int_s^t \eta_s(r)\,F(r,X(r,s,x))\mathrm{d}r + \\ \qquad + \displaystyle\int_s^t \eta_s(r)\big(G(r,X(r,s,x)),\mathrm{d}w(r)\big) + \\ \qquad + \displaystyle\int_{\mathbb{R}^m_*\times]s,t]} \eta_s(r)\,H(z,r,X(r,s,x))\tilde{p}(\mathrm{d}z,\mathrm{d}r), \end{cases} \tag{7.128}$$

where $X(t,s,x)$ is the solution of (7.119).

From now on, we denote this process $\eta_s(t)$ by $\eta^t(s,x)$, because we want to study the dependency on (s,x). Under conditions (7.116), (7.117), (7.118) and (7.127), the unique solution of problem (7.128) is given by

$$\begin{cases} \eta^t(s,x) = \exp\Big\{\displaystyle\int_s^t \big(G(r,X(r,s,x)),\mathrm{d}w(r)\big) + \\ \quad + \displaystyle\int_s^t \big[F(r,X(r,s,x)) - \tfrac{1}{2}|G(r,X(r,s,x))|^2\big]\,\mathrm{d}r + \\ \quad + \displaystyle\int_{\mathbb{R}^m_*\times]s,t]} H(z,r,X(r,s,x))\tilde{p}(\mathrm{d}z,\mathrm{d}r) + \\ \quad + \displaystyle\int_s^t \mathrm{d}r \int_{\mathbb{R}^m_*} \big[H(z,r,X(r,s,x)) - \ln\big(1+H(z,t,X(r,s,x))\big)\big]\pi(\mathrm{d}z)\Big\}. \end{cases} \tag{7.129}$$

We want to compute the backward Itô differential of $\eta^t(s,x)$ in s. To this end, let us introduce the following operators

$$\begin{cases} \mathcal{L}_s^{F,G,H}\psi(x) = \mathcal{L}_s\psi(x) + \big(\sigma(s,x)G(s,x),D\psi(x)\big) + \\ \quad + \displaystyle\int_{\mathbb{R}^m_*} H(z,s,x)[\psi(x+\gamma(z,s,x)) - \psi(x)]\pi(\mathrm{d}z) + F(s,x)\psi(x), \end{cases} \tag{7.130}$$

and

$$\begin{cases} \mathcal{M}_s^G\psi(x) = \mathcal{M}_s\psi(x) + G(s,x)\psi(x), \\ \\ \mathcal{N}_s^H(z)\psi(x) = \mathcal{N}_s(z)\psi(x) + H(z,s,x)\psi(x+\gamma(z,s,x)), \end{cases} \tag{7.131}$$

for ψ in $C_b^2(\mathbb{R}^d)$. Then, by using the same technique as in the proof of Proposition 7.12, we can prove the following result:

Proposition 7.16. *Let us assume that conditions* (7.116), (7.117), (7.118) *and* (7.127) *hold. Then we have*

$$\begin{cases} \eta^t(s,x) = 1 + \displaystyle\int_{\mathbb{R}_*^m\times[s,t[} \mathcal{N}_r^H(z)\eta^t(r,\cdot)(x)\tilde{p}(\mathrm{d}z,\hat{\mathrm{d}}r) + \\ \qquad + \displaystyle\int_s^t \big(\mathcal{M}_r^G\eta^t(r,\cdot)(x),\hat{\mathrm{d}}w_r\big) + \int_s^t \mathcal{L}_r^{F,G,H}\eta^t(r,\cdot)(x)\mathrm{d}r, \end{cases} \tag{7.132}$$

for any $0 \le s < t \le T$, *and* x *in* $\mathbb{R}^d$.

Proof. For this end we note first that $\eta^t(s,x)$ verifies the following identity

$$\eta^t(s,x) = \eta^t(r, X(r,s,x))\ \eta^r(s,x), \tag{7.133}$$

where $s \le r \le t$.

As usual, the first step is to prove the result without the small jumps, actually we approximate π, p and $\tilde{p}$ by π_n, p_n and $\tilde{p}_n$ as before.

Considering $X_n(t)$ given by (7.123) and

$$\begin{aligned} \eta_n^t(s,x) = 1 &+ \int_s^t \eta_n^r(s,x)\, F(r, X_n(r,s,x))\mathrm{d}r + \\ &+ \int_s^t \eta_n^r(s,x)\big(G(r, X_n(r,s,x)), \mathrm{d}w(r)\big) + \\ &+ \int_{\mathbb{R}_*^m\times]s,t]} \eta_n^r(s,x)\, H_n(z,r,X_n(r,s,x))\tilde{p}_n(\mathrm{d}z,\mathrm{d}r), \end{aligned}$$

which can be rewrite as

$$\begin{aligned} \eta_n^t(s,x) = 1 &+ \int_s^t \eta_n^r(s,x)\, F(r, X_n(r,s,x))\mathrm{d}r + \\ &+ \int_s^t \eta_n^r(s,x)\big(G(r, X_n(r,s,x)), \mathrm{d}w(r)\big) + \\ &+ \sum_i \eta_n^{\tau_i}(s,x)\, H(z_i,\tau_i, X_n(\tau_i,s,x))\mathbb{1}_{\{s<\tau_i\le t\}} - \\ &- \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \eta_n^r(s,x)\, H(z,r,X_n(r,s,x))\pi_n(\mathrm{d}z), \end{aligned}$$

then we have

$$\begin{aligned} \eta_n^t(s,x) - 1 = &\sum_i \{\eta_n^t(s\vee\tau_{i-1},x) - \eta_n^t(s\vee\tau_i-,x)\} + \\ &+ \sum_i \{\eta_n^t(s\vee\tau_i-,x) - \eta_n^t(s\vee\tau_i,x)\}. \end{aligned}$$

The first sum can be dealt as a continuous process, that is like (7.125) and

$$\begin{aligned} \eta_n^t(s,x) = 1 &+ \int_s^t \eta_n^r(s,x)\, \tilde{F}_n(r, X_n(r,s,x))\mathrm{d}r + \\ &+ \int_s^t \eta_n^r(s,x)\big(G(r, X_n(r,s,x)), \mathrm{d}w(r)\big) \end{aligned}$$

where

$$\tilde{F}_n(t,x) = F(t,x) - \int_{\mathbb{R}^m_*} H(z,t,x)\pi_n(\mathrm{d}z).$$

Then, let $\sigma = \{s = s_0 < s_1 < \ldots < s_N = t\}$ in Σ_{st}. Then we have

$$\eta_n^t(s,x) - 1 = \sum_{k=1}^N \left[\eta_n^t(s_{k-1},x) - \eta_n^t(s_k,x)\right].$$

Now we note that, in view of (7.133), we have

$$\eta_n^t(s_{k-1},x) - \eta_n^t(s_k,x) = \eta_n^t(s_k, X_n(s_k,s_{k-1},x))\,\eta_n^{s_k}(s_{k-1},x) - \eta_n^t(s_k,x),$$

and then we can write $\eta_n^t(s_{k-1},x) - \eta_n^t(s_k,x) = J_1^k + J_2^k + J_3^k$, where

$$\begin{aligned}
J_1^k &= \eta_n^t(s_k, X_n(s_k,s_{k-1},x)) - \eta_n^t(s_k,x),\\
J_2^k &= \eta_n^t(s_k,x)\big(\eta_n^{s_k}(s_{k-1},x) - 1\big),\\
J_3^k &= \big[\eta_n^t(s_k, X_n(s_k,s_{k-1},x)) - \eta_n^t(s_k,x)\big]\big(\eta_n^{s_k}(s_{k-1},x) - 1\big).
\end{aligned}$$

Considering $J_1 = \sum_{k=1}^N J_1^k$ and arguing as in preceding section, we can write

$$\begin{aligned}
J_1 = \sum_{k=1}^N \big[\mathcal{L}^0_{s_k}\eta_n^t(s_k,\cdot)(x) - \int_{\mathbb{R}^m_*} (\gamma(z,s_k,x), D\eta_n^t(s_k,x))\pi_n(\mathrm{d}z)\big](s_k - s_{k-1})\\
+ \sum_{k=1}^N \big(\mathcal{M}_{s_k}\eta_n^t(s_k,\cdot)(x), w(s_k) - w(s_{k-1})\big) + R'(|\sigma|).
\end{aligned}$$

Analogously, for $J_2 = \sum_{k=1}^N J_2^k$, we can write

$$\begin{aligned}
J_2 = \sum_{k=1}^N \eta_n^t(s_k,x)\big[F(s_k,x) - \int_{\mathbb{R}^m_*} H(z,s_k,x)\pi_n(\mathrm{d}z)\big](s_k - s_{k-1}) +\\
+ \sum_{k=1}^N \eta_n^t(s_k,x)\big(G(s_k,x), w(s_k) - w(s_{k-1})\big) + R''(|\sigma|).
\end{aligned}$$

Finally, for $J_3 = \sum_{k=1}^N J_3^k$, we can write

$$J_3 = \sum_{k=1}^N \big(\mathcal{M}_{s_k}\eta_n^t(s_k,\cdot)(x), G(s_k,x)\big)(s_k - s_{k-1}) + R'''(|\sigma|).$$

Taking into account that $X_n(t,s,x)$ fulfills equation (7.119), and that $\eta_n^t(s,x)$ is the solution to (7.128), we have, arguing as in the proof of Itô formula,

$$R(|\sigma|) = R'(|\sigma|) + R''(|\sigma|) + R'''(|\sigma|) \to 0,$$

in probability as $|\sigma| \to 0$. Since $\eta_n^t(s,x) - 1 = J_1 + J_2 + J_3$, letting $|\sigma|$ tend to 0, we have the following contribution to $\eta_n^t(s,x)$ from the first sum

$$
\begin{aligned}
1 + \int_s^t \mathcal{L}_r^0 \eta_n^t(r,x)\mathrm{d}r - \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big(\gamma(z,r,x), D\eta_n^t(r,x)\big)\pi_n(\mathrm{d}z) + \\
+ \int_s^t \big(\mathcal{M}_r \eta_n^t(r,\cdot)(x), G(r,x)\big)\mathrm{d}r + \\
+ \int_s^t \eta_n^t(r,x)\big[F(r,x) - \int_{\mathbb{R}_*^m} H(z,r,x)\pi_n(\mathrm{d}z)\big]\mathrm{d}r \\
+ \int_s^t \big(\mathcal{M}_r \eta_n^t(r) + \eta_n^t(r,x)\, G(r,x), \hat{\mathrm{d}}w(r)\big).
\end{aligned}
$$

For the second sum, i.e., the jumps, we have to use the analogous of identity (7.133), that is

$$
\eta_n^t(\tau_i-,x) = \eta_n^t(\tau_i, X(\tau_i,\tau_i-,x))\eta_n^{\tau_i}(\tau_i-,x).
$$

