1-1-2010

Asymptotic Properties Of Markov Modulated Sequences With Fast And Slow Time Scales

Son Luu Nguyen
Wayne State University

ASYMPTOTIC PROPERTIES OF MARKOV MODULATED SEQUENCES WITH FAST AND SLOW TIME SCALES

by

SON LUU NGUYEN

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

2010

MAJOR: MATHEMATICS

Approved by:

____
Advisor Date

____

____

____

DEDICATION

To my grandmother

ACKNOWLEDGEMENTS

It is a pleasure to thank those who made this dissertation possible.

I do not have enough words to express my deepest gratitude to my advisor, Professor George Yin, who introduces me to the beautiful world of stochastic processes, proposes me exciting problems, teaches me lots of attractive topics in maths, and constantly helps as well as encourages me for all 5 years I study at Wayne Sate University. This dissertation would not have been completed without his guidance and endless support.

I am taking this opportunity to thank Professor Rafail Khasminskii, Professor Boris Mordukhovich, Professor Tze Chien Sun, and Professor Le Yi Wang for serving in my committee.

I am grateful to Professor Nguyen H. V. Hung, who recommended me to Wayne State University and helped me a lot when I worked in Vietnam.

I owe my thanks to Ms. Mary Klamo, Ms. Patricia Bonesteel, Professor Chenggui Yuan, Dr. Chao Zhu, Professor Choon-Jai Rhee, Professor John Breckenridge, Mr. James Veneri, and Ms. Tiana Fluker, who have made available their support in a number of ways, and to Professor Rafail Khasminskii, Professor Alex Korostelev, Professor Bertram Schreiber, Professor Nguyen Duy Tien, Professor Nguyen Huu Du, and Professor Dang Hung Thang, who have contributed to my solid knowledge in Probability and Stochastic Processes Theory.

I would like to share this moment with my family. I am indebted to my parents, my sister and grandmothers for their unconditional and unlimited love and support since I was born. My special gratitude goes to my wife for her love, encouragement and sharing with me everything.

Finally, but not least, I would like to express my appreciation to the entire Department of Mathematics for their hospitality and services. I enjoyed the warm and friendly atmosphere in the department, and I appreciated the support I received during my study at Wayne State University.

TABLE OF CONTENTS

1 Introduction

This dissertation concerns Markov modulated random sequences. It focuses on the asymptotic behavior of suitably scaled processes. Our motivation stems from a wide variety of applications in communication networks, stochastic hybrid systems, queueing systems, control and optimization, economic systems, production planning, actuarial science, and financial engineering. Owing to the increasing complexity of the real-world applications, one is often forced to deal with large-scale systems. Due to the uncertainty of the random environment, there is a growing interest in modeling, analysis, and optimization of large-scale systems using an additional random factor in addition to the usual dynamic systems. In the past few years, increasing and resurgent efforts have been devoted to treating regime-switching processes; see for example, [2, 25, 26] for communication networks, [3] for computer models, [19, 20, 27] for queueing systems, [9] for stochastic hybrid systems, [1] for option pricing under random environment, [8] for economic systems, [24] for state aggregations, [26] for wireless communications, and [29] for Markowitz's portfolio optimization under Markov modulation.

To further our understanding, we focuses on the study of non-Markov random sequences in discrete time in which the primary sequence is modulated by a switching process. The modulating force, representing random environment and other stochastic factors, is modeled by a Markov chain α_k with a finite state space $\mathcal{M}$ with all states being recurrent. We are concerned with asymptotic properties of the process

$\{X(k, \alpha_k)\}$, where for each $\alpha \in \mathcal{M}$, $\{X(k, \alpha)\}$ is the primary random sequence, and α_k is a Markov chain. To visualize the movement of the resulting random sequence, suppose for instance, initially, the Markov chain resides in a state α. It sojourns in that state for an exponentially distributed random duration until time τ_1, the first jump time of α_k. The process takes the form $\{X(k, \alpha) : 0 \le k < \tau_1\}$. Then at τ_1, the chain switches to a new state $\beta \ne \alpha$ and stays there for a random duration until τ_2 the second jump time. During this period, the process becomes $\{X(k, \beta) : \tau_1 \le k < \tau_2\}$ and so on.

