1-1-2014

Stability And Controls For Stochastic Dynamic Systems

Zhixin (harriet) Yang
Wayne State University,

STABILITY AND CONTROLS FOR STOCHASTIC DYNAMIC SYSTEMS

by

ZHIXIN YANG

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: MATHEMATICS

Approved by:

Advisor	Date

DEDICATION

To my family and teachers

ACKNOWLEDGEMENTS

I would like to take this special opportunity to express my appreciation on all of my friends who helped me excel myself. In particular, I owe my deepest gratitude to my advisor, Professor George Yin for the immeasurable support and guidance he has provided to me in the past five years. As the Chinese saying goes, He who teaches me for one day is my father for life. It goes without saying that I would not be able to launch my career in the United States without Prof. Yin's inspiring guidance, valuable suggestions and endless care. It is a huge blessing for me to have him as my lifelong mentor, friend and father.

I want to thank Professor Tze-Chien Sun, Professor Boris Mordukhovich, Professor Le Yi Wang and Professor Paul Chow for being my committee members, and for the guidance and help for my PhD study. I enjoyed chatting with Professor Sun so much, which helped me know more about life outside campus.

During my graduate study, Professors Rafail Khasminskii, Professor Alex Korostelev, Professor Guozhen Lu, Professor William Cohn, Professor Ualbai Umirbaev, and Professor John Klein have taught me lessons. I appreciate their help. Professor Dan Isaksen has spared his valuable time to help me prepare the preliminary exam, I still remember his free treat and warmly encouragement. Ms. Mary Klamo, Ms. Patricia Bonesteel, Dr. Choon-Jai Rhee, Dr. John Breckenridge, Ms. Barbara Malicke, Ms. Joyce Wynn, and Ms. Tiana Fluker, have trained me and supported me on teaching and many other aspects. I appreciate you all.

I am also grateful to my friends Aodan, Yuehai Xu, Xinye Wu, Fengping Dong, Qi He, Guangliang Zhao, Dan Wang, Xiaolong Han, Yayuan Xiao, Wenhui Han, Yu Sun, Hongwei Mei, Xiaoyue Cui, Wei Ouyang, Yuan Tian, Ren Zhao, Hailong Guo, Janice Gayle Veenstra, Shirley Lamp, Patricia Ann Rees, Pat Raburn, Carolyn Hochstadt, Matthew Buckman, and Matthew Zabka, Because of you all, my stay in Detroit has been colorful and memorable.

Finally, I thank my parents, Hongwei Yang and Qinglan Yin, for the inspiring education and endless love they provided me. My thanks go to my parents and also to my parents-inlaw, Shuncai Yuan and Yuyan Shi, for offering my son a second home when I am busy, and for supporting me in so many ways. My special thanks go to my husband Quan and my son Jacob, for sharing my hopes and dreams.

TABLE OF CONTENTS

1 Introduction

1.1 Background and Main Issues

This dissertation focuses on stabilities analysis and optimal controls for stochastic dynamic systems. It encompasses an in-depth study of stability for multi-dimensional jump diffusions, Markov switching jump diffusions, and regime-switching jump diffusions as well as stability of the associated numerical solutions. In addition, the dissertation treats nearly optimal mean-variance problem. We examine the mean-variance control problem under two-time-scale and hidden Markov chain scenarios. In what follows, we present the background and main issues of these problems.

Since systems often run for an extended time, stability is of critical importance. As a result, much effort has been devoted to the stability analysis in the literature; see [6] for stability of diffusion processes, [12] for Markovian switching diffusions, and [20] for switching diffusions in which the switching depends on the diffusion parts. In practice, closed-form solutions are difficult to obtain. Numerical methods are more viable or even the only possible alternative to solve the problems. Starting with a practical problem, an immediate question is: If the system of interest is stable, what can be said about the corresponding numerical approximation?

Because of the importance, there has been much work on numerics of diffusions, jump diffusions, and their regime-switching counterparts. General survey and classical treatments can be found in [7, 13] and references therein. In [5], almost sure exponential stability of Euler-Maruyama (E-M) algorithm as well as that of exponential p-stability were treated for diffusion systems. In [15], almost sure exponential stability and exponential p-stability for E-

M algorithms were studied for Markovian switching diffusions. In [2], asymptotical stability in the large of E-M algorithm was examined for jump diffusion systems. In [3], mean square stability and asymptotical stability in the large of stochastic theta methods were presented. In [4], split-step backward Euler method and compensated split-step backward Euler method were analyzed and strong convergence results were obtained under certain assumptions for nonlinear jump diffusion systems. In [12], mean square stability was treated for Markovian switching diffusions. In lieu of the Brownian increments, i.i.d. sequences were used and path-wise convergence rates for diffusions were dealt with in [14] by consideration of re-embedded sequences. Given the key roles of jump-diffusions played in networked systems, our work is devoted to answering stability questions of numerical solutions to jump diffusions. Although there have been many excellent works on numerical solutions of stochastic differential equations, the study on numerical methods of almost sure exponential stability and exponential p-stability for jump diffusions has not been done yet to the best of our knowledge. One intuitive thought might be: Perhaps one can repeat the success in the numerical approximation to diffusions, in which the techniques used were asymptotic expansions (using an asymptotic series of expansion of moments of Brownian motion). A scrutiny, however, shows that such an approach is not going to work. The essential reason is that a Gaussian distribution is completely determined by the first and the second moments, whereas for a Poisson random variable, the mean and variance are the same. Thus using expansions of the Poisson increments will not produce higher powers in terms of the small step size in contrast to the case of the Brownian increments. This rules out the possibility of using existing techniques in the current problem. To illustrate, let us start with an algorithm with step size $\varepsilon > 0$. The increment of a standard Brownian motion $\Delta w \sim N(0, \varepsilon)$ satisfies $E(\Delta w)^{2n} = \frac{(2n)!}{n!2^n}\varepsilon^n$

and all odd moments of increment of Brownian motion are 0. Thus it is advantageous to use series expansions since the higher the moment, the higher order of ε. In contrast, unlike the increments of Brownian motion, the increments of a Poisson process behave very differently. In fact, since $\Delta N \sim$Poisson$(\lambda\varepsilon)$, we have $E(\Delta N)^n = \sum_{i=1}^{n} (\lambda\varepsilon)^i \frac{1}{i!} \sum_{j=0}^{i} (-1)^{i-j} \begin{pmatrix} i \\ j \end{pmatrix} j^n$. A moment of reflection reveals that in the nth moment of ΔN, the leading term is $\lambda\varepsilon$ for all n. That is, higher moments do not yield higher order of ε (in terms of order of magnitude estimates), which rules out the possibility of using series expansion methods. Our question is: Passing from the original systems to that of the numerical solutions, under what conditions, stability will be preserved. In the traditional approach for numerical methods of stochastic differential equations, one often has to use Taylor expansions. For Poisson processes, since the mean and variance are the same, the Taylor expansions do not really help. We use techniques from the stochastic approximation toolbox, which enables us to resolve the problem and obtain convergence and stability. Using our definitions of stability for the numerical algorithms, stability of numerical algorithms will imply that of SDEs. Not only are these questions important from a theoretical point of view, but also they provide crucial practical insight for actual computing. To get the insight and to make comparisons, we first begin with one-dimensional benchmark models. We then further our study for considering multi-dimensional cases and systems with switching.

The key model appears in our work in regime switching model. Randomly-varying switching systems have drawn increasing attention recently owing to their ability to model complex systems, which can be used in a wide range of applications in consensus controls, distributed computing, autonomous or semi-autonomous vehicles, multi-agent systems, tele-medicine, smart grids, and financial engineering etc. Regime-switching diffusions consist of a number of diffusions coupled by a switching process, which reflects the feature of the coexistence of dynamics described by solutions of stochastic differential equations and discrete events whose values belong to a finite set. The usual formulation in the traditional dynamic system setup described by differential or difference equations alone becomes not suitable to describe such features. A class of models naturally replacing the traditional setup is a process with two components in which one of them delineates the dynamics that may be represented as a solution of a differential equation and the other portraits the discrete event movements. Recently, there are growing interests on formulating complex systems by use of regime-switching processes, which largely enriched the applicability of the dynamic models; see [20] and many references therein. To take into consideration of possible inclusion of Poisson type random processes, we consider jump diffusion processes with random switching. One of the pioneering study on stability is conducted by [56]. In recent years, stability of switching stochastic systems have received much attention; see [12, 36, 50] and references therein for a systematic treatment on Markov modulated switching diffusions; see also [51] for stability of switching diffusions with delays. In addition, switching diffusion with continuous dependence on initial data were treated in [49]. Concerning jump diffusions, we refer the reader to [33, 37, 44] for the study on such properties as ergodicity and stability. Switching jump diffusions with state dependent switching have also been examined in [19, 20, 47] etc., in which stability in probability, asymptotic stability in probability, and almost surely exponential stability were dealt with. Our aims are to establish a number of results on different modes of stability that have not been studied for switching jump diffusions to date to the best of our knowledge.

The other part of our work is mean-variance optimization problem, which can be traced back to the Nobel-prize-winning work of Markowitz [60]. The salient feature of the model is that, in the context of finance, it enables an investor to seek highest return after specifying the acceptable risk level quantified by the variance of the return. The mean-variance approach has become the foundation of modern finance theory and has inspired numerous extensions and applications. Using the stochastic linear-quadratic (LQ) control framework, Zhou and Li [86] studied the mean-variance problem for a continuous-time model. Note that the problem becomes fundamentally different from the traditional LQ problem studied in literature. In the classical time-honored LQ theory, the matrix related to the control (known as control weight) needs to be positive definite. In the mean variance setup for linear systems, the control weight is non-positive definite. In [87], the mean-variance problems for switching diffusion models were treated and a number of results including optimal portfolio selection, efficient frontier, and mutual fund theory were discovered. Inspired by platoon controls of networked systems, we consider a mean-variance control problem, in which the network topology or the environment is modeled as a continuous-time Markov chain. We assume that the Markov chain has a large state space in order to deal with complex systems. To treat the platoon problems, we could in principle apply the results in [87]. Nevertheless, the large state space of the Markov chain renders a straightforward implementation of the mean-variance control strategy obtained in [87] practically infeasible. The computational complexity becomes a major concern. Inspired by the idea in the work [80], to exploit the hierarchical structure of the underlying systems, and to fully utilize the near decomposability [74, 79] by means of considering fast and slow switching modes, the work [85] treated near-optimal control problems of LQG with regime switching. Another point is that only positive definite control weights were allowed in the usual quadratic control criteria. In our current setup, the control weights are indefinite, so the main assumptions in [85] do not hold. This two-time-scale scenario provides an opportunity to reduce computational complexity for the Markov chain. The main idea is a decomposition of the large space into sub-clusters and aggregation of states in each sub-cluster. That is, we partition the state space of the Markov chain into subspaces (or sub-groups or sub-clusters). Then, in each of the sub-clusters, we aggregate all the states into one super state. Thus the total number of discrete states is substantially reduced.

Next, we further extend the mean-variance methods to incorporate hidden Markov chains. In particular, the underlying system is modeled as a controlled switching diffusion modulated by a finite-state Markov chain representing the system modes. We consider the case that a function of the chain with additive noise is observable. In networked systems, such measurement can be obtained with the addition of a sensor. The underlying problem is a stochastic control problem with partial observation. Given the target expectation of the state variable at the terminal time, the objective is to minimize the variance at the terminal time. We use the mean-variance approach to treat the problem and aim at developing feasible numerical methods for solutions of the associated control problems. To solve the problem, we resort to the Wonham filter method to estimate the state. The original system is converted into a completely observable one. In stochastic control literature, a suboptimal filter for linear systems with hidden Markov switching coefficients was considered in [53] in connection with a quadratic cost control problem. Given that our problem cannot be solved in closed form, our main effort is devoted to developing numerical methods. We use ideas in the Markov chain approximation methods of Kushner and Dupuis [58]. Nevertheless, the methods in [58] cannot be directly adopted since there are switching processes involved and the problem is only partially observable. Different from the numerical methods for controlled regime-switching diffusions [64] and [71], in addition to the partially observed system, the variance is control dependent. Therefore, extra care must be taken to address such control dependence.

1.2 Outline of the Dissertation

The remainder of the dissertation is arranged as follows. In Chapter 2, stability issue is dealt with. We obtain asymptotic stability in distribution for Markov switching jump diffusions. Then we further examine the even more difficult case for x-dependent switching jump diffusions. In addition, asymptotic stability in the large, exponential p-stability are carried out. In addition, we obtain stability results for both jump diffusion systems and the associated numerical approximations, in which the traditional treatment for Euler-Maurayama (E-M) algorithm breaks down. In Chapter 3 and Chapter 4, we studies stochastic optimization and controls. The motivation stems from the Nobel prize-winning work of Markowitz. Specifically, our work is mean-variance type control problems. In Chapter 3, with motivations from earlier work on singularly perturbed Markovian systems [79, 83, 84], we use a two-time-scale formulation to treat the underlying systems and obtain a limit problem. Using the limit problem as a guide, we construct controls for the original problem, and show that the controls so constructed are nearly optimal. In Chapter 4, we consider the scenario that instead of having access to full information of the switching process, we know a noisy observation of switching process. We still focus on minimizing the variance subject to a fixed terminal expectation. Using the Wonham filter, we convert the partially observed system to a completely observable one first. Since closed-form solutions are virtually impossible be obtained, a Markov chain approximation method is used to devise a computational scheme. Convergence of the algorithm is obtained.

1.3 Notation Index

Before proceeding further, we compile the following list of notation index to be used in the entire dissertation.

$\mathbb{R}^{r\times d}$	$r \times d$-dimensional Euclidean space, where r and d are positive integers
$\mathbb{R}^r$	r dimensional row vector.
$\lvert x\rvert$	Euclidean norm of $x \in \mathbb{R}^n$
z'	transpose of $z \in \mathbb{R}^{l_1\times l_2}$
$\mathrm{tr}(A)$	trace of $A \in \mathbb{R}^{n\times n}$
∇f	gradient of $f(x)$ w.r.t. x
Hf	Hessian of $f(x)$ w.r.t. x
$\mathbb{R}_+$	positive real number.
$\Lambda_{max}(A)$	the largest eigenvalue of the symmetric matrix A
$\Lambda_{min}(A)$	the smallest eigenvalue of the symmetric matrix A
w.p.1	with probability 1
$\mathcal{M}$	$\mathcal{M} = \{1, 2, \ldots, m\}$
K	K is a generic constant whose value can be different in different context.
$\mathcal{L}X$	the generator of a diffusion.
$\mathbb{1}$	$= (1, \ldots, 1)' \in \mathbb{R}^m$

2 Stability of Jump Diffusions and Their Numerical Methods

2.1 Stability of One Dimensional Jump Diffusions and Their Numerical Methods

In this chapter, we consider stability of one dimensional jump diffusions, Markov switching jump diffusions first, then we consider the multi-dimensional jump diffusions and finally we consider the regime switching jump diffusions. To proceed with our analysis, we first give definitions of stability. We make the following definitions by adopting the terminologies of [20].

Definition 2.1. The equilibrium point $x = 0$ of dynamic system is said to be

(i) *asymptotically stable in the large*, if it is stable in probability and $P\{\lim_{t\to\infty} X^{x,\alpha}(t) = 0\} = 1,$ for any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$;

(ii) *exponentially p-stable*, if for some positive constants K and k, $E|X^{x,\alpha}(t)|^p \le K|x|^p e^{-kt}$, for any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$;

(iii) *almost surely exponential stable*, if for any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$, $\limsup_{t\to\infty} \frac{1}{t} \ln(|X^{x,\alpha}(t)|) < 0$ w.p.1.

Remark 2.2. The exponential p-stability can also be stated as

$$\limsup_{t\to\infty} \frac{1}{t} \ln E|X^x(t)|^p < 0 \quad \text{for any initial value} \quad x.$$

Here we used definition 5.7 in [12, p. 166] and we will use these definitions interchangeably, whichever is more convenient in what follows.

We first consider the benchmark test model

$$\begin{aligned} dX(t) &= bX(t)dt + \sigma X(t)dw(t) + \gamma X(t^-)dN(t) \\ X(0) &= x, \end{aligned} \tag{2.1}$$

where b, σ, and γ are real constants, $w(t)$ is a scalar Brownian motion, and $N(t)$ is a scalar Poisson process independent of the Brownian motion. We denotes the solution of (2.1) as $X^x(t)$ to emphasize its initial data x dependence. It is easy to see that 0 is the only equilibrium point of the dynamic system.

Lemma 2.3. *For the jump diffusion given by* (2.1)*, the pth moment Lyapunov exponent is*

$$\limsup_{t\to\infty} \frac{1}{t} \ln(E|X^x(t)|^p) = bp + \frac{1}{2}p(p-1)\sigma^2 + \lambda(|1+\gamma|^p - 1).$$

Therefore, the equilibrium point of the system is exponentially p-stable if and only if

$$bp + \tfrac{1}{2}p(p-1)\sigma^2 + \lambda(|1+\gamma|^p - 1) < 0. \tag{2.2}$$

Proof. It is well known that the explicit solution of (2.1) is given by

$$X(t) = x\exp((b - \frac{1}{2}\sigma^2)t + \sigma w(t))(1+\gamma)^{N(t)}. \tag{2.3}$$

Note that

$$\begin{aligned} E\exp(p(b-\tfrac{1}{2}\sigma^2)t+p\sigma w(t)) &= \exp(p(b-\frac{1}{2}\sigma^2)t+\frac{1}{2}p^2\sigma^2 t) \\ &= \exp(pbt+\frac{1}{2}\sigma^2 p(p-1)t), \end{aligned}$$

that

$$E(1+\gamma)^{pN(t)} = \exp(t\lambda(|1+\gamma|^p-1)),$$

and that the Brownian motion $w(\cdot)$ is independent of the Poisson process $N(\cdot)$,

$$\begin{aligned} \limsup_{t\to\infty}\frac{1}{t}\ln E|X^x(t)|^p &= \lim_{t\to 0}\frac{p\ln|x|}{t}+pb+\frac{1}{2}\sigma^2 p(p-1)+\lambda(|1+\gamma|^p-1) \\ &= bp+\frac{1}{2}p(p-1)\sigma^2+\lambda(|1+\gamma|^p-1). \end{aligned} \tag{2.4}$$

The proof is complete.

To numerically solve (2.1), we choose $\varepsilon>0$ as the step size. Now we define the increment of Brownian motion as Δw_n and Process process as ΔN_n, which are a bit different from before to illustrate the depedence of iteration number n. Define

$$\begin{aligned} \Delta w_n &= w(\varepsilon(n+1))-w(\varepsilon n), \\ \Delta N_n &= N(\varepsilon(n+1))-N(\varepsilon n). \end{aligned} \tag{2.5}$$

We will also use $\Delta^2 w_n = (\Delta w_n)^2$.

$$\begin{aligned} & x_{n+1} = x_n+\varepsilon b x_n+\sigma x_n\Delta w_n+\gamma x_n\Delta N_n \\ & x_0 = X(0) = x. \end{aligned} \tag{2.6}$$

Recall that the sequence $\{x_n : \varepsilon > 0, n < \infty\}$ generated by algorithm (2.6) is said to be tight or bounded in probability, if for any $\eta>0$, there is a $K_\eta>0$ such that for all n, $P(|x_n|\geq K_\eta)<\eta$.

Lemma 2.4. *Assume that* (2.2) *holds. Then the sequence* $\{x_n : \varepsilon > 0, n < \infty\}$ *is tight.*

Proof. The proof uses a Lyapunov function argument. Note that in (2.2), when $p = 2$, we have

$$\hat{\xi}^* = 2b + \sigma^2 + \lambda(|1+\gamma|^2 - 1) < 0. \tag{2.7}$$

To obtain the tightness, we first demonstrate $E|x_n|^2 < \infty$ as follows. Define a Lyapunov function $V(x) = x^2$. Note that

$$E_n V(x_{n+1}) - V(x_n) = E_n V_x(x_n)[x_{n+1} - x_n] + E_n|x_{n+1} - x_n|^2, \tag{2.8}$$

Detailed calculation yields that

$$\begin{aligned} E_n V_x(x_n)[x_{n+1} - x_n] &= 2E_n x_n(x_{n+1} - x_n) \\ &= 2E_n x_n\big(\varepsilon b x_n + \sigma x_n \Delta w_n + \gamma x_n \Delta N_n\big) \\ &= 2b\varepsilon V(x_n) + 2\gamma x_n^2 \lambda\varepsilon \\ &= \varepsilon V(x_n)(2b + \lambda 2\gamma). \end{aligned} \tag{2.9}$$

and that

$$\begin{aligned} E_n|x_{n+1} - x_n|^2 &= E_n(b x_n \varepsilon + \sigma x_n \Delta w_n + \gamma x_n \Delta N_n)^2 \\ &\leq \varepsilon V(x_n)(\sigma^2 + \gamma^2\lambda) + \varepsilon^2 K(1 + V(x_n)). \end{aligned} \tag{2.10}$$

Combing the above two inequalities, we have

$$\begin{aligned} &E_n V(x_{n+1}) - V(x_n) \\ &\leq \varepsilon V(x_n)\hat{\xi}^* + \varepsilon^2 K(1 + V(x_n)). \end{aligned} \tag{2.11}$$

Taking expectation on both sides of the above expression, we have

$$EV(x_{n+1}) \le (1+\varepsilon\hat{\xi}^*)EV(x_n) + K\varepsilon^2 + K\varepsilon^2 EV(x_n). \tag{2.12}$$

Define

$$b_{nk} = (1+\hat{\xi}^*\varepsilon)^{n-k}.$$

Detailed calculation yields

$$EV(x_{n+1}) \le b_{n0}EV(x_1) + K\sum_{k=1}^{n} b_{nk}\varepsilon^2 + K\varepsilon^2 \sum_{k=1}^{n} b_{nk}EV(x_k).$$

Observe that $\sum_{k=1}^{n} b_{nk}\varepsilon^2 \le K$ since $\hat{\xi}^* < 0$. Then Gronwall's inequality leads to $EV(x_n)$ is bounded. The tightness of the sequence $\{x_n : \varepsilon > 0, n < \infty\}$ then follows. □

To proceed, we define the continuous time interpolations as

$$x^\varepsilon(t) = x_n \text{ for } t \in [n\varepsilon, n\varepsilon+\varepsilon),$$

and denote

$$\begin{aligned} &\widetilde{x}^\varepsilon(\cdot) = x^\varepsilon(\cdot + t_\varepsilon) \ \text{ for any } \ t_\varepsilon \to \infty \ \text{ as } \ \varepsilon \to 0, \\ &\widetilde{x}^\varepsilon_T(\cdot) = \widetilde{x}^\varepsilon(\cdot - T) = x^\varepsilon(\cdot + t_\varepsilon - T) \ \text{ for any } \ 0 < T < \infty. \end{aligned} \tag{2.13}$$

We infer that

$$\widetilde{x}^\varepsilon_T(T) = \widetilde{x}^\varepsilon(0) = x^\varepsilon(t_\varepsilon). \tag{2.14}$$

What we are using is a standard technique in stochastic approximation; see [8, Section IV, pp. 179-180]; see also [9, Section 5]. Note that in view of the definition and notation above, we work with a shifted sequence $x^\varepsilon(\cdot + t_\varepsilon)$ that effectively "start" at large and arbitrary real time. Note that weak convergence along does not imply that $\widetilde{x}^\varepsilon(\cdot)$ converges to the equilibrium point. The tightness in Lemma 2.4 is crucial. To study the stability of numerical algorithm, we need to consider the limit $\varepsilon \to 0$, $n \to \infty$, and $\varepsilon n \to \infty$. As in the analysis of stochastic approximation methods [10], this is equivalent to the study of $\widetilde{x}^\varepsilon(\cdot)$. For future use, for arbitrary $0 < T < \infty$, we shall work with the pair of processes $(\widetilde{x}^\varepsilon(\cdot), \widetilde{x}^\varepsilon_T(\cdot))$.

Note that $x^\varepsilon(\cdot) \in D([0,\infty) : \mathbb{R})$. By using the weak convergence methods [10, 18], it can be shown that $x^\varepsilon(\cdot)$ is tight and converges weakly to $X(\cdot)$, which is the solution of (2.1) (Note that (2.1) has a unique solution because it is linear). Moreover, we can also show that $\{\widetilde{x}^\varepsilon(\cdot)\}$ is also tight in $D([0,\infty);\mathbb{R})$ and all weakly convergent subsequence satisfy (2.1) as well. For an arbitrary $T < \infty$, we work with the pair $(\widetilde{x}^\varepsilon(\cdot), \widetilde{x}^\varepsilon_T(\cdot))$. It can be shown that $\{\widetilde{x}^\varepsilon(\cdot), \widetilde{x}^\varepsilon_T(\cdot)\}$ is tight in $D([0,\infty);\mathbb{R}\times\mathbb{R})$. Extract a weakly convergent subsequence and denote the limit by $(\widetilde{x}(\cdot), \widetilde{x}_T(\cdot))$. In view of the Skorohod representation [10, p. 230]we may regard that the convergence is in the sense of w.p.1 for the sake of convenience. The convergence is uniform on any bounded interval. It can be shown that $\widetilde{x}^\varepsilon(\cdot)$ still have the same limit $X(\cdot)$. By the construction, $\widetilde{x}(0) = \widetilde{x}_T(T)$. By virtue of Lemma 2.4, the set of all possible values of $\{\widetilde{x}_T(0)\}$ is tight (over all T and convergent subsequences). It follows that $\widetilde{x}(0) = \widetilde{x}_T(T)$ and

$$\widetilde{x}_T(T) = \widetilde{x}_T(0)\exp((b - \frac{1}{2}\sigma^2)T + \sigma w(T))(1+\gamma)^{N(T)}. \tag{2.15}$$

We now present the following lemma based on the above analysis. The detailed proof is omitted. For a more complex case, the reader is referred to the weak convergence in [17].

Lemma 2.5. *We have that $x^\varepsilon(\cdot)$ converges weakly to $X(\cdot)$ the unique solution to* (2.1).

Moreover, $\{\widetilde{x}^{\varepsilon}(\cdot), \widetilde{x}^{\varepsilon}_T(\cdot)\}$ *is tight in* $D([0,\infty); \mathbb{R}\times\mathbb{R})$ *such that the weak limit of* $\widetilde{x}_T(T)$ *whose dynamic is described by* (2.15)*.*

In view of Lemma 2.5, we proceed to define the exponentially p-stable for the approximation using the interpolated process. Using the auxiliary process $\widetilde{x}_T(\cdot)$ given by (2.15), we can take the logarithm of its pth moment. This motivates the definition of exponential p-stability for the numerical schemes. The definition is used not only for the benchmark example, but also for other cases considered in this paper.

Definition 2.6. Algorithm (2.6) associated with (2.1) is said to be exponentially p-stable if for any $t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$\limsup_{T\to\infty} \frac{1}{T} \lim_{\varepsilon\to 0} \ln E|\widetilde{x}^{\varepsilon}_T(T)|^p < 0. \tag{2.16}$$

Remark 2.7. For the benchmark model (one-dimensional linear scalar jump diffusions), taking limsup and taking the limit do not make difference. However, we use limsup to be consistent with the case of multi-dimensional nonlinear systems in the following paragraphs.

Theorem 2.8. *Suppose that*

$$pb + \frac{1}{2}p(p-1)\sigma^2 + \lambda(|1+\gamma|^p - 1) < 0.$$

Then, the iterates generated by algorithm (2.6) *satisfy*

$$\limsup_{T\to\infty} \frac{1}{T} \lim_{\varepsilon\to 0} \ln E|\widetilde{x}^{\varepsilon}_T(T)|^p = pb + \frac{1}{2}p(p-1)\sigma^2 + \lambda(|1+\gamma|^p - 1) < 0. \tag{2.17}$$

Thus, the algorithm is exponentially p-stable.

Proof. To begin, we have Lemma 2.5 in force. By using (2.4), (2.15), and the dominated convergence theorem, we have

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \lim_{\varepsilon\to 0} \ln E|\widetilde{x}^{\varepsilon}_T(T)|^p \\
&= \lim_{T\to\infty} \frac{1}{T} \ln E|\widetilde{x}_T(T)|^p \\
&= bp + \frac{1}{2}p(p-1)\sigma^2 + \lambda(|1+\gamma|^p - 1).
\end{aligned} \tag{2.18}$$

The desired pth moment stability then follows. □

Remark 2.9. Note that in obtaining the pth-moment (also in what follows the almost sure stability) of the numerical algorithms, Lemma 2.4 is crucial. Without the tightness, the stability and even the convergence cannot be guaranteed. Throughout the paper, for simplicity, we have assumed that the initial data x_0 is independent of the step size ε. If one wishes to let $x_0 = x^{\varepsilon}_0$, then a condition of tightness of x^{ε}_0 (or x^{ε}_0 converges in distribution to x_0 for some finite x_0) needs to be used. Such a condition is used extensively in stochastic approximation literature; see [10, Chapter 8.5]. In [5, p. 594], a motivating example is given, in which the system has a Lyapunov exponent being negative w.p.1, but the numerical algorithm blows up in finite time. In addition to instability, the algorithm is not even convergent. The main problem is that the initial condition is chosen to be inversely proportional to $\sqrt{\varepsilon}$. Thus, x^{ε}_0 is not tight, neither does x^{ε}_0 converges to x_0. Further discussion will be provided in the example section.

Lemma 2.10. *For the jump diffusion* (2.1), *the Lyapunov exponent is given by*

$$\limsup_{t\to\infty} \frac{1}{t} \ln |X^x(t)| = b - \frac{1}{2}\sigma^2 + \lambda \ln |1+\gamma| \quad \textit{w.p.1.} \tag{2.19}$$

Thus, the equilibrium point of the system is almost surely exponentially stable if and only if

$$b - \frac{1}{2}\sigma^2 + \lambda \ln |1+\gamma| < 0 \qquad \textit{w.p.1.}$$

Proof. Using the explicit solution for system (2.1), by the law of large numbers for local martingales [11], we can obtain the result. A detailed proof can also be found in [2]. □

Next we demonstrate that the numerical algorithm is also almost surely exponential stability by virtue of the Borel-Cantelli lemma. We shall replace T by a positive integer n in Definition 2.6.

Definition 2.11. Algorithm (2.6) associated with (2.1) is said to be almost surely exponentially stable if for some K_0, $K_1 > 0$, and any $t_\varepsilon \to \infty$ as $\varepsilon \to 0$,

$$|\widetilde{x}_n^\varepsilon(n)| \le K_1 \exp(-K_0 n) \ \ \text{w.p.1.} \tag{2.20}$$

Theorem 2.12. *Under the conditions of Theorem* 3.4, *the numerical algorithm is almost surely exponentially stable.*

Proof. In view of Theorem 3.4, use the definition of $\widetilde{x}_T^\varepsilon(\cdot)$ but replace T with n. Then for sufficiently large n and sufficiently small ε, we have that

$$E|\widetilde{x}_n^\varepsilon(n)|^p \leq K\exp(-\widetilde{\lambda}pn),$$

where

$$-\widetilde{\lambda} = b + \frac{1}{2}(p-1)\sigma^2 + \frac{1}{p}\lambda(|1+\gamma|^p - 1). \tag{2.21}$$

Then by the Markov inequality, for a positive Δ_0,

$$\begin{aligned} P(|\widetilde{x}_n^\varepsilon(n)| \geq \exp(-\Delta_0 n)) &\leq \frac{E|\widetilde{x}_n^\varepsilon(n)|^p}{\exp(-\Delta_0 pn)} \\ &\leq K\exp(-(\widetilde{\lambda} - \Delta_0)pn). \end{aligned} \tag{2.22}$$

Choose Δ_0 small enough so that $\widetilde{\lambda} - \Delta_0 > 0$. (2.22) leads to

$$P(|\widetilde{x}_n^\varepsilon(n)| \geq \exp(-\Delta_0 n)) \leq K\exp(-(\widetilde{\lambda} - \Delta_0)pn).$$

Clearly,

$$\sum_{n=0}^{\infty} \exp(-(\widetilde{\lambda} - \Delta_0)pn) < \infty.$$

The Borel-Cantelli lemma then yields that

$$P(|\widetilde{x}_n^\varepsilon(n)| \geq \exp(-\Delta_0 n) \text{ i.o.}) = 0.$$

Thus, as $n \to \infty$,

$$P(|\widetilde{x}_n^\varepsilon(n)| \leq \exp(-\Delta_0 n)) = 1. \tag{2.23}$$

The desired almost sure stability then follows. $\square$

A moment of reflection reveals that we can recapture the Lyapunov exponent (in the almost sure sense) (2.19) in the continuous-time equation. We wish to choose $\Delta_0 > 0$ so that $-\widetilde{\lambda} + \Delta_0 < 0$ and that is close to $b - \frac{1}{2}\sigma^2 + \lambda \ln |1+\gamma|$. Note that

$$\lim_{p\to 0} \frac{|1+\gamma|^p - 1}{p} = \ln |1+\gamma|.$$

So for sufficiently small $p > 0$, we can choose our $\Delta_0 > 0$ so that

$$-\widetilde{\lambda} + \Delta_0 = b - \frac{1}{2}\sigma^2 + \lambda \ln |1+\gamma| + \theta < 0, \tag{2.24}$$

for some small enough $\theta > 0$. Thus, we arrive at the following corollary.