Hence,

$$
\begin{aligned}
\sum_i \{\eta_n^t(s\vee\tau_i-,x)) - \eta_n^t(s\vee\tau_i,x))\} = \\
= \sum_i \{\eta_n^t(\tau_i-,x) - \eta_n^t(\tau_i,x)\}\mathbb{1}_{\{s<\tau_i\le t\}} = \\
= \sum_i \{\eta_n^t(\tau_i, X(\tau_i,\tau_i-,x))\eta_n^{\tau_i}(\tau_i-,x) - \eta_n^t(\tau_i,x)\}\mathbb{1}_{\{s<\tau_m\le t\}} = \\
= \sum_i \{\eta_n^t(\tau_i, X(\tau_i,\tau_i-,x)) - \eta_n^t(\tau_i,x) + \\
+\eta_n^t(\tau_i, X(\tau_i,\tau_i-,x))[\eta_n^{\tau_i}(\tau_i-,x) - 1]\}\mathbb{1}_{\{s<\tau_m\le t\}}
\end{aligned}
$$

Writing this sum as an integral, that is

$$
\begin{aligned}
\int_{\mathbb{R}_*^m\times[s,t)} \big[\eta_n^t(r-, x+\gamma(z,r,x)) - \eta_n^t(r,x)\big]p_n(\mathrm{d}z,\hat{\mathrm{d}}r) + \\
+ \int_{\mathbb{R}_*^m\times[s,t)} \big[\eta_n^t(r-, x+\gamma(z,r,x))H(z,r,x)\big]p_n(\mathrm{d}z,\hat{\mathrm{d}}r),
\end{aligned}
$$

we arrive to write it in the final form

$$
\begin{aligned}
\sum_i \{\eta_n^t(s\vee\tau_i-,x)) - \eta_n^t(s\vee\tau_i,x))\} = \\
= \int_{[s,t)\times\mathbb{R}_*^m} \big[\eta_n^t(r-, x+\gamma(z,r,x)) - \eta_n^t(r,x)\big]\tilde{p}_n(\mathrm{d}z,\hat{\mathrm{d}}r) + \\
+ \int_s^t \int_{\mathbb{R}_*^m} \big[\eta_n^t(r-, x+\gamma(z,r,x)) - \eta_n^t(r,x)\big]\pi_n(\mathrm{d}z)\,\mathrm{d}r + \\
+ \int_{[s,t)\times\mathbb{R}_*^m} \big[\eta_n^t(r-, x+\gamma(z,r,x))H(z,r,x)\big]\tilde{p}_n(\mathrm{d}z,\hat{\mathrm{d}}r) + \\
+ \int_s^t \int_{\mathbb{R}_*^m} \big[\eta_n^t(r-, x+\gamma(z,r,x))H(z,r,x)\big]\pi_n(\mathrm{d}z)\,\mathrm{d}r.
\end{aligned}
$$

Thus, we have established (7.122) for $\eta_n^t(s,x)$. Now, we can pass to the limit for $n \to \infty$ in the same way as in the preceding section. □

7.7.3 Generalized Girsanov Formula

Using the notation for the operators $\mathcal{L}_s^{F,G,H}$, $\mathcal{M}_s^G$ and $\mathcal{N}_s^H(z)$, defined by (7.130) and (7.131), we have the following theorem

Theorem 7.17. *Assume that conditions (7.116), (7.117), (7.118) and (7.127) hold. Let φ in $C_b^2(\mathbb{R}^d)$ and denote*

$$y^t(s,x) = \varphi(X(t,s,x))\eta^t(s,x).$$

Then $y^t(s,x)$ satisfies the backward stochastic partial differential equation

$$\begin{cases} y^t(s,x) = \varphi(x) + \displaystyle\int_s^t \mathcal{L}_r^{F,G,H} y^t(r,\cdot)(x)\mathrm{d}r + \\ \quad + \displaystyle\int_s^t \big(\mathcal{M}_r^G y^t(r,\cdot)(x), \hat{\mathrm{d}}w_r\big) + \int_{\mathbb{R}_*^m \times [s,t[} \mathcal{N}_r^H(z) y^t(r,\cdot)(x)\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r), \end{cases} \tag{7.134}$$

where the operators in (7.134) are defined in the previous section. Here recall that $\hat{\mathrm{d}}w(t)$ and $\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}t)$ mean the backward stochastic integration.

Proof. We point out that we do not need any change of probability space to obtain a Girsanov formula.

The proof of Theorem 7.17 uses the following Lemma 7.18 as follows. First we apply the following lemma with $u(s) = \varphi(X(t,s,x))$ and $v(s) = \eta^t(s,x)$. Recalling (7.122) and (7.132), we easily arrive to the conclusion. □

Lemma 7.18. *Let us suppose that $u(s)$, $v(s)$ be of the following form*

$$\begin{aligned} u(s) - u(t) &= \int_s^t u_1(r)\mathrm{d}r + \int_s^t u_2(r)\,\hat{\mathrm{d}}w_r + \int_{\mathbb{R}_*^m \times [s,t)} u_3(z,r)\,\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r), \\ v(s) - v(t) &= \int_s^t v_1(r)\mathrm{d}r + \int_s^t v_2(r)\,\hat{\mathrm{d}}w_r + \int_{\mathbb{R}_*^m \times [s,t)} v_3(z,r)\,\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r). \end{aligned}$$

Then we have

$$\begin{cases} u(s)\,v(s) - u(t)\,v(t) = \\ \quad = \displaystyle\int_s^t \Big(u_1(r)\,v(r) + u(r)\,v_1(r) + u_2(r)\,v_2(r) + \\ \qquad + \displaystyle\int_{\mathbb{R}_*^m} u_3(z,r)v_3(z,r)\pi(\mathrm{d}z)\Big)\mathrm{d}r + \\ \quad + \displaystyle\int_s^t \Big(u_2(r)\,v(r) + u(r)\,v_2(r)\Big)\hat{\mathrm{d}}w_r + \\ \quad + \displaystyle\int_{[s,t)\times\mathbb{R}_*^m} \Big(u_3(z,r)\,v(r) + u(r)\,v_3(z,r) + \\ \qquad + u_3(z,r)v_3(z,r)\Big)\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r) \end{cases} \tag{7.135}$$

Proof. The procedure is just the same as before. As usual, the first step is to prove the result without the small jumps, actually we approximate π, p and $\tilde{p}$ by π_n, p_n and $\tilde{p}_n$ as before, i.e., $\pi_n(B) = \pi(\{z \in B \,:\, n|z| \geq 1\})$, $n = 1, 2, \dots$ and B any Borel subset of $\mathbb{R}^m_*$.

Thus, let us consider the following processes

$$\begin{aligned} u^{(n)}(s) - u^{(n)}(t) &= \int_s^t u_1(r)\mathrm{d}r + \int_s^t u_2(r)\,\hat{\mathrm{d}}w_r + \\ &\quad + \int_{[s,t)\times\mathbb{R}^m_*} u_3(z,r)\,\tilde{p}_n(\mathrm{d}z, \hat{\mathrm{d}}r) \\ v^{(n)}(s) - v^{(n)}(t) &= \int_s^t v_1(r)\mathrm{d}r + \int_s^t v_2(r)\,\hat{\mathrm{d}}w_r + \\ &\quad + \int_{[s,t)\times\mathbb{R}^m_*} v_3(z,r)\,\tilde{p}_n(\mathrm{d}z, \hat{\mathrm{d}}r), \end{aligned}$$

then, there exists a sequence of stopping times (where one jump occurs) $0 < \tau_1 < \tau_2 < \cdots$ and jumps $z_1, z_2, \dots$, such that we can write the preceding equations as

$$\begin{aligned} u^{(n)}(s) - u^{(n)}(t) &= \int_s^t u_1(r)\mathrm{d}r + \int_s^t u_2(r)\,\hat{\mathrm{d}}w_r + \\ &\quad + \sum_i u_3(z_i,\tau_i)\mathbb{1}_{\{s\leq\tau_i<t\}} - \int_s^t \mathrm{d}r \int_{\mathbb{R}^m_*} u_3(z,r)\,\pi_n(\mathrm{d}z), \\ v^{(n)}(s) - v^{(n)}(t) &= \int_s^t v_1(r)\mathrm{d}r + \int_s^t v_2(r)\,\hat{\mathrm{d}}w_r + \\ &\quad + \sum_i v_3(z_i,\tau_i)\mathbb{1}_{\{s\leq\tau_i<t\}} - \int_s^t \mathrm{d}r \int_{\mathbb{R}^m_*} v_3(z,r)\,\pi_n(\mathrm{d}z), \end{aligned}$$

then we have

$$\begin{aligned} &u^{(n)}(s)v^{(n)}(s) - u^{(n)}(t)v^{(n)}(t) = \\ &\quad = \sum_i \{u^{(n)}(s\vee\tau_{i-1})v^{(n)}(s\vee\tau_{i-1}) - u^{(n)}(s\vee\tau_i-)v^{(n)}(s\vee\tau_i-)\} + \\ &\qquad + \sum_i \{u^{(n)}(s\vee\tau_i-)v^{(n)}(s\vee\tau_i-) - u^{(n)}(s\vee\tau_i)v^{(n)}(s\vee\tau_i)\}. \end{aligned}$$

The first sum can be dealt as a continuous process, that is (we drop the index n to simplify notation)

$$\begin{aligned} u(s) - u(t) &= \int_s^t \bar{u}_1(r)\mathrm{d}r + \int_s^t u_2(r)\,\hat{\mathrm{d}}w_r, \\ v(s) - v(t) &= \int_s^t \bar{v}_1(r)\mathrm{d}r + \int_s^t v_2(r)\,\hat{\mathrm{d}}w_r \end{aligned}$$

with

$$\bar{u}_1(r) = u_1(r) - \int_{\mathbb{R}_*^m} u_3(z,r)\,\pi_n(\mathrm{d}z),$$
$$\bar{v}_1(r) = v_1(r) - \int_{\mathbb{R}_*^m} v_3(z,r)\,\pi(\mathrm{d}z).$$

Hence, let us write

$$\begin{aligned} u(s)\,v(s) - u(t)\,v(t) &= \sum_{k=1}^{N} \Big(u(s_{k-1})\,v(s_{k-1}) - u(s_k)\,v(s_k) \Big) \\ &= \sum_{k=1}^{N} (u(s_{k-1}) - u(s_k))\,v(s_k) + \sum_{k=1}^{N} u(s_k)\,(v(s_{k-1}) - v(s_k)) + \\ &\qquad + \sum_{k=1}^{N} (u(s_{k-1}) - u(s_k))\,(v(s_{k-1}) - v(s_k)). \end{aligned}$$

Now we can write

$$u(s_{k-1}) - u(s_k) = \bar{u}_1(s_k)(s_k - s_{k-1}) + u_2(s_k)(w(s_k) - w(s_{k-1})) + L_k^1,$$

$$v(s_{k-1}) - v(s_k) = \bar{v}_1(s_k)(s_k - s_{k-1}) + v_2(s_k)(w(s_k) - w(s_{k-1})) + L_k^2,$$

and

$$(u(s_{k-1}) - u(s_k))\,(v(s_{k-1}) - v(s_k)) = u_2(s_k)v_2(s_k)(s_k - s_{k-1}) + L_k^3.$$