Because of the practical needs, the underlying Markov chain often has a large state space (i.e., $|\mathcal{M}|$, the cardinality of $\mathcal{M}$, is large). Apparently, corresponding to a large state space $\mathcal{M}$, there are large number of sequences $\{X(k, \alpha)\}$ to be considered (in fact $|\mathcal{M}|$ sequences). The complexity becomes a real issue. It is important to reduce the complexity. We note that although the Markov chain has a large number of states, the transition rates among different states are not the same. A typical situation is that transitions among some of the states are changing rapidly, whereas others are varying slowly. The state space can often be split into smaller subspaces such that within each subspace the transitions are about the same rate, and from one subspace to another, the transitions happen relatively rarely. Such a model is known as having nearly completely decomposable structure [3, 24] in the literature. From a mathematical point of view, it can be setup as a two-time-scale model. A systematic study of the related Markovian models has been taken recently [28].

For the random sequences under consideration, there are two main issues. The first one is not much structure of the sequence $\{X(k,\alpha_k)\}$ is known. The second one is that as alluded to in the previous paragraph, for a large $|\mathcal{M}|$, there are $|\mathcal{M}|$ sequences of $\{X(k,\alpha) : \alpha \in \mathcal{M}\}$ to be dealt with. Suppose the state space $\mathcal{M}$ admits the representation $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_{l_0}$ so that $\mathcal{M}_i$ for $i = 1, \ldots, l_0$ can be considered as subspaces, where the $\mathcal{M}_i$'s are not isolated. There are weak interactions among the $\mathcal{M}_i$'s, and $\mathcal{M}$ is not completely decomposable but only "nearly completely decomposable." The precise form of transition probabilities will be specified later.

In this dissertation, we examine the random sequence $\{X(k,\alpha_k)\}$ and aim to reveal the intrinsic features of the underling processes. The sequences of interest are formulated as two-time-scale processes to achieve the goal of reduction of complexity. Under suitable conditions, we obtain invariance principles in the sense of weak convergence. There are many well-known treaties of weak convergence methods for stochastic processes and their applications. These include techniques based on operator semigroup convergence theorems for Markov processes, martingale characterization of limit processes, and representation of the limit as solutions of stochastic equations; for example, [7], [15], [21], and many references therein. Here, we use a martingale averaging approach. Due to the interactions of the switching components, the primary process and the modulating process are intertwined and tangled together, which makes the existing results not directly applicable to our problem. However, using stochastic analysis techniques and by careful examination of the underlying processes, we are able to overcome the difficulties and to obtain the desired results. Dealing with mixing type processes, we carry out careful analysis for the coupled system. There are really two averages are involved. One is the average of the two-time-scale Markov chain leading to a reduced Markov chain with a much smaller state space, and the other is an average of the mixing process leading to diffusion processes. However, the primary sequences and the modulating sequence are intertwined making the averaging analysis a nontrivial task. In the literature, effort has been made to treat evolution of systems in random media; see for example [14] and references therein. In this reference, semi-Markov processes in general Banach spaces are treated. In our setup, the primary sequence is non-Markov. The limit does not have a Gaussian distribution but Gaussian mixtures.

Using two-time scales in the formulation, we introduce a small parameter $\varepsilon > 0$ into the transition probabilities so as to highlight the different rates of transitions. Thus, we can write the Markov chain as α_k^ε and write the sequence as $X(k, \alpha_k^\varepsilon)$. The significance of our results can be illustrated from the following example. Considered an optimal control problem. Let Γ be a compact set of a multi-dimensional Euclidean space, and $u(\cdot) = \{u(x, \alpha) \in \mathbb{R}^d \times \mathcal{M}\}$ be a function such that $u(x, \alpha) \in \Gamma$ for all $(x, \alpha) \in \mathbb{R}^d \times \mathcal{M}$. Then $u(\cdot)$ is said to be an admissible control and the collection of all such functions is denoted by $\mathcal{A}$, termed admissible control set. We wish to find the optimal control of

$$J(x, \alpha, u(\cdot)) = E_{x,\alpha} \sum_{k=0}^{\infty} (1 - \beta\varepsilon)^k L(X_k, \alpha_k^\varepsilon, u(\cdot)),$$

where $0 < \beta < 1$ is a discount factor, $L(x, \alpha, u)$ is a suitable cost function, and $X(0, \alpha_0^\varepsilon) = (x, \alpha)$. This is an analogue of the so-called Markov decision process; see