Corollary 2.13. *Assume that the conditions of Theorem* 2.12 *hold, with the choice of* θ *given by* (2.24)*. Then the almost sure Lyapunov exponent is given by*

$$\limsup_{n\to\infty} \frac{1}{n} \lim_{\varepsilon\to 0} \ln |\widetilde{x}_n^{\varepsilon}(n)| \le b - \frac{1}{2}\sigma^2 + \lambda \ln |1+\gamma| + \theta < 0 \quad \textit{w.p.1.} \tag{2.25}$$

Remark 2.14. Note that we can also achieve the almost sure exponential stability in an alternative way. We listed the key ideas below. First, by the similar techniques involved in Lemma 2.4, considering Lyapunov function $V(x) = x^2$, we can show $EV(x_{n+1}) \le (1 + \varepsilon\hat{\xi}^*)EV(x_n) + o(\varepsilon)$. Here the definition of $\hat{\xi}^*$ can be found in (2.7). Detailed calculation leads to $E|\widetilde{x}^{\varepsilon}(t)|^2 = |x_0|^2(1+\varepsilon\hat{\xi}^*)^{\frac{t+t_\varepsilon}{\varepsilon}} + \frac{t+t_\varepsilon}{\varepsilon}o(\varepsilon)$. Next, we can choose $\varepsilon^* \in (0,1)$ so small that for all $0 < \varepsilon < \varepsilon^*$, $-1 < \varepsilon\hat{\xi}^* < 0$. Recalling the definition of continuous-time interpolation $x^\varepsilon(\cdot)$ and using the notation (2.13), we will see that mean square stable is achieved if $\hat{\xi}^* < 0$. Finally, by virtue of Borel-Cantelli lemma, similar to the procedure in Theorem 2.12, almost sure exponential stability can be guaranteed.

Remark 2.15. From the results of Theorem 3.4, Theorem 2.12, and Corollary 2.13, we conclude that for linear systems of the simple form (2.1), if the jump diffusions is stable and the step size is small enough, the numerical algorithm is also stable. The tightness given in Lemma 2.4 in fact is crucial. For the benchmark example, in the process of deriving the desired results, we have obtained the tightness. In more general setup, some sufficient conditions are needed to ensure the tightness.

2.2 Stability of Markovian Jump Diffusions and Their Numerical Methods

In this subsection, we focus on system given by

$$\begin{aligned} &dX(t) = b(\alpha(t))X(t)dt + \sigma(\alpha(t))X(t)dw(t) + \gamma(\alpha(t^-))X(t^-)dN(t) \\ &X(0) = x, \end{aligned} \tag{2.26}$$

where $\alpha(t)$ is a Markov chain and $\alpha(t) \in \mathcal{M}$ with generator Q . We assume that $w(t)$, $N(t)$, and $\alpha(t)$ are independent throughout the paper, and we further assume that the Markov chain $\alpha(t)$ is irreducible. Under this condition, $\alpha(t)$ has a unique stationary distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_m) \in \mathbb{R}^{1\times m}$. We proceed with the study on almost sure exponential stability.

Lemma 2.16. *For the jump diffusion* (2.26)*, the Lyapunov exponent is given by*

$$\limsup_{t\to\infty} \frac{1}{t} \ln |X^x(t)| = \sum_{i=1}^{m} \pi_i \left(b(i) - \frac{\sigma^2(i)}{2} + \lambda \ln |1 + \gamma(i)| \right) \ \textit{w.p.1.}$$

Therefore, the equilibrium point of the system is almost surely exponentially stable if and only if

$$\sum_{i=1}^{m} \pi_i \left(b(i) - \frac{\sigma^2(i)}{2} + \lambda \ln |1 + \gamma(i)| \right) < 0 \qquad \textit{w.p.1.}$$

Proof. First, we have

$$X(t) = x \exp \left(\int_0^t [b(\alpha(s)) - \frac{1}{2}\sigma^2(\alpha(s))]ds + \int_0^t \sigma(\alpha(s))dw(s) + \int_0^t \ln |1 + \gamma(\alpha(s))| dN(s) \right). \tag{2.27}$$

Therefore,

$$\begin{aligned} \lim_{t\to\infty} \frac{\ln |X(t)|}{t} = \lim_{t\to\infty} \Big(& \frac{\ln |x|}{t} + \frac{\int_0^t [b(\alpha(s)) - \frac{1}{2}\sigma^2(\alpha(s))]ds}{t} + \frac{\int_0^t \sigma(\alpha(s))dw(s)}{t} \\ & + \frac{\int_0^t \ln |1 + \gamma(\alpha(s))| d\tilde{N}(s)}{t} + \lambda \frac{\int_0^t \ln |1 + \gamma(\alpha(s))| ds}{t} \Big). \end{aligned} \tag{2.28}$$

Note that the quadratic variation for the term involving the Brownian motion is given by

$$\left\langle \int_0^t \sigma(\alpha(s))dw(s), \int_0^t \sigma(\alpha(s))dw(s) \right\rangle = \int_0^t \sigma^2(\alpha(s))ds \le \max_{i\in\mathcal{M}} \sigma^2(i)t.$$

By the law of large numbers for local martingales [11], we have

$$\frac{1}{t}\int_0^t \sigma(\alpha(s))dw(s) \to 0 \quad \text{w.p.1 as} \;\; t\to\infty.$$

Similarly, we get

$$\left\langle \int_0^t \ln|1+\gamma(\alpha(s))|d\tilde{N}(s), \int_0^t \ln|1+\gamma(\alpha(s))|d\tilde{N}(s) \right\rangle \le \lambda \max_{i\in\mathcal{M}}(\ln|1+\gamma(i)|)^2 t$$

and thus

$$\frac{1}{t}\int_0^t \ln|1+\gamma(\alpha(s))|d\tilde{N}(s) \to 0 \;\; \text{w.p.1 as } t\to\infty.$$

Then by the ergodicity of the Markov chain, we have

$$\lim_{t\to\infty}\frac{1}{t}\ln|X^x(t)| = \sum_{i=1}^m \pi_i\left(b(i)-\frac{\sigma^2(i)}{2}+\lambda\ln|1+\gamma(i)|\right) \quad \text{w.p.1.}$$

□

Next, we study the exponential p-stability.

Lemma 2.17. *For the jump diffusion* (2.26)*, the pth moment Lyapunov exponent is*

$$\limsup_{t\to\infty}\frac{1}{t}\ln(E|X^x(t)|^p) = \sum_{i=1}^m \pi_i\left(pb(i)+\frac{1}{2}p(p-1)\sigma^2(i)+\lambda(|1+\gamma(i)|^p-1)\right).$$

Therefore, the equilibrium point of the system is exponentially p-stable if and only if

$$\sum_{i=1}^m \pi_i\left(b(i)p+\frac{1}{2}p(p-1)\sigma^2(i)+\lambda(|1+\gamma(i)|^p-1)\right) < 0.$$

Proof. Note that

$$
\begin{aligned}
E|X(t)|^p = |x|^p E \exp \Big[p \int_0^t [b(\alpha(s)) - \frac{1}{2}\sigma^2(\alpha(s))] ds \\
+ \int_0^t p\sigma(\alpha(s)) dw(s) + p \int_0^t \ln|1 + \gamma(\alpha(s))| dN(s) \Big].
\end{aligned} \tag{2.29}
$$

To get the desired result, first note that

$$
\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t} \ln E \exp \left(p \int_0^t \sigma(\alpha(s)) dw(s) \right) \\
&\quad = \lim_{t\to\infty} \frac{1}{t} \ln E \exp \left(\frac{1}{2} p^2 \int_0^t \sigma^2(\alpha(s)) ds \right) \\
&\quad = \lim_{t\to\infty} \frac{1}{t} \ln E \exp \left(\frac{1}{2} p^2 \sum_{i=1}^m \sigma^2(i) \int_0^t (I_{\{\alpha(s)=i\}} - \pi_i) ds \right) \\
&\qquad\qquad + \lim_{t\to\infty} \frac{1}{t} \ln \exp \left(\frac{1}{2} p^2 \int_0^t \sum_{i=1}^m \sigma^2(i) \pi_i ds \right) \\
&\quad = \frac{p^2}{2} \sum_{i=1}^m \pi_i \sigma^2(i).
\end{aligned} \tag{2.30}
$$

Similarly, we have

$$
\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t} \ln E e^{p \int_0^t \ln|1+\gamma(\alpha(s))| dN(s)} \\
&\quad = \lim_{t\to\infty} \frac{1}{t} \ln E e^{\lambda \int_0^t (|1+\gamma(\alpha(s))|^p - 1) ds} \\
&\quad = \lambda \sum_{i=1}^m \pi_i [|1 + \gamma(i)|^p - 1].
\end{aligned}
$$

Also, we have

$$
\lim_{t\to\infty} \frac{1}{t} \ln E \exp \Big(p \int_0^t [b(\alpha(s)) - \frac{1}{2}\sigma^2(\alpha(s))] ds \Big) = p \sum_{i=1}^m \pi_i \Big(b(i) - \frac{1}{2}\sigma^2(i) \Big).
$$

Combing the above estimates,

$$\begin{aligned}\limsup_{t\to\infty}\frac{1}{t}\ln E|X^x(t)|^p &= \sum_{i=1}^m \pi_i\left(pb(i)-\frac{p}{2}\sigma^2(i)+\lambda[|1+\gamma(i)|^p-1]+\frac{p^2}{2}\sigma^2(i)\right)\\ &= \sum_{i=1}^m \pi_i\left(b(i)p+\frac{p(p-1)}{2}\sigma^2(i)+\lambda[|1+\gamma(i)|^p-1]\right).\end{aligned}$$

Remark 2.18. When Markov switching is involved, to numerically solve the equation, we use the similar algorithm in which the α_n can be constructed by using a one-step transition matrix $\exp(\varepsilon Q)$ or alternatively, as observed in [14], instead of the discrete-time Markov chain, we can use the so-called ε-skeleton $\alpha_n = \alpha(\varepsilon n)$, where $\alpha(\cdot)$ is the original continuous-time Markov chain. The stability analysis for numerical algorithms can be carried out similar to that of Theorem 2.12 and Theorem 3.4 with proper utilization of the irreducibility of the Markov chain.

Theorem 2.19. *Suppose that*

$$\sum_{i=1}^m \pi_i\left(pb(i)+\frac{1}{2}p(p-1)\sigma^2(i)+\lambda[|1+\gamma(i)|^p-1]\right)<0.$$

Then numerical algorithm is exponentially p-stable and

$$\limsup_{T\to\infty}\frac{1}{T}\lim_{\varepsilon\to 0}\ln E|\widetilde{x}^{\varepsilon}_T(T)|^p = \sum_{i=1}^m \pi_i\left(pb(i)+\frac{1}{2}p(p-1)\sigma^2(i)+\lambda[|1+\gamma(i)|^p-1]\right). \quad (2.31)$$

Theorem 2.20. *Suppose that the numerical algorithm is exponentially p-stable, using n in lieu of T, as $n\to\infty$,* (2.23) *holds. As a result, for small enough $\theta\in(0,1)$, the numerical* *algorithm is almost surely exponentially stable and with the property that*

$$\limsup_{n\to\infty}\frac{1}{n}\lim_{\varepsilon\to 0}\ln|\widetilde{x}_n^{\varepsilon}(n)| \le \sum_{i=1}^{m}\pi_i\left(b(i)-\frac{\sigma^2(i)}{2}+\lambda\ln|1+\gamma(i)|\right)+\theta<0 \quad w.p.1. \tag{2.32}$$

2.3 Stability of Multi-Dimensional Jump Diffusions and Their Numerical Methods

In this section, we consider the nonlinear r-dimensional jump diffusion systems

$$\begin{aligned} dX(t) &= b(X(t))dt + \sigma(X(t))dw(t) + g(X(t^-),\gamma)dN(t) \\ X(0) &= x, \end{aligned} \tag{2.33}$$

where $b(\cdot):\mathbb{R}^r\mapsto\mathbb{R}^r$, $\sigma(\cdot):\mathbb{R}^r\mapsto\mathbb{R}^{r\times d}$, $g(\cdot,\cdot):\mathbb{R}^r\times\mathbb{R}^r\mapsto\mathbb{R}^r$, and $w(\cdot)$ is a d-dimensional Brownian motion in which each $w_j(t)$ is a scalar Brownian motion. $N(\cdot)$ is a one dimensional Poisson process as in (2.1). Our interest lies in the case that the systems can be linearized. We further assume that the jump diffusion equation has a unique strong solution for each initial condition.

(A1) There exist $b\in\mathbb{R}^{r\times r}$, $\sigma_l\in\mathbb{R}^{r\times r}$, and $G(\gamma)\in\mathbb{R}^{r\times r}$ for $l=1,2,\ldots,d$ such that as $x\to 0$,

$$\begin{aligned} b(x) &= bx + o(|x|) \\ \sigma(x) &= (\sigma_1 x, \sigma_2 x, \ldots, \sigma_d x) + o(|x|) \\ g(x,\gamma) &= G(\gamma)x + o(|x|), \end{aligned} \tag{2.34}$$

where $\gamma\in\Gamma$ and Γ is a subset of $\mathbb{R}^r-\{0\}$ that is the range space of the impulses jumps.

To proceed, we first derive sufficient conditions for exponential p-stability and almost sure exponential stability, then we study the numerical part accordingly.

Denoting

$$\rho(A) = \begin{cases} \Lambda_{\max}(A), & p \geq 2 \\ \Lambda_{\min}(A), & p < 2. \end{cases}$$

We present a sufficient condition to guarantee system (2.33) being exponentially p-stable.

Theorem 2.21. *Assume* (A1) *and*

$$\xi^* = \Lambda_{\max}(\frac{b+b'}{2}) + \frac{1}{2}\Lambda_{\max}(\sum_{j=1}^{d} \sigma_j'\sigma_j) + \frac{p-2}{2}\rho^2(\frac{\sum_{j=1}^{d}(\sigma_j + \sigma_j')}{2}) + \frac{\lambda}{p}[|I + G(\gamma)|^p - 1] < 0. \tag{2.35}$$

Then system (2.33) *is exponentially* p*-stable.*

Proof. By the Dynkin formula, we have the following expression

$$\begin{aligned}
&E|X(t)|^p - |x|^p \\
&\quad = E\int_0^t \{p|X(s)|^{p-2}X'(s)[bX(s) + o(|X(s)|)] + \lambda[|X(s) + g(X(s))|^p - |X(s)|^p] \\
&\qquad + \frac{1}{2}\mathrm{tr}\big[(\sum_{j=1}^{d} \sigma_j X(s)X'(s)\sigma_j' + o(|X(s)|^2)\big) \\
&\qquad \cdot \big(p|X(s)|^{p-2}I + p(p-2)|X(s)|^{p-4}X(s)X'(s)\big)\big]\}ds \\
&\quad \leq E\int_0^t p|X(s)|^p\big(\frac{X'(s)bX(s)}{|X(s)|^2} + \frac{\lambda}{p}[|I + G(\gamma)|^p - 1] + \frac{1}{2}\frac{\sum_{j=1}^{d} X'(s)\sigma_j'\sigma_j X(s)}{|X(s)|^2} \\
&\qquad + \frac{1}{2}(p-2)\frac{\sum_{j=1}^{d}(X'(s)\sigma_j' X(s))^2}{|X(s)|^4} + o(1)\big)ds \\
&\quad \leq \int_0^t pE|X(s)|^p\big(\Lambda_{\max}(\frac{b+b'}{2}) + \frac{1}{2}\Lambda_{\max}(\sum_{j=1}^{d} \sigma_j'\sigma_j) + \frac{p-2}{2}\rho^2(\frac{\sum_{j=1}^{d}(\sigma_j + \sigma_j')}{2}) \\
&\qquad + \frac{\lambda}{p}[|I + G(\gamma)|^p - 1]\big)ds \\
&\quad \leq p\xi^* \int_0^t E|X(s)|^p ds.
\end{aligned} \tag{2.36}$$

The Gronwall inequality leads to

$$E|X(t)|^p \leq |x|^p e^{p\xi^* t}.$$

Therefore, system (2.33) is exponentially p-stable if $\xi^* < 0$. □

Theorem 2.22. *Assume* (A1). *If*

$$\xi = \Lambda_{\max}(\frac{b+b'}{2}) + \frac{1}{2}\Lambda_{\max}(\sum_{j=1}^{d} \sigma_j' \sigma_j) - \Lambda_{\max}^2(\frac{\sum_{j=1}^{d}(\sigma_j + \sigma_j')}{2}) + \lambda \ln|I + G(\gamma)| < 0, \quad (2.37)$$

then system (2.33) *is almost surely exponentially stable.*

Proof. We consider the Lyapunov function $V(x) = \ln|x|$. By the generalized Itô formula, we get

$$\begin{aligned}
\ln|X(t)| - \ln|x| = & \int_0^t \frac{X'(s)}{|X(s)|^2}[bX(s) + o(|X(s)|)]ds + M_1(t) \\
& + \lambda \int_0^t [\ln|X(s^-) + g(X(s^-))| - \ln|X(s^-)|]ds + M_2(t) \\
& + \frac{1}{2}\int_0^t \mathrm{tr}[(\frac{I}{|X(s)|^2} - 2\frac{X(s)X'(s)}{|X(s)|^4})(\sum_{j=1}^{d}(\sigma_j X(s)X'(s)\sigma_j') + o(|X(s)|^2))]ds,
\end{aligned} \quad (2.38)$$

in which

$$\begin{aligned}
M_1(t) &= \int_0^t \frac{X'(s)}{|X(s)|^2}\sigma(X(s))dw(s), \\
M_2(t) &= \int_0^t [\ln|X(s^-) + g(X(s^-))| - \ln|X(s^-)|]d\tilde{N}(s).
\end{aligned}$$

Note that the quadratic variation of $M_1(t)$ is given by

$$\begin{aligned}\langle M_1, M_1\rangle(t) &\leq \int_0^t \frac{|X'(s)\sigma(X(s))|^2}{|X(s)|^4} ds \\ &\leq \int_0^t \frac{\sum_{j=1}^d X'(s)\sigma_j'\sigma_j X(s) + o(|X(s)|^2)}{|X(s)|^2} ds \leq \Lambda_{\max}(\sum_{j=1}^d \sigma_j'\sigma_j)t.\end{aligned}$$

The law of large numbers for local martingales [11] yields that $\frac{M_1(t)}{t} \to 0$ w.p.1. as $t \to \infty$.

For the term $M_2(t)$, the corresponding quadratic variation is as follows

$$\begin{aligned}\langle M_2, M_2\rangle(t) &= \lambda \int_0^t [\ln |X(s^-) + g(X(s^-))| - \ln |X(s^-)|]^2 ds \\ &\leq \lambda \int_0^t (\ln \frac{|X(s^-) + g(X(s^-))|}{|X(s^-)|})^2 ds \leq \lambda \int_0^t \frac{|g(X(s^-))|^2}{|X(s^-)|^2} ds \\ &\leq \lambda \int_0^t \frac{|G(\gamma)X(s^-)|^2 + o(|X(s^-)|^2)}{|X(s^-)|^2} ds \leq \lambda G^2(\gamma)t.\end{aligned}$$

Similarly, the law of large numbers for local martingales implies that $\frac{M_2(t)}{t} \to 0$ w.p.1 as $t \to \infty$.

Now let us work on the rest of the terms. We have that w.p.1.,

$$\begin{aligned}&\limsup_{t\to\infty} \frac{1}{t} \int_0^t \frac{X'(s)}{|X(s)|^2}[bX(s) + o(|X(s)|)]ds \\ &= \limsup_{t\to\infty} \frac{1}{t} \int_0^t [\frac{X'(s)bX(s)}{|X(s)|^2} + \frac{X'(s)o(|X(s)|)}{|X(s)|^2}]ds \\ &\leq \limsup_{t\to\infty} \frac{1}{t} \int_0^t (\Lambda_{\max}(\frac{b+b'}{2}) + o(1))ds \\ &\leq \limsup_{t\to\infty} \frac{1}{t}\Lambda_{\max}(\frac{b+b'}{2})t \\ &= \Lambda_{\max}(\frac{b+b'}{2}),\end{aligned}$$

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{\lambda}{t}\int_0^t[\ln|X(s^-)+g(X(s^-))|-\ln|X(s^-)|]ds\\
&=\limsup_{t\to\infty}\frac{\lambda}{t}\int_0^t\ln\frac{|X(s^-)+g(X(s^-))|}{|X(s^-)|}ds\\
&\le\limsup_{t\to\infty}\frac{\lambda}{t}\ln|I+G(\gamma)|t=\lambda\ln|I+G(\gamma)|,
\end{aligned}$$

and

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{1}{t}\int_0^t\mathrm{tr}[(\frac{I}{|X(s)|^2}-2\frac{X(s)X'(s)}{|X(s)|^4})(\sum_{j=1}^d\sigma_jX(s)X'(s)\sigma_j'+o(|X(s)|^2))]ds\\
&=\limsup_{t\to\infty}\frac{1}{t}\int_0^t[\sum_{j=1}^d(\frac{X'(s)\sigma_j'\sigma_jX(s)}{|X(s)|^2}-2\frac{(X'(s)\sigma_j'X(s))^2}{|X(s)|^4})+o(1)]ds\\
&\le\limsup_{t\to\infty}[\frac{1}{t}\int_0^t\sum_{j=1}^d\frac{X'(s)\sigma_j'\sigma_jX(s)}{|X(s)|^2}ds-\frac{2}{t}\int_0^t\sum_{j=1}^d\frac{(X'(s)\sigma_j'X(s))^2}{|X(s)|^4}ds]\\
&\le\Lambda_{\max}(\sum_{j=1}^d\sigma_j'\sigma_j)-2\Lambda^2_{\min}(\frac{\sum_{j=1}^d(\sigma_j+\sigma_j')}{2}).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\limsup_{t\to\infty}\frac{1}{t}\ln|X(t)|\le\Lambda_{\max}(\frac{b+b'}{2})+\frac{1}{2}\Lambda_{\max}(\sum_{j=1}^d\sigma_j'\sigma_j)-\Lambda^2_{\max}(\frac{\sum_{j=1}^d(\sigma_j+\sigma_j')}{2})\\
+\lambda\ln|I+G(\gamma)|=\xi<0\quad\text{w.p.1.}
\end{aligned}$$

□

We use the following algorithm to approximate the solution of (2.33)

$$\begin{aligned}
x_{n+1}&=x_n+b(x_n)\varepsilon+\sigma(x_n)\Delta w_n+g(x_n)\Delta N_n\\
x_0&=X(0)=x.
\end{aligned}\tag{2.39}$$

To proceed, we demonstrate that the sequence $\{x_n:\varepsilon>0,n<\infty\}$ is tight under suitable conditions.

Lemma 2.23. *For algorithm* (2.39)*, assume* (2.35) *holds. Then the sequence* $\{x_n:\varepsilon>$

$0, n < \infty\}$ *is tight.*

The proof of the lemma is similar to that of Lemma 2.4; thus the detailed calculations are omitted.

Theorem 2.24. *Assume* (A1) *and Lemma* 2.23 *hold, and suppose*

$$\xi^* = \Lambda_{\max}(\frac{b+b'}{2}) + \frac{1}{2}\Lambda_{\max}(\sum_{j=1}^{d}\sigma_j'\sigma_j) + \frac{p-2}{2}\rho^2(\frac{\sum_{j=1}^{d}(\sigma_j+\sigma_j')}{2}) + \frac{\lambda}{p}[|I+G(\gamma)|^p - 1] < 0.$$

Then algorithm (2.39) *is exponentially p-stable, and has property that*

$$\limsup_{T\to\infty}\frac{1}{T}\lim_{\varepsilon\to 0}\ln E|\widetilde{x}_T^{\varepsilon}(T)|^p \le p\xi^* < 0.$$

Proof. We obtain

$$\begin{aligned} &\limsup_{T\to\infty}\frac{1}{T}\lim_{\varepsilon\to 0}\ln E|\widetilde{x}_T^{\varepsilon}(T)|^p \\ &\quad = \limsup_{T\to\infty}\frac{1}{T}\ln E|\widetilde{x}_T(T)|^p \\ &\quad \le \lim_{T\to\infty}\frac{p\ln E|\widetilde{x}_T(0)|}{T} + p\xi^* = p\xi^* < 0. \end{aligned} \tag{2.40}$$

Theorem 2.25. *Assume* (A1) *and Lemma* 2.23 *hold, and suppose*

$$\xi = \Lambda_{\max}(\frac{b+b'}{2}) + \frac{1}{2}\Lambda_{\max}(\sum_{j=1}^{d}\sigma_j'\sigma_j) - \Lambda_{\max}^2(\frac{\sum_{j=1}^{d}(\sigma_j+\sigma_j')}{2}) + \lambda\ln|I+G(\gamma)| < 0.$$

Then algorithm (2.39) *is almost surely exponentially stable and for small enough* θ*,*

$$\limsup_{n\to\infty}\frac{1}{n}\lim_{\varepsilon\to 0}\ln|\widetilde{x}_n^{\varepsilon}(n)| \le \xi + \theta < 0 \quad \textit{w.p.1.}$$

The proof is similar to that of Theorem 2.12 with the use of (2.37) and (2.35). The details are omitted for brevity.

Remark 2.26. In view of Theorem 2.24 and Theorem 2.25, for nonlinear systems, as long as they can be linearized with appropriate conditions together with the tightness of the iterates, stability of stochastic jump diffusions will lead to stability of the corresponding numerical algorithms.

2.4 Stability of Nonlinear Regime Switching Jump Diffusions

Now we consider the dynamic system given by

$$\begin{aligned} &dX(t) = b(X(t), \alpha(t))dt + \sigma(X(t), \alpha(t))dw(t) + dJ(t), \\ &J(t) = \int_0^t \int_\Gamma g(X(s^-), \alpha(s^-), \gamma) N(ds, d\gamma), \\ &X(0) = x, \, \alpha(0) = \alpha, \end{aligned} \tag{2.41}$$

where the switching process $\alpha(\cdot)$ obeys the transition rule

$$P\{\alpha(t + \Delta t) = j | \alpha(t) = i, X(s), \alpha(s), s \le t\} = q_{ij}(X(t))\Delta t + o(\Delta t), \text{ for } i \ne j, \tag{2.42}$$

$w(t)$ is a d-dimensional standard Brownian motion, and $N(\cdot, \cdot)$ is a Poisson measure such that the jump process $N(\cdot, \cdot)$ is independent of the Brownian motion $w(\cdot)$. Equation (2.41) can be written as integral form:

$$X(t) = x + \int_0^t b(X(s), \alpha(s))ds + \int_0^t \sigma(X(s), \alpha(s))dw(s) + \int_0^t \int_\Gamma g(X(s^-), \alpha(s^-), \gamma)N(ds, d\gamma).$$

Here we have used a setup similar to [19]. When we wish to emphasize the initial data dependence in the sequel, we write the process as $(X^{x,\alpha}(t), \alpha^{x,\alpha}(t))$. Note that although the two-component process $(X(t), \alpha(t))$ is Markov, $\alpha(t)$ generally is not a Markov chain due to the dependence of the state x in the generator. The transition rule indicates that $\alpha(t)$ depends on the jump diffusion component. Thus the setup we consider is more general than that of considered in the literature, whereas in the past work it was often assumed that $\alpha(t)$ itself is a Markov chain and $w(t)$ and $\alpha(t)$ are independent.

For future use, we define a compensated or centered Poisson measure as

$$\widetilde{N}(t, B) = N(t, B) - \lambda t \pi(B) \text{ for } B \subset \Gamma,$$

where $0 < \lambda < \infty$ is known as the jump rate and $\pi(\cdot)$ is the jump distribution (a probability measure). With this centered Poisson measure, we can rewrite $J(t)$ as

$$J(t) = \int_0^t \int_\Gamma g(X(s^-), \alpha(s^-), \gamma)\widetilde{N}(ds, d\gamma) + \lambda \int_0^t \int_\Gamma g(X(s^-), \alpha(s^-), \gamma)\pi(d\gamma)ds,$$

which is the sum of a martingale and an absolute continuous process provided certain conditions are satisfied for the function $g(\cdot)$.

Note that the evolution of the discrete component $\alpha(\cdot)$ can be represented by a stochastic integral with respect to a Poisson measure (e.g., [31]). Define a function $h : \mathbb{R}^r \times \mathcal{M} \times \mathbb{R} \mapsto \mathbb{R}$ by

$$h(x, i, z) = \sum_{j=1}^{m} (j - i) I_{\{z \in \Delta_{ij}(x)\}}. \tag{2.43}$$

That is, with the partition $\{\Delta_{ij}(x) : i, j \in \mathcal{M}\}$ used and for each $i \in \mathcal{M}$, if $z \in \Delta_{ij}(x)$,

$h(x,i,z) = j - i$; otherwise $h(x,i,z) = 0$. Then we may write the switching process as a stochastic integral

$$d\alpha(t) = \int_{\mathbb{R}} h(X(t), \alpha(t-), z) N_1(dt, dz), \tag{2.44}$$

where $N_1(dt, dz)$ is a Poisson random measure with intensity $dt \times \widetilde{m}(dz)$, and $\widetilde{m}(\cdot)$ is the Lebesgue measure on $\mathbb{R}$. The Poisson random measure $N_1(\cdot,\cdot)$ is independent of the Brownian motion $w(\cdot)$ and the Poisson measure $N(\cdot,\cdot)$. For subsequent use, we define another centered Poisson measure as

$$\mu(dt, dz) = N_1(dt, dz) - dt \times \widetilde{m}(dz).$$

The generator $\mathcal{G}$ associated with the process $(X(t), \alpha(t))$ is defined as follows: For each $i \in \mathcal{M}$, and for any twice continuously differentiable function $f(\cdot, i)$,

$$\mathcal{G}f(x,\cdot)(i) = \mathcal{L}f(x,\cdot)(i) + \lambda \int_{\Gamma} [f(x + g(x,i,\gamma), i) - f(x,i)]\pi(d\gamma),$$

where $\mathcal{L}$ is the operator for a switching diffusion process given by

$$\begin{aligned} \mathcal{L}f(x,\cdot)(i) &= \frac{1}{2}\sum_{k,l=1}^{r} a_{kl}(x,i)\frac{\partial^2 f(x,i)}{\partial x_k \partial x_l} + \sum_{k=1}^{r} b_k(x,i)\frac{\partial f(x,i)}{\partial x_k} + Q(x)f(x,\cdot)(i) \\ &= \frac{1}{2}\mathrm{tr}(a(x,i)Hf(x,i)) + b'(x,i)\nabla f(x,i) + Q(x)f(x,\cdot)(i), \quad i \in \mathcal{M}, \end{aligned} \tag{2.45}$$

where $x \in \mathbb{R}^r$, $a(x,i) = \sigma(x,i)\sigma'(x,i)$. In what follows, we often write $\mathcal{L}f(x,\cdot)(i)$ as $\mathcal{L}f(x,i)$ and $\mathcal{G}f(x,\cdot)(i)$ as $\mathcal{G}f(x,i)$ for convenience whenever there is no confusion.