Setting

$$L(|\sigma|) = \sum_{k=1}^{n} \big\{ v(s_k)L_k^1 + u(s_k)L_k^2 + L_k^3 \big\},$$

it follows that $L(|\sigma|) \to 0$ in probability as $|\sigma| \to 0$. Hence

$$\begin{aligned} u^{(n)}(s)\,v^{(n)}(s) - u^{(n)}(t)\,v^{(n)}(t) &= \\ &= \int_s^t \Big(u_1(r)\,v(r) + u(r)\,v_1(r) + u_2(r)\,v_2(r) - \\ &\quad - \int_{\mathbb{R}_*^m} u(r)v_3(z,r)\,\pi(\mathrm{d}z) - \int_{\mathbb{R}_*^m} u_3(z,r)v(r)\,\pi(\mathrm{d}z) \Big)\mathrm{d}r + \\ &\quad + \int_s^t \Big(u_2(r)\,v(r) + u(r)\,v_2(r) \Big)\,\hat{\mathrm{d}}w_r \end{aligned}$$

For the second sum we have simply

$$\begin{aligned}
&\sum_i \{u^{(n)}(s\vee\tau_i-)v^{(n)}(s\vee\tau_i-) - u^{(n)}(s\vee\tau_i)v^{(n)}(s\vee\tau_i)\} = \\
&= \sum_i \big\{u^{(n)}(\tau_i-)v^{(n)}(\tau_i-) - u^{(n)}(\tau_i)v^{(n)}(\tau_i)\big\}\mathbb{1}_{\{s<\tau_i\le t\}} = \\
&= \int_{[s,t)\times\mathbb{R}^m_*} \big[u^{(n)}(r)v_3(z,r) + u_3(z,r)v^{(n)}(r) + \\
&\qquad\qquad + u_3(z,r)v_3(z,r)\big]p_n(\mathrm{d}z,\mathrm{d}r) = \\
&= \int_{[s,t)\times\mathbb{R}^m_*} \big[u^{(n)}(r)v_3(z,r) + u_3(z,r)v^{(n)}(r) + \\
&\qquad\qquad + u_3(z,r)v_3(z,r)\big]\tilde{p}_n(\mathrm{d}z,\mathrm{d}r) + \\
&\quad + \int_s^t \mathrm{d}r \int_{\mathbb{R}^m_*} \big[u^{(n)}(r)v_3(z,r) + u_3(z,r)v^{(n)}(r) + \\
&\qquad\qquad + u_3(z,r)v_3(z,r)\big]\pi_n(\mathrm{d}z).
\end{aligned}$$

Thus, we have established (7.135) for $u^{(n)}(s)v^{(n)}(s)$. Now, we can pass to the limit for $n\to\infty$ in the same way as in the preceding section. □

Corollary 7.19. *Assume that conditions* (7.116), (7.117), (7.118) *and* (7.127) *hold, G is bounded and F is bounded above. If φ belongs to $C^2_b(\mathbb{R}^d)$, then the function*

$$u(s,x) = \mathbb{E}\big\{y^t(s,x)\big\}, \quad 0<s<t,$$

is a solution to the (backward) Kolmogorov equation

$$\begin{cases} \partial_s u(s,x) + \mathcal{L}^{F,G,H}_s u(s,\cdot)(x) = 0, \quad \forall s<t, \\ u(t,x) = \varphi(x), \quad \forall x\in\mathbb{R}^d. \end{cases} \tag{7.136}$$

in the whole space $\mathbb{R}^d$. □

First, let us consider the case of $F=H=0$. Here we want to find an expression for the transition semigroup corresponding to the following the stochastic differential equation, with t in $[s,T]$,

$$\begin{cases} Y(t) = x + \displaystyle\int_s^t [g(r,Y(r)) + \sigma(r,Y(r)G(r,Y(r))]\mathrm{d}r + \\ \qquad + \displaystyle\int_s^t \sigma(r,Y(r))\mathrm{d}w(r) + \int_{\mathbb{R}^m_*\times]s,t]} \gamma(z,r,Y(r))\tilde{p}(\mathrm{d}z,\mathrm{d}r), \end{cases} \tag{7.137}$$

in terms of $X(t)$, the solution to the "simpler" equation

$$\begin{cases} X(t) = x + \displaystyle\int_s^t g(r,X(r))\,\mathrm{d}r + \int_s^t \sigma(r,X(r))\mathrm{d}w(r) + \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^m_*\times]s,t]} \gamma(z,r,X(r))\tilde{p}(\mathrm{d}z,\mathrm{d}r), \end{cases} \tag{7.138}$$

The following result can be considered as a formulation of Girsanov's theorem.

Theorem 7.20. *Assume that conditions (7.116), (7.117) and (7.127) hold with $F = 0$, $H = 0$, with G bounded. Let φ be in $C_b(\mathbb{R}^d)$, and let $Y(t,s,x)$ and $X(t,s,x)$ be the solutions to (7.137), and (7.138) respectively. Then we have, for every $0 \le s \le t \le T$ and $x \in \mathbb{R}^d$,*

$$\mathbb{E}\big\{\varphi(Y(t,s,x)\big\} = \mathbb{E}\big\{\varphi(X(t,s,x))\eta^t(s,x)\big\}, \tag{7.139}$$

where $\eta^t(s,x)$ is given by

$$\eta^t(s,x) = \exp\Big\{\int_s^t \big(G(X(r,s,x)), \mathrm{d}w(r)\big) - \frac{1}{2}\int_s^t |G(X(r,s,x))|^2 \mathrm{d}r$$

Proof. Actually, first under the regularity assumption (7.118) and assuming φ in $C_b^2(\mathbb{R}^d)$, we only need to prove that both sides of identity (7.139), considered as functions of (s,x) fulfill the Kolmogorov equation

$$\begin{cases} \partial_s u(s,x) + \mathcal{L}_s^{0,G,0} u(s,\cdot)(x) = 0, \quad \forall s < t,\ x \in \mathbb{R}^d, \\ u(t,x) = \varphi(x), \quad \forall x \in \mathbb{R}^d. \end{cases}$$

This is obviously true for the left hand side. For as the right hand side is concerned it is enough to apply formula (7.136), choosing $F = 0$ and $H = 0$. Finally, we approximate the coefficients and φ to conclude. □

Now consider $0 \le s < T$ fixed and the stochastic differential equation associated with the integro-differential operator $\mathcal{L}_s^{0,G,H}$, see (7.130) and (7.131), considered in the variable r, i.e., the probability P_{sx} in the canonical space such that $P_{sx}(\{\omega \in D([s,T],\mathbb{R}^d) : \omega(s) = x\}) = 1$ and the process

$$M_\varphi(t) = \varphi\big(\omega(t)\big) + \int_s^t \mathcal{L}_r^{0,G,H}\varphi(\omega(r))\mathrm{d}r, \quad \forall t \in [s,T],$$

is a P_{sx}-martingale for every smooth function φ.

Here we want to find an expression for the transition semigroup corresponding to the above probability P_{sx} in terms of the same $X(t)$ as in (7.138). The following result can be considered as a formulation of Girsanov's theorem.

Theorem 7.21. *Assume that conditions (7.116), (7.117), and (7.127) hold with $F = 0$, with G and H bounded. Let φ be in $C_b(\mathbb{R}^d)$, and let $X(t,s,x)$ be the solution to (7.138) and P_{sx} be the unique martingale solution associated with the integro differential operator $\mathcal{L}_r^{0,G,H}$. Then we have, for every $0 \le s \le t \le T$ and $x \in \mathbb{R}^d$,*

$$\mathbb{E}_{sx}\big\{\varphi\big(\omega(t)\big)\big\} = \mathbb{E}\big\{\varphi(X(t,s,x))\eta^t(s,x)\big\}, \tag{7.140}$$

where $\mathbb{E}_{sx}$ denotes the expectation with respect to P_{sx} and $\eta^t(s,x)$ is given by

$$\begin{aligned} \eta^t(s,x) = \exp\Big\{ & \int_s^t \big(G(X(r,s,x)), \mathrm{d}w(r)\big) - \frac{1}{2}\int_s^t |G(X(r,s,x))|^2 \mathrm{d}r + \\ & + \int_{\mathbb{R}_*^m \times]s,t]} H(z,r,X(r,s,x))\tilde{p}(\mathrm{d}z,\mathrm{d}r) + \\ & + \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big[H(z,r,X(r,s,x)) - \ln\big(1 + H(z,t,X(r,s,x))\big)\big]\pi(\mathrm{d}z)\Big\}. \end{aligned}$$

Proof. This is essentially the same of the case $F = 0$ and $H = 0$, the only point to notice is that now, the right-hand side of (7.140) is used to define a probability which is the solution of the desired martingale problem. □

Note that in the previous case, the stochastic equations for Y and X could be set (but non necessarily) in the same probability space with the same Wiener process and Poisson measure. However, when the jumps are involved, one may have the stochastic equations for X set in an arbitrary probability space, and by taken the image, one may suppose that (7.138) is really set on the canonical probability space $D([0,\infty),\mathbb{R}^d)$ where the canonical process $\omega(t) = X(t,\omega)$ solves the stochastic equation under the initial probability measure. Then, under the new probability measure P_{sx}, the same canonical process $\omega(t)$ solves the equivalent of equation (7.137), which is given as a martingale problem or by specifying the characteristics of the canonical process under P_{sx}, namely, the drift and diffusion terms are as in (7.137), but the integer measure ν (associated with the jumps of ω) has

$$\nu^p(B\times]a,b]) = \int_{]a,b]} \mathrm{d}r \int_{\left\{z\in\mathbb{R}^m_*:\gamma(z,r,\omega(r-))\in B\right\}} (1 + H(z,r,\omega(r-)))\pi(\mathrm{d}z),$$

for every B in $\mathcal{B}(\mathbb{R}^d_*)$, $0 \le a < b$, as its predictable compensator, i.e., the Lévy measure or kernel has changed from $\pi\{z \in \mathbb{R}^m_* : \gamma(z,r,\cdot) \in B\}\,\mathrm{d}s$ into

$$\mathtt{M}(B,s,\cdot)\,\mathrm{d}s = \Big(\int_{\{z\in\mathbb{R}^m_*:\gamma(z,r,\cdot)\in B\}} (1 + H(z,r,\cdot))\pi(\mathrm{d}z)\Big)\mathrm{d}s,$$

as expected.

7.7.4 Feynman-Kac Formula

We have

Theorem 7.22. *Assume that conditions (7.116), (7.117), (7.118) and (7.127) hold, with $G = 0$ and $H = 0$, and that F is bounded above. If φ is in $C^2_b(\mathbb{R}^d)$, then the function*

$$u(s,x) = \mathbb{E}\big\{y^t(s,x)\big\}, \quad 0 < s < t,$$

where $y^t(s,x) = \varphi(X(t,s,x))\eta^t(s,x)$, is a solution to the (backward) Kolmogorov equation

$$\begin{cases} u_s(s,x) + \mathcal{L}_s u(s,\cdot)(x) + F(s,x)\,u(s,x) = 0, \quad \forall s < t,\ x \in \mathbb{R}^d, \\ u(t,x) = \varphi(x), \quad x \in \mathbb{R}^d \end{cases}$$

where $\eta^t(s,x)$ is given by (7.128) or (7.129).