[28, Chapter 8]. However, the process $X(k,\alpha_k^\varepsilon)$ is non-Markov. We only assume that for each α, $\{X(k,\alpha)\}$ is mixing, and $\{\alpha_k^\varepsilon\}$ is a discrete Markov chain with nearly completely decomposable structure. Due to the lack of structure of the process, the problem is difficult to solve. The near decomposability however enables us to write

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cdots \mathcal{M}_{l_0}.$$

That is, we decompose the state space into subspaces although these subspaces are not isolated but weakly connected. Using the idea of aggregation, we lump the states of the Markov chain in each $\mathcal{M}_i$ into one state for $i = 1, \ldots, l_0$ to get an aggregate process $\overline{\alpha}_k^\varepsilon$. Corresponding to this, we consider a new sequence $\{\overline{X}(k,\overline{\alpha}_k^\varepsilon)\}$. Effectively, we use a single sequence $\{\overline{X}(k,i)\}$ as a representative for the sequences $\{X(k,\alpha)\}$ for all $\alpha \in \mathcal{M}_i$. Using the idea to be resented in this paper, it can be shown that this new sequence leads to a limit under suitable interpolations. Then one may construct optimal control of the limit process and use it to that of the original system leading to a near-optimal strategy. Similar approach may be taken to treat related optimization problems. Note that the original modulating Markov chain has a large state space, which renders the optimization problem computationally infeasible, whereas the limit process uses aggregated states with a much less computation needed. The original coupled sequence has little structure known to us and is difficult to handle. The limit process, however, is a switching diffusion with a well-defined operator. Thus, it is relatively easier to treat the associated limit system. Denote the original state space and the state space of the limit by $\mathcal{M}$ and $\overline{\mathcal{M}}$, respectively. If $|\mathcal{M}| \gg |\overline{\mathcal{M}}|$, a substantial reduction of computational complexity will be achieved when one treats control and optimization problems.

Next, we consider the case that the Markov chain is ergodic, but has a large number of states (i.e., the cardinality $|\mathcal{M}|$ is large). Owing to ergodicity, for certain optimization problems, instead of treating each sequence $\{X(k,i)\}$ independently, we can consider an effective sequence, namely, the average with respect to the ergodic measure of all the states. The message is that we can replace the large number of sequences $\{X(k,i) : i = 1, \ldots, |\mathcal{M}|\}$ by an aggregated average, whose precise definition will be given later. To facilitate the use of the average mentioned above, and dealing with many optimization problems, one frequently needs to answer an important question after the replacement mentioned above. The question is how good the approximation is. To answer the question, one needs to provide precise error bound. IN the next part of our work in this dissertation, under simple conditions, we establish strong approximation results for a centered and scaled sequence, which justifies the replacement and ascertains the error bounds.

The rest of the dissertation is arranged as follows. Chapter 2 begins with the two-time scale formulation and mixing property. Chapter 3 presents the precise formulation of the problem under consideration and takes up the weak convergence issue under a simplified setup. Careful analysis is provided leading to the desired limit system. Further results and ramifications will be also presented in this chapter. To facilitate the reading, a section is given at the end of the chapter to provide the proof of a technical result. Chapter 4 develops strong approximation for the sequence of interest. The coupling of the mixing random process $X(k,i)$ and the Markov chain α_k, makes the analysis difficult. We divide our task of analysis into several subtasks and use step-by-step approximation to reach our goal. In addition, an example of an optimization problem is provided as a demonstration. The proofs of a number of technical results are gathered and placed in the last section of Chapter 4. A few further remarks are made in Chapter 5.