To proceed, we need the following assumptions.

(A1) The functions $b(\cdot,i)$, $\sigma(\cdot,i)$, and $g(\cdot,i,\gamma)$ satisfy $b(0,i) = 0$, $\sigma(0,i) = 0$, and $g(0,i,\gamma) = 0$ for each $i \in \mathcal{M}$; $\sigma(x,i)$ vanishes only at $x = 0$ for each $i \in \mathcal{M}$.

(A2) There exists a positive constant K_0 such that for each $i \in \mathcal{M}, x, y \in \mathbb{R}^r$ and $\gamma \in \Gamma$,

$$\begin{aligned}
&|b(x,i) - b(y,i)| + |\sigma(x,i) - \sigma(y,i)| \leq K_0|x-y|, \\
&|g(x,i,\gamma) - g(y,i,\gamma)| \leq K_0|x-y|.
\end{aligned}$$

(A3) There exists $g^*(i)$ satisfying

$$|g(x,i,\gamma)| \leq g^*(i)|x| \text{ for each } x \in \mathbb{R}^r, i \in \mathcal{M}, \text{ and each } \gamma \in \Gamma.$$

We elaborate on the conditions briefly. Condition (A1) indicates that 0 is an equilibrium point; (A2) is a Lipschitz condition on the functions. It together with the equilibrium point 0 implies that the functions grow at most linearly. Several of our results to follow are concerned with equilibrium point of the switching jump diffusions. To proceed, As a preparation, we first recall a lemma, which indicates that the equilibrium $(0, \alpha)$ is inaccessible in that starting with any $x \neq 0$, the system will not reach the origin with probability one. The proof of this lemma can be found in [19, Lemma 2.10].

Lemma 2.27. $P\{X^{x,\alpha}(t) \neq 0, t \geq 0\} = 1$*, for any* $x \neq 0$ *and* $\alpha \in \mathcal{M}$*.*

To proceed, we first recall two lemmas. The detailed proof can be found in [19].

Lemma 2.28. *Let* $D \subset \mathbb{R}^r$ *is a neighborhood of* 0*. Suppose that for each* $i \in \mathcal{M}$*, there exists a nonnegative Lyapunov function* $V(\cdot, i): D \mapsto \mathbb{R}$ *such that*

(i) $V(\cdot, i)$ *is continuous in* D *and vanishes only at* $x = 0$*;*

(ii) $V(\cdot, i)$ *is twice continuously differentiable in* $D - \{0\}$ *and satisfies* $\mathcal{G}V(x,i) \leq 0$ *for all* $x \in D - \{0\}$*.*

Then the equilibrium point $x = 0$ is stable in probability.

Define

$$\tau_{\rho,\varsigma} := \inf\{t \geq 0 : |X(t)| = \rho \text{ or } |X(t)| = \varsigma\}, \tag{2.46}$$

for any $0 < \rho < \varsigma$ and any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$ with $\rho < |x| < \varsigma$.

Lemma 2.29. *Assume that the conditions of Lemma 2.28 hold, and that for any sufficiently small $0 < \varrho < \varsigma$ and any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$ with $\varrho < |x| < \varsigma$, $P\{\tau_{\varrho,\varsigma} < \infty\} = 1$. Then the equilibrium point $x = 0$ is asymptotically stable in probability.*

Theorem 2.30. *Assume that the conditions of Lemma 2.29 hold, and that $V_\varsigma := \inf\limits_{\substack{|x|\geq\varsigma \\ i\in\mathcal{M}}} V(x,i) \to \infty$ as $\varsigma \to \infty$. Then the equilibrium point $x = 0$ is asymptotically stable in the large.*

Proof. For any $\varepsilon > 0$, $i \in \mathcal{M}$, and $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$, there exists a $\varsigma > |x|$ large enough such that $\inf\limits_{\substack{|X|\geq\varsigma \\ i\in\mathcal{M}}} V(X,i) \geq 2V(x,\alpha)/\varepsilon$.

Let τ_ς be the stopping time $\tau_\varsigma := \inf\{t \geq 0 : |X(t)| \geq \varsigma\}$ and $t_\varsigma = \tau_\varsigma \wedge t$. Then it follows from Dynkin's formula that

$$EV(X(t_\varsigma), \alpha(t_\varsigma)) - V(x, \alpha) = E \int_0^{t_\varsigma} \mathcal{G}V(X(u), \alpha(u))du \leq 0.$$

Consequently, $EV(X(t_\varsigma), \alpha(t_\varsigma)) \leq V(x, \alpha)$. Then we have

$$E[V(X(\tau_\varsigma), \alpha(\tau_\varsigma))I_{\{\tau_\varsigma < t\}}] \leq V(x, \alpha).$$

Hence, $\frac{2V(x,\alpha)}{\varepsilon} P(\tau_\varsigma < t) \leq V(x, \alpha)$. So $P(\tau_\varsigma < t) \leq \varepsilon/2$. Let $t \to \infty$, $P(\tau_\varsigma < \infty) \leq \varepsilon/2$. Then it follows from Lemma 2.52 that, for any $\varrho > 0$ with $\varrho < |x| < \varsigma$ we have

$$1 = P(\tau_{\varrho,\varsigma} < \infty) \leq P(\tau_\varrho < \infty) + P(\tau_\varsigma < \infty),$$

in which τ_ϱ is the stopping time $\tau_\varrho := \inf\{t \geq 0 : |X(t)| \leq \varrho\}$, where $\tau_{\varrho,\varsigma}$ was defined in (2.87). Consequently, $P(\tau_\varrho < \infty) \geq 1 - \varepsilon/2$. This implies that $P\{\inf_{t\geq 0} |X(t)| \leq \varrho\} \geq 1 - \varepsilon/2$. Since $\varrho > 0$ can be arbitrarily small, $P\{\inf_{t\geq 0} |X(t)| = 0\} \geq 1 - \varepsilon/2$.

Now we can follow the same techniques in [20, Lemma 7.6] and obtain $P\{\lim_{t\to\infty} X(t) = 0\} \geq 1 - \varepsilon/2$. That is, the equilibrium point $x = 0$ is asymptotically stable in the large as desired.

For application, it is important to be able to handle linearized systems. Similar to (2.34), we pose the following condition.

(A4) For each $i \in \mathcal{M}$, there exist $b(i), \sigma_l(i) \in \mathbb{R}^{r\times r}$ for $l = 1, 2, \ldots, d$, and a generator of a continuous-time Markov chain $\widehat{Q} = (\hat{q}_{ij})$ with the corresponding Markov chain denoted by $\widehat{\alpha}(t)$ such that as $x \to 0$,

$$\begin{aligned} b(x,i) &= b(i)x + o(|x|),\\ \sigma(x,i) &= (\sigma_1(i)x, \sigma_2(i)x, ..., \sigma_d(i)x) + o(|x|),\\ Q(x) &= \widehat{Q} + o(1). \end{aligned}$$

Moreover, $\widehat{Q}$ is irreducible.

Assumption (A4) indicates that near the origin, the coefficients are locally linear. By choosing a Lyapunov function properly, we have the same sufficient condition for asymptotically stable in the large as that of asymptotically stable in probability. The result is provided below, and the proof is omitted. The method involved is similar to [19, Theorem 3.5].

Corollary 2.31. *Under assumptions* (A1)-(A4)*, the equilibrium point* $x = 0$ *of the system given by* (2.41) *and* (2.42) *is asymptotically stable in the large if*

$$\sum_{i\in\mathcal{M}} \mu_i \left(\Lambda_{\max}(\frac{b(i)+b'(i)}{2}) + \frac{1}{2}\Lambda_{\max}\left(\sum_{l=1}^{d} \sigma_l(i)\sigma_l'(i)\right) + \lambda g^*(i) \right) < 0.$$

In which $\mu = (\mu_1, \mu_2, \cdots, \mu_m) \in \mathbb{R}^{1\times m}$ is the stationary distribution of $\widehat{\alpha}(t)$ and $\Lambda_{\max}(A)$ denotes the largest eigenvalue of the symmetric part of A.

We first recall a lemma in bleow, which indicates that the process $(X(t), \alpha(t))$ has no finite explosion time, also known as regular. The proof of this lemma can be found in [19, Lemma 2.8].

Lemma 2.32. *Under assumptions* (A1)-(A3)*, the switching jump diffusion* $(X(t), \alpha(t))$ *is regular.*

Lemma 2.33. *Let* $D \subset \mathbb{R}^r$ *be a neighborhood of* 0*. Assume that the conditions of Lemma* 2.32 *hold and assume that for each* $i \in \mathcal{M}$*, there exists a nonnegative Lyapunov function* $V(\cdot, i)$*:* $D \mapsto \mathbb{R}$ *such that* $V(\cdot, i)$ *is twice continuously differentiable in* $D - \{0\}$*, and satisfies the following conditions:*

$$k_1|x|^p \le V(x,i) \le k_2|x|^p, \quad x \in D, \tag{2.47}$$

$$\mathcal{G}V(x,i) \le -kV(x,i) \text{ for all } x \in D - \{0\}, \tag{2.48}$$

for some positive constants k_1, k_2 *and* k*. Then the equilibrium point* $x = 0$ *is exponential* p*-stable.*

Remark 2.34. Under certain conditions, we can also obtain the result of almost surely exponential stability by similar argument in [6, Theorem 5.8.1].

Theorem 2.35. *Under assumptions* (A1)-(A3)*, exponential p-stability implies almost surely exponential stability.*

The proof details can be found in [90].

One of the important properties of a diffusion processes is the continuous and smooth dependence on the initial data. This property is preserved for the switching diffusion processes with state-dependent switching; however much work is needed. We will show that this property is also preserved for the switching jump diffusion processes. The results are stated for multi-dimensional cases, whereas the proofs are carried out for a one-dimensional process for the sake of convenience. Let $(X(t), \alpha(t))$ denote the switching jump process with initial condition (x, α) and $(\widetilde{X}(t), \widetilde{\alpha}(t))$ be the process starting from $(\widetilde{x}, \alpha)$, let $\Delta \neq 0$ be small and denote $\tilde{x} = x + \Delta$ in the sequel.

Lemma 2.36. *Under conditions* (A1)-(A3), we have for $0 \leq t \leq T$ and any positive constant ι, $E|X(t)|^{\iota} \leq |x|^{\iota} e^{\kappa t} \leq C$, for $x \neq 0$, $\alpha \in \mathcal{M}$, where $\kappa = \kappa(\iota, K_0, m, g^*(i))$ and $C = C(\kappa, T)$.

Proof. For each $i \in \mathcal{M}$ and $x \neq 0$, define $V(x, i) = |x|^{\iota}$ for any $\iota \in \mathbb{R}_+ - \{0\}$. Then for any $\Delta > 0$ and $|x| > \Delta$,

$$\begin{aligned}\mathcal{G}|x|^{\iota} = \iota|x|^{\iota-2}x'b(x,i) + \lambda \int_{\Gamma}(|x+g(x,i,\gamma)|^{\iota} - |x|^{\iota})\pi(d\gamma) \\ +\frac{1}{2}\text{tr}[\sigma(x,i)\sigma'(x,i)\iota|x|^{\iota-4}(|x|^2 I + (\iota-2)xx')].\end{aligned}$$

Since 0 is an equilibrium point, Cauchy-Schwartz inequality implies $|x'b(x,i)| \leq |x||b(x,i)| \leq$

$K_0|x|^2$,

$$\begin{aligned}
\operatorname{tr}(\sigma\sigma') &= |\sigma|^2 \le K_0|x|^2, \\
\operatorname{tr}(\sigma\sigma' x x') &= x'\sigma\sigma' x \le |x|^2|\sigma|^2 \le K_0|x|^4.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
|\mathcal{G}|x|^\iota| \le\ & K_0\iota|x|^\iota + \frac{1}{2}K_0\iota|x|^{\iota-2}(|x|^2 + (\iota-2)|x|^2) \\
& +\lambda|x|^\iota(|1+g^*(i)|^\iota - 1) \le \kappa|x|^\iota.
\end{aligned}$$

Define the stopping time $\tau_\Delta := \inf\{t \ge 0, |X(t)| \le \Delta\}$. Then by the generalized Itô lemma, we obtain

$$\begin{aligned}
E|X(\tau_\Delta \wedge t)|^\iota &= |x|^\iota + E\int_0^{\tau_\Delta \wedge t} \mathcal{G}|X(u)|^\iota du \\
&\le |x|^\iota + \kappa E\int_0^{\tau_\Delta \wedge t} |X(u)|^\iota du \\
&\le |x|^\iota + \kappa E\int_0^{t} |X(u \wedge \tau_\Delta)|^\iota du.
\end{aligned}$$

By Gronwall's inequality, it follows that

$$E|X(\tau_\Delta \wedge t)|^\iota \le |x|^\iota e^{\kappa t}.$$

Letting $\Delta \to 0$, by virtue of non-zero property of $X(t)$ shown in Lemma 2.27, we have

$$E|X(t)|^\iota \le |x|^\iota e^{\kappa t}.$$

For $0 \le t \le T$, we further have

$$E|X(t)|^\iota \le |x|^\iota e^{\kappa t} \le |x|^\iota e^{\kappa T} = C.$$

.

Theorem 2.37. *Under the conditions of Lemma* 2.36*, define*

$$\begin{aligned}\phi^{\Delta}(t) = &\frac{1}{\Delta}\int_0^t [b(\widetilde{X}(s),\widetilde{\alpha}(s)) - b(\widetilde{X}(s),\alpha(s))]ds \\ &+\frac{1}{\Delta}\int_0^t [\sigma(\widetilde{X}(s),\widetilde{\alpha}(s)) - \sigma(\widetilde{X}(s),\alpha(s))]dw(s) \\ &+\frac{1}{\Delta}\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma) - g(\widetilde{X}(s^-),\alpha(s^-),\gamma)]N(ds,d\gamma).\end{aligned} \tag{2.49}$$

Then we have $\lim\limits_{\Delta\to 0} E \sup\limits_{0\le t\le T} |\phi^{\Delta}(t)|^2 = 0.$

Proof. It can be verified that

$$\begin{aligned}E\sup_{0\le t\le T}|\phi^{\Delta}(t)|^2 = &\frac{K}{\Delta^2}E\int_0^T |b(\widetilde{X}(s),\widetilde{\alpha}(s)) - b(\widetilde{X}(s),\alpha(s))|^2 ds \\ &+\frac{K}{\Delta^2}E\sup_{0\le t\le T}|\int_0^t [\sigma(\widetilde{X}(s),\widetilde{\alpha}(s)) - \sigma(\widetilde{X}(s),\alpha(s))]dw(s)|^2 \\ &+\frac{K}{\Delta^2}E\int_0^T \int_\Gamma |g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma) - g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma) \\ &+\frac{K}{\Delta^2}E\sup_{0\le t\le T}|\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma) - g(\widetilde{X}(s^-),\alpha(s^-),\gamma)]\widetilde{N}(ds,d\gamma)|^2.\end{aligned} \tag{2.50}$$

Let us first consider the next to the last line of (2.50). By choosing $\eta = \Delta^{\gamma_0}$ with $\gamma_0 > 2$ and partition the interval $[0, T]$ by η we obtain

$$\begin{aligned}
&E\int_0^T\int_\Gamma |g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&= E\sum_{k=0}^{\lfloor\frac{T}{\eta}\rfloor-1}\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&= KE\sum_{k=0}^{\lfloor\frac{T}{\eta}\rfloor-1}[\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(k\eta),\widetilde{\alpha}(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&\qquad +\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&\qquad +\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)].
\end{aligned}\tag{2.51}$$

For the third line of (2.51), we have the following bound by virtue of (A2) and [32, Theorem 3.7.1],

$$\begin{aligned}
&E\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(k\eta),\widetilde{\alpha}(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&\le K\int_{k\eta}^{k\eta+\eta}E\left|\widetilde{X}(s^-)-\widetilde{X}(k\eta)\right|^2 ds\\
&\le K\int_{k\eta}^{k\eta+\eta}(s-k\eta)ds\le K\eta^2.
\end{aligned}\tag{2.52}$$

We can derive the upper bound for the last line of (2.51) similarly,

$$E\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\le O(\eta^2).$$

To treat the term on the next to the last line of (2.51), note that

$$\begin{aligned}
&E\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\\
&\le KE\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(k\eta),\widetilde{\alpha}(k\eta),\gamma)|^2 ds\pi(d\gamma)\\
&+KE\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\widetilde{\alpha}(k\eta),\gamma)-g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma).
\end{aligned}\tag{2.53}$$

For the term on the second line of (2.53) and $k = 0, 1, \cdots, \lfloor \frac{T}{\eta} \rfloor - 1$,

$$\begin{aligned}
&E \int_{k\eta}^{k\eta+\eta} \int_{\Gamma} |g(\widetilde{X}(k\eta), \widetilde{\alpha}(s^-), \gamma) - g(\widetilde{X}(k\eta), \widetilde{\alpha}(k\eta), \gamma)|^2 ds \pi(d\gamma) \\
&\quad = E \int_{k\eta}^{k\eta+\eta} \int_{\Gamma} |g(\widetilde{X}(k\eta), \widetilde{\alpha}(s^-), \gamma) - g(\widetilde{X}(k\eta), \widetilde{\alpha}(k\eta), \gamma)|^2 I_{\{\widetilde{\alpha}(s^-) \neq \widetilde{\alpha}(k\eta)\}} ds \pi(d\gamma) \\
&\quad = E \sum_{i \in \mathcal{M}} \sum_{j \neq i} \int_{k\eta}^{k\eta+\eta} \int_{\Gamma} |g(\widetilde{X}(k\eta), j, \gamma) - g(\widetilde{X}(k\eta), i, \gamma)|^2 I_{\{\widetilde{\alpha}(s^-) = \widetilde{\alpha}(s) = j\}} I_{\{\widetilde{\alpha}(k\eta) = i\}} ds \pi(d\gamma) \\
&\quad \leq KE \sum_{i \in \mathcal{M}} \sum_{j \neq i} \int_{k\eta}^{k\eta+\eta} [1 + |\widetilde{X}(k\eta)|^2] I_{\{\widetilde{\alpha}(k\eta) = i\}} \times E[I_{\{\widetilde{\alpha}(s) = j\}} | \widetilde{X}(k\eta), \widetilde{\alpha}(k\eta) = i] ds \\
&\quad \leq KE \sum_{i \in \mathcal{M}} \int_{k\eta}^{k\eta+\eta} [1 + |\widetilde{X}(k\eta)|^2] I_{\{\widetilde{\alpha}(k\eta) = i\}} \times [\sum_{j \neq i} q_{ij}(\widetilde{X}(k\eta))(s - k\eta) + o(s - k\eta)] ds \\
&\quad \leq K \int_{k\eta}^{k\eta+\eta} O(\eta) ds \leq K\eta^2.
\end{aligned} \tag{2.54}$$

In the above, we employed the fact that the time of jump of $X(t)$ does not coincide with that of switching part $\alpha(t)$ in [47, Proposition 2.2,]. Also, Lemma 2.36 and boundedness of $Q(x)$ are involved. Now let us deal with the last line of (2.53) by using the basic coupling techniques [24, p. 11]. Consider the measure

$$\Lambda((x, j), (\widetilde{x}, i)) = |x - \widetilde{x}| + d(j, i), \quad \text{where} \quad d(j, i) = \begin{cases} 0 & \text{if } j = i, \\ 1 & \text{if } j \neq i. \end{cases}$$

Let $(\alpha(t), \widetilde{\alpha}(t))$ be a random process with a finite state space $\mathcal{M} \times \mathcal{M}$ such that

$$\begin{aligned}
&P[(\alpha(t+h), \widetilde{\alpha}(t+h)) = (j, i) | (\alpha(t), \widetilde{\alpha}(t)) = (k, l), (X(t), \widetilde{X}(t)) = (x, \widetilde{x})] \\
&= \begin{cases} \widetilde{q}_{(k,l)(j,i)}(x, \widetilde{x}) h + o(h), & \text{if } (k, l) \neq (j, i), \\ 1 + \widetilde{q}_{(k,l)(k,l)}(x, \widetilde{x}) h + o(h), & \text{if } (k, l) = (j, i), \end{cases}
\end{aligned}$$

where $h \to 0$, and the matrix $(\widetilde{q}_{(k,l)(j,i)}(x, \widetilde{x}))$ is the basic coupling of matrices $Q(x) = (q_{kl}(x))$

and $Q(\widetilde{x}) = (q_{kl}(\widetilde{x}))$ satisfying

$$\begin{aligned}\widetilde{Q}(x,\widetilde{x})\widetilde{f}(k,l) &= \sum_{(j,i)\in\mathcal{M}\times\mathcal{M}} \widetilde{q}_{(k,l)(j,i)}(x,\widetilde{x})(\widetilde{f}(j,i)-\widetilde{f}(k,l))\\ &= \sum_j (q_{kj}(x)-q_{lj}(\widetilde{x}))^+(\widetilde{f}(j,l)-\widetilde{f}(k,l))\\ &\quad+\sum_j (q_{lj}(\widetilde{x})-q_{kj}(x))^+(\widetilde{f}(k,j)-\widetilde{f}(k,l))\\ &\quad+\sum_j (q_{kj}(x)\wedge q_{lj}(\widetilde{x}))(\widetilde{f}(j,j)-\widetilde{f}(k,l))\end{aligned} \tag{2.55}$$

for any function $\widetilde{f}(\cdot,\cdot)$ defined on $\mathcal{M}\times\mathcal{M}$. Then we have

$$\begin{aligned}&E[I_{\{\alpha(s)=j\}}|\alpha(k\eta)=i_1,\widetilde{\alpha}(k\eta)=i,X(k\eta)=x,\widetilde{X}(k\eta)=\widetilde{x}]\\ &=\sum_{l\in\mathcal{M}} E[I_{\{\alpha(s)=j\}}I_{\{\widetilde{\alpha}(s)=l\}}|\alpha(k\eta)=i_1,\widetilde{\alpha}(k\eta)=i,X(k\eta)=x,\widetilde{X}(k\eta)=\widetilde{x}]\\ &=\sum_{l\in\mathcal{M}} \widetilde{q}_{(i_1,i)(j,l)}(x,\widetilde{x})(s-k\eta)+o(s-k\eta)=O(\eta).\end{aligned} \tag{2.56}$$

Therefore, for $k=1,\cdots,\lfloor\frac{T}{\eta}\rfloor-1$, we have

$$\begin{aligned}&E\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),\widetilde{\alpha}(k\eta),\gamma)-g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)|^2 ds\pi(d\gamma)\\ &=E\sum_{i\in\mathcal{M}}\sum_{j\neq i}\int_{k\eta}^{k\eta+\eta}\int_\Gamma |g(\widetilde{X}(k\eta),i,\gamma)-g(\widetilde{X}(k\eta),j,\gamma)|^2 I_{\{\alpha(s)=\alpha(s^-)=j\}}I_{\{\widetilde{\alpha}(k\eta)=i\}}ds\pi(d\gamma)\\ &\le KE\sum_{i,i_1\in\mathcal{M}}\sum_{j\neq i}\int_{k\eta}^{k\eta+\eta}[1+|\widetilde{X}(k\eta)|^2]I_{\{\widetilde{\alpha}(k\eta)=i,\alpha(k\eta)=i_1\}}\\ &\quad\times E[I_{\{\alpha(s)=j\}}|\alpha(k\eta)=i_1,\widetilde{\alpha}(k\eta)=i,X(k\eta)=x,\widetilde{X}(k\eta)=\widetilde{x}]ds=O(\eta^2).\end{aligned} \tag{2.57}$$

For $k=0$, recall that $\alpha(0)=\widetilde{\alpha}(0)=\alpha$, $X(0)=x$ and $\widetilde{X}(0)=\widetilde{x}$, we have

$$\begin{aligned}
&E\int_0^\eta\int_\Gamma |g(\widetilde{X}(0),\widetilde{\alpha}(0),\gamma)-g(\widetilde{X}(0),\alpha(s),\gamma)|^2 ds\pi(d\gamma)\\
&\quad = E\int_0^\eta\int_\Gamma\sum_{j\neq\alpha}|g(\widetilde{x},\alpha,\gamma)-g(\widetilde{x},j,\gamma)|^2 I_{\{\alpha(s)=j\}}ds\pi(d\gamma)\\
&\quad \le K\sum_{j\neq\alpha}\int_0^\eta[1+\tilde{x}^2]E[I_{\{\alpha(s)=j\}}|\alpha(0)=\alpha,\widetilde{X}(0)=\widetilde{x}]ds\\
&\quad \le K\int_0^\eta\sum_{j\neq\alpha}[q_{\alpha j}(\widetilde{x})s+o(s)]ds\le K\eta^2.
\end{aligned}\tag{2.58}$$

Thus, for $k=0,1,\cdots,\lfloor\frac{T}{\eta}\rfloor-1$,

$$E\int_{k\eta}^{k\eta+\eta}\int_\Gamma|g(\widetilde{X}(k\eta),\widetilde{\alpha}(k\eta),\gamma)-g(\widetilde{X}(k\eta),\alpha(s^-),\gamma)|^2ds\pi(d\gamma)\le K\eta^2.\tag{2.59}$$

Now we can obtain

$$E\int_0^T\int_\Gamma|g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)|^2ds\pi(d\gamma)\le\sum_{k=0}^{\lfloor\frac{T}{\eta}\rfloor-1}K\eta^2\le K\eta.$$

Likewise, we also obtain the bound for the martingale part

$$E\sup_{0\le t\le T}|\int_0^t\int_\Gamma[g(\widetilde{X}(s^-),\widetilde{\alpha}(s^-),\gamma)-g(\widetilde{X}(s^-),\alpha(s^-),\gamma)]\widetilde{N}(ds,d\gamma)|^2\le K\eta.$$

For the drift and diffusion parts involved, the argument in [49, Lemma 4.3] leads to

$$\begin{aligned}
&E\int_0^T|(b(\widetilde{X}(s),\widetilde{\alpha}(s))-b(\widetilde{X}(s),\alpha(s))|^2ds\le K\eta,\\
&E\sup_{0\le t\le T}|\int_0^t[\sigma(\widetilde{X}(s),\widetilde{\alpha}(s))-\sigma(\widetilde{X}(s),\alpha(s))]dw(s)|^2\le K\eta.
\end{aligned}$$

Therefore, we obtain

$$E \sup_{0\le t\le T} |\phi^{\Delta}(t)|^2 \le K \frac{\eta}{\Delta^2} = K\Delta^{\gamma_0 - 2} \to 0 \text{ as } \Delta \to 0. \tag{2.60}$$

This concludes the proof.

Lemma 2.38. *Under the conditions of Theorem* 2.37, $E[\sup_{0\le t\le T} |\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t)|^2] \le C|\widetilde{x} - x|^2$, *where the constant* C *satisfies* $C = C(K_0, T)$.

Proof. Let $T > 0$ be fixed and recall that $\Delta = \widetilde{x} - x$, then we have $\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t) = \Delta + A(t) + B(t)$, where

$$\begin{aligned} A(t) &= \int_0^t [b(\widetilde{X}(s), \widetilde{\alpha}(s)) - b(\widetilde{X}(s), \alpha(s))]ds \\ &\quad + \int_0^t [\sigma(\widetilde{X}(s), \widetilde{\alpha}(s)) - \sigma(\widetilde{X}(s), \alpha(s))]dw(s) \\ &\quad + \int_0^t \int_\Gamma [g(\widetilde{X}(s^-), \widetilde{\alpha}(s^-), \gamma) - g(\widetilde{X}(s^-), \alpha(s^-), \gamma)]N(ds, d\gamma) \\ &= \Delta\phi^{\Delta}(t), \end{aligned} \tag{2.61}$$

$$\begin{aligned} B(t) &= \int_0^t [b(\widetilde{X}(s), \alpha(s)) - b(X(s), \alpha(s))]ds \\ &\quad + \int_0^t [\sigma(\widetilde{X}(s), \alpha(s)) - \sigma(X(s), \alpha(s))]dw(s) \\ &\quad + \int_0^t \int_\Gamma [g(\widetilde{X}(s^-), \alpha(s^-), \gamma) - g(X(s^-), \alpha(s^-), \gamma)]N(ds, d\gamma). \end{aligned} \tag{2.62}$$

Hence

$$\sup_{t\in[0,T]} |\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t)|^2 \le 3\Delta^2 + 3\sup_{t\in[0,T]} |A(t)|^2 + 3\sup_{t\in[0,T]} |B(t)|^2.$$

It follows from (2.60) that

$$E[\sup_{t\in[0,T]} |A(t)|^2] \le \Delta^2 E[\sup_{t\in[0,T]} |\phi^\Delta(t)|^2] \le K\Delta^{r_0} = o(\Delta^2).$$

By the Hölder inequality and the Lipschitz continuity, we have

$$E[\sup_{t\in[0,T]} |\int_0^t [b(\widetilde{X}(s),\alpha(s)) - b(X(s),\alpha(s))]ds|^2] \le K\int_0^T E|\widetilde{X}(s) - X(s)|^2 ds$$

and

$$\begin{aligned}&E[\sup_{t\in[0,T]} |\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\alpha(s^-),\gamma) - g(X(s^-),\alpha(s^-),\gamma)]ds\pi(d\gamma)|^2]\\ &\le K\int_0^T E|\widetilde{X}(s^-) - X(s^-)|^2 ds.\end{aligned}$$

Then the basic properties of stochastic integrals (w.r.t. $w(\cdot)$ and $\widetilde{N}(\cdot)$) together with the Lipschitz continuity lead to

$$E[\sup_{t\in[0,T]} |\int_0^t [\sigma(\widetilde{X}(s),\alpha(s)) - \sigma(X(s),\alpha(s))]dw(s)|^2] \le K\int_0^T E|\widetilde{X}(s) - X(s)|^2 ds$$

and

$$\begin{aligned}&E[\sup_{t\in[0,T]} |\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\alpha(s^-),\gamma) - g(X(s^-),\alpha(s^-),\gamma)]\widetilde{N}(ds,d\gamma)|^2]\\ &\le K\int_0^T E|\widetilde{X}(s^-) - X(s^-)|^2 ds.\end{aligned}$$

So,

$$E[\sup_{t\in[0,T]} |\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t)|^2] \le 3\Delta^2 + K\int_0^T E[\sup_{u\in[0,T]} |\widetilde{X}(u) - X(u)|^2]du + o(\Delta^2). \tag{2.63}$$

Now, by Gronwall's inequality

$$E[\sup_{t\in[0,T]} |\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t)|^2] \leq 3\Delta^2 \exp(KT) + o(\Delta^2) \leq K|\widetilde{x} - x|^2.$$

Thus, we have completed the proof.