Proof. Since $G = H = 0$, we have from (7.129)

$$\eta^t(s,x) = \exp\Big\{\int_s^t F(X(r,s,x))\mathrm{d}r\Big\}.$$

Moreover formula (7.134) becomes

$$y^t(s,x) = \varphi(x) + \int_s^t \big[\mathcal{L}_r y^t(r,\cdot)(x) + F(r,x)y^t(r,x)\big]\mathrm{d}r + \\ + \int_s^t \big(\mathcal{M}_r y^t(r,x), \hat{\mathrm{d}}w_r\big) + \int \mathcal{N}_r(z) y^t(r,x)\tilde{p}(\mathrm{d}z, \hat{\mathrm{d}}r).$$

Taking expectation, the conclusion follows. □

Now we take a look at a stochastic version of the Feynman-Kac formula. Assume that we are given two probability filtered spaces $(\Omega_1, \mathcal{F}^{(1)}, P_1)$ and $(\Omega_2, \mathcal{F}^{(2)}, P_2)$, with respectively a pair of a k_1-dimensional Wiener process w_1 and a (Poisson) random measure $p_1(\mathrm{d}z_1, \mathrm{d}t)$ (with compensator $\pi_1(\mathrm{d}z_1)$), with jump values in $\mathbb{R}^{m_1}_*$, both adapted to a filtration $\{\mathcal{F}^{(1)}_t\}_{t\geq 0}$, and a pair of a k_2-dimensional Wiener process w_2 and a (Poisson) random measure $p_2(\mathrm{d}z_2, \mathrm{d}t)$ (with compensator $\pi_2(\mathrm{d}z_2)$), with jump values in $\mathbb{R}^{m_2}_*$, both adapted to a filtration $\{\mathcal{F}^{(2)}_t\}_{t\geq 0}$.

Let us consider $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}^{(1)} \otimes \mathcal{F}^{(2)}$, $w = (w_1, w_2)$ and that $G = (0, G_2)$ where G_2 is k_2-dimensional and the matrix $\sigma = (\sigma_1, \sigma_2)$ where σ_1 is $n \times k_1$ matrix and σ_2 is $n \times k_2$ matrix. Moreover, we can define on $\mathbb{R}^m_*$, where $m = m_1 + m_2$, the random measure

$$p(\mathrm{d}z, \mathrm{d}t) = p_1(\mathrm{d}z_1, \mathrm{d}t)\,\delta(\mathrm{d}z_2) + \delta(\mathrm{d}z_1)\,p_2(\mathrm{d}z_2, \mathrm{d}t)$$

with associated compensator

$$\pi(\mathrm{d}z) = \pi_1(\mathrm{d}z_1)\,\delta(\mathrm{d}z_2) + \delta(\mathrm{d}z_1)\,\pi_2(\mathrm{d}z_2),$$

and $\gamma(z,t,x) = \gamma_1(z_1,t,x) + \gamma_2(z_2,t,x)$; we can suppose that $\gamma_i(0,t,x) = 0$. hence, we consider the solution $X(t) = X(t,s,x)$ to the equation (7.119), that can be written as

$$X(t) = x + \int_s^t g(r, X(r))\,\mathrm{d}r + \\ + \int_s^t \sigma_1(r, X(r))\mathrm{d}w_1(r) + \int_s^t \sigma_2(r, X(r))\mathrm{d}w_2(r) + \\ + \int_{\mathbb{R}^{m_1}_* \times]s,t]} \gamma_1(z_1, r, X(r))\tilde{p}_1(\mathrm{d}z_1, \mathrm{d}r) + \int_{\mathbb{R}^{m_2}_* \times]s,t]} \gamma_2(z_2, r, X(r))\tilde{p}_2(\mathrm{d}z_2, \mathrm{d}r).$$

We are here concerned with the following stochastic partial differential equation

$$u(s,x) = \varphi(x) + \int_s^t \Big\{\mathcal{L}_r u(r,\cdot)(x) + \sigma_2(r,x)G_2(r,x), Du(r,\cdot)(x)\big) + \\ + \int_{\mathbb{R}^{m_2}_*} H_2(z_2,r,x)[u(r, x+\gamma_2(z,r,x)) - u(r,x)]\pi_2(\mathrm{d}z_2)\Big\}\mathrm{d}r + \\ + \int_s^t \big(\sigma_2^*(r,x)Du(\cdot)(x) + G_2(r,x)u(r,x), \hat{\mathrm{d}}w_2(r)\big) + \\ + \int_{\mathbb{R}^{m_2}_* \times [s,t)} [u(r, x+\gamma_2(z_2,r,x)) - u(r,x) + H_2(z_2,r,x)u(r,x)]\tilde{p}_2(\mathrm{d}z_2, \hat{\mathrm{d}}r). \tag{7.141}$$

In our case, we can write $\mathcal{L}_r = \mathcal{L}_r^0 + \mathcal{L}_r^{\gamma_1} + \mathcal{L}_r^{\gamma_2}$ where for $i = 1, 2$

$$\mathcal{L}_s^{\gamma_i}\varphi(x) = \int_{\mathbb{R}_*^{m_i}} [\varphi(x + \gamma_i(z, s, x)) - \varphi(x) - (\gamma_i(z, s, x), D\varphi(x))]\pi_i(\mathrm{d}z).$$

This problem arises in studying Filtering Theory, e.g., see Pardoux [194] or Rozovskii [218].

Let us consider

$$\eta^t(s, x) = \exp\Big\{ \int_s^t \big(G_2(X(r, s, x)), dw_2(r)\big) - \frac{1}{2}\int_s^t |G_2(X(r, s, x))|^2)\mathrm{d}r + \\ + \int_{\mathbb{R}_*^{m_2}\times]s,t]} H_2(z_2, r, X(r, s, x))\tilde{p}_2(\mathrm{d}z_2, \mathrm{d}r) + \\ + \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^m} \big[H_2(z_2, r, X(r, s, x)) - \ln\big(1 + H_2(z_2, t, X(r, s, x))\big)\big]\pi_2(\mathrm{d}z_2)\Big\}.$$

where $X(t, s, x)$ is the process indicated above.

The following result gives a representation formula for the solution to (7.141).

Theorem 7.23. *Assume that conditions (7.116), (7.117), (7.118) and (7.127) hold with $F = 0$, and let φ be in $C_b^2(\mathbb{R}^d)$. Then function (here we denote by $\mathbb{E}_1$ the expectation value with respect to first variable ω_1)*

$$u(s, x) = \mathbb{E}_1\big\{\varphi(X(t, s, x))\eta^t(s, x)\big\}, \quad \forall s \in [0, t],$$

is a solution to the stochastic Kolmogorov equation (7.141).

Proof. Consider $y^t(s, x) = \varphi(X(t, s, x))\eta^t(s, x)$. This process satisfies the equation

$$y^t(s, x) = \varphi(x) + \int_s^t \big[\mathcal{L}_r y^t(r, \cdot)(x) + \big(\sigma_2(r, x)G_2(r, x), Dy^t(r, \cdot)(x)\big)\big]\mathrm{d}r + \\ + \int_s^t \mathrm{d}r \int_{\mathbb{R}_*^{m_2}} H_2(z_2, r, x)[y^t(r, x + \gamma_2(z_2, r, x)) - y^t(r, x)]\pi_2(\mathrm{d}z_2) + \\ + \int_s^t \big(Dy^t(r, \cdot)(x), \sigma_1(r, x)\hat{\mathrm{d}}w_1(r)\big) + \\ + \int_s^t \big(\sigma_2^*(r, x)Dy^t(r, \cdot)(x) + G_2(r, x)y^t(r, x), \hat{\mathrm{d}}w_2(r)\big) + \\ + \int_{\mathbb{R}_*^{m_1}\times[s,t)} [y^t(r, x + \gamma_1(z_1, r, x)) - y^t(r, x)]\,\tilde{p}_1(\mathrm{d}z_1, \hat{\mathrm{d}}r) + \\ + \int_{\mathbb{R}_*^{m_2}\times[s,t)} [y^t(r, x + \gamma_2(z_2, r, x)) - y^t(r, x) + \\ + H_2(z_2, r, x)y^t(r, x)]\,\tilde{p}_2(\mathrm{d}z_2, \hat{\mathrm{d}}r).$$

Taking conditional expectation $\mathbb{E}_1$, the conclusion follows. □

∴ ...

Bibliography

[1] N.U. Ahmed. *Semigroup Theory with Applications to Systems and Control*. Pitman Research Notes in Mathematics Series No 246. Longman, Harlow, 1991. 100

[2] R.F. Anderson. Diffusions with second order boundary conditions I and II. *Indiana Univ. J.*, 25:367–395 and 403–443, 1975. 694

[3] S.V. Anulova. On process with Lévy generating operators in a half-space. *Math. USSR–Izv*, 13:9–51, 1979. 486, 670, 694

[4] S.V. Anulova. On stochastic differential equations with boundary conditions in a half-space. *Math. USSR–Izv*, 18:423–437, 1982. 486, 670, 694

[5] S.V. Anulova and H. Pragarauskas. Weak Markov solutions of stochastic equations. *Litowskii Matematicheskii Sbornik*, 17:5–26, 1977. 478, 629

[6] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge, 2004. 256, 418, 490, 491, 511

[7] L. Arnold. *Stochastic Differential Equations*. Wiley, New York, 1974. 64, 488, 551, 562

[8] C. Baiocchi and A. Capelo. *Variational and Quasi-Variational Inequalities*. Wiley, New York, 1984. 119, 123

[9] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkäuser, Boston, 1997. vii, 542

[10] Ole E. Barndorff-Nielsen and Albert Shiryaev. *Change of time and change of measure*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2010. 584, 587

[11] E.N. Barron, R. Jensen, and J.L. Menaldi. Optimal control and differential games with measures. *Nonlinear Analysis*, 21:241–268, 1993. 510, 513, 572

[12] R.F. Bass. *Probabilistic Techniques in Analysis*. Springer-Verlag, New York, 1995. 89

[13] R.F. Bass. *Diffusions and Elliptic Operators*. Springer-Verlag, New York, 1997. 89

[14] R.F. Bass. *Stochastic Processes*. Cambridge University Press, Cambridge, 2011. 71

[15] H. Bauer. *Probability Theory and Elements of Measure Theory*. Holt, Rinehart and Winston, New York, 1972. 83

[16] A. Bensoussan and J.L. Lions. *Applications of Variational Inequalities in Stochastic Control*. North-Holland, Amsterdam, 1982. (pages are referred to the French edition). 117, 123, 502, 605, 627, 632, 670

[17] A. Bensoussan and J.L. Lions. *Impulse Control and Quasi-Variational Inequalities*. Gauthier-Villars, Paris, 1984. (pages are referred to the French edition). 61, 117, 146, 149, 409, 630, 632

[18] A. Bensoussan and J. L. Menaldi. Stochastic hybrid control. *J. Math. Anal. Appl.*, 249(1):261–288, 2000. Special issue in honor of Richard Bellman. 489, 519

[19] A. Bensoussan and A. Rascanu. Stochastic variational inequalities in infinite-dimensional spaces. *Numer. Funct. Anal. Optim.*, 18:19–54, 1997. 670

[20] A. Benveniste and J. Jacod. Systems de Lévy des processus de Markov. *Inventiones Math.*, 21:183–198, 1973. 282