2 Preminaries

This chapter is devoted to the two-time scale Markov chains and the concept of $\phi-$ mixing. In what follows, we first focus on asymptotic properties of Markov chains with two-time scale and then present some useful inequalities for mixing random variables.

2.1 Two-time Scale Formulation

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Throughout this dessertation, we use C to denote a generic positive constant with the convention $CC = C$ and $C + C = C$ used.

2.1.1 Recurrence Case

Let $\varepsilon > 0$ and α_k^ε be a time-homogeneous Markov chain on $(\Omega, \mathcal{F}, P)$ with state space $\mathcal{M}$ containing m_0 states and transition matrix $P_\varepsilon = P + \varepsilon Q$, where $P = (p^{ij})$ is a transition probability matrix and $Q = (q^{ij})$ is a generator of a continuous-time Markov chain (i.e., $p^{ij} \geq 0$ and $\sum_{j=1}^{m_0} p^{ij} = 1$; $q^{ij} \geq 0$ for $i \neq j$ and $\sum_{j=1}^{m_0} q^{ij} = 0$ for each i). Assume that the state space $\mathcal{M}$ can be written as

$$\begin{aligned} \mathcal{M} &= \{s_{11}, \ldots, s_{1,m_1}\} \cup \{s_{21}, \ldots, s_{2,m_2}\} \cup \cdots \cup \{s_{l_0 1}, \ldots, s_{l_0,m_{l_0}}\} \\ &= \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_{l_0}, \end{aligned} \tag{2.1}$$

with $m_0 = m_1 + m_2 + \cdots + m_{l_0}$ and $P = \text{diag}[P^1, P^2, \ldots, P^{l_0}]$, where P^i, $i \leq l_0$, are also transition matrices themselves. The subspace $\mathcal{M}_i$ for each $i = 1, 2, \ldots, l_0$, consists of recurrent states belonging to the ith ergodic class. We also assume that for $i \leq l_0$, P^i is irreducible and aperiodic.

Let p^ε_k be the probability vector $p^\varepsilon_k = (P(\alpha^\varepsilon_k = s_{ij})) \in \mathbb{R}^{1\times m_0}$, and ν^i the stationary distribution corresponding to the transition matrix P_i. Assume that $p^\varepsilon_0 = (P(\alpha^\varepsilon_0 = s_{ij})) = p_0$ and define an aggregated process $\overline{\alpha}^\varepsilon_k$ of α^ε_k by

$$\overline{\alpha}^\varepsilon_k = i \text{ if } \alpha^\varepsilon_k \in \mathcal{M}_i \quad \text{for} \quad i = 1, \dots, l_0, \; \overline{\alpha}^\varepsilon(t) = \overline{\alpha}^\varepsilon_k \quad \text{for} \quad t \in [\varepsilon k, \varepsilon(k+1)).$$

Before proceeding further, we present a result on asymptotic expansions of the probability vector p^ε_k and the k-step transition matrix $(P_\varepsilon)^k$ as well as the aggregated process. Part (a) and (b) can be found in [28, Theorem 4.1], whereas part (c) is in [28, Theorem 4.3].

Proposition 2.1. *The following assertions hold:*

(a) *For the probability distribution vector* $p^\varepsilon_k \in \mathbb{R}^{1\times m_0}$ *we have*

$$p^\varepsilon_k = \theta(\varepsilon k)\mathrm{diag}(\nu^1, \dots, \nu^{l_0}) + O(\varepsilon + \lambda^k) \tag{2.2}$$

for some λ *with* $0 < \lambda < 1$*, where* $\theta(t) = (\theta^1(t), \dots, \theta^{l_0}(t)) \in \mathbb{R}^{1\times l_0}$ *satisfies*

$$\frac{d\theta(t)}{dt} = \theta(t)\overline{Q}, \quad \theta(0) = p_0\widetilde{\mathbb{1}},$$

where

$$\overline{Q} = \mathrm{diag}(\nu^1, \dots, \nu^{l_0})Q\tilde{\mathbb{1}}, \; \tilde{\mathbb{1}} = \mathrm{diag}(\mathbb{1}_{m_1}, \dots, \mathbb{1}_{m_{l_0}}), \; \mathbb{1}_l = (1, \dots, 1)' \in \mathbb{R}^{l\times 1}. \tag{2.3}$$