Let us introduce some notations to proceed. Recall that a vector $\beta = (\beta_1, \beta_2, \cdots, \beta_r)$ with nonnegative integer component is referred to as a multi-index. Put $|\beta| = \beta_1 + \beta_2 + \cdots + \beta_r$, we define D_x^β as

$$D_x^\beta = \frac{\partial^\beta}{\partial x^\beta} = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial_{x_r}^{\beta_r}}.$$

Recall that $\Delta = \widetilde{x} - x$ and define

$$Z^\Delta(t) = \frac{\widetilde{X}^{\widetilde{x},\alpha}(t) - X^{x,\alpha}(t)}{\Delta}. \tag{2.64}$$

Then we have the following expression:

$$\begin{aligned} Z^\Delta(t) = 1 + \phi^\Delta(t) &+ \frac{1}{\Delta}\int_0^t [b(\widetilde{X}(s), \alpha(s)) - b(X(s), \alpha(s))]ds \\ &+ \frac{1}{\Delta}\int_0^t [\sigma(\widetilde{X}(s), \alpha(s)) - \sigma(X(s), \alpha(s))]dw(s) \\ &+ \frac{1}{\Delta}\int_0^t \int_\Gamma [g(\widetilde{X}(s^-), \alpha(s^-), \gamma) - g(X(s^-), \alpha(s^-), \gamma)]N(ds, d\gamma), \end{aligned} \tag{2.65}$$

where $\phi^\Delta(t)$ is defined in (2.49).

Lemma 2.39. *Under the conditions of Theorem* 2.38*, assume that for each $i \in \mathcal{M}$, $b(\cdot, i)$, $\sigma(\cdot, i)$ and $g(, i, \gamma)$ have continuously partial derivatives with respect to the variable x up to* *the second order and that*

$$|D_x^\beta b(x,i)| + |D_x^\beta \sigma(x,i)| + |D_x^\beta g(x,i,\gamma)| \le K(1+|x|^\rho),$$

where K and ρ are positive constants and β is a multi-index with $|\beta| \le 2$. Then $X^{x,\alpha}(t)$ is twice continuously differentiable in mean square with respect to x.

Proof. Given the definition of $Z^\Delta(t)$ above and Theorem 2.37, we just need to consider the last three terms of (2.65). First, note that

$$\begin{aligned}
&\frac{1}{\Delta}\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\alpha(s^-),\gamma) - g(X(s^-),\alpha(s^-),\gamma)] ds\pi(d\gamma) \\
&\quad = \frac{1}{\Delta}\int_\Gamma \int_0^t \int_0^1 \frac{d}{d\nu} g(X(s^-) + \nu(\widetilde{X}(s^-) - X(s^-)),\alpha(s^-),\gamma) d\nu ds\pi(d\gamma) \\
&\quad = \int_\Gamma \int_0^t [\int_0^1 g_x(X(s^-) + \nu(\widetilde{X}(s^-) - X(s^-)),\alpha(s^-),\gamma) d\nu] Z^\Delta(s^-) ds\pi(d\gamma),
\end{aligned} \tag{2.66}$$

where $g_x(\cdot)$ denotes the partial derivative of $g(\cdot,i,\gamma)$ with respect to x. It follows from Lemma 2.38 that for any $s \in [0,T]$, $\widetilde{X}(s^-) - X(s^-) \to 0$ in probability as $\Delta \to 0$. This implies that

$$\int_0^1 g_x(X(s^-) + \nu(\widetilde{X}(s^-) - X(s^-)),\alpha(s^-),\gamma) d\nu \to g_x(X(s^-),\alpha(s^-),\gamma) \tag{2.67}$$

in probability as $\Delta \to 0$. Therefore, we have

$$\begin{aligned}
&\frac{1}{\Delta}\int_0^t \int_\Gamma [g(\widetilde{X}(s^-),\alpha(s^-),\gamma) - g(X(s^-),\alpha(s^-),\gamma)] ds\pi(d\gamma) \\
&\to \int_0^t \int_\Gamma g_x(X(s^-),\alpha(s^-),\gamma) Z^\Delta(s^-) ds\pi(d\gamma).
\end{aligned} \tag{2.68}$$

Similarly, we have

$$\frac{1}{\Delta}\int_0^t [b(\widetilde{X}(s),\alpha(s)) - b(X(s),\alpha(s))]ds \to \int_0^t b_x(X(s),\alpha(s))Z^{\Delta}(s)ds \tag{2.69}$$

in probability as $\Delta \to 0$ and

$$\frac{1}{\Delta}\int_0^t [\sigma(\widetilde{X}(s),\alpha(s)) - \sigma(X(s),\alpha(s))]dw(s) \to \int_0^t \sigma_x(X(s),\alpha(s))Z^{\Delta}(s)dw(s) \tag{2.70}$$

in probability as $\Delta \to 0$. $b_x(\cdot)$ and $\sigma_x(\cdot)$ denote the partial derivative of $b(\cdot,i)$ and $\sigma(\cdot,i)$ with respect to x, respectively. Recall the definition of $Z^{\Delta}(t)$ in equation (2.64), Theorem 2.37, (2.67)-(2.70), and [26, Theorem 5.5.2] yield

$$E|Z^{\Delta}(t) - \varsigma(t)|^2 \to 0 \text{ as } \Delta \to 0. \tag{2.71}$$

where

$$\begin{aligned}\varsigma(t) = 1 &+ \int_0^t b_x(X(s),\alpha(s))\varsigma(s)ds + \int_0^t \sigma_x(X(s),\alpha(s))\varsigma(s)dw(s)\\ &+ \int_0^t\int_{\Gamma} g_x(X(s^-),\alpha(s^-),\gamma)\varsigma(s^-)N(ds,d\gamma)\end{aligned} \tag{2.72}$$

and $\varsigma(t) = \varsigma^{x,\alpha}(t)$ is mean square continuous with respect to x. Therefore, $\frac{\partial}{\partial x}X^{x,\alpha}(t)$ exists in the mean square sense and $\varsigma(t) = \frac{\partial}{\partial x}X^{x,\alpha}(t)$. Likewise, we can show $\frac{\partial^2}{\partial x^2}X^{x,\alpha}(t)$ exists in the mean square sense and is mean square continuous with respect to x.

Lemma 2.40. *Under the assumptions of Lemma* 2.39*, we have* $\sup_{t\in[0,T]} E|\varsigma(t)|^2 \le K = K(x,\widetilde{x},T,K_0) < \infty$.

Proof. For any $t \in [0,T]$, $E|\varsigma(t)|^2 \le 2E|\varsigma(t) - Z^{\Delta}(t)|^2 + 2E|Z^{\Delta}(t)|^2$. By (2.71), it suffices to consider the last term above. In fact,

$$\begin{aligned}E|Z^{\Delta}(t)|^2 \leq K + 5E|\phi^{\Delta}(t)|^2 + 5E|\frac{1}{\Delta}\int_0^t [b(\widetilde{X}(u),\alpha(u)) - b(X(u),\alpha(u))]du|^2\\ +5E|\frac{1}{\Delta}\int_0^t [\sigma(\widetilde{X}(u),\alpha(u)) - \sigma(X(u),\alpha(u))]dw(u)|^2\\ +5E|\frac{1}{\Delta}\int_0^t \int_\Gamma [g(\widetilde{X}(u^-),\alpha(u^-),\gamma) - g(X(u^-),\alpha(u^-),\gamma)]N(du,d\gamma)|^2,\end{aligned}$$

so

$$\begin{aligned}E|Z^{\Delta}(t)|^2 &\leq K + 5t\frac{1}{|\Delta|^2}E\int_0^t |b(\widetilde{X}(u),\alpha(u)) - b(X(u),\alpha(u))|^2 du\\ &+5\frac{1}{|\Delta|^2}E\int_0^t |\sigma(\widetilde{X}(u),\alpha(u)) - \sigma(X(u),\alpha(u))|^2 du\\ &+5t\frac{1}{|\Delta|^2}E\int_0^t \int_\Gamma |g(\widetilde{X}(u^-),\alpha(u^-),\gamma) - g(X(u^-),\alpha(u^-),\gamma)|^2 du\pi(d\gamma)+\\ &+5\frac{1}{|\Delta|^2}E\int_0^t \int_\Gamma |g(\widetilde{X}(u^-),\alpha(u^-),\gamma) - g(X(u^-),\alpha(u^-),\gamma)|^2 du\pi(d\gamma)\\ &\leq K + 5K_0(T+1)\frac{1}{|\Delta|^2}E\int_0^t |\widetilde{X}(u) - X(u)|^2 du\\ &+5K_0(T+1)\frac{1}{|\Delta|^2}E\int_0^t |\widetilde{X}(u^-) - X(u^-)|^2 du \leq K = K(x,\widetilde{x},T,K_0).\end{aligned} \tag{2.73}$$

Hence the proof is completed.

Lemma 2.41. *Assume the conditions of Lemma* 2.40 *hold. Then the function* $E|X^{x,\alpha}(t)|^p$ *is twice continuously differentiable with respect to the variable* x*, except possibly at* $x = 0$.

Proof. In what follows, let $u(t,x,\alpha) = E[\phi(X(t),\alpha(t))] = E|X^{x,\alpha}(t)|^p$, then

$$\begin{aligned}\frac{u(t,\tilde{x},\alpha) - u(t,x,\alpha)}{\Delta} &= \frac{1}{\Delta}E[|\widetilde{X}(t)|^p - |X(t)|^p]\\ &= \frac{1}{\Delta}E\int_0^1 \frac{d}{dv}|X(t) + v(\widetilde{X}(t) - X(t))|^p dv\\ &= E[Z^{\Delta}(t)\int_0^1 |X(t) + v(\widetilde{X}(t) - X(t))|_x^p dv],\end{aligned}$$

where $|\cdot|^p_x$ denotes the partial derivative of $\phi(\cdot,i)=|\cdot|^p$ with respect to x. Consider

$$\begin{aligned}
&|\frac{1}{\Delta}E[|\widetilde{X}(t)|^p-|X(t)|^p]-E[|X(t)|^p_x\varsigma(t)]| \\
&\leq |E\int_0^1[|X(t)+v(\widetilde{X}(t)-X(t))|^p_x dvZ^\Delta(t)]-E|X(t)|^p_x\varsigma(t)| \\
&\leq E\int_0^1|\left[|X(t)+v(\widetilde{X}(t)-X(t))|^p_x dv-|X(t)|^p_x\right]Z^\Delta(t)|+E||X(t)|^p_x[Z^\Delta(t)-\varsigma(t)]|.
\end{aligned} \tag{2.74}$$

For the second part of last line of (2.74), by Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
E\left||X(t)|^p_x[Z^\Delta(t)-\varsigma(t)]\right| &\leq E^{\frac{1}{2}}|X(t)|^{2p}_x E^{\frac{1}{2}}[Z^\Delta(t)-\varsigma(t)]^2 \\
&\leq KE^{\frac{1}{2}}[Z^\Delta(t)-\varsigma(t)]^2\to 0 \text{ as } \Delta\to 0.
\end{aligned}$$

Here we used Lemma 2.36 and (2.71). Similarly, we can show the first term of last line of (2.74) goes to 0 as $\Delta\to 0$. Thus $E|X^{x,\alpha}(t)|^p$ is differentiable with respect to the variable x. Likewise, we can also see it is twice continuously differentiable with respect to the variable x. As a nice application of the smooth dependence on the initial data, we obtain a Lyapunov converse theorem, namely, necessary conditions for exponential p stability.

Theorem 2.42. *Assume that the conditions of Lemma* 2.41 *hold and that the equilibrium point* 0 *is exponentially p-stable. Then for each $i\in\mathcal{M}$, there exists a function $V(\cdot,i)\in C^2(\mathbb{R}^r:\mathbb{R}_+)$ such that*

$$\begin{aligned}
&k_1|x|^p\leq V(x,i)\leq k_2|x|^p\ x\in D, \\
&\mathcal{G}V(x,i)\leq -k_3|x|^p\ \ \textit{for all}\ \ x\in D-\{0\}, \\
&\left|\frac{\partial V}{\partial x_j}(x,i)\right|\leq k_4|x|^{p-1}, \\
&\left|\frac{\partial^2 V}{\partial x_j\partial x_l}(x,i)\right|\leq k|x|^{p-2}.
\end{aligned}$$

for all $1 \le j, l \le r, x \in D - \{0\}$*, and for some positive constants* k, k_1, k_2, k_3 *and* k_4*, where* D *is a neighborhood of* 0.

Proof. For each $i \in \mathcal{M}$, consider the function

$$V(x,i) = \int_0^T E|X^{x,i}(u)|^p du.$$

It follows from Lemma 2.41, $V(x,i)$ is twice continuously differentiable with respect to x except possibly at 0. The equilibrium point 0 is exponential p-stable, therefore there is a $\kappa > 0$ such that

$$V(x,i) = \int_0^T E|X^{x,i}(u)|^p du \le \int_0^T K|x|^p e^{-\kappa u} du \le k_2|x|^p.$$

For the function $|x|^p$, we have $|\mathcal{G}|x|^p| \le K|x|^p$ for some positive real number K. An application of generalized Itô's formula leads to

$$E|X^{x,i}(T)|^p - |x|^p = E\int_0^T \mathcal{G}|X^{x,i}(u)|^p du \ge -KE\int_0^T |X^{x,i}(u)|^p du = -KV(x,i).$$

Again recall that equilibrium point $x = 0$ is exponential p-stable, we can choose T such that $E|X^{x,i}(T)|^p \le \frac{1}{2}|x|^p$, and therefore, we have $V(x,i) \ge \frac{|x|^p}{2K} = k_1|x|^p$. Notice that

$$\mathcal{G}V(x,i) = \int_0^T \mathcal{G}E|X^{x,i}(u)|^p du.$$

Let $u(t,x,i) = E|X^{x,i}(t)|^p$, by the similar argument in step 1 and step 2 of [20, Theorem

7.10], we obtain

$$\begin{aligned}\mathcal{G}V(x,i) &= \int_0^T \mathcal{G}E|X^{x,i}(u)|^p du = u(T,x,i) - u(0,x,i) \\ &= E|X^{x,i}(T)|^p - E|X^{x,i}(0)|^p = E|X^{x,i}(T)|^p - |x|^p \\ &\le -\frac{1}{2}|x|^p = -k_3|x|^p.\end{aligned}$$

Note that

$$\frac{\partial E|X^{x,i}(t)|^p}{\partial x_j} = pE|X^{x,i}(t)|^{p-1}\mathrm{sgn}(X^{x,i}(t))\frac{\partial X^{x,i}(t)}{\partial x_j},$$

so

$$\begin{aligned}\left|\frac{\partial E|X^{x,i}(t)|^p}{\partial x_j}\right| &= pE\left(|X^{x,i}(t)|^{p-1}\left|\frac{\partial X^{x,i}(t)}{\partial x_j}\right|\right) \\ &\le pE^{\frac{1}{2}}|X^{x,i}(t)|^{2p-2}E^{\frac{1}{2}}\left|\frac{\partial X^{x,i}(t)}{\partial x_j}\right|^2 \\ &\le K(|x|^{2p-2}e^{-\kappa t})^{\frac{1}{2}} = K|x|^{p-1}e^{-\kappa t/2}.\end{aligned}$$

For the last line above, we used the Lemma 2.36 and Lemma 2.40. Consequently, we have

$$\left|\frac{\partial V(x,i)}{\partial x_j}\right| = \left|\int_0^T \frac{\partial}{\partial x_j}E|X^{x,i}(u)|^p du\right| \le \int_0^T K|x|^{p-1}e^{-\kappa u/2}du \le k_4|x|^{p-1}.$$

We can have estimate of the second derivative of $V(x,i)$ by similar argument, the theorem is thus proved. □

For practical systems, frequently, we do not have information regarding the equilibria of the systems. Nevertheless, the systems still possesses certain kind of stability properties. Thus it is necessary to extend our definition to consider the so-called the asymptotic stability in distribution. To proceed, let us first give two definitions.

Definition 2.43. The dynamic system is asymptotically stable in distribution if, there exists such a probability measure $\nu(\cdot\times\cdot)$ on $\mathbb{R}^r\times\mathcal{M}$ that the transition probability $p(t,x,\alpha,dy\times\{i\})$

of $(X(t), \alpha(t))$ converges weakly to $\nu(dy \times \{i\})$ as $t \to \infty$ for every $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$.

Definition 2.44. The definitions of (P1) and (P2) are as follows.

- The switching jump diffusion process given by (2.41) and (2.42) is said to have property (**P1**) if, for any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$ and any $\varepsilon > 0$, there exists a constant $R > 0$ such that
$$P\{|X^{x,\alpha}(t)| \geq R\} < \varepsilon, \text{ for any } t \geq 0.$$
- The switching jump diffusion process given by (2.41) and (2.42) is said to have property (**P2**) if, for any $\varepsilon > 0$ and any compact subset $\widehat{\mathsf{C}}$ of $\mathbb{R}^r$, there exists a $T = T(\varepsilon, \widehat{\mathsf{C}}) > 0$ such that
$$P(|X^{x_0,i_0}(t) - X^{y_0,i_0}(t)| \leq \varepsilon) \to 1 \text{ as } t \to \infty,$$
whenever $(x_0, y_0, i_0) \in \widehat{\mathsf{C}} \times \widehat{\mathsf{C}} \times \mathcal{M}$.

In this section, we first establish asymptotic stability in distribution of the process $(X(t), \alpha(t))$ in which $\alpha(t)$ is a Markov chain that is independent of the Brownian motion, which is referred as Markov switching jump diffusions. Then we further extend the results to state-dependent switching process.

Proposition 2.45. *Suppose that* (A2) *is satisfied, that $b(\cdot, i)$, $\sigma(\cdot, i)$, and $g(\cdot, i, \gamma)$ grow at most linearly for each $i \in \mathcal{M}$ and $\gamma \in \Gamma$, that conditions* (P1) *and* (P2) *hold, and that the generator of the Markov chain Q is irreducible. Then the switching jump diffusion process $(X(t), \alpha(t))$ is stable in distribution.*

Proof. We note that [50, Theorem 3.1] in fact works not only for Markov switching diffusion processes but also for more general Markov processes. In our current setup, $(X(t), \alpha(t))$ is a Markov process. So we can use essentially the same steps as in the aforementioned reference to show the process is stable in distribution. The verbatim argument is omitted.

Our next task is to find sufficient conditions that ensure conditions (P1) and (P2) are in force. The result is stated in the next theorem.

Theorem 2.46. *Assume that for each $i \in \mathcal{M}$, there exists function $V(\cdot, i) \in C^2(\mathbb{R}^r : \mathbb{R}_+)$ satisfying the following two conditions: There exists a positive real number β such that*

$$\mathcal{G}V(x,i) \le -\beta V(x,i), \tag{2.75}$$

$$V_R := \inf_{\substack{|x|\ge R \\ i\in\mathcal{M}}} V(x,i) \to \infty \text{ as } R \to \infty. \tag{2.76}$$

Then (P1) *and* (P2) *hold.*

Proof. Let us first verify (P1). Define the stopping time

$$\tau_R := \inf\{t \ge 0 : |X(t)| \ge R\}.$$

Consider $V(x,i)e^{\beta t}$ and let $t_R = \tau_R \wedge t$. By virtue of Dynkin's formula, we have

$$\begin{aligned} E_{x,\alpha}[V(X(t_R), \alpha(t_R))e^{\beta t_R}] - V(x,\alpha) &= E_{x,\alpha}\int_0^{t_R} e^{\beta s}\mathcal{G}V(X(s),\alpha(s))ds \\ &\quad +\beta E_{x,\alpha}\int_0^{t_R} e^{\beta s}V(X(s),\alpha(s))ds, \end{aligned}$$

where $E_{x,\alpha}$ denotes the expectation with $X(0) = x$ and $\alpha(0) = \alpha$.

Hence, by virtue of (2.81), $E_{x,\alpha} V(X(t_R), \alpha(t_R)) \le V(x,\alpha)e^{-\beta t_R}$. We further have

$$V_R P\{\tau_R \le t\} \le E_{x,\alpha}[V(X(\tau_R), \alpha(\tau_R)) I_{\{\tau_R \le t\}}] \le V(x,\alpha)e^{-\beta \tau_R}.$$

Note that $\tau_R \le t$ if and only if $\sup\limits_{0\le u\le t} |X(u)| \ge R$. Therefore, it follows that

$$P\{\sup_{0\le u\le t} |X^{x,\alpha}(u)| \ge R\} \le \frac{V(x,\alpha)e^{-\beta\tau_R}}{V_R} \le \frac{V(x,\alpha)}{V_R}.$$

Then upon using (2.82), $P\{|X^{x,\alpha}(t)| \ge R\} \to 0$ as $R \to \infty$, for all $t \ge 0$. To guarantee (P2) hold, similar technique is involved here. But now we need to consider the difference between two solutions of equation (2.41) starting from different initial values in compact set $\widehat{\mathsf{C}}$. Namely, (x,α) and (y,α).

$$\begin{aligned}
&X^{x,\alpha}(t) - X^{y,\alpha}(t)\\
&\quad = x - y + \int_0^t [b(X^{x,\alpha}(s), \alpha(s)) - b(X^{y,\alpha}(s), \alpha(s))]ds\\
&\quad\quad + \int_0^t [\sigma(X^{x,\alpha}(s), \alpha(s)) - \sigma(X^{y,\alpha}(s), \alpha(s))]dw(s)\\
&\quad\quad + \int_0^t \int_\Gamma [g(X^{x,\alpha}(s^-), \alpha(s^-), \gamma) - g(X^{y,\alpha}(s^-), \alpha(s^-), \gamma)]N(ds, d\gamma).
\end{aligned}$$

Let $Z^{x,y,\alpha}(t) = X^{x,\alpha}(t) - X^{y,\alpha}(t)$, so $Z(0) = z = x - y$. Then

$$\begin{aligned}
dZ^{x,y,\alpha}(t) = {} & [b(X^{x,\alpha}(t), \alpha(t)) - b(X^{y,\alpha}(t), \alpha(t))]dt\\
& + [\sigma(X^{x,\alpha}(t), \alpha(t)) - \sigma(X^{y,\alpha}(t), \alpha(t))]dw(t)\\
& + \int_\Gamma [g(X^{x,\alpha}(t^-), \alpha(t^-), \gamma) - g(X^{y,\alpha}(t^-), \alpha(t^-), \gamma)]N(dt, d\gamma).
\end{aligned}$$

Define a stopping time $\tau_\varepsilon := \inf\{t \ge 0, |X^{x,\alpha}(t) - X^{y,\alpha}(t)| \ge \varepsilon\}$ and let $t_\varepsilon = \tau_\varepsilon \wedge t$. Then we have

$$
\begin{aligned}
E_{z,\alpha}V(Z(t_\varepsilon),\alpha(t_\varepsilon)) - V(z,\alpha) &= E_{z,\alpha}\int_0^{t_\varepsilon} \mathcal{G}V(Z(s),\alpha(s))ds \\
&\le -\beta \int_0^{t_\varepsilon} E_{z,\alpha}V(Z(s),\alpha(s))ds.
\end{aligned}
$$

Given $s \le \tau_\varepsilon \wedge t$, we have $s \wedge \tau_\varepsilon = s$. As a result,

$$
E_{z,\alpha}V(Z(t\wedge\tau_\varepsilon),\alpha(t\wedge\tau_\varepsilon)) - V(z,\alpha) \le -\beta\int_0^t E_{z,\alpha}V(Z(s\wedge\tau_\varepsilon),\alpha(s\wedge\tau_\varepsilon))ds.
$$

By applying Gronwall's inequality, we obtain

$$
E_{z,\alpha}V(Z(\tau_\varepsilon\wedge t),\alpha(\tau_\varepsilon\wedge t)) \le V(z,\alpha)e^{-\beta t}.
$$

Hence,

$$
V_\varepsilon P(\tau_\varepsilon \le t) \le E_{z,\alpha}[V(Z(\tau_\varepsilon),\alpha(\tau_\varepsilon))I_{\{\tau_\varepsilon\le t\}}] \le V(z,\alpha)e^{-\beta t},
$$

in which $V_\varepsilon = \inf\{V(z,i), z\in\mathbb{R}^r\backslash B_\varepsilon, i\in\mathcal{M}\}$ and $B_\varepsilon = \{z\in\widehat{\mathsf{C}}, |z|<\varepsilon\}$, so $V_\varepsilon > 0$. Note that $\tau_\varepsilon \le t$ if and only if $\sup\limits_{0\le u\le t}|Z(u)| \ge \varepsilon$. Therefore, it follows that $P\{\sup\limits_{0\le u\le t}|Z(u)|\ge\varepsilon\} \le \frac{V(z,\alpha)e^{-\beta t}}{V_\varepsilon}$, so $P(|Z(t)|\ge\varepsilon)\to 0$ as $t\to\infty$. That is, $P(|X^{x,\alpha}(t)-X^{y,\alpha}(t)|\le\varepsilon)\to 1$ as $t\to\infty$. Thus, the proof is concluded.

Now, let us consider the case when the generator of the discrete component $\alpha(t)$ is x dependent. In this case, the switching part is no longer a Markov chain. Because of the interplays between $\alpha(t)$ and $X(t)$, we need more complex notations. We use the same notations and technique as that of [23]. Switching diffusions were treated in [23], whereas we deal with switching jump diffusions. Define

$$\begin{aligned}\widetilde{X}(t) &= \left[X'(t)I_{\{\alpha(t)=1\}}, X'(t)I_{\{\alpha(t)=2\}}, \cdots, X'(t)I_{\{\alpha(t)=m\}}\right]',\\ S &= \bigcup_{i\in\mathcal{M}} 0_{r(i-1)} \times \mathbb{R}^r \times 0_{r(m-i)},\end{aligned} \tag{2.77}$$

Here and in the sequel $0_{k_1\times k_2}$ is a $\mathbb{R}^{k_1\times k_2}$ zero matrix , 0_k denotes the k-dimensional zero column vector. It is seen that $S \subseteq \mathbb{R}^{mr}$ and $\widetilde{X}(t)$ is an S-valued process. For $i \in \mathcal{M}, x \in \mathbb{R}^r$, define

$$\begin{aligned}\tilde{x}^i &= 0_{r(i-1)} \times x \times 0_{r(m-i)} \in S.\\ \Xi &= \bigcup_{\substack{i,j\in\mathcal{M}\\ i<j}} 0_{r(i-1)} \times \mathbb{R}^r \times 0_{r(j-i-1)} \times \mathbb{R}^r \times 0_{r(m-j)}.\end{aligned} \tag{2.78}$$

Then $\Xi \subseteq \mathbb{R}^{mr}$ and $\widetilde{X}^{x_0,i_0}(t) - \widetilde{X}^{y_0,j_0}(t)$ is a $\Xi \cup S$-valued process. For $x, y \in \mathbb{R}^r, i, j \in \mathcal{M}$,

$$\tilde{x}^i - \tilde{y}^j = \begin{cases} [0'_{r(i-1)}, x' - y', 0'_{r(m-i)}]' \in S \text{ for } \ i = j,\\ [0'_{r(i-1)}, x', 0'_{r(j-i-1)}, -y', 0'_{r(m-j)}]' \in \Xi \text{ for } \ i < j,\\ [0'_{r(j-1)}, -y', 0'_{r(i-j-1)}, x', 0'_{r(m-i)}]' \in \Xi \text{ for } \ i > j.\end{cases}$$

Similar to the conditions we mentioned in the previous part, under the condition (P1) and (P2'), we can obtain stability in distribution similar to the approach in [23]. Now let us give condition (P2').

Definition 2.47. The switching jump diffusion given by (2.41) and (2.42) is said to satisfy condition (P2') if, for any $\varepsilon > 0$ and any compact subset $\widehat{\mathsf{C}}$ of $\mathbb{R}^r$, there exists a $T = T(\varepsilon, \widehat{\mathsf{C}}) > 0$ such that

$$E|\widetilde{X}^{x_0,i_0}(t) - \widetilde{X}^{y_0,j_0}(t)| < \varepsilon \text{ for all } t \geq T,$$

whenever $(x_0, i_0, y_0, j_0) \in \widehat{\mathsf{C}} \times \mathcal{M} \times \widehat{\mathsf{C}} \times \mathcal{M}$.

We can obtain (P2) from (P2'). To continue, we focus on obtaining sufficient conditions for conditions (P1) and (P2'). From [23, Theorem 3.8] we can see these two properties imply asymptotic stability in distribution. So it is necessary to establish sufficient criteria for the two properties. To proceed, we need to introduce the following notations.

The generator $\tilde{\mathcal{G}}$ associated with the process $\tilde{x}^i - \tilde{y}^j$ is defined as follows: For each $i, j \in \mathcal{M}$, and for any twice continuously differentiable function f,

$$\tilde{\mathcal{G}} f(\tilde{x}^i - \tilde{y}^j) = \tilde{\mathcal{L}} f(\tilde{x}^i - \tilde{y}^j) + \lambda \int_\Gamma [f(\tilde{x}^i + \widetilde{g}(x,i,\gamma) - \tilde{y}^j - \widetilde{g}(y,j,\gamma)) - f(\tilde{x}^i - \tilde{y}^j)]\pi(d\gamma),$$

where $\tilde{\mathcal{L}}$ is the operator for a switching diffusion process given by

$$\begin{aligned}\tilde{\mathcal{L}} f(\tilde{x}^i - \tilde{y}^j) = &\frac{1}{2}\mathrm{tr}(\tilde{a}(\tilde{x}^i, \tilde{y}^j) H f(\tilde{x}^i - \tilde{y}^j)) + (\tilde{b}(x,i) - \tilde{b}(y,j))' \nabla f(\tilde{x}^i - \tilde{y}^j) \\ &+ \sum_{k=1}^m q_{ik}(x) f(\tilde{x}^k - \tilde{y}^j) + \sum_{k=1}^m q_{jk}(x) f(\tilde{x}^i - \tilde{y}^k) + \sum_{\substack{k=1\\k\neq i}}^m \sum_{\substack{l=1\\l\neq j}}^m \widetilde{m}(\Delta_{ik}(x) \cap \Delta_{jl}(y)) \\ &\times [f(\tilde{x}^k - \tilde{y}^l) - f(\tilde{x}^i - \tilde{y}^l) - f(\tilde{x}^k - \tilde{y}^j) + f(\tilde{x}^i - \tilde{y}^j)],\end{aligned} \tag{2.79}$$

in which

$$\begin{aligned}\tilde{b}(x,i) &= \left[0'_{r(i-1)}, b'(x,i), 0'_{r(m-i)}\right]', \\ \tilde{\sigma}(x,i) &= \left[0'_{r(i-1)\times d}, \sigma'(x,i), 0'_{r(m-i)\times d}\right]', \\ \tilde{g}(x,i,\gamma) &= \left[0'_{r(i-1)}, g'(x,i,\gamma), 0'_{r(m-i)}\right]', \\ \tilde{a}(\tilde{x}^i, \tilde{y}^j) &= (\tilde{\sigma}(x,i) - \tilde{\sigma}(y,j)) \times (\tilde{\sigma}(x,i) - \tilde{\sigma}(y,j))',\end{aligned}$$

where $0_{l_1\times l_2}$ is an $l_1 \times l_2$ matrix with all entries being 0, $b(x,i)$ and $g(x,i,\gamma) \in \mathbb{R}^r$, and $\sigma(x,i) \in \mathbb{R}^{r\times d}$. Recall that $\Delta_{ik}(x)$ are the intervals having length $q_{ik}(x)$; $\widetilde{m}$ is the Lebesgue measure on $\mathbb{R}$ such that $dt \times \widetilde{m}(dz)$ is the density of Poisson measure with which we can represent the discrete component $\alpha(t)$ by a stochastic integral.

Theorem 2.48. *Assume the conditions of Theorem* 2.46 *hold and assume that for each* $i, j \in \mathcal{M}$*, there exists a Lyapunov function* $V(z) = z'z \in C^2(\mathbb{R}^{mr} : \mathbb{R}_+)$ *satisfying the following condition: There exists a positive real number* $\widetilde{\beta}$ *such that*

$$\tilde{\mathcal{G}}V(\tilde{x}^i - \tilde{y}^j) \leq -\widetilde{\beta}V(\tilde{x}^i - \tilde{y}^j), \tag{2.80}$$

then (P1) and (P2') hold.