[21] J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996. 52, 53, 61, 256, 264, 266, 647, 664

[22] M. Bertoldi and L. Lorenzi. *Analytical Methods for Markov Semigroups*. Chapman & Hall / CRC Press, Boca Raton (FL), 2007. 127

[23] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, 2nd, two-volume set edition, 2001. vii

[24] P. Besala. On the existence of a fundamental solution for a parabolic differential equation with unbounded coefficients. *Ann. Polonici Math.*, 29:403–409, 1975. 627

[25] K. Bichteler. *Stochastic Integration with jumps*. Cambridge University Press, Cambridge, 2002. 35, 66, 196, 198, 206, 301, 309, 324, 372, 440, 490

[26] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. 64, 65, 72

[27] P. Billingsley. *Probability and Measure*. Wiley, New York, 2nd edition, 1986. 29

[28] R.M. Blumental and R.K. Getoor. *Markov Processes and Potential Theory*. Academic Press, New York, 1968. 45, 49, 50, 55, 229, 263, 276, 281

[29] A. Bobrowski. *Functional analysis for probability and stochastic processes*. Cambridge University Press, Cambridge, 2005. 99

[30] A. N. Borodin and P. Salminen. *Handbook of Brownian Motion - Facts and Formulae*. Birkäuser, Boston, 1996. 175

[31] L. Breiman. *Probability*. Society for Industrial and Applied Mathematics, Philadelphia, 1992. 13

[32] P. Bremaud. *Point Processes and Queues*. Springer-Verlag, New York, 1981. 61, 186, 193, 195, 243, 266, 284, 332

[33] P. Bremaud. *Markov Chains*. Springer-Verlag, New York, 1998. 37, 243

[34] H. Brezis. *Operateurs Maximaux Monotones*. North-Holland, Amsterdam, 1973. 106, 117, 123

[35] H. Brezis. *Analyse Fonctionnelle*. Masson, Paris, 1983. 94

[36] S. Cerrai. *Second Order PDE's in Finite and Infinite Dimensions. A probabilistic Approach*. Lecture Notes in Math. No. 1762. Springer-Verlag, New York, 2001. 540, 627, 632

[37] M. Chaleyat-Maurel, N. El Karoui, and B. Marchal. Reflexion discontinue et systemes stochastiques. *Ann. Probab.*, 8:1049–1067, 1980. 694

[38] T.M. Chao and C.S. Chou. Anticipative stochastic integrals equations driven by semimartingales. *Stochastic Anal. Appl.*, 20(4):755–782, 2002. 434

[39] L. Chaumont and M. Yor. *Exercises in probability*. Cambridge University Press, Cambridge, second edition, 2012. A guided tour from measure theory to random processes, via conditioning. 82

[40] Y.Z. Chen and L.C. Wu. *Second Order Elliptic Equations and Elliptic Systems*. Amer. Math. Soc., Providence, Rhode Island, 1998. 119

[41] T. S. Chihara. *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York, 1978. 318

[42] Pao-Liu Chow. *Stochastic partial differential equations*. Chapman & Hall/CRC, Boca Raton, FL, 2007. 518

[43] K.L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, New York, 2nd edition, 1980. 37, 38, 63, 193

[44] K.L. Chung. *From Markov Processes to Brownian Motion*. Springer-Verlag, New York, 1982. 50, 58, 59, 283

[45] K.L. Chung and R.J. Williams. *Introduction to Stochastic Integration*. Birkhäuser, Boston, 2nd edition, 1990. 200, 301, 373, 374, 381, 396, 398, 409, 488, 490, 583

[46] E. Çınlar. *Probability and stochastics*. Springer, New York, 2011. 22, 50, 330

[47] E. Cinlar and J. Jacod. *Representation of semimartingale Markov processes in terms of Wiener processes and Poisson random measures*, volume 1 of *Seminar on Stochastic Processes (Evanston, Illinois, 1981)*, pages 159–242. Birkhäuser, Boston, Massachussetts, 1981. 393

[48] J. Conway. *A Course in Functional Analysis*. Springer-Verlag, New York, 1990. 94

[49] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1 and 2. Wiley, New York, 1989. 94

[50] G. Da Prato, J.L. Menaldi, and L. Tubaro. Some applications of backward Itô formula. *Probability Survey*, 00:000–000, 2006 HOPE! 763

[51] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimension*. Cambridge University Press, Cambridge, 1992. 30, 79, 96, 116, 253, 307, 488

[52] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. Cambridge University Press, Cambridge, 1996. 71, 117, 127, 136, 146, 257, 556

[53] G. Da Prato and J. Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces*. Cambridge University Press, Cambridge, 2002. 518, 542

[54] G. Dal Maso and F. Rampazzo. On systems of ordinary differential equations with measures as controls. *Differential Integral Equations*, 4(4):739–765, 1991. 510, 513

[55] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York, 1998. 243

[56] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993. 195, 229, 232, 233, 234, 243, 244, 249, 251, 252, 286

[57] B. De Finetti. *Theory of Probability*, volume 1 and 2. Wiley, London, 1975. 2

[58] C. Dellacherie and P.A. Meyer. *Probabilités et Potentiel*. Hermann, Paris. Chapter I–IV, 1976, Chapters V–VIII (Théories des martingales), 1980, Chapters IX–XI (Théories discréte du potentiel), 1983, Chapters XI–XVI (Théories des processus de Markov), 1987, Chapters XVII-XXIV (with B. Maisonneuve, Processus de Markov - fin). 2, 28, 50, 108, 109, 127, 128, 193, 205, 211, 217, 223, 225, 227, 272, 276, 283, 300, 316, 323, 357, 370, 375, 381, 409, 426, 647

[59] J.L. Doob. *Stochastic Processes*. Wiley, New York, 1953. 11, 29, 36, 43, 57, 72, 193, 288, 294, 355, 374, 388

[60] J.L. Doob. *Classical Potential Theory and its Probabilistic Counterpart*. Springer-Verlag, New York, 1984. 5, 17, 20, 49, 51, 83, 89, 189, 196, 229, 632

[61] J.L. Doob. *Measure Theory*. Springer-Verlag, New York, 1984. 193, 200, 207

[62] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, 2002. 11, 64, 197, 301, 316, 462

[63] N. Dunford and J.T. Schwartz. *Linear Operators, Three Volumes*. Wiley, New York, 1988. 94, 125

[64] P. Dupuis and H. Ishii. SDEs with oblique reflection on nonsmooth domains. *Ann. Probab.*, 21:554–580, 1993. 694

[65] P. Dupuis and K. Ramanan. Convex duality and the Skorokhod problem I and II. *Probab. Theory Related Fields*, 115:153–195 and 197–236, 1999. 694

[66] R. Durrett. *Brownian Motion and Martingale in Analysis*. Wadsworth, Belmont, 1984. 89

[67] R. Durrett. *Stochastic Calculus. A practical introduction*. CRC Press, Boca Raton, 1996. 32, 87, 219, 301, 382

[68] G. Duvaut and J.L. Lions. *Inequations en Mecanique et en Physique*. Springer-Verlag, New York, 1976. 117

[69] E.B. Dynkin. *Foundations of the Theory of Markov Processes*. Prentice-Hall, Englewood Cliffs (NJ), 1961. 50

[70] E.B. Dynkin. *Markov Processes*, volume 1 and 2. Springer-Verlag, New York, 1965. 50, 55, 89, 131, 227, 229

[71] D. Edmunds and W.D. Evans. *Spectral Theory and Differential Operators*. Oxford Univ. Press, New York, 1987. 119

[72] S.D. Eidelman. *Parabolic Systems*. North Holland, Amsterdam, 1969. 120, 632, 633, 730, 733, 739

[73] R.J. Elliott. *Stochastic Calculus and Applications*. Springer-Verlag, New York, 1982. 205, 266, 301

[74] K.J. Engel and R. Nagel. *One-parameter Semigroups for Linear Evolution Equations*. Springer-Verlag, New York, 2000. 89, 93, 103, 109, 125, 632

[75] M. Errami, F. Russo, and P. Vallois. Itô's formula for $C^{1,\lambda}$-functions of a càdlàg process and related calculus. *Probab. Theory Related Fields*, 122:191–221, 2002. 434

[76] S.N. Ethier and T.G. Kurtz. *Markov Processes*. Wiley, New York, 1986. 50, 55, 68, 70, 72, 233, 276, 284, 602, 607, 625

[77] L.C. Evans. *Partial Differential Equations*. Amer. Math. Soc., Providence, Rhode Island, 1998. 118, 119

[78] H.O. Fattorini. *The Cauchy Problem*. Addison-Wesley, Reading, 1983. 89, 117

[79] H. Federer. *Geometric Measure Theory*. Springer-Verlag, New York, 1996. 316

[80] W. Feller. The parabolic differential equations and the associated semigroups of transformations. *Ann. of Math.*, 55:468–519, 1952. 173

[81] W. Feller. *An Introduction to Probability Theory and its Applications*, volume Vols I and II. Wiley, New York, 1960 and 1966. 11, 38, 40, 284

[82] D.L. Fisk. The parabolic differential equations and the associated semigroups of transformations. *Trans. of Am. Math. Soc.*, 120:369–389, 1967. 434

[83] W.H. Fleming and R.W. Rishel. *Deterministic and Stochastic Control Theory*. Springer-Verlag, New York, 1975. vii, 488

[84] W.H. Fleming and H.M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York, 1992. vii

[85] G.B. Folland. *Real Analysis*. Wiley, New York, 2nd edition, 1999. 11, 125

[86] H. Föllmer. *Calcul d'Itó sans probabilités. Séminaire de Proba. XV*, volume 850 of *Lectures Notes in Mathematics*, pages 144–150. Springer-Verlag, Berlin, 1981. 434

[87] M. Freidlin. *Functional Integration and Partial Differential Equations*. Princeton University Press, Princeton, 1985. 488, 632

[88] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliff, 1964. 729, 730

[89] A. Friedman. *Partial Differential Equations*. Holt, Rinehart and Winston, New York, 1969. 117, 120, 633

[90] A. Friedman. *Stochastic Differential Equations and Applications*, volume I and II. Academic Press, New York, 1975 and 1976. 257, 488, 632

[91] A. Friedman. *Variational Principles and Free-Boundary Problems*. Krieger, Malabar, 1988. 117, 123

[92] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and Symmetric Markov Processes*. De Gruyter, Berlin, 1994. 89, 107, 115, 123, 138, 140, 281, 285, 642

[93] M.G. Garroni and J.L. Menaldi. *Green Functions for Second Order Integral-Differential Problems*. Pitman Research Notes in Mathematics Series No 275. Longman, Harlow, 1992. 146, 155, 156, 181, 184, 607, 632, 633, 695, 739, 743, 749

[94] M.G. Garroni and J.L. Menaldi. *Second Order Elliptic Integro-Differential Problems*. Research Notes in Mathematics Series No 430. Chapman & Hall / CRC Press, Boca Raton (FL), 2002. 146, 155, 156, 181, 184, 607, 632, 633, 636, 658, 695, 739, 749

[95] M.G. Garroni and V.A. Solonnikov. On parabolic oblique derivative problem with Hölder continuous coefficients. *Comm. Partial Diff. Eq.*, 9:1323–1372, 1984. 732