(b) *For* $k \le T/\varepsilon$ *with some fixed* T*, the* k*-step transition matrix* $(P_\varepsilon)^k$ *satisfies*

$$(P_\varepsilon)^k = \Phi(\varepsilon k) + \varepsilon\hat{\Phi}(\varepsilon k) + \Psi(k) + \varepsilon\hat{\Psi}(k) + O(\varepsilon^2), \tag{2.4}$$

where

$$\Phi(t)=\tilde{\mathbb{1}}\Theta(t)\mathrm{diag}(\nu^1,\ldots,\nu^{l_0}),\quad \frac{d\Theta(t)}{dt}=\Theta(t)\bar{Q},\ \Theta(0)=I. \tag{2.5}$$

Moreover, $\Phi(\varepsilon k)$ *and* $\hat{\Phi}(\varepsilon k)$ *are uniformly bounded in* $[0,T]$ *and* $\Psi(k)$ *and* $\hat{\Psi}(k)$ *decay exponentially, i.e.,* $|\Psi(k)|+|\hat{\Psi}(k)|\le K\lambda^k$ *for some* $K>0$ *and some* $0<\lambda<1$.

(c) *The aggregated process* $\overline{\alpha}^{\varepsilon}(\cdot)$ *converges weakly to* $\overline{\alpha}(\cdot)$ *that is a continuous-time Markov chain generated by* $\overline{Q}$.

Remark 2.2. (i) In view of the asymptotic expansion, we have $(P_\varepsilon)^k=\Phi(\varepsilon k)+O(\varepsilon+\lambda^k)$.

(ii) The matrix $\Psi(k)$ is selected so that $\Phi(0)+\Psi(0)=I$ and $\Psi(k)=\Psi(0)(P)^k$. In view of (2.5),

$$\Psi(k)=\mathrm{diag}\big((I_{m_1}-\mathbb{1}_{m_1}\nu^1)(P^1)^k,\ldots,(I_{m_{l_0}}-\mathbb{1}_{m_{l_0}}\nu^{l_0})(P^{l_0})^k\big), \tag{2.6}$$

where I_{m_i} is the $m_i\times m_i$ identity matrix. Thus $\Psi(k)$ is again of block-diagonal form. Taking this into account, the fact that $\lim_{\varepsilon\to 0}\theta_{i_1i_2}(\varepsilon k)=0$ for $i_1\neq i_2$, $1\le i_1,i_2\le l_0$ (where $\Theta(t)=(\theta_{i_1i_2}(t))$) implies

$$\lim_{\varepsilon\to 0}P(\alpha_k^\varepsilon=s_{i_2j_2}|\alpha_0^\varepsilon=s_{i_1j_1})=0 \tag{2.7}$$

for all $1\le i_1\neq i_2\le l_0$, $1\le j_1\le m_{i_1}$, $1\le j_2\le m_{i_2}$.

(iii) For $k=0,\ldots,T/\varepsilon$, $i=1,\ldots,l_0$, $j=1,\ldots,m_i$, denote $\pi_k^{\varepsilon,ij}=\varepsilon\sum_{l=0}^{k-1}\Big(I(\alpha_l^\varepsilon=$

$s_{ij}) - \nu^{ij} I(\overline{\alpha}_l^\varepsilon = i)\Big)$. Then by [28, Theorem 4.5], we have

$$\sup_{0\le k\le T/\varepsilon} E|\pi_k^{\varepsilon,ij}|^2 = O(\varepsilon) \quad \text{and} \sup_{0\le k\le T/\varepsilon} E\Big[\varepsilon \sum_{l=0}^{k-1} \Big| I(\alpha_l^\varepsilon = s_{ij}) - \nu^{ij} I(\overline{\alpha}_l^\varepsilon = i)\Big|\Big]^2 = O(\varepsilon) \tag{2.8}$$

for $i = 1, \dots, l_0$, $j = 1, \dots, m_i$.