Proof. We need only verify (P2'). Let $\widehat{\mathsf{C}}$ be any compact subset of $\mathbb{R}^r$, and fix any $x_0, y_0 \in \widehat{\mathsf{C}}$, $i_0, j_0 \in \mathcal{M}$. Define

$$\begin{aligned}
\zeta_N &= \inf\{t \geq 0, |\widetilde{X}^{x_0,i_0}(t) - \widetilde{X}^{y_0,j_0}(t)| > N\},\\
\widetilde{\zeta}_R &= \inf\{t \geq 0, |\widetilde{X}^{x_0,i_0}(t)|^2 + |\widetilde{X}^{y_0,j_0}(t)|^2 > R\}.
\end{aligned}$$

Let $\zeta = \zeta_N \wedge \widetilde{\zeta}_R$.

By virtue of the generalized Itô formula, we have

$$E|\widetilde{X}^{x_0,i_0}(t\wedge\zeta) - \widetilde{X}^{y_0,j_0}(t\wedge\zeta)|^2 = |\tilde{x}_0^{i_0} - \tilde{y}_0^{j_0}|^2 + \int_0^{t\wedge\zeta} E\tilde{\mathcal{G}}|\widetilde{X}^{x_0,i_0}(u) - \widetilde{X}^{y_0,j_0}(u)|^2 du.$$

Given the fact that for $u \leq t\wedge\zeta$, we have $u \wedge \zeta = u$. As a result,

$$E|\widetilde{X}^{x_0,i_0}(t\wedge\zeta) - \widetilde{X}^{y_0,j_0}(t\wedge\zeta)|^2 = |\tilde{x}_0^{i_0} - \tilde{y}_0^{j_0}|^2 + \int_0^{t} E\tilde{\mathcal{G}}|\widetilde{X}^{x_0,i_0}(u\wedge\zeta) - \widetilde{X}^{y_0,j_0}(u\wedge\zeta)|^2 du.$$

Then

$$\frac{dE}{dt}|\widetilde{X}^{x_0,i_0}(t\wedge\zeta)-\widetilde{X}^{y_0,j_0}(t\wedge\zeta)|^2 = E\widetilde{\mathcal{G}}|\widetilde{X}^{x_0,i_0}(t\wedge\zeta)-\widetilde{X}^{y_0,j_0}(t\wedge\zeta)|^2$$
$$\leq -\widetilde{\beta}E|\widetilde{X}^{x_0,i_0}(t\wedge\zeta)-\widetilde{X}^{y_0,j_0}(t\wedge\zeta)|^2.$$

Solving the differential inequality above leads to

$$E|\tilde{X}^{x_0,i_0}(t\wedge\zeta)-\tilde{X}^{y_0,j_0}(t\wedge\zeta)|^2 \leq e^{-\widetilde{\beta}t}|\tilde{x}_0^{i_0}-\tilde{y}_0^{j_0}|^2.$$

Let $N\to\infty, R\to\infty$, we obtain

$$E|\tilde{X}^{x_0,i_0}(t)-\tilde{X}^{y_0,j_0}(t)|^2 \leq e^{-\widetilde{\beta}t}|\tilde{x}_0^{i_0}-\tilde{y}_0^{j_0}|^2.$$

Condition (P2') is thus verified. □

Theorem 2.49. *Assume that for each $i\in\mathcal{M}$, there exists function $V(\cdot,i)\in C^2(\mathbb{R}^r:\mathbb{R}_+)$ satisfying the following two conditions: There exists a positive real number β such that*

$$\mathcal{G}V(x,i)\leq -\beta V(x,i), \tag{2.81}$$

$$V_R := \inf_{\substack{|x|\geq R\\ i\in\mathcal{M}}} V(x,i)\to\infty \text{ as } R\to\infty. \tag{2.82}$$

Then the Markovian switching jump diffusion is asymptotic stability in distribution.

The above theorem takes care of the case of Markovian switching jump diffusion. For x depending on switching jump diffusion, one of the difficulties is the interplays between $x(t)$

and $\alpha(t)$. We redefine $\widetilde{X}(t)$ and the set S as

$$\begin{aligned}
&\widetilde{X}(t) = \big[X'(t)I_{\{\alpha(t)=1\}}, X'(t)I_{\{\alpha(t)=2\}}, \cdots, X'(t)I_{\{\alpha(t)=m\}}\big]', \\
&S = \bigcup_{i\in\mathcal{M}} 0_{r(i-1)} \times \mathbb{R}^r \times 0_{r(m-i)},
\end{aligned} \tag{2.83}$$

Here and in the sequel $0_{k_1\times k_2}$ is a $\mathbb{R}^{k_1\times k_2}$ zero matrix, 0_k denotes the k-dimensional zero column vector. It is seen that $S \subseteq \mathbb{R}^{mr}$ and $\widetilde{X}(t)$ is an S-valued process. For $i \in \mathcal{M}, x \in \mathbb{R}^r$, define

$$\begin{aligned}
&\widetilde{x}^i = 0_{r(i-1)} \times x \times 0_{r(m-i)} \in S. \\
&\Xi = \bigcup_{\substack{i,j\in\mathcal{M}\\ i<j}} 0_{r(i-1)} \times \mathbb{R}^r \times 0_{r(j-i-1)} \times \mathbb{R}^r \times 0_{r(m-j)}.
\end{aligned} \tag{2.84}$$

Then $\Xi \subseteq \mathbb{R}^{mr}$ and $\widetilde{X}^{x_0,i_0}(t) - \widetilde{X}^{y_0,j_0}(t)$ is a $\Xi \cup S$-valued process. For $x, y \in \mathbb{R}^r, i, j \in \mathcal{M}$,

$$\widetilde{x}^i - \widetilde{y}^j = \begin{cases} [0'_{r(i-1)}, x' - y', 0'_{r(m-i)}]' \in S \text{ for } \ i = j, \\ [0'_{r(i-1)}, x', 0'_{r(j-i-1)}, -y', 0'_{r(m-j)}]' \in \Xi \text{ for } \ i < j, \\ [0'_{r(j-1)}, -y', 0'_{r(i-j-1)}, x', 0'_{r(m-i)}]' \in \Xi \text{ for } \ i > j. \end{cases}$$

The generator $\widetilde{\mathcal{G}}$ associated with the process $\widetilde{x}^i - \widetilde{y}^j$ is defined as follows: For each $i, j \in \mathcal{M}$, and for any twice continuously differentiable function f,

$$\widetilde{\mathcal{G}}f(\widetilde{x}^i - \widetilde{y}^j) = \widetilde{\mathcal{L}}f(\widetilde{x}^i - \widetilde{y}^j) + \lambda \int_\Gamma [f(\widetilde{x}^i + \widetilde{g}(x,i,\gamma) - \widetilde{y}^j - \widetilde{g}(y,j,\gamma)) - f(\widetilde{x}^i - \widetilde{y}^j)]\pi(d\gamma),$$

where $\widetilde{\mathcal{L}}$ is the operator for a switching diffusion process given by

$$\begin{aligned}\widetilde{\mathcal{L}}f(\widetilde{x}^i-\widetilde{y}^j) &= \frac{1}{2}\mathrm{tr}(\widetilde{a}(\widetilde{x}^i,\widetilde{y}^j)Hf(\widetilde{x}^i-\widetilde{y}^j)) + (\widetilde{b}(x,i)-\widetilde{b}(y,j))'\nabla f(\widetilde{x}^i-\widetilde{y}^j)\\ &\quad+\sum_{k=1}^{m} q_{ik}(x)f(\widetilde{x}^k-\widetilde{y}^j)+\sum_{k=1}^{m} q_{jk}(x)f(\widetilde{x}^i-\widetilde{y}^k)+\sum_{\substack{k=1\\k\neq i}}^{m}\sum_{\substack{l=1\\l\neq j}}^{m}\widetilde{m}(\Delta_{ik}(x)\cap\Delta_{jl}(y))\\ &\quad\times[f(\widetilde{x}^k-\widetilde{y}^l)-f(\widetilde{x}^i-\widetilde{y}^l)-f(\widetilde{x}^k-\widetilde{y}^j)+f(\widetilde{x}^i-\widetilde{y}^j)],\end{aligned} \tag{2.85}$$

in which

$$\begin{aligned}\widetilde{b}(x,i) &= \left[0'_{r(i-1)}, b'(x,i), 0'_{r(m-i)}\right]',\\ \widetilde{\sigma}(x,i) &= \left[0'_{r(i-1)\times d}, \sigma'(x,i), 0'_{r(m-i)\times d}\right]',\\ \widetilde{g}(x,i,\gamma) &= \left[0'_{r(i-1)}, g'(x,i,\gamma), 0'_{r(m-i)}\right]',\\ \widetilde{a}(\widetilde{x}^i,\widetilde{y}^j) &= (\widetilde{\sigma}(x,i)-\widetilde{\sigma}(y,j))\times(\widetilde{\sigma}(x,i)-\widetilde{\sigma}(y,j))',\end{aligned}$$

where $\Delta_{ik}(x)$ are the intervals having length $q_{ik}(x)$; $\widetilde{m}$ is the Lebesgue measure on $\mathbb{R}$ such that $dt\times\widetilde{m}(dz)$ is the density of Poisson measure with which we can represent the discrete component $\alpha(t)$ by a stochastic integral. Then we have the following theorem:

Theorem 2.50. *Assume that the conditions of Theorem* 2.49 *hold and and assume that for each $i,j\in\mathcal{M}$, there exists a Lyapunov function $V(z)=z'z\in C^2(\mathbb{R}^{mr}:\mathbb{R}_+)$ satisfying the following condition: There exists a positive real number $\widetilde{\beta}$ such that*

$$\tilde{\mathcal{G}}V(\tilde{x}^i-\tilde{y}^j)\leq-\widetilde{\beta}V(\tilde{x}^i-\tilde{y}^j), \tag{2.86}$$

then Then regime switching jump diffusion is asymptotic stable in distribution.

Together with our results in asymptotic stable in the large, exponential p-stable as below, the stability study for regime switching jump diffusion is complete now.

Lemma 2.51. *Let $D \subset \mathbb{R}^r$ is a neighborhood of* 0. *Suppose that for each $i \in \mathcal{M}$, there exists a nonnegative Lyapunov function $V(\cdot, i) : D \mapsto \mathbb{R}$ such that*

(i) *$V(\cdot, i)$ is continuous in D and vanishes only at $x = 0$;*

(ii) *$V(\cdot, i)$ is twice continuously differentiable in $D - \{0\}$ and satisfies $\mathcal{G}V(x, i) \leq 0$ for all $x \in D - \{0\}$.*

Then the equilibrium point $x = 0$ is stable in probability.

Define

$$\tau_{\rho,\varsigma} := \inf\{t \geq 0 : |X(t)| = \rho \text{ or } |X(t)| = \varsigma\}, \tag{2.87}$$

for any $0 < \rho < \varsigma$ and any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$ with $\rho < |x| < \varsigma$.

Lemma 2.52. *Assume that the conditions of Lemma* 2.51 *hold, and that for any sufficiently small $0 < \rho < \varsigma$ and any $(x, \alpha) \in \mathbb{R}^r \times \mathcal{M}$ with $\rho < |x| < \varsigma$, $P\{\tau_{\rho,\varsigma} < \infty\} = 1$. Then the equilibrium point $x = 0$ is asymptotically stable in probability.*

Theorem 2.53. *Assume that the conditions of Lemma* 2.52 *hold, and that $V_\varsigma := \inf\limits_{\substack{|x| \geq \varsigma \\ i \in \mathcal{M}}} V(x, i) \to \infty$ as $\varsigma \to \infty$. Then the equilibrium point $x = 0$ is asymptotically stable in the large.*

3 Nearly Optimal Controls of Mean Variance Problems

3.1 Formulation

The origin of the mean-variance optimization problem can be traced back to the Nobel-prize-winning work of Markowitz [60]. The mean-variance approach has become the foundation of modern finance theory and has inspired numerous extensions and applications. In our work, we consider the mean variance optimization problem of switching process. Our objective is to find an admissible control $u(\cdot)$ among all the admissible controls given that the expected terminal value (wealth or things we want to focus) of the whole system is $Ex(T) = z$ for some given $z \in \mathbb{R}$ so that the risk measured by the variance at the terminal of the flow is minimized. Specifically, we have the following performance measure

$$\begin{aligned} &\min \left\{ J(x, \alpha, u(\cdot)) = E[x(T) - z]^2 \right\} \\ &\text{subject to } Ex(T) = z. \end{aligned} \tag{3.1}$$

Note that in this case, the objective function does not involve control u. Therefore, it is a LQG problem with indefinite control weights. By stating that mean variance control problem of switching process we are interested on solving the classical mean variance problem in which switching process is embedded in certain ways. suppose that switching process $\alpha(t)$ is continuous-time Markov chain with state space $\mathcal{M} = \{1, 2, \ldots, m\}$ The new feature considered here is that the state space of the discrete event process $\alpha(\cdot)$ is large. Obtaining the optimal strategy in such a large-scale system involves high computational complexity, optimal control a difficult task. To reduce the computational complexity, we note that in the Markov chain, some groups of states vary rapidly whereas others change slowly. Based on this feature, we decompose the state space $\mathcal{M}$ into subspaces $\mathcal{M} = \cup_{i=1}^{l} \mathcal{M}_i$ such that within each $\mathcal{M}_i$, the transitions happen frequently and among different clusters the transitions are relatively infrequent. To reflect the different transition rates, we let $\alpha(t) = \alpha^{\varepsilon}(t)$ where $\varepsilon > 0$ is a small parameter so that the generator of the Markov chain is given by

$$Q^{\varepsilon} = \frac{\widetilde{Q}}{\varepsilon} + \widehat{Q}. \tag{3.2}$$

Suppose that $x_i^{\varepsilon}(\cdot)$ are real-valued functions with $i = 0, \ldots, d_1$ such that

$$\begin{aligned} dx_0^{\varepsilon}(t) &= r(t, \alpha(t)) x_0^{\varepsilon}(t) dt \\ x_0^{\varepsilon}(0) &= x_0,\ \alpha(0) = \alpha \end{aligned} \tag{3.3}$$

for $\alpha(t) \in \{1, 2, \ldots, m\}$. The flows of the other d_1 nodes follow geometric Brownian motion:

$$\begin{aligned} dx_i^{\varepsilon}(t) &= x_i^{\varepsilon}(t) r_i(t, \alpha(t)) dt + x_i^{\varepsilon}(t) \tilde{\sigma}_i(t, \alpha(t)) dw(t) \\ x_i^{\varepsilon}(0) &= x_i,\ \alpha(0) = \alpha \text{ for } i = 1, 2, \ldots, d_1,\ \alpha \in \mathcal{M}, \end{aligned} \tag{3.4}$$

where $\tilde{\sigma}_i(t, \alpha(t)) = (\tilde{\sigma}_{i1}(t, \alpha(t)), \tilde{\sigma}_{i2}(t, \alpha(t)), \ldots, \tilde{\sigma}_{id}(t, \alpha(t))) \in \mathbb{R}^{1\times d}$. In the finance application, $x_0^{\varepsilon}(\cdot)$ represents an investor's bank account value, whereas $x_i^{\varepsilon}(\cdot)$ for each $i = 1, \ldots, d_1$ is his wealth devoted to the ith stock or risky asset. The motivation of our work is from network system where we use $x_i^{\varepsilon}(\cdot)$ to represent the flows of the ith node. We can represent the total flows of the entire system as $x^{\varepsilon}(t)$ and we need to decide the proportion $n_i(t)$ of flow $x_i^\varepsilon(t)$ to put on node i, i.e.,

$$x^\varepsilon(t) = \sum_{i=0}^{d_1} n_i(t) x_i^\varepsilon(t).$$

By assuming that the interaction among these $d_1 + 1$ nodes occurs continuously, we have

$$\begin{aligned} dx^\varepsilon(t) &= \sum_{i=0}^{d_1} n_i(t) dx_i^\varepsilon(t) \\ &= [r(t,\alpha(t))x^\varepsilon(t) + B(t,\alpha(t))u(t)]dt + u'(t)\tilde{\sigma}(t,\alpha(t))dw(t) \\ x^\varepsilon(0) &= x = \sum_{i=1}^{d_1} n_i(0)x_i, \ \alpha(0) = \alpha, \ \text{for } 0 \le t \le T, \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} &B(t,\alpha(t)) = (r_1(t,\alpha(t)) - r(t,\alpha(t)), r_2(t,\alpha(t)) - r(t,\alpha(t)), \dots, \\ &\qquad\qquad\qquad\qquad\qquad\qquad r_{d_1}(t,\alpha(t)) - r(t,\alpha(t))), \\ &\tilde{\sigma}(t,\alpha(t)) = (\tilde{\sigma}_1(t,\alpha(t)), \dots, \tilde{\sigma}_{d_1}(t,\alpha(t)))' \in \mathbb{R}^{d_1 \times d}, \\ &u(t) = (u_1(t), \dots, u_{d_1}(t))' \in \mathbb{R}^{d_1 \times 1}, \end{aligned}$$

and $u_i(t) = n_i(t)x_i^\varepsilon(t)$ is the total amount of flow for node i at time t for $i = 1, 2, \dots, d_1$. throughout this paper that all the functions $r(t,i)$, $B(t,i)$, and $\sigma(t,i)$ are measurable and uniformly bounded in t. We also assume the non-degeneracy condition is satisfied, i.e., there is a $\delta > 0$ such that $a(t,i) = \tilde{\sigma}(t,i)\tilde{\sigma}'(t,i) \ge \delta I$ for any $t \in [0,T]$ and $i \in \mathcal{M}$. We denote by $L^2_{\mathcal{F}}(0,T;\mathbb{R}^{l_0})$ the set of all $\mathbb{R}^{l_0}$-valued, measurable stochastic processes $f(t)$ adapted to $\{\mathcal{F}_t\}_{t\ge 0}$ such that $E\int_0^T |f(t)|^2 dt < +\infty$.

Let $\mathcal{U}$ be the set of controls which is a compact set in $\mathbb{R}^{d_1\times 1}$. The $u(\cdot)$ is said to be admissible if $u(\cdot) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^{d_1})$ and the equation (3.5) has a unique solution $x^\varepsilon(\cdot)$ corresponding to $u(\cdot)$. In this case, we call $(x^\varepsilon(\cdot), u(\cdot))$ an admissible (total flow, flow distribution) pair. To find the minimum of $J(x, \alpha, u(\cdot), \lambda)$, it suffices to choose $u(\cdot)$ so that $E(x^\varepsilon(T)+\lambda-z)^2$ is minimized. We regard this part as $J(x, \alpha, u(\cdot))$ in what follows. Let $v^\varepsilon(x, \alpha) = \inf_{u(\cdot)} J^\varepsilon(x, \alpha, u(\cdot))$ be the value function to show the dependence on the parameter ε.

$$\rho(t,i) = B(t,i)[\sigma(t,i)\sigma'(t,i)]^{-1}B'(t,i), i \in \{1,2,\ldots,m\}. \tag{3.6}$$

Consider the following two systems of ODEs for $i = 1, 2, \ldots, m$:

$$\begin{aligned} &\dot{P}^\varepsilon(t,i) = P^\varepsilon(t,i)[\rho(t,i) - 2r(t,i)] - \sum_{j=1}^{m} q_{ij}^\varepsilon P^\varepsilon(t,j) \\ &P^\varepsilon(T,i) = 1. \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} \dot{H}^\varepsilon(t,i) = {} & H^\varepsilon(t,i)r(t,i) - \frac{1}{P^\varepsilon(t,i)} \sum_{j=1}^{m} q_{ij}^\varepsilon P^\varepsilon(t,j) H^\varepsilon(t,j) \\ & + \frac{H^\varepsilon(t,i)}{P^\varepsilon(t,i)} \sum_{j=1}^{m} q_{ij}^\varepsilon P^\varepsilon(t,j), \\ H^\varepsilon(T,i) = {} & 1. \end{aligned} \tag{3.8}$$

The existence and uniqueness of solutions to the above two systems of equations are evident as both are linear with uniformly bounded coefficients. Applying the generalized Itô's formula to

$$v^\varepsilon(t, x^\varepsilon(t), i) = P^\varepsilon(t,i)(x^\varepsilon(t) + (\lambda - z)H^\varepsilon(t,i))^2,$$

by employing the completing square techniques, we obtain

$$
\begin{aligned}
&dP^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]^2\\
&\quad=2P^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]dx^\varepsilon(t)+P^\varepsilon(t,i)(dx^\varepsilon(t))^2\\
&\qquad+\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,j)]^2dt\\
&\qquad+\dot{P}^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]^2dt+2P^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)](\lambda-z)\dot{H}^\varepsilon(t,i)dt.
\end{aligned}
\tag{3.9}
$$

Therefore, by plugging in the dynamic equation satisfied by $P(t,i)$ and $H(t,i)$, we have the following expression:

$$
\begin{aligned}
&dP^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]^2\\
&=P^\varepsilon(t,i)\{u'(t)\sigma(t,i)\sigma'(t,i)u(t)+2u'(t)B'(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]\\
&\quad+2r(t,i)x^\varepsilon(t)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]\}dt-\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]^2dt\\
&\quad+2P^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)](\lambda-z)\{H^\varepsilon(t,i)r(t,i)-\frac{1}{P^\varepsilon(t,i)}\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)H^\varepsilon(t,j)\\
&\quad+\frac{H^\varepsilon(t,i)}{P^\varepsilon(t,i)}\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)\}dt+\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,j)]^2dt\\
&\quad+[\rho(t,i)-2r(t,i)]P^\varepsilon(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)]^2dt+(\cdots)dw(t)\\
&=P^\varepsilon(t,i)\{(u(t)+(\sigma(t,i)\sigma'(t,i))^{-1}B'(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)])'[\sigma(t,i)\sigma'(t,i)]\\
&\quad\times(u(t)+(\sigma(t,i)\sigma'(t,i))^{-1}B'(t,i)[x^\varepsilon(t)+(\lambda-z)H^\varepsilon(t,i)])\}dt\\
&\quad+(\lambda-z)^2\sum_{j=1}^m q_{ij}^\varepsilon P^\varepsilon(t,j)[H^\varepsilon(t,j)-H^\varepsilon(t,i)]^2dt+(\cdots)dw(t).
\end{aligned}
\tag{3.10}
$$

Integrating both sides of the above equation from 0 to T and taking expectation, we obtain

$$
\begin{aligned}
&E[x^{\varepsilon}(T)+\lambda-z]^2\\
&\quad = P^{\varepsilon}(0,\alpha)[x+(\lambda-z)H^{\varepsilon}(0,\alpha)]^2\\
&\quad +E\int_0^T (\lambda-z)^2\sum_{j=1}^{m} q_{ij}^{\varepsilon}P^{\varepsilon}(t,j)[H^{\varepsilon}(t,j)-H^{\varepsilon}(t,i)]^2dt\\
&\quad +E\int_0^T P^{\varepsilon}(t,i)(u(t)-u^{\varepsilon,*}(t))'(\sigma(t,i)\sigma'(t,i))(u(t)-u^{\varepsilon,*}(t))dt.
\end{aligned}
\tag{3.11}
$$

Thus, the optimal control u^* has the form

$$
u^{\varepsilon,*}(t,\alpha^{\varepsilon}(t),x^{\varepsilon}(t)) = -(\sigma(t,\alpha^{\varepsilon}(t))\sigma'(t,\alpha^{\varepsilon}(t)))^{-1}B'(t,\alpha^{\varepsilon}(t))[x^{\varepsilon}(t)+(\lambda-z)H^{\varepsilon}(t,\alpha^{\varepsilon}(t))].
\tag{3.12}
$$

3.2 Key Results and Proofs

Note that when $|\mathcal{M}|=m$ is large, although we can get the optimal solution of the mean-variance control problem. For a large-scale system, solving this problem is still computationally intensive and practically unattractive. As a viable alternative, we focus on an decomposition-aggregation approach. Assume that $\widetilde{Q}$ is of the block-diagonal form $\widetilde{Q} = \mathrm{diag}(\widetilde{Q}^1,\ldots,\widetilde{Q}^l)$ in which $\widetilde{Q}^k \in \mathbb{R}^{m_k\times m_k}$ are irreducible for $k=1,2,\ldots,l$ and $\sum_{k=1}^l m_k = m$, and $\widetilde{Q}^k$ denotes the kth block matrix in $\widetilde{Q}$. Let $\mathcal{M}_k=\{s_{k1},s_{k2},\ldots,s_{km_k}\}$ denote the states corresponding to $\widetilde{Q}^k$ and let $\mathcal{M}=\mathcal{M}_1\cup\mathcal{M}_2\ldots\cup\mathcal{M}_l=\{s_{11},s_{12},\ldots,s_{1m_1},\ldots,s_{l1},s_{l2},\ldots,s_{lm_l}\}$. The slow and fast components are coupled through weak and strong interactions in the sense that the underlying Markov chain fluctuates rapidly within a single group $\mathcal{M}_k$ and jumps less frequently among groups $\mathcal{M}_k$ and $\mathcal{M}_j$ for $k\neq j$.

By aggregating the states in $\mathcal{M}_k$ as one state k, we can obtain an aggregated process

$\overline{\alpha}^{\varepsilon}(\cdot)$. That is, $\overline{\alpha}^{\varepsilon}(t)=k$ when $\alpha^{\varepsilon}(t)\in\mathcal{M}_k$. By virtue of [83, Theorem7.4], $\overline{\alpha}^{\epsilon}(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$ whose generator is given by

$$\overline{Q}=\operatorname{diag}(\mu^1,\mu^2,\dots,\mu^l)\widehat{Q}\operatorname{diag}(\mathbb{1}_{m_1},\mathbb{1}_{m_2},\dots,\mathbb{1}_{m_l}), \tag{3.13}$$

where μ^k is the stationary distribution of $\widetilde{Q}^k, k=1,2,\dots,l$, and $\mathbb{1}_n=(1,1,\dots,1)\in\mathbb{R}^n$. Define an operator $\mathcal{L}^{\varepsilon}$ by

$$\begin{aligned}\mathcal{L}^{\varepsilon}f(x,t,\iota)=&\frac{\partial f(x,t,\iota)}{\partial t}+[r(t,\iota)x+B(t,\iota)u(t)]\frac{\partial f(x,t,\iota)}{\partial x}\\&+\frac{1}{2}[u'(t)\sigma(t,\iota)\sigma'(t,\iota)u(t)]\frac{\partial^2 f(x,t,\iota)}{\partial x^2}+Q^{\varepsilon}f(x,t,\iota),\ \iota\in\mathcal{M},\end{aligned} \tag{3.14}$$

where

$$Q^{\varepsilon}f(x,t,\cdot)(\iota)=\sum_{\ell\neq\iota}q^{\varepsilon}_{\iota\ell}(f(x,t,\ell)-f(x,t,\iota)), \tag{3.15}$$

and for each $\iota\in\mathcal{M}$, $f(\cdot,\cdot,\iota)\in C^{2,1}$ (that is, $f(\cdot)$ has continuous derivatives up to the second order with respect to x and continuous derivative with respect to t up to the first order). Define

$$\begin{aligned}\overline{\mathcal{L}}f(x,t,k)=&\frac{\partial f(x,t,k)}{\partial t}+[\overline{r}(t,k)x+\overline{B}(t,k)u(t)]\frac{\partial f(x,t,k)}{\partial x}\\&+\frac{1}{2}[u'(t)\overline{\sigma}(t,k)\overline{\sigma}'(t,k)u(t)]\frac{\partial^2 f(x,t,k)}{\partial x^2}+\overline{Q}f(x,t,k),\ k\in\overline{\mathcal{M}},\end{aligned} \tag{3.16}$$

where $\overline{Q}$ is defined in (3.13) and

$$\begin{aligned}
\overline{r}(t,k) &= \sum_{j=1}^{m_k} \mu_j^k r(t, s_{kj}),\\
\overline{B}(t,k) &= \sum_{j=1}^{m_k} \mu_j^k B(t, s_{kj}),\\
\overline{\sigma}^2(t,k) &= \sum_{j=1}^{m_k} \mu_j^k \sigma^2(t, s_{kj}).
\end{aligned}$$

The following theorems are concerned with the weak convergence of a pair of processes.

Theorem 3.1. *Suppose that the martingale problem with operator $\overline{\mathcal{L}}$ defined in (3.16) has a unique solution for each initial condition. Then the pair of processes $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(x(\cdot), \overline{\alpha}(\cdot))$, which is the solution of the martingale problem with operator $\overline{\mathcal{L}}$.*

Proof. The proof is divided into the following steps. First, we prove the tightness of $x^\varepsilon(\cdot)$. Once the tightness is verified, we proceed to obtain the convergence by using a martingale problem formulation. We first show that a priori bound holds.

Lemma 3.2. *Let $x^\varepsilon(t)$ denote flow of system corresponding to $\alpha^\varepsilon(t)$. Then*

$$\sup_{0\le t\le T} E|x^\varepsilon(t)|^2 = O(1).$$

Proof. Recall that

$$\begin{aligned}
dx^\varepsilon(t) &= [r(t,\alpha^\varepsilon(t))x^\varepsilon(t) - \rho(t,\alpha^\varepsilon(t))x^\varepsilon(t) - \rho(t,\alpha^\varepsilon(t))(\lambda - z)\overline{H}(t,\overline{\alpha}^\varepsilon(t))]dt\\
&\quad + \sum_{i=1}^{d} \sqrt{\left(\sum_{n=1}^{d_1} u_n^{\varepsilon,*}(t, x^\varepsilon(t), \alpha^\varepsilon(t))\sigma_{ni}(t,\alpha^\varepsilon(t))\right)^2} dw_i(t)\\
x^\varepsilon(0) &= x.
\end{aligned}$$

So,

$$
\begin{aligned}
E|x^\varepsilon(t)|^2 &\le K|x|^2 + E\left|\int_0^t (r(\nu,\alpha^\varepsilon(\nu)) + \rho(\nu,\alpha^\varepsilon(\nu)))x^\varepsilon(\nu))d\nu\right|^2 \\
&\quad + KE\int_0^t \left(\sum_{n=1}^{d_1} u_n^{\varepsilon,*}(\nu, x^\varepsilon(\nu), \alpha^\varepsilon(\nu))\sigma_{ni}(\nu,\alpha^\varepsilon(\nu))\right)^2 d\nu \\
&\le K + KE\int_0^t |x^\varepsilon(\nu)|^2 d\nu.
\end{aligned}
$$

Here, recall that $\sigma(t,\alpha^\varepsilon(t)) = (\sigma_{ni}(t,\alpha^\varepsilon(t))) \in \mathbb{R}^{d_1\times d}$ and note that $u_n^{\varepsilon,*}$ is the nth component of the d_1 dimensional variable. Using properties of stochastic integrals, Hölder inequality, and boundedness of $r(\cdot), B(\cdot), \sigma(\cdot)$, by Gronwall's inequality, we obtain the second moment bound of $x^\varepsilon(t)$ as desired. $\square$

Lemma 3.3. *$\{x^\varepsilon(\cdot)\}$ is tight in $D([0,T]:\mathbb{R})$.*

Proof. Denote $\mathcal{F}_t^\varepsilon$ as the σ-algebra generated by $\{w(s), \alpha^\varepsilon(s) : s \le t\}$ and E_t^ε as the conditional expectation w.r.t. $\mathcal{F}_t^\varepsilon$. For any $T<\infty$, any $0\le t\le T$, any $s>0$, and any $\delta>0$ with $0<s\le\delta$, by properties of stochastic integral and boundedness of coefficients,

$$
\begin{aligned}
E_t^\varepsilon|x^\varepsilon(t+s) - x^\varepsilon(t)|^2 &\le KE_t^\varepsilon\int_t^{t+s} |(r(\nu,\alpha^\varepsilon(\nu)) + \rho(\nu,\alpha^\varepsilon(\nu)))x^\varepsilon(\nu)|^2 d\nu \\
&\quad + KE_t^\varepsilon\int_t^{t+s} (\sum_{n=1}^{d_1} u_n^{\varepsilon,*}(\nu, x^\varepsilon(\nu), \alpha^\varepsilon(\nu))\sigma_{ni}(\nu,\alpha^\varepsilon(\nu)))^2 d\nu \\
&\le Ks + KE_t^\varepsilon\int_t^{t+s} |x^\varepsilon(\nu)|^2 d\nu.
\end{aligned}
$$

Thus we have

$$
\lim_{\delta\to0}\limsup_{\varepsilon\to0}\sup_{0\le s\le\delta}\left\{E[E_t^\varepsilon|x^\varepsilon(t+s) - x^\varepsilon(t)|^2]\right\} = 0.
$$

Then the tightness criterion [76, Theorem 3] yields that process $x^\varepsilon(\cdot)$ is tight. Now we describe the limit process. Since $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ is tight, we can extract a weakly convergent subsequence. For notional simplicity, we still denote the subsequence by $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ with limit $(x(\cdot), \overline{\alpha}(\cdot))$. By Skorohod representation with no change of notation, we may assume $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges to $(x(\cdot), \overline{\alpha}(\cdot))$ w.p.1. We next show that the limit $(x(\cdot), \overline{\alpha}(\cdot))$ is a solution of the martingale problem with operator $\overline{\mathcal{L}}$ defined by (3.16).