[96] I. Gelfand and N.Ya. Vilenkin. *Generalized Functions*, volume 4. Academic Press, New York, 1964. 94

[97] I.I. Gihman and A.V. Skorohod. *The Theory of Stochastic Processes*, volume I, II and III. Springer-Verlag, New York, 1974, 1975 and 1979. 2, 632

[98] I. Gikhman and A. Skorokhod. *Introduction to the Theory of Random Process*. Dover Publications, New York, 1996. 29, 30, 31, 34, 288, 294

[99] I.I. Gikhman and A.V. Skorokhod. *Stochastic Differential Equations*. Springer-Verlag, Berlin, 1972. 146, 149, 409, 488

[100] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, New York, 2nd edition, 1983. 98, 118, 632

[101] B.V. Gnedenko. *The Theory of Probability*. Chelsea, New York, 1989. 11, 40, 42, 73

[102] S. Goldberg. *Unbounded Linear Operators*. Dover, New York, 1966. 103

[103] R.A. Gordon. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*. Am. Math. Soc., Providence, Rhode Island, 1994. 199

[104] P.R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974. 11, 187

[105] S.W. He, J.G. Wang, and J.A. Yan. *Semimartingale theory and stochastic calculus*. Kexue Chubanshe (Science Press), Beijing, 1992. 23, 83, 227, 324, 329, 426, 602

[106] G. Hellwig. *Partial Differential Equations: An Introduction*. Blaisdel Publ., New York, 1964. 118

[107] T. Hida, H. Kuo, J. Potthoff, and L. Streit. *White Noise: An Infinite-Dimensional Calculus*. Kluwer Academic Publishers, Dordrecht, 1993. 308

[108] E. Hille and R.S. Phillips. *Functional Analysis and Semi-Groups*. Am. Math. Soc., Providence, Rhode Island, 1957. 89

[109] H. Holden, B. Oksendal, J. Uboe, and T. Zhang. *Stochastic Partial Differential Equations*. Birkhäuser, Boston, 1996. 74

[110] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland, Amsterdam, 2nd edition, 1989. 68, 219, 364, 391, 488, 577, 600, 622, 646, 668, 669, 767

[111] K. Itô. *Introduction to Probability Theory*. Cambridge University Press, Cambridge, 1978. 13

[112] K. Itô. *Stochastic Processes*. Springer-Verlag, New York, 2004. 256

[113] K. Itô and H.P. McKean. *Diffusion Processes and their Sample Paths*. Springer-Verlag, New York, 1965. 58, 169, 488

[114] S.D. Ivasišen. Green's matrices of boundary value problems for Petrovskii parabolic systems of general form, I and II. *Math. USSR Sbornik*, 42:93–144, and 461–489, 1982. 157, 632, 730, 732, 739

[115] N. Jacob. *Pseudo-Differential Operators and Markov Processes*, volume 1: Fourier Analysis and Semigroups, 2: Generators and Potential Theory, 3: Markov Processes and Applications. Imperial College Press - World Scientific Publishing, Singapur, 2001, 2002 and 2005. 73, 82, 89, 106, 107, 123, 129, 131, 163

[116] J. Jacod and P. Protter. *Probability Essentials*. Springer-Verlag, New York, 2nd edition, 2004. 28

[117] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, 1987. 61, 67, 72, 194, 203, 205, 217, 224, 277, 324, 326, 332, 363, 364, 366, 375, 376, 381, 390, 409, 412, 418, 426, 435, 450, 451, 457, 469, 568, 569, 580, 605, 609, 626, 757

[118] A. Jakubowski. A non-Skorohod topology on the Skorohod space. *Electronic J. Probab.*, 2:1–21, 1997. 463, 467

[119] A. Jakubowski, J. Mémin, and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace ^{1} de skorokhod. *Probab. Theory Related Fields*, 81:111–137, 1989. 467

[120] O. Kallenberg. *Random Measures*. Akademie-Verlag, Berlin, 1983. 324

[121] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 2nd edition, 2002. 14, 34, 51, 72, 82, 83, 221, 223, 333, 414, 426, 438, 490, 583, 596, 602

[122] G. Kallianpur and R.L. Karandikar. *White Noise Theory of Prediction, Filtering and Smothing*. Gordon and Breach, London, 1988. 64, 308

[123] G. Kallianpur and J. Xiong. *Stochastic Differential Equations in Infinite Dimensional Spaces*, volume 26 of *IMS Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1995. 81, 97, 209, 293, 309, 518, 754

[124] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, 2nd edition, 1991. 17, 30, 51, 172, 217, 339, 383, 390, 398, 577, 597, 600, 647, 663, 664

[125] S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York, 1975. 193

[126] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, New York, 2nd edition, 1981. 193

[127] A.F. Karr. *Probability*. Springer-Verlag, New York, 1993. 191

[128] T. Kato. *Perturbations Theory for Linear Operators*. Springer-Verlag, New York, 2nd edition, 1976. 89, 93, 94

[129] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976. 37

[130] R.Z. Khasminskii. *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, The Netherlands, 1980. 35, 138, 503, 616

[131] D. Kinderlehrer and G. Stampacchia. *An introduction to Variational Inequalities and their Applications*. Academic Press, New York, 1980. 119, 123, 670

[132] L. Kleinrock. *Queueing Systems*, volume 1 and 2. Wiley, New York, 1975. 37

[133] F.B. Knight. *Essentials of Brownian Motion and Diffusions*. American Mathematical Society, Providence, R.I., 1981. 59

[134] V.N. Kolokoltsov. *Semiclasical Analysis for Diffusions and Stochastic Processes*, volume 1724 of *Lecture Notes in Mathematics*. Springer-Varlag, Berlin, 2000. 163

[135] T. Komatsu. Markov processes associated with certain integro-differential operator. *Osaka J. Math.*, 10:271–303, 1973. 163

[136] T. Komatsu. Uniform estimates for fundamental solutions associated with non-local Dirichlet form. *Osaka J. Math.*, 32:833–860, 1995. 163

[137] N.V. Krylov. On the selection of a Markov process from a system of processes. *Math. USSR Izvestija*, 37:691–708, 1973. 628

[138] N.V. Krylov. Some estimates of the probability density of a stochastic integral. *Math. USSR Izvestija*, 38:228–248, 1974. 472

[139] N.V. Krylov. *Contolled Diffusion Processes*. Springer-Verlag, New York, 1980. 472, 479, 629

[140] N.V. Krylov. *Nonlinear Elliptic and Parabolic Equations of Second Order*. Reidel, Dordrecht, 1987. 472

[141] N.V. Krylov. *Introduction to the Theory of Diffusion Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 2, 58, 257, 288, 294, 539

[142] N.V. Krylov. *Introduction to the Theory of Random Processes*. American Mathematics Society, Providence, Rhode Island, 1995. 255, 618

[143] H. Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1990. 488, 556, 595, 762, 771

[144] H.H. Kuo. *White Noise Distribution Theory*. CRC Press, Boca Raton, 1996. 308

[145] H.H. Kuo. *Introduction to Stochastic Integration*. Springer-Verlag, New York, 2006. 301

[146] G.S. Ladde and V. Lakshmikantham. *Random Differential Inequalities*. Academic Press, New York, 1980. 248, 487

[147] O.A. Ladyzhenskaya, V.A Solonnikov, and N.N. Uraltseva. *Linear and Quasilinear Equations of Parabolic Type*. Am. Math. Soc., Providence, Rhode Island, 1968. 98, 120, 632, 633, 729

[148] O.A. Ladyzhenskaya and N.N. Uraltseva. *Linear and Quasilinear Elliptic Equations*. Academic Press, New York, 1968. 98, 118

[149] J. Lamperti. *Stochastic Processes. A survey of the mathematical theory*. Springer-Verlag, New York, 1977. 63

[150] E.M. Landis. *Second Order Equations of Elliptic and Parabolic Type*. Am. Math. Soc., Providence, Rhode Island, 1998. 118

[151] W. Laukajtys and L. Slominski. Penalization methods for reflecting stochastic differential equations with jumps. *Stochastics and Stochastics Reports*, 75:275–293, 2003. 464

[152] G. Letta. *Martingales et intégration stochastique*. Quaderni della Scuola Normale Superiore, Pisa, 1984. 266

[153] G.M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific Publishing, Singapur, 1997. 98, 120, 632

[154] M. Lieberstein. *Theory of Partial Differential Equations*. Academic Press, New York, 1972. 119

[155] J.L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*, volume Vol. 1. Dunod, Paris, 1968. 114, 123

[156] P.L. Lions and J.L. Menaldi. Optimal control of stochastic integrals and Hamilton-Jacobi-Bellman equations. *SIAM J Control Optim.*, 20:58–95, 1982. 658

[157] P.L. Lions and A.S. Sznitman. Stochastic differential equations with reflecting boundary conditions. *Comm. Pure Appl. Math.*, 37:511–537, 1984. 694

[158] R.S. Liptser and A.N. Shiryayev. *Theory of Martingales*. Kluwer Academic Publ., Dordrecht, 1986. 70, 222, 223, 224, 225, 227, 310, 375, 584

[159] M. Loève. *Probability Theory*, volume 1 and 2. Springer-Verlag, New York, 4rd edition, 1977 and 1978. 2

[160] A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Boston, 1995. 89, 104, 105, 117, 121

[161] Z. Ma and M. Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Springer-Verlag, Berlin, 1992. 107, 111, 112, 115, 123, 138, 140, 281, 642

[162] P. Malliavin. *Integration and Probability*. Springer-Verlag, New York, 1995. 20, 24, 72, 125, 304

[163] P. Malliavin. *Stochastic Analysis*. Springer-Verlag, New York, 1997. 51

[164] P. Mandl. *Analytical Treatment of One-Dimensional Markov Processes*. Springer-Verlag, New York, 1968. 175

[165] X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishing, Chichester, 1997. 488, 504, 546, 749

[166] M.B. Marcus and J. Rosen. *Markov processes, Gaussian processes, and local times*. Cambridge University Press, Cambridge, 2006. 50, 55, 64, 256, 649

[167] H.P. McKean. *Stochastic Integral*. Academic Press, New York, 1971. 333, 339, 563

[168] Péter Medvegyev. *Stochastic integration theory*. Oxford University Press, Oxford, 2007. 301, 490

[169] J.L. Menaldi. On the optimal impulse control problem for degenerate diffusions. *SIAM J Control Optim.*, 18:722–739, 1980. 658

[170] J.L. Menaldi. On the optimal stopping time problem for degenerate diffusions. *SIAM J Control Optim.*, 18:697–721, 1980. 658

[171] J.L. Menaldi. Stochastic variational inequality for reflected diffusions. *Indiana Univ. Math. J.*, 32:733–744, 1983. 670

[172] J.L. Menaldi. Stochastic hybrid optimal control models. *Aportaciones Matemáticas*, 16:205–250, 2001. 132, 489, 519

[173] J.L. Menaldi and M. Robin. Reflected diffusion processes with jumps. *Ann. Probab.*, 13:319–341, 1985. 670

[174] J.L. Menaldi and M. Robin. Invariant measure for diffusions with jumps. *Appl. Math. Optim.*, 40:105–140, 1999. 627

[175] J.L. Menaldi and S.S. Sritharan. Stochastic 2-d Navier-Stokes equation. *Appl. Math. Optim.*, 46:31–53, 2002. 133

[176] J.L. Menaldi and S.S. Sritharan. Impulse control of stochastic Navier-Stokes equations. *Nonlinear Anal.*, 52:357–381, 2003. 133

[177] J.L. Menaldi and L. Tubaro. Green and Poisson functions with Wentzell boundary conditions. *J. Differential Equations*, 237:77–115, 2007. 156, 696

[178] M. Métivier. *Semimartingales: A course on Stochastic Processes.* De Gruyter, Berlin, 1982. 209, 218, 310, 752

[179] P.A. Meyer. *Probability and potentials.* Blaisdell Publ. Co., Waltham, Mass., 1966. 3, 214

[180] P.A. Meyer. *Un cours sus le intégrales stochastiques. Séminaire de Proba. X*, volume 511 of *Lectures Notes in Mathematics*, pages 246–400. Springer-Verlag, Berlin, 1976. 434

[181] R. Mikulevicius and H. Pragarauskus. On the Cauchy problem for certain integro-differential operators in Sobolev and Hölder spaces. *Liet. Matem. Rink*, 32:299–331, 1992. 163

[182] R. Mikulevicius and H. Pragarauskus. On the martingale problem associated with non-degenerate Lévy operators. *Lithunian Math. J.*, 32:297–311, 1992. 163

[183] R. Mikulevicius and B.L. Rozovskii. *Martingale problems for stochastic PDE's*, volume 64 of *Mathematical Surveys and Monographs, eds: R.A. Carmona and B.L. Rozovskii*, pages 243–325. Amer. Math. Soc., Providence, 1999. 163

[184] H. Morimoto. *Stochastic Control and Mathematical Modeling.* Cambridge Univ Press, New York, 2010. vii, 488

[185] M. Motta and F. Rampazzo. Dynamic programming for nonlinear systems driven by ordinary and impulsive controls. *SIAM J. Control Optim.*, 34:199–225, 1996. 572

[186] J. Nečas. *Les méthodes directes en théorie des équations elliptiques*. Czech. Acad. Sci., Praga, 1967. 119

[187] R. Nelson. *Probability, Stochastic Processes and Queueing Theory*. Springer-Verlag, New York, 1995. 37, 193

[188] J. Neveu. *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris, 1970. 29, 187, 189

[189] J. Neveu. *Discrete-Parameters Martingales*. North Holland, Amsterdam, 1975. 193, 217

[190] B. Oksendal. *Stochastic Differential Equations*. Springer-Verlag, New York, 3rd edition, 1992. 488

[191] B. Oksendal and A. Sulem. *Applied stochastic control of jump diffusions*. Springer-Verlag, Berlin, 2005. vii

[192] O.A. Oleinik and E.V. Radkevic. *Second Order Equations with Nonnegative Characteristic Form*. AMS, Providence, Rhode Island, 1973. 119, 658

[193] R. Pallu de la Barrière. *Optimal Control Theory*. Dover, New York, 1967. 53, 64

[194] E. Pardoux. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics*, 6:127–167, 1979. 783

[195] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967. 316

[196] A. Pazy. *Semigroups and Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag, New York, 2nd edition, 1983. 89, 90, 99, 101, 103, 105, 109, 117, 121, 124, 131

[197] S. Peszat and J. Zabczyk. *Stochastic partial differential equations with Lévy noise. An evolution equation approach*. Cambridge University Press, Cambridge, 2007. 490, 518

[198] R.G. Pinsky. *Positive Harmonic Functions and Diffusion*. Cambridge Univ. Press, Cambridge, 1995. 89

[199] D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, New York, 1984. 64

[200] D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, 2002. 82

[201] P.R. Popivanov and D.K. Palagachev. *The Degenerate Oblique Derivative Problem for Elliptic and Parabolic Equations*. Akademie Verlag, Berlin, 1997. 119

[202] S.C. Port and C.J. Stone. *Brownian Motion and Classical Potential Theory*. Academic Press, New York, 1978. 89

[203] N.I. Portenko. *Generalized Diffusion Processes*. Amer. Math. Soc., Providence, Rhode Island, 1990. 156

[204] Narahari Prabhu. *Topics in probability*. World Scientific Publishing, Singapur, 2011. 11

[205] E. Priola. On a class of markov type semigroups in spaces of uniformly continuous and bounded functions. *Studia Math.*, 136(3):271–295, 1999. 135

[206] P. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, 1990. 53, 60, 61, 195, 205, 218, 223, 225, 266, 301, 370, 381, 426, 488, 490, 574

[207] M.M. Rao. *Conditional Measures and Applications*. Marcel Dekker, New York, 1993. 18

[208] M.M. Rao. *Stochastic Processes: General Theory*. Kluwer Academic, Dordrecht, 1995. 2

[209] A. Rascanu. Deterministic and stochastic differential equations in Hilbert spaces involving multivalued maximal monotone operators. *Panamer. Math. J.*, 6:83–119, 1996. 670

[210] M. Reed and B. Simon. *Methods of modern mathematical physics*, volume I: Functional Analysis, II: Fourier Analysis, Sef-Adjointness. Academic Press, 1972 and 1975. 93, 110, 112, 113

[211] M. Renardy and R.C. Rogers. *Introduction to Partial Differential Equations*. Springer-Verlag, New York, 1992. 93, 94, 125

[212] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, New York, 3rd edition, 1999. 35, 51, 54, 217, 223, 243, 310, 382, 664

[213] F. Riesz and B.S. Nagy. *Functional analysis*. Dover, New York, 1990. 93

[214] L.C.G. Rogers and D. Williams. *Diffusions. Markov Processes and Martingales*, volume 1: Foundations, 2: Itô calculus. Cambridge Univ. Press, Cambridge, 2nd edition, 2000. 50, 51, 109, 185, 196, 205, 211, 214, 221, 225, 226, 227, 229, 281, 283, 370, 371, 375, 381, 580, 593, 596, 597

[215] SiTu Rong. *Reflecting Stochastic Differential Equations with Jumps and Applications*. Chapman & Hall / CRC Press, Boca Raton (FL), 2000. 527, 694

[216] SiTu Rong. *Theory of Stochastic Differential Equations with Jumps and Applications*. Springer-Verlag, New York, 2005. 694

[217] S. Ross. *Simulation*. Academic Press, San Diego, 1997. 15

[218] B.L. Rozovskii. *White Noise: An Infinite-Dimensional Calculus*. Stochastic Evolution Systems, Dordrecht, 1990. 771, 783

[219] B. Rudiger. Stochastic integration with respect to compensated Poisson random measures on separable Banach spaces. *Stochastics and Stochastics Reports*, 76:213–242, 2004. 312

[220] K.I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Univ. Press, Cambridge, 1999. 8, 52, 60, 61, 199, 256, 260, 265, 266, 272, 275, 276, 664

[221] M. Schechter. *Modern Methods in Partial Differential Equations*. McGraw-Hill, New York, 1977. 119

[222] W.W. Schmaedke. Optimal control theory for nonlinear vector differential equations containing measures. *SIAM J. Control*, 3:231–280, 1965. 510, 513, 572

[223] Z. Schuss. *Theory and Applications of Stochastic Differential Equations*. Wiley, New York, 1980. 487

[224] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1978. 63

[225] M. Sharpe. *General Theory of Markov Processes*. Academic Press, New York, 1988. 2, 282, 283, 284

[226] P.S. Shields. *The Ergodic Theory of Discretes Sample Paths*. Am. Math. Soc., Providence, Rhode Island, 1996. 38

[227] A.N. Shiryayev. *Probability*. Springer-Verlag, New York, 1984. 3, 52, 56, 187, 200

[228] M.L. Silverstein. *Symmetric Markov processes*. Lecture Notes in Mathematics 426. Springer-Verlag, 1974. 138

[229] A.V. Skorokhod. *Studies in the Theory of Random Processes*. Dover, New York, 1982. 444, 452

[230] A.V. Skorokhod. *Asymptotic Methods in the Theory of Stochastic Differential Equations*. English transl., Amer. Math. Soc., Providence, Rhode Island, 1989. 35, 503, 634

[231] A.L. Skubachevskii. *Elliptic Functional Differential Equations and Applications*. Birkhäuser, Boston, 1997. 156, 646, 670, 741

[232] V.A. Solonnikov. A priori estimates for second order equations of parabolic type. *Trudy Math. Inst. Steklov*, 70:133–212, 1964. 157, 632

[233] V.A. Solonnikov. *On Boundary Value Problems for Linear General Parabolic Systems of Differential Equations*. Trudy Math. Inst. Steklov. English transl., Amer. Math. Soc., Providence, Rhode Island, 1967. 157, 632, 730, 732, 733, 739

[234] W. Stannat. Dirichlet forms and Markov processes: A generalized framework including both elliptic and parabolic cases. *Potential Analysis*, 8:21–60, 1998. 143

[235] W. Stannat. (Nonsymmetric) dirichlet operators on l^1: existence, uniqueness and associated Markov processes. *Ann. Scuola Norm. Sup. Pisa Cl. Sci*, **(4)** 28:99–140, 1999. 138, 143

[236] W. Stannat. *The theory of generalized Dirichlet forms and its applications in analysis and stochastics*, volume 142, no. 678, 101 pp. of *Mem. Amer. Math. Soc.* Am. Math. Soc., Providence, Rhode Island, 1999. (PhD Thesis, University of Bielefeld 1996). 143

[237] K.R. Stromberg. *Probability for Analysts*. Chapman and Hall, New York, 1994. 2, 11

[238] D.W. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, Cambridge, 1999. revised edition. 2, 11, 254

[239] D.W. Stroock and S.R. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971. 286, 669

[240] D.W. Stroock and S.R. Varadhan. On degenerate elliptic-parabolic operators of second order and their associagted diffusions. *Comm. Pure Appl. Math.*, 25:651–714, 1972. 141, 286, 658, 669

[241] D.W. Stroock and S.R. Varadhan. *Multidimensional diffusion process*. Springer-Verlag, Berlin, 1979. 24, 71, 82, 89, 269, 286, 602, 607, 609, 627, 629, 633

[242] G. Szegö. *Orthogonal Polynomials*. Amer. Math. Soc., Providence, Rhode Island, 4th edition, 1975. 318

[243] K. Taira. *Diffusion Processes and Partial Differential Equations*. Academic Press, Boston, 1988. 128, 129, 274, 632, 646

[244] K. Taira. *Analytic Semigroups and Semilinear Initial Boundary Valued Problems*. Cambridge University Press, Cambridge, 1995. 89, 104

[245] K. Taira. *Partial Differential Equations and Markov Processes*. Springer Verlag, New York, 2013. 27, 50

[246] H. Tanabe. *Equations of evolution*. Pitman, London, 1979. 89, 106, 117

[247] H. Tanaka. Stochastic differential equations with reflecting boundary conditions. *Hiroshima Math. J.*, 9:162–177, 1979. 670

[248] J.C. Taylor. *An Introduction to Measure and Probability*. Springer-Verlag, New York, 1996. 2, 24