Lemma 3.4. *The process $x(\cdot)$ is the solution of the martingale problem with the operator $\overline{\mathcal{L}}$.*

Proof. To obtain the desirable result, we need to show

$$f(x(t), t, \overline{\alpha}(t)) - f(x, 0, \alpha) - \int_0^t \overline{\mathcal{L}} f(x(\nu), \nu, \overline{\alpha}(\nu)) d\nu \ \text{ is a martingale,}$$

This can be done by showing that for any integer $n > 0$, any bounded and measurable function $h_p(\cdot, \cdot)$ with $p \le n$, and any $t, s, t_p > 0$ with $t_p \le t < t + s \le T$,

$$\begin{aligned} E \prod_{p=1}^n h_p(x^\varepsilon(t_p), \overline{\alpha}^\varepsilon(t_p)) [f(x(t+s), t+s, \overline{\alpha}(t+s)) - f(x(t), t, \overline{\alpha}(t)) \\ - \int_t^{t+s} \overline{\mathcal{L}} f(x(\nu), \nu, \overline{\alpha}(\nu)) d\nu] = 0. \end{aligned}$$

We further deduce that

$$\begin{aligned} &\lim_{\varepsilon \to 0} E \prod_{p=1}^n h_p(x^\varepsilon(t_p), \overline{\alpha}^\varepsilon(t_p)) (f(x^\varepsilon(t+s), t+s, \overline{\alpha}^\varepsilon(t+s)) - f(x^\varepsilon(t), t, \overline{\alpha}^\varepsilon(t)) \\ &\quad = E \prod_{p=1}^n h_p(x(t_p), \overline{\alpha}(t_p)) (f(x(t+s), t+s, \overline{\alpha}(t+s)) - f(x(t), t, \overline{\alpha}(t)). \end{aligned} \tag{3.17}$$

Moreover,

$$\begin{aligned} &\lim_{\varepsilon \to 0} E \prod_{p=1}^n h_p(x^\varepsilon(t_p), \overline{\alpha}^\varepsilon(t_p)) \Big[\int_t^{t+s} \frac{\partial f(x^\varepsilon(\nu), \nu, \overline{\alpha}^\varepsilon(\nu))}{\partial \nu} d\nu \Big] \\ &\quad = E \prod_{p=1}^n h_p(x(t_p), \overline{\alpha}(t_p)) \Big[\int_t^{t+s} \frac{\partial f(x(\nu), \nu, \overline{\alpha}(\nu))}{\partial \nu} d\nu \Big] \end{aligned} \tag{3.18}$$

by the weak convergence of $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ and the Skorohod representation.

For any $f(\cdot)$ chosen above, define

$$\widehat{f}(x^\varepsilon(t), t, \alpha^\varepsilon(t)) = \sum_{i=1}^{l} f(x^\varepsilon(t), t, i) I_{\{\alpha^\varepsilon(t) \in \mathcal{M}_i\}}$$

since $(x^\varepsilon(t), \alpha^\varepsilon(t))$ is a Markov process, we have

$$\widehat{f}(x^\varepsilon(t), t, \alpha^\varepsilon(t)) - \widehat{f}(x, 0, \alpha) - \textstyle\int_0^t \mathcal{L}^\varepsilon \widehat{f}(x^\varepsilon(\nu), \nu, \alpha^\varepsilon(\nu)) d\nu$$

is a martingale. Consequently,

$$\begin{aligned} E \prod_{p=1}^{n} h_p(x^\varepsilon(t_p), \overline{\alpha}^\varepsilon(t_p)) (\widehat{f}(x^\varepsilon(t+s), t+s, \alpha^\varepsilon(t+s)) - \widehat{f}(x^\varepsilon(t), t, \alpha^\varepsilon(t)) \\ - \int_t^{t+s} \mathcal{L}^\varepsilon \widehat{f}(x^\varepsilon(\nu), \nu, \alpha^\varepsilon(\nu)) d\nu) = 0. \end{aligned}$$

Note that $\widehat{f}(x^\varepsilon(t), t, \alpha^\varepsilon(t)) = f(x^\varepsilon(t), t, \overline{\alpha}^\varepsilon(t))$.

Next we need to show that

$$\begin{aligned} \lim_{\varepsilon \to 0} E \prod_{p=1}^{n} h_p(x^\varepsilon(t_p), \overline{\alpha}^\varepsilon(t_p)) \int_t^{t+s} \mathcal{L}^\varepsilon \widehat{f}(x^\varepsilon(\nu), \nu, \alpha^\varepsilon(\nu)) d\nu \\ = E \prod_{p=1}^{n} h_p(x(t_p), \overline{\alpha}(t_p)) \int_t^{t+s} \overline{\mathcal{L}} f(x(\nu), \nu, \overline{\alpha}(\nu)) d\nu. \end{aligned}$$

Note that we can rewrite $E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}\mathcal{L}^\varepsilon\widehat{f}(x^\varepsilon(\nu),\nu,\alpha^\varepsilon(\nu))d\nu$ as

$$\begin{aligned}
&E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))[\int_t^{t+s}\sum_{k=1}^{l}\sum_{j=1}^{m_k} Q^\varepsilon\widehat{f}(x^\varepsilon(\nu),\nu,\cdot)(s_{kj})I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}d\nu\\
&+\int_t^{t+s}\sum_{k=1}^{l}\sum_{j=1}^{m_k}\frac{\partial\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x}I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}[r(\nu,s_{kj})x^\varepsilon(\nu)+B(\nu,s_{kj})u(\nu)]d\nu\\
&+\int_t^{t+s}\frac{1}{2}\sum_{k=1}^{l}\sum_{j=1}^{m_k}[u'(\nu)\sigma(\nu,s_{kj})\sigma'(\nu,s_{kj})u(\nu)]\frac{\partial^2\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x^2}I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}]d\nu.
\end{aligned}$$

Since $\widetilde{Q}^k\mathbb{1}_{m_k}=0$, we have

$$Q^\varepsilon\widehat{f}(x^\varepsilon(t),t,\cdot)(s_{kj})=\widehat{Q}\widehat{f}(x^\varepsilon(t),t,\cdot)(s_{kj}).$$

We decompose

$$E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}\mathcal{L}^\varepsilon\widehat{f}(x^\varepsilon(\nu),\nu,\alpha^\varepsilon(\nu))d\nu$$

as $H_1^\varepsilon(t+s,t)+H_2^\varepsilon(t+s,t)$. In which

$$\begin{aligned}
H_1^\varepsilon(t+s,t)=&\ E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\\
&\times\Bigg[\sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s}\mu_j^k\frac{\partial\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x}I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}}[r(\nu,s_{kj})x^\varepsilon(\nu)+B(\nu,s_{kj})u(\nu)]d\nu\\
&+\frac{1}{2}\sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s}\mu_j^k[u'(\nu)\sigma(\nu,s_{kj})\sigma'(\nu,s_{kj})u(\nu)]\frac{\partial^2\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x^2}I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}}d\nu\\
&+\sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s}\mu_j^k\widehat{Q}\widehat{f}(x^\varepsilon(\nu),\nu,\cdot)(s_{kj})I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}}d\nu\Bigg]
\end{aligned}$$

and $H_2^\varepsilon(t+s,t)$ can be represented as

$$\begin{aligned}
&H_2^\varepsilon(t+s,t)\\
&= E\prod_{p=1}^n h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\Big(\sum_{k=1}^l\sum_{j=1}^{m_k}\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\frac{\partial\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x}\times\\
&\quad [r(\nu,s_{kj})x^\varepsilon(\nu)+B(\nu,s_{kj})u(\nu)]d\nu+\sum_{k=1}^l\sum_{j=1}^{m_k}\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\widehat{Q}\times\\
&\quad \widehat{f}(x^\varepsilon(\nu),\nu,\cdot)(s_{kj})d\nu+\frac{1}{2}\sum_{k=1}^l\sum_{j=1}^{m_k}\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\times\\
&\quad [u'(\nu)\sigma(\nu,s_{kj})\sigma'(\nu,s_{kj})u(\nu)]\frac{\partial^2\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x^2}d\nu\Big).
\end{aligned}$$

By virtue of Lemma 3.6, [83, Theorem7.14], Cauchy-Schwartz inequality, boundedness of $h_p(\cdot)$, $r(\cdot)$ and $B(\cdot)$, for each $k=1,2,\ldots,l; j=1,2,\ldots,m_k$, as $\varepsilon\to 0$

$$\begin{aligned}
E|\prod_{p=1}^n h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\frac{\partial\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x}\\
\times[r(\nu,s_{kj})x^\varepsilon(\nu)+B(\nu,s_{kj})u(\nu)]d\nu|^2\to 0.
\end{aligned}$$

Similarly as $\varepsilon\to 0$,

$$\begin{aligned}
E|\prod_{p=1}^n h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\\
\times[u'(\nu)\sigma(\nu,s_{kj})\sigma'(\nu,s_{kj})u(\nu)]\frac{\partial^2\widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x^2}d\nu|^2\to 0,
\end{aligned}$$

and

$$E|\prod_{p=1}^n h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}(I_{\{\alpha^\varepsilon(\nu)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}})\widehat{Q}\widehat{f}(x^\varepsilon(\nu),\nu,\cdot)(s_{kj})d\nu|^2\to 0.$$

Therefore, $H_2^\varepsilon(t+s,t)$ converges to 0 in probability. On the other hand, we obtain

$$
\begin{aligned}
&E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p)) \sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s} \mu_j^k \frac{\partial \widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x}[r(\nu,s_{kj})x^\varepsilon(\nu)+B(\nu,s_{kj})u(\nu)] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}}d\nu \\
&\to \sum_{k=1}^{l}\sum_{j=1}^{m_k} E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p)) \int_t^{t+s} \mu_j^k \frac{\partial f(x(\nu),\nu,\overline{\alpha}(\nu))}{\partial x}[r(\nu,s_{kj})x(\nu)+B(\nu,s_{kj})u(\nu)] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times I_{\{\overline{\alpha}(\nu)=k\}}d\nu \\
&= \sum_{k=1}^{l} E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p)) \int_t^{t+s} \frac{\partial f(x(\nu),\nu,\overline{\alpha}(\nu))}{\partial x}[\overline{r}(\nu,\overline{\alpha}(\nu))x(\nu)+\overline{B}(\nu,\overline{\alpha}(\nu))u(\nu))] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times I_{\{\overline{\alpha}(\nu)=k\}}d\nu \\
&= E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p)) \int_t^{t+s} \frac{\partial f(x(\nu),\nu,\overline{\alpha}(\nu))}{\partial x}[\overline{r}(\nu,\overline{\alpha}(\nu))x(\nu)+\overline{B}(\nu,\overline{\alpha}(\nu))u(\nu)]d\nu.
\end{aligned}
\tag{3.19}
$$

Similarly,

$$
\begin{aligned}
&E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p)) \sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s} \mu_j^k [u'(\nu)\sigma^2(\nu,s_{kj})u(\nu)] \frac{\partial^2 \widehat{f}(x^\varepsilon(\nu),\nu,s_{kj})}{\partial x^2} I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}}d\nu \\
&\quad\to\ E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p)) \int_t^{t+s} \frac{\partial^2 f(x(\nu),\nu,\overline{\alpha}(\nu))}{\partial x^2}[u'(\nu)\overline{\sigma}^2(\nu,\overline{\alpha}(\nu))u(\nu)]d\nu.
\end{aligned}
\tag{3.20}
$$

Note that

$$
\sum_{k=1}^{l}\sum_{j=1}^{m_k}\int_t^{t+s} \mu_j^k I_{\{\overline{\alpha}^\varepsilon(\nu)=k\}} \widehat{Q}\widehat{f}(x^\varepsilon(\nu),\nu,\cdot)(s_{kj})d\nu = \int_t^{t+s} \overline{Q}f(x^\varepsilon(\nu),\nu,\cdot)(\overline{\alpha}^\varepsilon(\nu))d\nu.
$$

So as $\varepsilon \to 0$,

$$
\begin{aligned}
&E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p)) \int_t^{t+s} \overline{Q}f(x^\varepsilon(\nu),\nu,\cdot)(\overline{\alpha}^\varepsilon(\nu))d\nu \\
&\quad\to E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p)) \int_t^{t+s} \overline{Q}f(x(\nu),\nu,\cdot)(\overline{\alpha}(\nu))d\nu.
\end{aligned}
\tag{3.21}
$$

Combining the results from (3.19) to (3.21), we have

$$\begin{aligned}&\lim_{\varepsilon\to 0} E\prod_{p=1}^{n} h_p(x^\varepsilon(t_p),\overline{\alpha}^\varepsilon(t_p))\int_t^{t+s}\mathcal{L}^\varepsilon\widehat{f}(x^\varepsilon(\nu),\nu,\alpha^\varepsilon(\nu))d\nu\\&= E\prod_{p=1}^{n} h_p(x(t_p),\overline{\alpha}(t_p))\int_t^{t+s}\overline{\mathcal{L}}f(x(\nu),\nu,\overline{\alpha}(\nu))d\nu\end{aligned}\tag{3.22}$$

Finally, we complete the proof by combining all the previous results.

Theorem 3.5. *For $k=1,2,\ldots,l$ and $j=1,2,\ldots,m_k$, $P^\varepsilon(t,s_{kj})\to\overline{P}(t,k)$ and $H^\varepsilon(t,s_{kj})\to\overline{H}(t,k)$ uniformly on $[0,T]$ as $\varepsilon\to 0$, where $\overline{P}(t,k)$ and $\overline{H}(t,k)$ are the unique solutions of the following differential equations for $k=1,2,\ldots,l$,*

$$\begin{aligned}\dot{\overline{P}}(t,k) &= (\overline{\rho}(t,k)-2\overline{r}(t,k))\overline{P}(t,k)-\bar{Q}\overline{P}(t,\cdot)(k)\\\overline{P}(T,k)&=1.\end{aligned}\tag{3.23}$$

and

$$\begin{aligned}\dot{\overline{H}}(t,k) &= \overline{r}(t,k)\overline{H}(t,k)-\frac{1}{\overline{P}(t,k)}\bar{Q}\overline{P}(t,\cdot)\overline{H}(t,\cdot)(k)+\frac{\overline{H}(t,k)}{\overline{P}(t,k)}\bar{Q}\overline{P}(t,\cdot)(k)\\\overline{H}(T,k)&=1.\end{aligned}\tag{3.24}$$

Proof. We prove the convergence of P^ε (the proof of H^ε is similar). it is easy to see that $P^\varepsilon(t,s_{kj})$ is equicontinuous and uniformly bounded, it follows from Arzela-Ascoli theorem that, for each sequence of $\varepsilon\to 0$, a further subsequence exists (we still use the index ε for the sake of simplicity) such that $P^\varepsilon(t,s_{kj})$ converges uniformly on $[0,T]$ to a continuous function, say, $P^0(t,s_{kj})$. First, we show $P^0(t,s_{kj})$ is independent of j. Given that

$$P^\varepsilon(t,s_{kj}) = 1-\int_t^T[P^\varepsilon(s,s_{kj})(\rho(s,s_{kj})-2r(s,s_{kj}))-Q^\varepsilon P^\varepsilon(s,\cdot)(s_{kj})]ds.$$

Multiplying both sides of above equation by ε yields that

$$0 = \lim_{\varepsilon\to 0}\int_t^T \widetilde{Q}^k P^\varepsilon(s,\cdot)(s_{kj})ds = \int_t^T \widetilde{Q}^k P^0(s,\cdot)(s_{kj})ds.$$

Thus, in view of the continuity of $P^0(t,\cdot)(s_{kj})$, we obtain

$$\widetilde{Q}^k P^0(t,\cdot)(s_{kj}) = 0 \text{ for } t \in [0,T]. \tag{3.25}$$

Given the fact that $\widetilde{Q}^k$ is irreducible, we have $P^0(t,s_{kj}) = P^0(t,k)$ which is independent of j. Now let us multiply $P^\varepsilon(t,s_{kj})$ by μ_j^k and then add the index j. Recall the definition of $\overline{F}(t,k)$, we have the following equation

$$\sum_{j=1}^{m_k}\mu_j^k P^\varepsilon(t,s_{kj}) = 1 - \sum_{j=1}^{m_k}\mu_j^k\int_t^T [P^\varepsilon(s,s_{kj})(\rho(s,s_{kj}) - 2r(s,s_{kj})) - Q^\varepsilon P^\varepsilon(s,\cdot)(s_{kj})]ds.$$

Letting $\varepsilon \to 0$ and noting that uniform convergence of $P^\varepsilon(t,s_{kj}) \to P^0(t,k)$ and μ^k is the stationary distribution corresponding to $\widetilde{Q}^k$, we have

$$(\sum_{j=1}^{m_k}\mu_j^k\widehat{Q}\mathbb{1}_{m_k})P^0(t,\cdot)(k) = \overline{Q}P^0(t,\cdot)(k).$$

Therefore, we obtain

$$P^0(t,k) = 1 - \int_t^T \Big(P^0(s,k)(\overline{\rho}(s,k) - 2\overline{r}(s,k) - \overline{Q}P^0(s,\cdot)(k)\Big)ds$$

Then the uniqueness of solution of the Riccati equation implies $P^0(s,k) = \overline{P}(s,k)$. Therefore,

$P^{\varepsilon}(t, s_{kj}) \to \overline{P}(t, k)$ and the proof is thus concluded. We thus have $v^{\varepsilon}(t, s_{kj}, x) \to \overline{v}(t, k, x)$ as $\varepsilon \to 0$, in which $\overline{v}(t, k, x) = \overline{P}(t, k)(x + (\lambda - z)\overline{H}(t, k))^2$, where $\overline{v}(t, k, x)$ corresponds to the value function of a limit problem. Let $\mathcal{U}$ denote the control set for the limit problem: $\mathcal{U} = \{U = (U^1, U^2, \dots, U^l) : U^k = (u^{k1}, u^{k2}, \dots, u^{km_k}), u^{kj} \in \mathbb{R}^{d_1}\}$. Define

$$f(t, x, k, U) = \sum_{j=1}^{m_k} \mu_j^k r(t, s_{kj})x + \sum_{j=1}^{m_k} \mu_j^k B(t, s_{kj}) u^{kj}(t) \quad \text{and}$$
$$g(t, k, U) = (g_1(t, k, U), \dots, g_d(t, k, U)) \text{ with } g_i(t, k, U) = \sqrt{\sum_{j=1}^{m_k} \mu_j^k (\sum_{n=1}^{d_1} u_n^{kj} \sigma_{ni}(t, \alpha^{\varepsilon}(t)))^2}.$$

Recall that $\sigma(t, \alpha^{\varepsilon}(t)) = (\sigma_{ni}(t, \alpha^{\varepsilon}(t))) \in \mathbb{R}^{d_1 \times d}$ and note that u_n^{kj} is the nth component of the d_1-dimensional variable. The corresponding dynamic system of the state is

$$dx(t) = f(t, x(t), \overline{\alpha}(t), U(t))dt + \sum_{i=1}^{d} g_i(t, \overline{\alpha}(t), U(t))dw_i(t). \tag{3.26}$$

where $\overline{\alpha}(\cdot) \in \{1, 2, \dots, l\}$ is a Markov chain generated by $\overline{Q}$ with $\overline{\alpha}(0) = \alpha$. Calculation similar to (3.9) and (3.10) shows that the optimal control for this limit problem is

$$U^*(t) = (U^{1*}(t, x), U^{2*}(t, x), \dots, U^{l*}(t, x)), \quad \text{with } U^{k*}(t, x) = (u^{k1*}(t, x), u^{k2*}(t, x), \dots, u^{km_k*}(t, x)),$$
$$u^{kj*}(t, x) = -(\sigma(t, s_{kj})\sigma'(t, s_{kj}))^{-1} B'(t, s_{kj})[x + (\lambda - z)\overline{H}(t, k)].$$

In the following, we denote nth component of the optimal control for this limit system as $u_n^{kj*}(t, x)$ Using such controls, we construct

$$u^{\varepsilon}(t, \alpha^{\varepsilon}(t), x) = \sum_{k=1}^{l} \sum_{j=1}^{m_k} I_{\{\alpha^{\varepsilon}(t) = s_{kj}\}} u^{kj*}(t, x) \tag{3.27}$$

for the original problem. This control can also be written as if $\alpha^\varepsilon(t) \in \mathcal{M}_k, u^\varepsilon(t, \alpha^\varepsilon(t), x) = -(\sigma(t, \alpha^\varepsilon(t))\sigma'(t, \alpha^\varepsilon(t)))^{-1}B'(t, \alpha^\varepsilon(t))[x + (\lambda - z)\overline{H}(t, \overline{\alpha}^\varepsilon(t))]$. To proceed, we present the following lemmas first.

Lemma 3.6. *For a positive T and any $k = 1, 2, \ldots, l, j = 1, 2, \ldots, m_k$,*

$$\sup_{0 \le t \le T} E \left| \int_0^t [I_{\{\alpha^\varepsilon(s) = s_{kj}\}} - \mu_j^k I_{\{\overline{\alpha}^\varepsilon(s) = k\}}] x^\varepsilon(s) r(s, s_{kj}) ds \right|^2 \to 0 \text{ as } \varepsilon \to 0. \tag{3.28}$$

The proof is omitted for brevity.

Lemma 3.7. *For any $k = 1, 2, \ldots, l, j = 1, 2, \ldots, m_k$,*

$$E(I_{\{\overline{\alpha}^\varepsilon(s) = k\}} - I_{\{\overline{\alpha}(s) = k\}})^2 \to 0 \text{ as } \varepsilon \to 0. \tag{3.29}$$

Proof. Similar to [83, Theorem 7.30], we can show that $(I_{\{\overline{\alpha}^\varepsilon(\cdot)=1\}}, \ldots, I_{\{\overline{\alpha}^\varepsilon(\cdot)=l\}})$ converges weakly to $(I_{\{\bar{\alpha}(\cdot)=1\}}, \ldots, I_{\{\bar{\alpha}(\cdot)=l\}})$ in $(D[0, T] : \mathbb{R}^l)$ as $\varepsilon \to 0$. By means of Cramér-Word's device, for each $i \in \mathcal{M}$, $I_{\{\overline{\alpha}^\varepsilon(\cdot)=i\}}$ converges weakly to $I_{\{\bar{\alpha}(\cdot)=i\}}$. Then by virtue of the Skorohod representation (with a slight abuse of notation), we may assume $I_{\{\overline{\alpha}^\varepsilon(\cdot)=i\}} \to I_{\{\bar{\alpha}(\cdot)=i\}}$ w.p.1. without change of notation. Now by dominance convergence theorem, we can conclude the proof. □

Theorem 3.8. *The control $u^\varepsilon(t)$ defined in* (3.27) *is nearly optimal in that*

$$\lim_{\varepsilon \to 0} |J^\varepsilon(\alpha, x, u^\varepsilon(\cdot)) - v^\varepsilon(\alpha, x)| = 0.$$

Proof. Recall the definition of $\rho(t, s_{kj})$ in (3.6) and note that the constructed control is given as $u^{\varepsilon}(t,x,\alpha^{\varepsilon}(t)) = -(\sigma(t,\alpha^{\varepsilon}(t))\sigma'(t,\alpha^{\varepsilon}(t)))^{-1}B'(t,\alpha^{\varepsilon}(t))[x+(\lambda-z)\overline{H}(t,\overline{\alpha}^{\varepsilon}(t))]$. Then $x^{\varepsilon}(t)$ follows

$$\begin{aligned} dx^{\varepsilon}(t) &= \sum_{k=1}^{l}\sum_{j=1}^{m_k}[r(t,s_{kj})x^{\varepsilon}(t) - \rho(t,s_{kj})x^{\varepsilon}(t) - \rho(t,s_{kj})(\lambda - z)\overline{H}(t,k)]I_{\{\alpha^{\varepsilon}(t)=s_{kj}\}}dt \\ &\quad + \sum_{i=1}^{d}\sqrt{\sum_{k=1}^{l}\sum_{j=1}^{m_k}(\sum_{n=1}^{d_1}u_n^{\varepsilon}(t,x^{\varepsilon}(t),\alpha^{\varepsilon}(t))\sigma_{ni}(t,\alpha^{\varepsilon}(t)))^2 I_{\{\alpha^{\varepsilon}(t)=s_{kj}\}}}dw_i(t). \\ x^{\varepsilon}(0) &= \hat{x}. \end{aligned}$$

The cost function $J^{\varepsilon}(\alpha,x,u^{\varepsilon}(\cdot)) = E[x^{\varepsilon}(T)+\lambda-z]^2$. Let $x^*(t)$ be the optimal trajectory of the limit problem. Recall the definition of $f(\cdot)$ and $g(\cdot)$ in the Theorem 3.5. Then

$$dx^*(t) = f(t,x^*(t),\overline{\alpha}(t),U^*(t))dt + \sum_{i=1}^{d} g_i(t,\overline{\alpha}(t),U^*(t))dw_i(t),\ \ x^*(0) = \hat{x}.$$

Similar to the methods in [83, Theorem 9.8], for all $\alpha \in \mathcal{M}_k$, and $k = 1, 2, \ldots, l$,

$$\lim_{\varepsilon\to 0} v^{\varepsilon}(x,\alpha) = \overline{v}(x,k).$$

Here $\overline{v}(x,k)$ is the value function of the limit problem. For any $\alpha \in \mathcal{M}_k, k = 1, 2, \ldots, l$,

$$0 \le |J^{\varepsilon}(x,u^{\varepsilon}(\cdot),\alpha) - v^{\varepsilon}(x,\alpha)| = |J^{\varepsilon}(x,u^{\varepsilon}(\cdot),\alpha) - \overline{v}(x,k) + \overline{v}(x,k) - v^{\varepsilon}(x,\alpha)|.$$

To establish the assertion, it suffices to show that

$$|J^{\varepsilon}(x,u^{\varepsilon}(\cdot),\alpha) - \overline{v}(x,k)| \to 0,$$

$$
\begin{aligned}
|J^\varepsilon(x,u^\varepsilon(\cdot),\alpha)-\overline{v}(x,\alpha)| &= |E[x^\varepsilon(T)+\lambda-z]^2-E[x^*(T)+\lambda-z]^2| \\
&= |E(x^\varepsilon(T))^2+2(\lambda-z)Ex^\varepsilon(T)-E(x^*(T))^2-2(\lambda-z)Ex^*(T)| \\
&\le CE^{\frac{1}{2}}[x^\varepsilon(T)-x^*(T)]^2
\end{aligned}
\tag{3.30}
$$

for some constant C. Here, Hölder inequality and finite second moment of $x^\varepsilon(T)$ and $x^*(T)$ are used. Note that we can write $E(x^\varepsilon(T)-x^*(T))^2$ as follows:

$$
\begin{aligned}
&E(x^\varepsilon(T)-x^*(T))^2 \\
&\le K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[r(s,s_{kj})x^\varepsilon(s)(I_{\{\alpha^\varepsilon(s)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(s)=k\}})]ds)^2 \\
&\quad +K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[\mu_j^k r(s,s_{kj})(x^\varepsilon(s)-x^*(s))I_{\{\overline{\alpha}^\varepsilon(s)=k\}}]ds)^2 \\
&\quad +K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[\mu_j^k r(s,s_{kj})x^*(s)(I_{\{\overline{\alpha}^\varepsilon(s)=k\}}-I_{\{\overline{\alpha}(s)=k\}})]ds)^2 \\
&\quad -K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T\rho(s,s_{kj})x^\varepsilon(s)(I_{\{\alpha^\varepsilon(s)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(s)=k\}})ds)^2 \\
&\quad +K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[\mu_j^k\rho(s,s_{kj})(x^\varepsilon(s)-x^*(s))I_{\{\overline{\alpha}^\varepsilon(s)=k\}}]ds)^2 \\
&\quad +K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T\mu_j^k\rho(s,s_{kj})x^*(s)(I_{\{\overline{\alpha}^\varepsilon(s)=k\}}-I_{\{\overline{\alpha}(s)=k\}})ds)^2 \\
&\quad -K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[\rho(s,s_{kj})(\lambda-z)\overline{H}(s,k)(I_{\{\alpha^\varepsilon(s)=k\}}-\mu_j^k I_{\{\overline{\alpha}^\varepsilon(s)=k\}})]ds)^2 \\
&\quad +K\sum_{k=1}^{l}\sum_{j=1}^{m_k}E(\int_0^T[\rho(s,s_{kj})(\lambda-z)\overline{H}(s,k)\mu_j^k(I_{\{\overline{\alpha}^\varepsilon(s)=k\}}-I_{\{\overline{\alpha}(s)=k\}})]ds)^2+D,
\end{aligned}
\tag{3.31}
$$

where

$$
\begin{aligned}
D = KE\Bigg[\int_0^T\sum_{i=1}^{d}\Big[&\sqrt{\sum_{k=1}^{l}\sum_{j=1}^{m_k}(\sum_{n=1}^{d_1}u_n^\varepsilon(s,x^\varepsilon(s),\alpha^\varepsilon(s))\sigma_{ni}(s,\alpha^\varepsilon(s)))^2 I_{\{\alpha^\varepsilon(s)=s_{kj}\}}} \\
&-\sqrt{\sum_{k=1}^{l}\sum_{j=1}^{m_k}\mu_j^k(\sum_{n=1}^{d_1}u_n^{kj*}(s,x^*(s),\overline{\alpha}(s))\sigma_{ni}(s,\alpha^\varepsilon(s)))^2 I_{\{\overline{\alpha}(s)=k\}}}\Big]dw_i(s)\Bigg]^2.
\end{aligned}
$$

First, we use Lemma 3.6, Lemma 3.7, and Hölder inequality repeatedly to handel the drift part. For the diffusion part, realizing that

$$
\begin{aligned}
D \le KE\int_0^T \sum_{i=1}^{d}\Bigg[&\sqrt{\sum_{k=1}^{l}\sum_{j=1}^{m_k}(\sum_{n=1}^{d_1} u_n^{\varepsilon}(s,x^{\varepsilon}(s),\alpha^{\varepsilon}(s))\sigma_{ni}(s,\alpha^{\varepsilon}(s)))^2[I_{\{\alpha^{\varepsilon}(s)=s_{kj}\}}-\mu_j^k I_{\{\overline{\alpha}^{\varepsilon}(s)=k\}}]} \\
&+\sqrt{\sum_{k=1}^{l}\sum_{j=1}^{m_k}\mu_j^k(\sum_{n=1}^{d_1} u_n^{kj*}(s,x^*(s),\overline{\alpha}(s))\sigma_{ni}(s,\alpha^{\varepsilon}(s)))^2[I_{\{\overline{\alpha}^{\varepsilon}(s)=k\}}-I_{\{\overline{\alpha}(s)=k\}}]} \\
&+(x^{\varepsilon}(s)-x^*(s))\Bigg]^2 ds.
\end{aligned}
$$

Here, we plugged in the control constructed in (3.27) for the last term above and utilized the non-degeneracy assumption mentioned in the previous section. Then we can use property of stochastic integral, dominance convergence theorem, similar techniques involved in dealing with the drift part and the finite second moment of $x^{\varepsilon}(\cdot)$ and $x^*(\cdot)$ to proceed with the diffusion part. Finally, after detailed calculation, we have $E(x^{\varepsilon}(T)-x^*(T))^2 \le o(\varepsilon)+K\int_0^T E(x^{\varepsilon}(s)-x^*(s))^2 ds$. Now with the help of Gronwall's inequality, we obtain $E(x^{\varepsilon}(T)-x^*(T))^2 \to 0$ as $\varepsilon \to 0$. The proof is thus concluded.