[249] M.E. Taylor. *Partial Differential Equations*, volume Vols. 1, 2 and 3. Springer-Verlag, New York, 1997. 103, 131

[250] G.M. Troianiello. *Elliptic Differential Equations and Obstacle Problems*. Plenum Press, New York, 1987. 119

[251] M. Tsuchiya. On the oblique derivative problem for diffusion processes and diffusion equations with Hölder continuous coefficients. *Trans. Amer. Math. Soc.*, 346(1):257–281, 1994. 156

[252] M. Tsuchiya. Supplement to the paper: "On the oblique derivative problem for diffusion processes and diffusion equations with Hölder continuous coefficients". *Ann. Sci. Kanazawa Univ.*, 31:1–52, 1994. 156

[253] D. Williams. *Probabiliy with Martingales*. Cambridge Univ. Press, Cambridge, 1991. 28, 193

[254] J. Wloka. *Partial Differential Equations*. Cambridge University Press, Cambridge, 1987. 94, 119

[255] E. Wong. *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York, 1971. 31, 56, 487

[256] E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems*. Springer-Verlag, New York, 1985. 487

[257] A.I. Yanushauskas. *The Oblique Derivative Problem of Potential Theory*. Consultants Bureau, New York, 1989. 119

[258] J. Yeh. *Martingales and Stochastic Analysis*. World Scientific Publishing, Singapur, 1995. 414

[259] G.G. Yin and Q. Zhang. *Continuous-Time Markov Chains and Applications*. Springer-Verlag, New York, 1998. 37

[260] J. Yin and R. SiTu. On solutions of forward-backward stochastic differential equations with poisson jumps. *Stochastic Anal. Appl.*, 21:1419–1448, 2003. 393

[261] J. Yong and X.Y. Zhou. *Stochastic Controls*. Springer-Verlag, New York, 1999. vii, 84, 488, 749

[262] K. Yosida. *Functional Analysis*. Springer-Verlag, New York, 1965. 89, 93, 94, 99, 146

[263] J. Zabczyk. *Mathematical Control Theory: An Introduction*. Birkhäuser, Boston, 1992. vii, 107

[264] E. Zeidler. *Nonlinear functional analysis and its applications*, volume 2A: Linear monotone operators, 2B: Nonlinear monotone operators. Springer-Verlag, New York, 1990. 106

Notation

Some Common Uses:

$\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$: natural, rational, real and complex numbers.

$\mathtt{i}$, $\Re(\cdot)$, I: imaginary unit, the real part of complex number and the identity (or inclusion) mapping or operator.

P, $\mathbb{E}\{\cdot\}$: for a given measurable space $(\Omega, \mathcal{F})$, P denotes a probability measure and $\mathbb{E}\{\cdot\}$ the expectation (or integration) with respect to P. As customary in probability, the random variable ω in Ω is seldom used in a explicit notation, this is understood from the context.

$\mathcal{F}(t)$, $\mathcal{F}_t$, $\mathcal{B}(t)$, $\mathcal{B}_t$: usually denote a family increasing in t of σ-algebra (also called σ-fields) of a measurable space $(\Omega, \mathcal{F})$. If $\{x_t : t \in T\}$ is a family of random variables (i.e., measurable functions) then $\sigma(x_t : t \in T)$ usually denotes the σ-algebra generated by $\{x_t : t \in T\}$, i.e., the smallest sub σ-algebra of $\mathcal{F}$ such that each function $\omega \to x_t(\omega)$ is measurable. Usually $\mathbb{F}$ denotes the family of σ-algebras $\{\mathcal{F}(t) : t \in T\}$, which is referred to as a filtration.

$X(t)$, X_t, $x(t)$, x_t: usually denote the same process in some probability space $(\Omega, \mathcal{F}, P)$. One should understand from the context when we refer to the value of the process (i.e., a random variable) or to the generic function definition of the process itself.

$\mathbb{1}_A$: usually denotes the characteristic function of a set A, i.e., $\mathbb{1}_A(x) = 1$ if x belongs to A and $\mathbb{1}_A(x) = 0$ otherwise. Sometimes the set A is given as a condition on a function τ, e.g., $\tau < t$, in this case $\mathbb{1}_{\tau<t}(\omega) = 1$ if $\tau(\omega) < t$ and $\mathbb{1}_{\tau<t}(\omega) = 0$ otherwise.

δ: most of the times this is the δ function or Dirac measure. Sometimes one write $\delta_x(\mathrm{d}y)$ to indicate the integration variable y and the mass concentrated at x. On certain occasions, δ denotes the jumps operator, defined be $\delta X(0) = 0$ and $\delta X = X(t+) - X(t-)$, $t > 0$, any process X without discontinuity of the second kind.

$\mathrm{d}\mu$, $\mu(\mathrm{d}x)$, $\mathrm{d}\mu(x)$: together with the integration sign, usually these expressions denote integration with respect to the measure μ. Most of the times $\mathrm{d}x$

means integration respect to the Lebesgue measure in the variable x, as understood from the context.

E^T**,** $\mathcal{B}(E^T)$**,** $\mathcal{B}^T(E)$**:** for E a Hausdorff topological (usually a separable complete metric, i.e., Polish) space and T a set of indexes, usually this denotes the product topology, i.e., E^T is the space of all function from T into E and if T is countable then E^T is the space of all sequences of elements in E. As expected, $\mathcal{B}(E^T)$ is the σ-algebra of E^T generated by the product topology in E^T, but $\mathcal{B}^T(E)$ is the product σ-algebra of $\mathcal{B}(E)$ or generated by the so-called cylinder sets. In general $\mathcal{B}^T(E) \subset \mathcal{B}(E^T)$ and the inclusion may be strict.

$C([0,\infty),\mathbb{R}^d)$ **or** $D([0,\infty),\mathbb{R}^d)$ canonical sample spaces of continuous or cadlag (continuous from the right having left-hand limit) and functions, with the locally uniform or the Skorokhod topology, respectively. Sometimes the notation $\mathbb{C}_d$ or $C([0,\infty[,\mathbb{R}^d)$ or $\mathbb{D}_d$ or $D([0,\infty[,\mathbb{R}^d)$ could be used.

Most Commonly Used Function Spaces:

$C(X)$**:** for X a Hausdorff topological (usually a separable complete metric, i.e., Polish) space, this is the space of real-valued (or complex-valued) continuous functions on X. If X is a compact space then this space endowed with sup-norm is a separable Banach (complete normed vector) space. Sometimes this space may be denoted by $C^0(X)$, $C(X,\mathbb{R})$ or $C(X,\mathbb{C})$ depending on what is to be emphasized.

$C_b(X)$**:** for X a Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the Banach space of real-valued (or complex-valued) continuous and bounded functions on X, with the sup-norm.

$C_0(X)$**:** for X a *locally compact* (but not compact) Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing at infinity on X, i.e., a continuous function f belongs to $C_0(X)$ if for every $\varepsilon > 0$ there exists a compact subset $K = K_\varepsilon$ of X such that $|f(x)| \leq \varepsilon$ for every x in $X \setminus K$. This is a proper subspace of $C_b(X)$ with the sup-norm.

$C_0(X)$**:** for X a *compact* subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing on the boundary of X, with the sup-norm. In particular, if $X = X_0 \cup \{\infty\}$ is the one-point compactification of X_0 then the boundary of X is only $\{\infty\}$ and $C_0(X) = C_0(X_0)$ via the zero-extension identification.

$C_0(X)$, $C_0^0(X)$**:** for X a proper *open* subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Fréchet (complete

locally convex vector) space of real-valued (or complex-valued) continuous functions with a compact support X, with the inductive topology of uniformly convergence on compact subset of X. When necessary, this Fréchet space may be denoted by $C_0^0(X)$ to stress the difference with the Banach space $C_0(X)$, when X is also regarded as a locally compact Hausdorff topological. Usually, the context determines whether the symbol represents the Fréchet or the Banach space.

$C_b^k(E)$, $C_0^k(E)$**:** for E a *domain* in the Euclidean space $\mathbb{R}^d$ (i.e, the closure of the interior of E is equal to the closure of E) and k a nonnegative integer, this is the subspace of either $C_b(E)$ or $C_0^0(E)$ of functions f such that all derivatives up to the order k belong to either $C_b(E)$ or $C_0^0(E)$, with the natural norm or semi-norms. For instance, if E is open then $C_0^k(E)$ is a separable Fréchet space with the inductive topology of uniformly convergence (of the function and all derivatives up to the order k included) on compact subset of E. If E is closed then $C_b^k(E)$ is the separable Banach space with the sup-norm for the function and all derivatives up to the order k included. Clearly, this is extended to the case $k = \infty$.

$B(X)$**:** for X a Hausdorff topological (mainly a Polish) space, this is the Banach space of real-valued (or complex-valued) Borel measurable and bounded functions on X, with the sup-norm. Note that $\mathcal{B}(X)$ denotes the σ-algebra of Borel subsets of X, i.e., the smaller σ-algebra containing all open sets in X, e.g., $B(\mathbb{R}^d)$, $\mathcal{B}(\mathbb{R}^d)$, or $B(E)$, $\mathcal{B}(E)$ for a Borel subset E of d-dimensional Euclidean space $\mathbb{R}^d$.

$L^p(X, m)$**:** for $(X, \mathcal{X}, m)$ a complete σ-finite measure space and $1 \le p < \infty$, this is the separable Banach space of real-valued (or complex-valued) $\mathcal{X}$-measurable (class) functions f on X such that $|f|^p$ is m-integrable, with the natural p-norm. If $p = 2$ this is also a Hilbert space. Usually, X is also a locally compact Polish space and m is a Radon measure, i.e., finite on compact sets. Moreover $L^\infty(X, m)$ is the space of all (class of) m-essentially bounded (i.e., bounded except in a set of zero m-measure) with essential-sup norm.

$L^p(\mathcal{O})$**,** $H_0^m(\mathcal{O})$**,** $H^m(\mathcal{O})$**:** for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, $1 \le p \le \infty$ and $m = 1, 2, \ldots$, these are the classic Lebesgue and Sobolev spaces. Sometimes we may use vector-valued functions, e.g., $L^p(\mathcal{O}, \mathbb{R}^n)$.

$\mathcal{D}(\mathcal{O})$**,** $\mathcal{S}(\mathbb{R}^d)$**,** $\mathcal{D}'(\mathcal{O})$**,** $\mathcal{S}'(\mathbb{R}^d)$**:** for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, these are the classic test functions (C^∞ functions with either compact support in $\mathcal{O}$ or rapidly decreasing in $\mathbb{R}^d$) and their dual spaces of distributions. These are separable Fréchet spaces with the inductive topology. Moreover, $\mathcal{S}(\mathbb{R}^d) = \cap_m H^m(\mathbb{R}^d)$ is a countable Hilbertian nuclear space. Thus its dual space $\mathcal{S}'(\mathbb{R}^d) = \cup_m H^{-m}(\mathbb{R}^d)$, where $H^{-m}(\mathbb{R}^d)$ is the dual space of $H^m(\mathbb{R}^d)$. Sometimes we may use vector-valued functions, e.g., $S(\mathbb{R}^d, \mathbb{R}^n)$.

Some ??:

??

Index