4 Mean-Variance Type Controls Involving a Hidden Markov Chain

4.1 Formulation

In this section, we consider the mean variance control problem but under the assumption that the switching process is given as a hidden Markov chain. Our objective is again to find an $\mathcal{F}_t$ admissible control $u(\cdot)$ in a compact set $\mathcal{U}$ under the constraint that the expected terminal flow is $Ex(T) = \kappa$ for some given $\kappa \in \mathbb{R}$, so that the risk measured by the variance of terminal flow at time T is minimized. Specifically, we have the following goal

$$\begin{aligned} &\min\ J(s, x, p, u(\cdot)) := E[x(T) - \kappa]^2 \\ &\text{subject to } Ex(T) = \kappa. \end{aligned} \tag{4.1}$$

where $x(t)$ is the total flows for the whole networked system and we have the same dynamics of $x(t)$ as previous chapter. Also,

$$x(t) = \sum_{l=1}^{d+1} N_l(t) x_l(t), t \geq s.$$

where $N_l(t)$ is the proportion that we need to put to that of the lth node at time t. However, different transition rates are considered in the previous chapter. In this chapter, instead of having full information of the Markov chain, we can only observe it in white noise. That is, we observe $y(t)$, whose dynamics is given by

$$\begin{aligned} dy(t) &= g(\alpha(t))dt + \sigma_0 dw_2(t), \\ y(s) &= 0, \end{aligned} \tag{4.2}$$

where $\sigma_0 > 0$ and $w_2(\cdot)$ is a standard scalar Brownian motion, where $w_2(\cdot)$, $w(\cdot)$, and $\alpha(\cdot)$ are independent. Moreover, the initial data $p(s) = p = (p^1, p^2, \ldots, p^m)$ in which $p^i = p^i(s) = P(\alpha(s) = i)$ is given for $1 \le i \le m$.

4.2 Key Results and Proofs

Note that one of the striking feature of our model is that we have no access to the value of Markov chain at a given time t, which makes the problem more difficult than [87]. Let $p(t) = (p^1(t), \ldots, p^m(t)) \in \mathbb{R}^{1\times m}$ with $p^i(t) = P(\alpha(t) = i|\mathcal{F}^y(t))$ for $i = 1, 2, \ldots, m$, with $\mathcal{F}^y(t) = \sigma\{y(\widetilde{s}) : s \le \widetilde{s} \le t\}$. It was shown in Wonham [65] that this conditional probability satisfies the following system of stochastic differential equations

$$\begin{aligned} dp^i(t) &= \sum_{j=1}^{m} q^{ji} p^j(t)dt + \frac{1}{\sigma_0} p^i(t)(g(i) - \overline{\alpha}(t))d\widehat{w}_2(t), \\ p^i(s) &= p^i, \end{aligned} \tag{4.3}$$

where $\overline{\alpha}(t) = \sum_{i=1}^{m} g(i)p^i(t)$ and $\widehat{w}_2(t)$ is the innovation process. It is easy to see that $\widehat{w}_2(\cdot)$ is independent of $w(\cdot)$. With the help of Wonham filter, given the independence conditions, we can find the best estimator for $r(t, \alpha(t))$, $B(t, \alpha(t))$, and $\tilde{\sigma}(t, \alpha(t))$ in the sense of least mean square prediction error and thus transform the partial observable system into completely observable system given as below:

$$dx(t) = [r\widehat{(t,\alpha(t))}x(t) + B\widehat{(t,\alpha(t))}u(t)]dt + u'(t)\tilde{\sigma}\widehat{(t,\alpha(t))}dw(t),$$

where

$$\begin{aligned}
r\widehat{(t,\alpha(t))} &\stackrel{\text{def}}{=} \sum_{i=1}^{m} r(t,i)p^i(t) \in \mathbb{R}^1, \\
B\widehat{(t,\alpha(t))} &\stackrel{\text{def}}{=} (\sum_{i=1}^{m}(b_2(t,i) - r(t,i))p^i(t), \ldots, \sum_{i=1}^{m}(b_{d_1+1}(t,i) - r(t,i))p^i(t)) \in \mathbb{R}^{1\times d_1}, \\
\tilde{\sigma}\widehat{(t,\alpha(t))} &\stackrel{\text{def}}{=} (\sum_{i=1}^{m} \tilde{\sigma}_{lj}(t,i)p^i(t))_{d_1\times d}.
\end{aligned} \tag{4.4}$$

In this way, by putting the two components $p(t)$ and $x(t)$ together, we get

$$(x(t), p(t)) = (x(t), p^1(t), ..., p^m(t)),$$

a completely observable system whose dynamics are as follows

$$\begin{aligned}
dx(t) &= [\sum_{i=1}^{m} r(t,i)p^i(t)x(t) + \sum_{l=2}^{d_1+1}\sum_{i=1}^{m}(b_l(t,i) - r(t,i))p^i(t)u_l(t)]dt \\
&\quad + \sum_{l=2}^{d_1+1}\sum_{j=1}^{d}\sum_{i=1}^{m} u_l(t)\tilde{\sigma}_{lj}(t,i)p^i(t)dw_1^j(t) \\
&= b(x(t), p(t), u(t))dt + \sigma(x(t), p(t), u(t))dw(t) \\
dp^i(t) &= \sum_{j=1}^{m} q^{ji}p^j(t)dt + \frac{1}{\sigma_0}p^i(t)(g(i) - \overline{\alpha}(t))d\widehat{w}_2(t), \text{ for } i = \{1, \ldots, m\} \\
x(s) &= x, \quad p^i(s) = p^i.
\end{aligned} \tag{4.5}$$

Let $W(s,x,p,u)$ be the objective function and let $E^u_{s,x,p}$ denote the expectation of functionals on $[s,T]$ conditioned on $x(s)=x, p(s)=p$ and the admissible control $u=u(\cdot)$.

$$W(s,x,p,u)=E^{u}_{s,x,p}(x(T)+\lambda-k)^2-\lambda^2 \tag{4.6}$$

and $V(s,x,p)$ be the value function

$$V(s,x,p)=\inf\nolimits_{u\in\mathcal{U}} W(s,x,p,u). \tag{4.7}$$

To proceed, for an arbitrary $r\in\mathcal{U}$ and $\phi(\cdot,\cdot,\cdot)\in C^{1,2,2}(\mathbb{R})$, we first define the differential operator $\mathcal{L}^r$ by

$$\begin{aligned}\mathcal{L}^r\phi(s,x,p)=&\frac{\partial\phi}{\partial s}+\frac{\partial\phi}{\partial x}b(x,p,r)+\frac{1}{2}\frac{\partial^2\phi}{\partial x^2}[\sigma(x,p,r)\sigma'(x,p,r)]\\&+\sum_{i=1}^{m}\frac{\partial\phi}{\partial p^i}\sum_{j=1}^{m}q^{ji}p^j+\frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2\phi}{\partial(p^i)^2}\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2.\end{aligned} \tag{4.8}$$

The value function is the solution of the following system of HJB equation

$$\inf\nolimits_{r\in\mathcal{U}}\mathcal{L}^rV(s,x,p)=0, \tag{4.9}$$

with boundary condition $V(T,x,p)=(x(T)+\lambda-\kappa)^2-\lambda^2$.

Note that there is little hope that we can find closed form solution for this problem. In our work, we work on finding the numerical solution for this question. Let $v^i(t)=\log p^i(t)$, by choosing the constant step size $h_2>0$ for time variable we can discrete the dynamic of

$v^i(t)$ as follows:

$$
\begin{aligned}
v_{n+1}^{h_2,i} &= v_n^{h_2,i} + h_2[\sum_{j=1}^{m} q^{ji}\frac{p_n^{h_2,j}}{p_n^{h_2,i}} - \frac{1}{2\sigma_0^2}(g(i)-\bar{\alpha}_n^{h_2})^2] + \sqrt{h_2}\frac{1}{\sigma_0}(g(i)-\bar{\alpha}_n^{h_2})\varepsilon_n,\\
v_0^{h_2,i} &= \log(p^i),\\
p_{n+1}^{h_2,i} &= \exp(v_{n+1}^{h_2,i}),\\
p_0^{h_2,i} &= p^i,
\end{aligned}
\tag{4.10}
$$

where $\bar{\alpha}_n^{h_2} = \sum_{i=1}^m g(i)p_n^{h_2,i}$ and $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables satisfying $E\varepsilon_n = 0$, $E\varepsilon_n^2 = 1$, and $E|\varepsilon_n|^{2+\gamma} < \infty$ for some $\gamma > 0$ with

$$
\varepsilon_n = \frac{\widehat{w}_2((n+1)h_2) - \widehat{w}_2(nh_2)}{\sqrt{h_2}}.
$$

Note that $p_n^{h_2,i}$ appeared as the denominator in (4.10) and we have focused on the case that $p_n^{h_2,i}$ stays away from 0. A modification can be made to take into consideration the case of $p_n^{h_2,i} = 0$. In that case, we can choose a fixed yet arbitrarily large positive real number M and use the idea of penalization to construct the approximation as below:

$$
\begin{aligned}
v_{n+1}^{h_2,i} &= v_n^{h_2,i} + h_2\{[\sum_{j=1}^{m} q^{ji}\frac{p_n^{h_2,j}}{p_n^{h_2,i}} - \frac{1}{2\sigma_0^2}(g(i)-\bar{\alpha}_n^{h_2})^2]I_{\{p_n^{h_2,i}\geq e^{-M}\}} - MI_{\{p_n^{h_2,i}<e^{-M}\}}\}\\
&\qquad + \sqrt{h_2}\frac{1}{\sigma_0}(g(i)-\bar{\alpha}_n^{h_2})\varepsilon_n,\\
v_0^{h_2,i} &= \log(p^i),\\
p_{n+1}^{h_2,i} &= \exp(v_{n+1}^{h_2,i}),\\
p_0^{h_2,i} &= p^i.
\end{aligned}
\tag{4.11}
$$

In what follows, we construct a discrete-time finite state Markov chain to approximate the controlled diffusion process, $x(t)$. Our construction of Markov chain takes care of time and state variables as follows. Recall that $h_2 > 0$ is the step size for time variable and let $N_{h_2} = (T-s)/h_2$ be an integer. Let $h_1 > 0$ be a discretization parameter for state variables and define $S_{h_1} = \{x : x = kh_1, k = 0, \pm 1, \pm 2, \ldots\}$. We use $u_n^{h_1,h_2}$ to denote the control action for the chain at discrete time n. Let $u^{h_1,h_2} = (u_0^{h_1,h_2}, u_1^{h_1,h_2}, \ldots)$ denote the sequence of $\mathcal{U}$-valued random variables that are the control actions at time $0, 1, \ldots$ and $p^{h_2} = (p_0^{h_2}, p_1^{h_2}, \ldots)$ are the corresponding posterior probability in which $p_n^{h_2} = (p_n^{h_2,1}, p_n^{h_2,2}, \ldots, p_n^{h_2,m})$. We define the difference $\Delta\xi_n^{h_1,h_2} = \xi_{n+1}^{h_1,h_2} - \xi_n^{h_1,h_2}$ and let $E_{x,p,n}^{h_1,h_2,r}$, $V_{x,p,n}^{h_1,h_2,r}$ denote the conditional expectation and variance given $\{\xi_k^{h_1,h_2}, u_k^{h_1,h_2}, p_k^{h_2}, k \le n, \xi_n^{h_1,h_2} = x, p_n^{h_2} = p, u_n^{h_1,h_2} = r\}$. By stating that $\{\xi_n^{h_1,h_2}, n < \infty\}$ is a controlled discrete-time Markov chain on a discrete time state space S_{h_1} with transition probabilities $p^{h_1,h_2}((x,y)|r,p)$, we mean that the transition probabilities are functions of a control variable r and posterior probability p. The sequence $\{\xi_n^{h_1,h_2}, n < \infty\}$ is said to be locally consistent with (4.5), if it satisfies

$$\begin{aligned}
&E_{x,p,n}^{h_1,h_2,r}\Delta\xi_n^{h_1,h_2} = b(x,p,r)h_2 + o(h_2),\\
&V_{x,p,n}^{h_1,h_2,r}\Delta\xi_n^{h_1,h_2} = \sigma(x,p,r)\sigma'(x,p,r)h_2 + o(h_2),\\
&\sup_n |\Delta\xi_n^{h_1,h_2}| \to 0, \text{ as } h_1, h_2 \to 0.
\end{aligned} \tag{4.12}$$

Let $\mathcal{U}^{h_1,h_2}$ denote the collection of ordinary controls, which is determined by a sequence of such measurable functions $F_n^{h_1,h_2}(\cdot)$ that $u_n^{h_1,h_2} = F_n^{h_1,h_2}(\xi_k^{h_1,h_2}, p_k^{h_2}, k \le n, u_k^{h_1,h_2}, k < n)$. We say that u^{h_1,h_2} is admissible for the chain if $u_n^{h_1,h_2}$ are $\mathcal{U}$ valued random variables and the

Markov property continues to hold under the use of the sequence $\{u_n^{h_1,h_2}\}$, namely,

$$\begin{aligned}
&P\{\xi_{n+1}^{h_1,h_2} = y | \xi_k^{h_1,h_2}, u_k^{h_1,h_2}, p_k^{h_2}, k \leq n\} \\
&= P\{\xi_{n+1}^{h_1,h_2} = y | \xi_n^{h_1,h_2}, u_n^{h_1,h_2}, p_n^{h_2}\} = p^{h_1,h_2}((\xi_n^{h_1,h_2}, y) | u_n^{h_1,h_2}, p_n^{h_2}).
\end{aligned}$$

Using the Markov chain given above, we can approximate the objective function defined in (4.6) by

$$W^{h_1,h_2}(s, x, p, u^{h_1,h_2}) = E_{s,x,p}^{u^{h_1,h_2}} (\xi_{N_{h_2}}^{h_1,h_2} + \lambda - k)^2 - \lambda^2. \tag{4.13}$$

Here, $E_{s,x,p}^{u^{h_1,h_2}}$ denotes the expectation given that $\xi_0^{h_1,h_2} = x, p_0^{h_2} = p$ and that an admissible control sequence $u^{h_1,h_2} = \{u_n^{h_1,h_2}, n < \infty\}$ is used. Now we need that the approximating Markov chain constructed above satisfies local consistency, which is one of the necessary conditions for weak convergence. To find a reasonable Markov chain that is locally consistent, we first suppose that control space has a unique admissible control $u^{h_1,h_2} \in \mathcal{U}^{h_1,h_2}$, so that we can drop inf in (4.9). We discrete (4.8) by the following finite difference method using step-size $h_1 > 0$ for state variables and $h_2 > 0$ for time variable as mentioned above:

$$V(t, x, p) \to V^{h_1,h_2}(t, x, p). \tag{4.14}$$

For the derivative with respect to time variable, we use

$$V_t(t, x, p) \to \frac{V^{h_1,h_2}(t + h_2, x, p) - V^{h_1,h_2}(t, x, p)}{h_2}. \tag{4.15}$$

For the first derivative with respect to x, we use one-side difference method

$$V_x(t,x,p) \to \begin{cases} \dfrac{V^{h_1,h_2}(t+h_2,x+h_1,p)-V^{h_1,h_2}(t+h_2,x,p)}{h_1} & \text{for } b(x,p,r)\geq 0 \\ \dfrac{V^{h_1,h_2}(t+h_2,x,p)-V^{h_1,h_2}(t+h_2,x-h_1,p)}{h_1} & \text{for } b(x,p,r)<0. \end{cases} \tag{4.16}$$

For the second derivative with respect to x, we have standard difference method

$$V_{xx}(t,x,p) \to \frac{V^{h_1,h_2}(t+h_2,x+h_1,p)+V^{h_1,h_2}(t+h_2,x-h_1,p)-2V^{h_1,h_2}(t+h_2,x,p)}{h_1^2}. \tag{4.17}$$

For the first and second derivatives with respect to the posterior probability p^i, we also have similar expressions as above. Let $V^{h_1,h_2}(t,x,p)$ denote the solution to the finite difference equation with x and p^i be an integral multiplier of h_1 and h_2. Plugging all the necessary expressions into (4.9), combining the like terms and multiplying all terms by h_2 yield the following expression:

$$\begin{aligned}
&V^{h_1,h_2}(nh_2,x,p) \\
&\quad = V^{h_1,h_2}(nh_2+h_2,x,p)[1-\frac{|b(x,p,r)|h_2}{h_1}-\frac{h_2\sigma(x,p,r)\sigma'(x,p,r)}{h_1^2}] \\
&\quad +V^{h_1,h_2}(nh_2+h_2,x+h_1,p)\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2+2h_1h_2b^+(x,p,r)}{2h_1^2} \\
&\quad +V^{h_1,h_2}(nh_2+h_2,x-h_1,p)\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2+2h_1h_2b^-(x,p,r)}{2h_1^2} \\
&\quad +\sum_{i=1}^m V^{h_1,h_2}(nh_2+h_2,x,p^i+h_1)\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2+2h_1(\sum_{j=1}^m q^{ji}p^j)^+h_2}{2h_1^2} \\
&\quad +\sum_{i=1}^m V^{h_1,h_2}(nh_2+h_2,x,p^i-h_1)\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2+2h_1(\sum_{j=1}^m q^{ji}p^j)^-h_2}{2h_1^2} \\
&\quad +\sum_{i=1}^m V^{h_1,h_2}(nh_2+h_2,x,p^i)[-\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2}{h_1^2}-\frac{h_2|\sum_{j=1}^m q^{ji}p^j|}{h_1}],
\end{aligned} \tag{4.18}$$

where $b^+(x,p,r)$, $(\sum_{j=1}^m q^{ji}p^j)^+$ and $b^-(x,p,r)$, $(\sum_{j=1}^m q^{ji}p^j)^-$ are positive and negative parts of $b(x,p,r)$ and $\sum_{j=1}^{m} q^{ji}p^{j}$, respectively and $nh_2 < T$. Note the sum of the coefficients of the first three lines in the above equation is unity. By choosing proper h_1 and h_2, we can reasonably assume that the coefficient

$$1 - \frac{|b(x,p,r)|h_2}{h_1} - \frac{h_2\sigma(x,p,r)\sigma'(x,p,r)}{h_1^2}$$

of term $V^{h_1,h_2}(nh_2 + h_2, x, p)$ is in $[0,1]$. Therefore, we can regard the coefficients as the transition functions of a Markov chain and define the transition probabilities in the following way,

$$\begin{aligned}
p^{h_1,h_2}((nh_2, nh_2+h_2))|x,p,r) &= 1 - \frac{|b(x,p,r)|h_2}{h_1} - \frac{h_2\sigma(x,p,r)\sigma'(x,p,r)}{h_1^2} \\
p^{h_1,h_2}((nh_2,x),(nh_2+h_2,x+h_1)|p,r) &= \frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^{+}(x,p,r)}{2h_1^2} \\
p^{h_1,h_2}((nh_2,x),(nh_2+h_2,x-h_1)|p,r) &= \frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^{-}(x,p,r)}{2h_1^2}.
\end{aligned} \tag{4.19}$$

Theoretically, we can find approximation of $V(s,x,p)$ in (4.7) by using (4.13) and

$$V^{h_1,h_2}(s,x,p) = \inf_{u^{h_1,h_2} \in \mathcal{U}^{h_1,h_2}} W^{h_1,h_2}(s,x,p,u^{h_1,h_2}). \tag{4.20}$$

Practically, with the transition probabilities defined above, we can compute $V^{h_1,h_2}(s,x,p)$

by the following iteration method

$$
\begin{aligned}
&V^{h_1,h_2}(nh_2,x,p)\\
&\quad = p^{h_1,h_2}((nh_2,x)(nh_2+h_2,x+h_1)|p,r)V^{h_1,h_2}(nh_2+h_2,x+h_1,p)\\
&\quad\; +p^{h_1,h_2}((nh_2,x),(nh_2+h_2,x-h_1)|p,r)V^{h_1,h_2}(nh_2+h_2,x-h_1,p)\\
&\quad\; +p^{h_1,h_2}((nh_2,nh_2+h_2)|x,p,r)V^{h_1,h_2}(nh_2+h_2,x,p)\\
&\quad\; +\sum_{i=1}^{m}V^{h_1,h_2}(nh_2+h_2,x,p^i+h_1)\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2+2h_1(\sum_{j=1}^m q^{ji}p^j)^+h_2}{2h_1^2}\\
&\quad\; +\sum_{i=1}^{m}V^{h_1,h_2}(nh_2+h_2,x,p^i-h_1)\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2+2h_1(\sum_{j=1}^m q^{ji}p^j)^-h_2}{2h_1^2}\\
&\quad\; +\sum_{i=1}^{m}V^{h_1,h_2}(nh_2+h_2,x,p^i)[-\frac{\frac{1}{\sigma_0^2}[p^i(g(i)-\overline{\alpha})]^2h_2}{h_1^2}-\frac{h_2|\sum_{j=1}^m q^{ji}p^j|}{h_1}].
\end{aligned}
\tag{4.21}
$$

Note that we used local transitions here, we can avoid the problem of “numerical noise” or “numerical viscosity” in this way, which appears in non-local transitions case, and is even more serious in higher dimension scenario, see [57] for more details. We can show that the Markov chain $\{\xi_n^{h_1,h_2}, n<\infty\}$ with transition probabilities $p^{h_1,h_2}(\cdot)$ defined in (4.19) is locally consistent with (4.5) by verifying the following equations:

$$
\begin{aligned}
&E^{h_1,h_2,r}_{x,p,n}\Delta\xi^{h_1,h_2}_n \\
&\quad = h_1\left(\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^+(x,p,r)}{2h_1^2}\right) \\
&\quad\quad -h_1\left(\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^-(x,p,r)}{2h_1^2}\right) \\
&\quad = b(x,p,r)h_2, \\
&V^{h_1,h_2,r}_{x,p,n}\Delta\xi^{h_1,h_2}_n \\
&\quad = h_1^2\left(\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^+(x,p,r)}{2h_1^2}\right) \\
&\quad\quad +h_1^2\left(\frac{\sigma(x,p,r)\sigma'(x,p,r)h_2 + 2h_1h_2b^-(x,p,r)}{2h_1^2}\right) \\
&\quad = \sigma(x,p,r)\sigma'(x,p,r)h_2 + O(h_1h_2).
\end{aligned}
\tag{4.22}
$$

It turns out that in convergence analysis, the classical control is inadequate so we need to enlarge our control class to include relaxed control $m(\cdot)$ and utilize the idea of martingale measure $M(\cdot)$ to proceed. Under certain conditions, we can actually rewrite our original system as

$$
\begin{aligned}
x(t) &= x + \int_s^t\int_{\mathcal{U}} b(x(z),p(z),c)m_z(dc)dz + \int_s^t\int_{\mathcal{U}}\sigma(x(z),p(z),c)M(dc,dz) \\
p^i(t) &= \int_s^t\sum_{j=1}^m q^{ji}p^j(z)dz + \int_s^t\frac{1}{\sigma_0}[p^i(z)(g(i)-\overline{\alpha}(z))]d\widehat{w}_2(z),\ \text{for } i=\{1,\dots,m\},
\end{aligned}
\tag{4.23}
$$

where

$$\sigma(x(z),p(z),c) = (\sigma_1(x(z),p(z),c),\dots,\sigma_d(x(z),p(z),c)) \in \mathbb{R}^{1\times d}.$$

Equation (4.23) represents our control system.

In order to approximate the continuous time process $(x(t),p(t),M(t),m(t))$, we use continuous-time interpolation. We define the piecewise constant interpolations by

$$\begin{aligned}
&\xi^{h_1,h_2}(t) = \xi_n^{h_1,h_2}, \quad p^{h_2}(t) = p_n^{h_2}, \bar{\alpha}^{h_1,h_2}(t) = \sum_{i=1}^{m} g(i)p_n^{h_2}, u^{h_1,h_2}(t) = u_n^{h_1,h_2}, \\
&z^{h_2}(t) = n, w_l^{h_1,h_2}(t) = \sum_{k=0}^{z^{h_2}(t)-1} \Delta w_{l,k}^{h_1,h_2}, \varepsilon^{h_1,h_2}(t) = \varepsilon_n^{h_1,h_2}, \quad \text{for} \quad t \in [nh_2, (n+1)h_2).
\end{aligned} \tag{4.24}$$

The following lemma demonstrate the fact that we can approximate $(x(t), p(t), M(t), m(t))$ by a quadruple satisfying

$$\begin{aligned}
\xi^{h_1,h_2}(t) &= x + \int_s^t \int_{\mathcal{U}} b(\xi^{h_1,h_2}(z), p^{h_2}(z), c) m_z^{h_1,h_2}(dc)dz \\
&\quad + \int_s^t \int_{\mathcal{U}} \sigma(\xi^{h_1,h_2}(z), p^{h_2}(z), c) M^{h_1,h_2}(dc, dz) + \varepsilon^{h_1,h_2}(t) \\
&= x + \int_s^t \sum_l b(\xi^{h_1,h_2}(z), p^{h_2}(z), c_l) m_z(C_l^{h_1,h_2}) dz \\
&\quad + \int_s^t \sum_l \sigma(\xi^{h_1,h_2}(z), p^{h_2}(z), c_l)(m_z(C_l^{h_1,h_2}))^{\frac{1}{2}} dw_l^{h_1,h_2}(z) + \varepsilon^{h_1,h_2}(t),
\end{aligned} \tag{4.25}$$

where $m^{h_1,h_2}(\cdot)$ is a piecewise constant and takes finitely many values and $M^{h_1,h_2}(\cdot)$ is represented in terms of a finite number of Wiener process. The idea is similar to the method used in [56, Theorem 8.1], we omit the detail here for brevity. With the notation of relaxed control given above, we can rewrite the value function as

$$W^{h_1,h_2}(s, x, p, m^{h_1,h_2}) = E_{s,x,p}^{m^{h_1,h_2}} (\xi^{h_1,h_2}(T) + \lambda - k)^2 - \lambda^2. \tag{4.26}$$

$$V^{h_1,h_2}(s, x, p) = \inf_{m^{h_1,h_2} \in \Gamma^{h_1,h_2}} W^{h_1,h_2}(s, x, p, m^{h_1,h_2}). \tag{4.27}$$

Theorem 4.1. *Under certain assumptions and letting the approximating chain $\{\xi_n^{h_1,h_2}, n < \infty\}$ be constructed with transition probability and $p_n^{h_2}$ defined properly. Let $\{u_n^{h_1,h_2}, n < \infty\}$*

be a sequence of admissible controls, $\xi^{h_1,h_2}(\cdot)$ *and* $p^{h_2}(\cdot)$ *be the continuous time interpolation defined in* (4.24)*,* $m^{h_1,h_2}(\cdot)$ *be the relaxed control representation of* $u^{h_1,h_2}(\cdot)$ (*continuous time interpolation of* $u_n^{h_1,h_2}$)*. Then* $\{\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), m^{h_1,h_2}(\cdot), M^{h_1,h_2}(\cdot)\}$ *is tight. Denoting the limit of a weakly convergent subsequence by* $\{x(\cdot), p(\cdot), m(\cdot), M(\cdot)\}$ *such that* (4.23) *is satisfied.*

Proof. Note that $m^{h_1,h_2}(\cdot)$ is tight due to the compactness of the relaxed control. Since $(\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot)) \in \mathbb{R}^{m+1}$, the tightness of $p^{h_2}(\cdot)$ can be obtained as in [68, Theorem 8.15]. Therefore, we just need to take care that of $\xi^{h_1,h_2}(\cdot)$ now. For the tightness of $\xi^{h_1,h_2}(\cdot)$, by assumption (A1), for $s \leq t \leq T$,

$$\begin{aligned} E_{s,x,p}^{m^{h_1,h_2}} |\xi^{h_1,h_2}(t) - x|^2 = E_{s,x,p}^{m^{h_1,h_2}} | &\int_s^t \int_{\mathcal{U}} b(\xi^{h_1,h_2}(z), p^{h_2}(z), c) m_z^{h_1,h_2}(dc)dz \\ &+ \int_s^t \int_{\mathcal{U}} \sigma(\xi^{h_1,h_2}(z), p^{h_2}(z), c) M^{h_1,h_2}(dc, dz) + \varepsilon^{h_1,h_2}(t)|^2 \\ &\leq Kt^2 + Kt + \varepsilon^{h_1,h_2}(t), \end{aligned} \tag{4.28}$$

where $\limsup_{h_1,h_2 \to 0} E|\varepsilon^{h_1,h_2}(t)| \to 0$ for any $s \leq t \leq T$. Similarly, we can guarantee $E_{s,x,p}^{m^{h_1,h_2}} |\xi^{h_1,h_2}(t+\delta) - \xi^{h_1,h_2}(t)|^2 = O(\delta) + \varepsilon^{h_1,h_2}(\delta)$. Therefore, the tightness of $\xi^{h_1,h_2}(\cdot)$ follows. By the compactness of set $\mathcal{U}$, we can see that $M^{h_1,h_2}(\cdot)$ is also tight. In view of the tightness, we can extract a weakly convergent subsequence, and denote its limit by $\{x(\cdot), p(\cdot), m(\cdot), M(\cdot)\}$. Next we show that the limit is the solution of SDE driven by $(p(\cdot), m(\cdot), M(\cdot))$.

For $\delta > 0$ and any process $\nu(\cdot)$ define the process $\nu^\delta(\cdot)$ by $\nu^\delta(t) = \nu(n\delta)$ for $t \in [n\delta, n\delta + \delta)$.

Then by the tightness of $\xi^{h_1,h_2}(\cdot)$ and $p^{h_2}(\cdot)$, (4.25) can be rewritten as

$$\begin{aligned}\xi^{h_1,h_2}(t) = x &+ \int_s^t \int_{\mathcal{U}} b(\xi^{h_1,h_2}(z), p^{h_2}(z), c) m_z^{h_1,h_2}(dc)dz \\ &+ \int_s^t \int_{\mathcal{U}} \sigma(\xi^{h_1,h_2,\delta}(z), p^{h_2,\delta}(z), c) M^{h_1,h_2}(dc, dz) + \varepsilon^{h_1,h_2,\delta}(t),\end{aligned} \tag{4.29}$$

where $\lim_{\delta\to 0}\limsup_{h_1,h_2\to 0} E|\varepsilon^{h_1,h_2,\delta}(t)| \to 0$. We further assume that the probability space is chosen as required by Skorohod representation. Therefore, we can assume the sequence $\{\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), m^{h_1,h_2}(\cdot), M^{h_1,h_2}(\cdot)\}$ converges to $(x(\cdot), p(\cdot), m(\cdot), M(\cdot))$ w.p.1 with a little bit abuse of notation. Taking limit as $h_1 \to 0$ and $h_2 \to 0$, the convergence of

$$\{\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), m^{h_1,h_2}(\cdot), M^{h_1,h_2}(\cdot)\}$$

to its limit w.p.1 implies that

$$E\left|\int_s^t \int_{\mathcal{U}} b(\xi^{h_1,h_2}(z), p^{h_2}(z), c) m_z^{h_1,h_2}(dc)dz - \int_s^t \int_{\mathcal{U}} b(x(z), p(z), c) m_z^{h_1,h_2}(dc)dz\right| \to 0$$

uniformly in t. Also, recall that $m^{h_1,h_2}(\cdot) \to m(\cdot)$ in the “compact weak” topology if and only if

$$\int_s^t \int_{\mathcal{U}} \phi(c, z) m^{h_1,h_2}(dc, dz) \to \int_s^t \int_{\mathcal{U}} \phi(c, z) m(dc, dz)$$

for any continuous and bounded function $\phi(\cdot)$ with compact support. Thus, weak convergence and Skorohod representation imply that

$$\int_s^t \int_{\mathcal{U}} b(x(z), p(z), c) m_z^{h_1,h_2}(dc)dz \to \int_s^t \int_{\mathcal{U}} b(x(z), p(z), c) m_z(dc)dz \text{ as } h_1, h_2 \to 0, \tag{4.30}$$

uniformly in t on any bounded interval w.p.1.

Recall that $M^{h_1,h_2}(\cdot)$ is a martingale measure with quadratic variation process $m^{h_1,h_2}(\cdot)$ and that $\xi^{h_1,h_2,\delta}(\cdot)$ and $p^{h_2,\delta}(\cdot)$ are piecewise constant functions, following from the probability one convergence, we have

$$\int_s^t \int_{\mathcal{U}} \sigma(\xi^{h_1,h_2,\delta}(z), p^{h_2,\delta}(z), c) M^{h_1,h_2}(dc, dz) \to \int_s^t \int_{\mathcal{U}} \sigma(x^\delta(z), p^\delta(z), c) M^{h_1,h_2}(dc, dz). \tag{4.31}$$

Recall that $M^{h_1,h_2}(\cdot) \to M(\cdot)$ in the "compact weak" topology if and only if

$$\int_s^t \int_{\mathcal{U}} f(c, z) M^{h_1,h_2}(dc, dz) \to \int_s^t \int_{\mathcal{U}} f(c, z) M(dc, dz) \text{ as } h_1, h_2 \to 0$$

for each bounded and continuous function $f(\cdot)$, we have

$$\int_s^t \int_{\mathcal{U}} \sigma(x^\delta(z), p^\delta(z), c) M^{h_1,h_2}(dc, dz) \to \int_s^t \int_{\mathcal{U}} \sigma(x^\delta(z), p^\delta(z), c) M(dc, dz),$$

uniformly in t on any bounded interval w.p.1; see [58, pp. 352]. Combining the above results, we have

$$x(t) = x + \int_s^t \int_{\mathcal{U}} b(x(z), p(z), c) m(dc, dz) + \int_s^t \int_{\mathcal{U}} \sigma(x^\delta(z), p^\delta(z), c) M(dc, dz) + \varepsilon^\delta(t), \tag{4.32}$$

where $\lim_{\delta \to 0} E|\varepsilon^\delta(t)| = 0$. Taking limit of the above equation as $\delta \to 0$ yields (4.23).

Theorem 4.2. *Under assumptions* (A1)-(A4), $V^{h_1,h_2}(s, x, p)$ *and* $V(s, x, p)$ *are value func-*

tions defined in (4.27) *and* (4.7) *respectively, we have*

$$V^{h_1,h_2}(s,x,p) \to V(s,x,p), \ \ as \ h_1 \to 0, h_2 \to 0. \tag{4.33}$$

Proof. For each h_1, h_2, let $\widehat{m}^{h_1,h_2}$ be an optimal relaxed control for $\{x^{h_1,h_2}(\cdot), p^{h_2}(\cdot)\}$. i.e.,

$$V^{h_1,h_2}(s,x,p) = W^{h_1,h_2}(s,x,p,\widehat{m}^{h_1,h_2}) = \inf_{m^{h_1,h_2} \in \Gamma^{h_1,h_2}} W^{h_1,h_2}(s,x,p,m^{h_1,h_2})$$

Choose a subsequence $\{\widetilde{h}_1, \widetilde{h}_2\}$ of $\{h_1, h_2\}$ such that

$$\liminf_{h_1,h_2 \to 0} V^{h_1,h_2}(s,x,p) = \lim_{\widetilde{h}_1,\widetilde{h}_2 \to 0} V^{\widetilde{h}_1,\widetilde{h}_2}(s,x,p) = \lim_{\widetilde{h}_1,\widetilde{h}_2 \to 0} W^{\widetilde{h}_1,\widetilde{h}_2}(s,x,p,\widehat{m}^{\widetilde{h}_1,\widetilde{h}_2}).$$

Note that we can assume that $\{\xi^{\widetilde{h}_1,\widetilde{h}_2}(\cdot), p^{\widetilde{h}_2}(\cdot), \widehat{m}^{\widetilde{h}_1,\widetilde{h}_2}(\cdot), \widehat{M}^{\widetilde{h}_1,\widetilde{h}_2}(\cdot)\}$ converges weakly to $\{x(\cdot), p(\cdot), m(\cdot), M(\cdot)\}$. Otherwise, take a subsequence of $\{\widetilde{h}_1, \widetilde{h}_2\}$ to assume its weak limit. Theorem 4.1, Skorohod representation and dominance convergence theorem imply that as $\widetilde{h}_1, \widetilde{h}_2 \to 0$,

$$E^{\widehat{m}^{\widetilde{h}_1,\widetilde{h}_2}}_{s,x,p}(\xi^{\widetilde{h}_1,\widetilde{h}_2}(T) + \lambda - k)^2 - \lambda^2 \to E^m_{s,x,p}(x(T) + \lambda - k)^2 - \lambda^2.$$

So

$$W^{\widetilde{h}_1,\widetilde{h}_2}(s,x,p,\widehat{m}^{\widetilde{h}_1,\widetilde{h}_2}) \to W(s,x,p,m) \geq V(s,x,p).$$

It follows that

$$\liminf_{h_1,h_2 \to 0} V^{h_1,h_2}(s,x,p) \geq V(s,x,p)$$

Next, we need to show $\limsup_{h_1,h_2\to 0} V^{h_1,h_2}(s,x,p) \le V(s,x,p)$ to complete the proof. Given any $\rho > 0$, there is a $\delta > 0$ so that we are able to approximate $(x(t), p(t), m(t), M(t))$ by a quadruple $(x^\delta(t), p^\delta(t), m^\delta(t), M^\delta(t))$ satisfying

$$x^\delta(t) = x + \int_s^t \int_{\mathcal{U}} b(x^\delta(z), p^\delta(z), c) m_z^\delta(dc) dz + \int_s^t \int_{\mathcal{U}} \sigma(x^\delta(z), p^\delta(z), c) M^\delta(dc, dz),$$

where $m^\delta(\cdot)$ is piecewise constant and takes finitely many values, $M^\delta(\cdot)$ is represented in terms of a finite number of d-dimensional Wiener processes and the controls are concentrated on the points $c_1, c_2, \ldots, c_N$ for all t. Let $\widehat{u}^\rho(\cdot)$ be the optimal control and $\widehat{m}^\rho(\cdot)$ be its relaxed control representation, and let $(\widehat{x}^\rho(\cdot), \widehat{p}^\rho(\cdot))$ be the associated solution process. Since $\widehat{m}^\rho(\cdot)$ is optimal in the chosen class of controls, we must have

$$W(s, x, p, \widehat{m}^\rho) \le V(s, x, p) + \tfrac{\rho}{3}. \tag{4.34}$$

Note that for each given integer ι, there is a measurable function $F_\iota^\rho(\cdot)$ such that

$$\widehat{u}^\rho(t) = F_\iota^\rho(w_l(s), p(s), s \le \iota\delta, l \le N)$$

on $[\iota\delta, \iota\delta + \delta)$. We next approximate $F_\iota^\rho(\cdot)$ by a function that depends only on the sample of $(w_l(\cdot), p(\cdot), l \le N)$ at a finite number of time points. Let $\theta < \delta$ such that δ/θ is an integer. Because the $\sigma-$ algebra determined by $\{w_l(\nu\theta), p(\nu\theta), \nu\theta \le \iota\delta, l \le N\}$ increases to the σ-algebra determined by $\{w_l(s), p(s), s \le \iota\delta, l \le N\}$, the martingale convergence theorem implies that for each δ, ι, there are measurable function $F_\iota^{\rho,\theta}(\cdot)$, such that as $\theta \to 0$,

$$F_\iota^{\rho,\theta}(w_l(\nu\theta), p(\nu\theta), \nu\theta \le \iota\delta, l \le N) = u_\iota^{\rho,\theta} \to \widehat{u}^\rho(\iota\delta) \quad \text{w.p.1.}$$

Here, we select $F_\iota^{\rho,\theta}(\cdot)$ such that there are N disjoint hyper-rectangles that cover the range of its arguments and that $F_\iota^{\rho,\theta}(\cdot)$ is constant on each hyper-rectangle. Let $m^{\rho,\theta}(\cdot)$ denote the relaxed control representation of the ordinary control $u^{\rho,\theta}(\cdot)$ which takes value $u_\iota^{\rho,\theta}$ on $[\iota\delta, \iota\delta + \delta)$, and let $(x^{\rho,\theta}(\cdot), p^{\rho,\theta}(\cdot))$ denote the associated solution. Then for small enough θ, we have

$$W(s, x, p, m^{\rho,\theta}) \le W(s, x, p, \widehat{m}^\rho) + \tfrac{\rho}{3}. \tag{4.35}$$

Next, we adapt $F_\iota^{\rho,\theta}(\cdot)$ such that it can be applied to $\{\xi_n^{h_1,h_2}\}$ and let $\bar{u}_n^{h_1,h_2}$ denote the ordinary admissible control to be used for the approximation chain $\{\xi_n^{h_1,h_2}\}$.

For n such that $nh_2 < \delta$, we can use any control. For $\iota = 1, 2, \ldots$ and n such that $nh_2 \in [\iota\delta, \iota\delta + \delta)$, we use the control defined by $\bar{u}_n^{h_1,h_2} = F_\iota^{\rho,\theta}(w_l^{h_1,h_2}(\nu\theta), p^{h_2}(\nu\theta), \nu\theta \le \iota\delta, l \le N)$. Recall that $\bar{m}^{h_1,h_2}(\cdot)$ denote the relaxed control representation of the continuous interpolation of $\bar{u}_n^{h_1,h_2}$, then

$$\begin{aligned}
&(\xi^{h_1,h_2}(\cdot), \bar{m}^{h_1,h_2}(\cdot), w_l^{h_1,h_2}(\cdot), F_\iota^{\rho,\theta}(w_l^{h_1,h_2}(\nu\theta), p^{h_2}(\nu\theta), \nu\theta \le \iota\delta, l \le N, \iota = 0, 1, 2, \ldots)) \\
&\quad \to (x^{\rho,\theta}(\cdot), m^{\rho,\theta}(\cdot), w_l(\cdot), F_\iota^{\rho,\theta}(w_l(\nu\theta), p(\nu\theta), \nu\theta \le \iota\delta, l \le N, \iota = 0, 1, 2, \ldots)).
\end{aligned}$$

Thus

$$W(s, x, p, \bar{m}^{h_1,h_2}) \le W(s, x, p, m^{\rho,\theta}) + \frac{\rho}{3}$$

Note that

$$V^{h_1,h_2}(s,x,p) \le W(s,x,p,\bar{m}^{h_1,h_2}).$$

Combing the above inequalities, we can see $\limsup_{h_1,h_2\to 0} V^{h_1,h_2}(s,x,p) \le V(s,x,p)$ for the chosen subsequence. By the tightness of $(\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), \bar{m}^{h_1,h_2}(\cdot))$ and arbitrary of ρ, we get

$$\limsup_{h_1,h_2\to 0} V^{h_1,h_2}(s,x,p) \le V(s,x,p)$$

and conclude the proof.

5 Concluding Remarks and Future Directions

In this dissertation, we have concentrated on stability and controls for stochastic dynamic systems. In Chapter 2, we first studied the benchmark linear scalar jump diffusions. Then exponential p stability and almost surely exponential stability for both SDEs and that of its numerical solutions are examined. The generalization to linear Markov switching jump diffusions and multi-dimensional jump diffusions are also discussed. For regime switching jump diffusions, under simple conditions, we derived sufficient conditions for asymptotic stability in the large and asymptotic stability in distribution. We also provided necessary and sufficient conditions for exponential stability. The connection between exponential stability and almost surely exponential stability was studied. Smooth dependence on the initial data was demonstrated as well.

One of future research efforts can be directed to the study of positive recurrence and egrodicity of regime-switching jump diffusions, which was called weak stability in [45] for diffusion processes. Another effort can be directed to studying stability of numerical algorithms for regime-switching jump diffusions for x dependent regime switching. For simplicity, throughout the the numerical analysis part, the Poisson process is assumed to be one-dimensional. For many applications, such a consideration is sufficient. An extension is to treat multi-dimensional counterparts. Another more delicate issue of much theoretical value is that the jump diffusions involve a more general Lévy process. This deserves a careful study and in-depth investigation.

In Chapter 3, we are interested on a mean variance control problem of Markov switching diffusions with large states. To reduce the computational complexity, we first used a two-time-scale formulation to relate the underlying problem with that of a limit problem then we demonstrated the near-optimal controls using two-time scale formulation and weak convergence techniques. In lieu of handling large dimensional systems, we need only solve a reduced set of limit equations that have much smaller dimensions. In Chapter 4, we keep working on the mean variance control problem under a quite different formulation in which we have a switching diffusion system with a hidden Markov chain. Using Markov chain approximation techniques combined with the Wonham filtering, a numerical scheme was developed. Our on-going effort will be directed to use the approach developed in this work to treat several networked systems that involve platoon controls with wireless communications and actuarial science.

REFERENCES

[1] L. Breiman, *Probability*, SIAM, Philadelphia, 1992.

[2] G. Chalmers and D. Higham, Asymptotic stability of a jump-diffusion equation and its numerical approximation, *SIAM J. Sci. Comp*, **31** (2008), 1141–1155.

[3] D. Higham, Mean-square and asymptotic stability of the stochastic theta method, *SIAM J. Num. Anal*, **38** (2000), 753–769.

[4] D. Higham and P. Kloeden, Numerical methods for nonlinear stochastic differential equations with jumps, *Numerische Mathematik*, **101** (2005), 101–119.

[5] D. Higham, X. Mao and C. Yuan, Almost sure and moment exponential stability in the numerical simulation of stochastic differential equations, *SIAM J. Num. Anal*, **45** (2007), 592–609.

[6] R. Khasminskii, *Stochastic Stability of Differential Equations.* Sijthoff and Noordhoff, Alphen aan den Rijin, The Netherlands, 1980.

[7] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1999.

[8] H.J. Kushner and A. Shwartz, Weak convergence and asymptotic properties of adaptive filters with constant gains, *IEEE Trans. Inform. Theory* **IT-30** (1984), 177–182.

[9] H.J. Kushner and G. Yin, Asymptotic properties of distributed and communicating stochastic approximation algorithms, *SIAM J. Control Optim.*, **25** (1987), 1266–1290.

[10] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, 2nd Ed., New York, 2003.

[11] R. Lipster, A strong law of large numbers for local martingales, *Stochastics*, **3** (1980), 217–228.

[12] X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, 2006.

[13] G.N. Milstein and M.V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.

[14] S.L. Nguyen and G. Yin, Pathwise convergence rate for numerical solutions of stochastic differential equations, *IMA J. Nume. Anal.*, **32** (2012), 701–723.

[15] S. Pang, F. Deng, and X. Mao, Almost sure and moment exponential stability of Euler–Maruyama discretizations for hybrid stochastic differential equations, *J. Comput. Appl. Math.*, **213**, (2008), 127–141.

[16] F. Xi and G. Yin, Almost sure stability and instability for switching-jump-diffusion systems with state-dependent switching, to appear in *J. Math. Analy. Appl.*

[17] G. Yin, X. Mao, C. Yuan, and D. Cao, Appriximation methods for hybrid diffusion systems with state-dependent switching processes: numerical algorithms and existence and uniqueness of solutions, *SIAM J. Math. Anal*, **41** (2010), 2335–2352.

[18] G. Yin, Q. Song, and Z. Zhang, Numerical solutions for jump-diffusions with regime switching, *Stochastics* , **77** (2005), 61–79.

[19] G. Yin and F. Xi, Stability of regime-switching jump diffusions, *SIAM J. Control Optim.*, **48** (2010), 4525–4549.

[20] G. Yin and C. Zhu, *Hybrid Switching Diffusions.* New York: Springer, 2010.

[21] G. Badowski and G. Yin, Stability of hybrid dynamic systems containing singularly perturbed random processes, *IEEE Trans. Automatic Control*, **47** (2002), 2021–2032.

[22] G. Barone-Adesi and R. Whaley, Efficient analytic approximation of American option values, *J. Finance*, **42** (1987), 301–320.

[23] G.K. Basak, A. Bisi and M.K. Ghosh, Stability of degenerate diffusions with state-dependent switching, *J. Math. Anal. Appl.*, **240** (1999), 219–248.

[24] M.-F. Chen, *From Markov Chains to Non-equilibrium Particle Systems*, 2nd ed., World Scientific, Singapore, 2004.

[25] M.-F. Chen, *Coupling methods for multidimensional diffusion processes*, *The Ann. Probab.*, Vol.17. (1989), 151–177.

[26] A. Friedman, *Stochastic Differential Equations and Applications*, Vol I, Academic Press, New York, 1975.

[27] J.P. Hespanha, A model for stochastic hybrid systems with application to communication networks, *Nonlinear Anal.*, **62** (2005), 1353–1383.

[28] A.M. Il'in, R.Z. Khasminskii, and G. Yin, Asymptotic expansion of solutions of integro-differential equations for transition densities of singularly perturbed switching diffusions: rapid switchings, *J. Math. Anal. Appl.*, **238** (1999), 516–539.

[29] S. Kar and J.M.F. Moura, Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise, *IEEE Trans. Signal Processing*, **57** (2009), no.1, 355–369.

[30] R.Z. Khasminskii, C. Zhu, and G. Yin, Stability of regime-switching diffusions, *Stochastic Process Appl.*, **117** (2007), 1037–1051.

[31] A.V.Skorohod , *Asymptotic Methods in the Theory of Stochastic Differential Equations*, Amer. Math. Soc., Providence,RI,1989.

[32] H.J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birhkhäuser, Boston,1990.

[33] C.W. Li, Z. Dong, and R. Situ, Almost sure stability of linear stochastic differential equations with jump, *Probab. Theory Related Fields*, **123** (2002), 121–155.

[34] Y. Liu, K. Passino, and M.M. Polycarpou, Stability analysis of M-dimensional asynchronous swarms with a fixed communication topology, *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 76-95, Jan. 2003.

[35] X. Mao, *Stochastic Differential Equations and Applications*, 2nd Ed., Horwood, Chichester, UK, 2007.

[36] X. Mao, Stability of stochastic differential equations with Markovian switching, *Stochastic Processes Appl.*, **79** (1999), 45–67.

[37] J.L. Menaldi and M. Robin, Invariant measure for diffusions with jumps, *Appl. Math. Optim.*, **40** (1999), 105–140.

[38] R. Olfati-Saber and R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays, *IEEE Trans. Automatic Control*, **49** (2004), 1520–1533.

[39] C.W. Reynolds, Flocks, herds, and schools: a distributed behavioral model, *Computer Graphics*, **21**(4): 25-34, July 1987.

[40] B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd Ed., Prentice Hall, 2001.

[41] A.V. Svishchuk and Yu.I. Kazmerchuk, Statiblity of stochastic delay equations of Itô form with jumps and Markovian switchings, and their applications in finance, *Theory Probab. Math. Statist.* **64** (2002), 167–178.

[42] J. Toner and Y. Tu, Flocks, herds, and schools: A quantitative theory of flocking, *Physical Review E*, **58** (4): 4828–4858, October, 1998.

[43] T. Viseck, A. Czirook, E. Ben-Jacob, O. Cohen, and I. Shochet, Novel type of phase transition in a system of self-deriven particles, *Physical Review Letters*, **75** (6): 1226–1229, August, 1995.

[44] I.S. Wee, Stability for multidimensional jump-diffusion processes, *Stochastic Process. Appl.*, **80** (1999), 193–299.

[45] W.M. Wonham, Liapunov criteria for weak stochastic stability, *J. Differential Eqs.*, **2** (1966), 195–207.

[46] S. Wu and Y. Zeng, The term structure of interest rates under regime shifts and jumps, *Economics Lett.*, **93** (2006), 215–221.

[47] F. Xi, Asymptotic properties of jump-diffusion processes with state-depedent switching, *Stochastic Process. Appl.*, **119** (2009), 2198–2221.

[48] G. Yin, Y. Sun, and L.Y. Wang, Asymptotic properties of consensus-type algorithms for networked systems with regime-switching topologies, *Automatica*, **47** (2011) 1366–1378.

[49] G.Yin and C. Zhu, Properties of solutions of stochastic differential equations with continuous-state-depedent switching, *J. Differential Eqs.*,**249** (2010), 2409–2439.

[50] C. Yuan and X. Mao, Asymptotic stability in distribution of stochastic differential equations with Markovian switching, *Stochastic Process. Appl.*, **103** (2003), 277–291.

[51] C. Yuan and X. Mao, Stability of stochastic delay hybrid systems with jumps, *Eur. J. Control*, **16** (2010), no.6,595-608.

[52] S. Chen, X. Li, and X.Y. Zhou, stochastic linar quadratic regulators with indefinite control weight costs, *SIAM J. Control Optim.*, **36** (1998), 1685–1702.

[53] F. Dufour and P. Bertrand, The filtering problem for continuous time linear systems with Markovian switching coefficients, *Syst. Control Lett.*, **23** (1994), 453–461.

[54] D. Duffie and H. Richardson, Mean-variance hedging in continuous time, *Ann. Appl. Probab.*, **1** (1991), 1–15.

[55] W.H. Fleming and M. Nisio, On stochastic relaxed control for partially observed diffusions, *Nagoya Math. J.*, **93** (1984), 71–108.

[56] H.J. Kushner, Numerical methods for stochastic control problems in continuous time, *SIAM J. Control Optim.*, **28** (1990), 999–1048.

[57] H.J. Kushner, Consistency issues for numerical methods for variance control with applications to optimization in finance, *IEEE. Trans. Automat. Control*, **44** (2000), 2283–2296.

[58] H.J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, New York, 2001.

[59] D. Li and W.L. Ng, Optimal dynamic portfolio selection: Multi-period mean-variance formulation, *Math. Finance.*, **10** (2000), 387–406.

[60] H. Markowitz, Portfolio selection, *J. Finance.*, **7** (1952), 77–91.

[61] P.A. Samuelson, Lifetime portfolio selection by dynamic stochastic programming, *Rev. Econ. Statist.*, **51** (1969), 239–246.

[62] S.R. Pliska, *Introduction to Mathematical Finance*, Basil Blackwell, Malden, UK, 1997.

[63] M. Schweizer, Approximation pricing and the variance optimal martingale measure, *Ann. Probab.*, **24** (1996), 206–236.

[64] Q.S. Song, G. Yin, and Z. Zhang, Numerical method for controlled regime-switching diffusions and regime-switching jump diffusions, *Automatica*, **42** (2006), 1147–1157.

[65] W.M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, *SIAM J. Control*, **2** (1965), 347–369.

[66] Z. Yang, G. Yin, L,Y, Wang, and H. Zhang, Near-Optimal mean-variance controls under two-time-scale formulations and applications, *Stochastics*, **85** (2013), 723–741.

[67] D. Yao, Q. Zhang, X.Y. Zhou, A regime switching model for European options, in *Stochastic Process, Optimization, and Control Theory*, H. Yan, G. Yin and Q. Zhang Eds., 281–300, Springer, 2006.

[68] G. Yin and Q. Zhang, *Discrete-Time Markov chains: Two-time-scale methods and application*, Springer, New York, 2005.

[69] G. Yin and X.Y. Zhou, Markowitz mean-variance portfolio selection with regime switching: from discrete-time models to their continuous-time limits, *IEEE Trans. Automatic Control.*, **49** (2004), 349–360.

[70] J. Yong and X.Y. Zhou, *Stochastic controls: Hamiltonian Sytems and HJB Equations.* New York: Springer, 1999.

[71] L. Yu, Q. Zhang and G. Yin, Asset allocation for regime switching market models under partial observation, to appear in *Dynamic Systems and Applications.*

[72] Q. Zhang, Stock trading: an optional selling rule, *SIAM J.Control Optim.*, **40** (2001), 64–87.

[73] S.B. Choi and J.K. Hedrick, Vehicle longitudinal control using an adaptive observer for automated highway systems, in *Proc. of ACC*, Seattle, 1995.

[74] P.J. Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, NY, 1977.

[75] J.K. Hedrick, D. McMahon, and D. Swaroop, Vehicle modeling and control for automated highway systems, PATH Research Report, UCB-ITS-PRR-93-24, 1993.

[76] H.J. Kushner, *Approximation and Weak Convergence Methods for Random Processess, with Application to Stochastic Systems Theory.* MIT Press, Cambridge, Massachusetts.,1984.

[77] C.Y. Liang and H. Peng, String stability analysis of adaptive cruise controlled vehicles, *JSME Internat. J. Series C.*, **43** (2000), 671–677.

[78] R. Rajamani, H.S. Tan, B. Law, and W.B. Zhang, Demonstration of Integrated Lateral and Longitudinal Control for the Operation of Automated Vehicles in Platoons, *IEEE Trans. Control Sys. Tech.*, **8** (2000), 695–708.

[79] S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser, Boston, MA, 1994.

[80] H.A. Simon and A. Ando, Aggregation of variables in dynamic systems, *Econometrica*, **29** (1961), 111–138.

[81] D. Swaroop and J. Hedrick, String Stability of Interconnected Systems, *IEEE Trans. Automat. Control*, **41** (1996), 349–357.

[82] W.S. Wong, Control communication complexity of distributed control systems, *SIAM J. Control Optim.*, **48** (2009), 1722–1742.

[83] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Pertubation Approach*, New York, Springer-Verlag, 1998.

[84] G. Yin, Q. Zhang, and G. Badowski, Asymptotic properties of a singularly perturbed markov chain with inclusion of transient states,*The Ann. Appl. Probab.*, **10** (2000), 549–572.

[85] Q. Zhang and G. Yin, On nearly optimal controls of hybrid LQG problems, *IEEE Trans. Automat. Control.*, **44** (1999), 2271–2282.

[86] X.Y. Zhou and D. Li, Continuous-time mean-variance portfolio selection: A stochastic LQ framework, *Appl. Math. Optim.* **42** (2000), 19–33.

[87] X.Y. Zhou and G. Yin, Markovitz's mean-variance porfolio selection with regime switching:a continuous time model, *SIAM J. Control Optim.*, **42** (2003), 1466–1482.

[88] S. Asmussen, and M. Taksar, Controlled diffusion models for optimal dividend pay-out, *Insurance Math. Econom.* **20** (1997), 1–15.

[89] Y.J. Liu, G. Yin, and X.Y. Zhou, Near-optimal controls of random-switching LQ with indefinite control weight, *Automatica*, **41** (2005), 1063–1070.

[90] Z. Yang and G. Yin, Stability of Nonlinear Regime-switching Jump Diffusions, *Nonlinear Analysis*, 75, (2012), 3854–3873.

ABSTRACT

STABILITY AND CONTROLS FOR STOCHASTIC DYNAMIC SYSTEMS

by

ZHIXIN YANG

August 2014

Advisor: Dr. George Yin

Major: Mathematics (Applied)

Degree: Doctor of Philosophy

This dissertation focuses on stability analysis and optimal controls for stochastic dynamic systems. It encompasses two parts. The first part of our work gives an in-depth study of stability of linear jump diffusions, linear Markovian jump diffusions, multi-dimensional jump diffusions, and regime-switching jump diffusions together with the associated numerical methods. The second part of our work treats controls for stochastic dynamic systems. We concentrate on mean variance types of control under different formulations. We obtain the nearly optimal mean-variance controls under both two-time-scale and hidden Markov chain formulations and convergence analysis for each case is carried out.

In Chapter 2, stability analysis of benchmark linear scalar jump diffusions is studied first. We present the conditions for exponential p stability and almost surely exponentially stability for SDEs and for numerical solutions. Note that due to the use of Poisson processes, using asymptotic expansions as in the literature for treating diffusion processes does not work. Different from the existing treatments of Euler-Maurayama methods for solutions of stochastic differential equations, techniques from stochastic approximation is employed in our work.

Then similar analysis is carried out for Markov jump diffusions and multi-dimensional jump diffusions. In addition, we carry out a thorough study on asymptotic stability in the large and exponential p-stability for regime-switching jump diffusions. Connection between almost surely exponential stability and exponential p-stability is exploited. Necessary conditions for exponential p-stability are derived and criteria for asymptotic stability in distribution are provided. In Chapter 3 We work on the well-known mean-variance problem with new a twist in which a switching process is embedded. We first use a two-time-scale formulation to treat the underlying system with the use of a small parameter. As the small parameter goes to 0, we obtain a limit problem. Using the limit problem as a guide, we construct controls for the original problem, and show that the control so constructed is nearly optimal. In chapter 4, we revisit the mean variance control problem in which the switching process is a hidden Markov chain. Instead of having full knowledge of switching process, we assume only the noisy observation of the switching process corrupted by white noise is available. We focus on minimizing the variance subject to a fixed terminal expectation. Using the Wonham filter, we convert the partially observable system to a completely observable one first. Because closed-form solutions are virtually impossible to obtain, our main effort is devoted to designing a numerical algorithm. Convergence of the algorithm is obtained.

AUTOBIOGRAPHICAL STATEMENT

Zhixin Yang

Education

Ph.D. in Applied Mathematics, Aug, 2014 (expected)
Wayne State University, Detroit, Michigan

M.A. in Mathematical Statistics, May, 2013
Wayne State University, Detroit, Michigan

B.S. in Mathematics, June 2005
Hunan Normal University, Changsha, China

Awards

C.E. Prins Award for outstanding achievement in the PhD program, Department of Mathematics, Wayne State University, 2014.

Maurice J. Zelonka Endowed Scholarship, Department of Mathematics, Wayne State University, 2014

3rd place poster in 2014 Graduate Exhibition, Graduate School, Wayne State University, 2014

The Alfred L. Nelson Award for outstanding achievement in the Masters Program, Department of Mathematics, Wayne State University, 2013

The Karl W. and Helen L. Folley Endowed Mathematics Scholarship, Department of Mathematics, Wayne State University, 2013 and 2012.

Graduate Student Professional Travel Award, Department of Mathematics, Wayne State University, 2012, 2013, 2014.

List of Publications and submitted

Z. Yang and G. Yin, Stability of Nonlinear Regime-switching Jump Diffusions, *Nonlinear Analysis*, 75, (2012), 3854–3873.

Z. Yang, G. Yin, L.Y. Wang, and H.W. Zhang, Near-Optimal Mean-Variance Controls under Two-time-scale Formulations and Applications, *Stochastics*, 85, (2013), 723–741.

Z. Yang, G. Yin and Q. Zhang, Mean-Variance Type Controls Involving a Hidden Markov Chain: Models and Numerical Approximation, to appear in *IMA Journal of mathematical control and information.*

Z. Yang, G. Yin, L.Y. Wang, and H.W. Zhang, A Mean-Variance Control Framework for Platoon Control Problems: Weak Convergence Results and Applications on Reduction of Complexity, to appear in *Communication in Information and Systems.*

Z. Yang, G. Yin, and H.B. Li, Stability of Numerical Methods for Jump Diffusions and Markovian Switching Jump Diffusions, submitted.