2.1.2 Ergodic Case

Let $\varepsilon > 0$ and α_k^ε be a time-homogeneous Markov chain on $(\Omega, \mathcal{F}, P)$ with state space $\mathcal{M} = \{1, 2, \dots, m\}$ and transition matrix

$$P^\varepsilon = P + \varepsilon Q, \tag{2.9}$$

where $P = (p^{ij})$ is a transition probability matrix and $Q = (q^{ij})$ is a generator of a continuous-time Markov chain (i.e., $p^{ij} \ge 0$ and $\sum_{j=1}^m p^{ij} = 1$; $q^{ij} \ge 0$ for $i \ne j$ and $\sum_{j=1}^m q^{ij} = 0$ for each i). Suppose that P is irreducible and aperiodic with the stationary distribution denoted by $\nu = (\nu_1, \nu_2, \dots, \nu_m) \in \mathbb{R}^{1\times m}$. Denote by p_k^ε the probability vector $p_k^\varepsilon = (P(\alpha_k^\varepsilon = 1), \cdots, P(\alpha_k^\varepsilon = m)) \in \mathbb{R}^{1\times m}$. Assume that the initial probability p_0^ε is independent of ε, i.e., $p_0^\varepsilon = p_0 = (p_0^1, p_0^2, \dots, p_0^m)$. Before proceeding further, we present a result on asymptotic expansions of the probability vector p_k^ε and the k-step transition matrix $(P^\varepsilon)^k$. The following lemma is a special case of Proposition 2.1.

Lemma 2.3. *Assume that P in* (2.9) *is irreducible and aperiodic. Then the following assertions hold:*

(a) *For the probability distribution vector p^ε_k, for some λ with $0 < \lambda < 1$,*

$$p^\varepsilon_k = \nu + O(\varepsilon + \lambda^k). \tag{2.10}$$

(b) *For $k \le T/\varepsilon$ and some fixed T, the k-step transition matrix $(P^\varepsilon)^k$ satisfies*

$$(P^\varepsilon)^k = \Phi + \varepsilon\widehat{\Phi}(\varepsilon k) + \Psi(k) + \varepsilon\widehat{\Psi}(k) + O(\varepsilon^2), \tag{2.11}$$

where $\Phi = (1, 1, \ldots, 1)'(\nu_1, \nu_2, \ldots, \nu_m)$, $\widehat{\Phi}(t)$ is uniformly bounded in $[0, T]$, and $\Psi(k)$ and $\widehat{\Psi}(k)$ satisfy $|\Psi(k)| + |\widehat{\Psi}(k)| \le K\lambda^k$ for some $K > 0$ and $0 < \lambda < 1$.

Remark 2.4. (i) Denote $\mathcal{F}^{\alpha^\varepsilon}_n = \sigma\{\alpha^\varepsilon_k :\ 0 \le k \le n\}$ for $n = 0, 1, \ldots$ From the above lemma, there exists a constant C not depending on ε, k, l such that for $k \ge l \ge 0$,

$$\begin{aligned}
&|P(\alpha^\varepsilon_k = i) - \nu_i| \le C(\varepsilon + \lambda^k),\\
&|P(\alpha^\varepsilon_k = i|\alpha^\varepsilon_l = j) - \nu_i| \le C(\varepsilon + \lambda^{k-l}),\\
&|E(I[\alpha^\varepsilon_k = i] - \nu_i\big|\mathcal{F}^{\alpha^\varepsilon}_l)| \le C(\varepsilon + \lambda^{k-l}).
\end{aligned} \tag{2.12}$$

(ii) In view of (2.10) and (2.11), for positive integers $p > k$ and $i, j \in \mathcal{M}$,

$$P\big(\alpha^\varepsilon_p = j, \alpha^\varepsilon_k = i\big) = P\big(\alpha^\varepsilon_p = j\big|\alpha^\varepsilon_k = i\big)P\big(\alpha^\varepsilon_k = i\big) = \big[\nu_j + \psi_{ij}(p-k)\big]\nu_i + O\big(\varepsilon + \lambda^k\big).$$

Hence

$$E\big[I(\alpha^\varepsilon_k = i) - \nu_i\big]\big[I(\alpha^\varepsilon_p = j) - \nu_j\big] = \nu_i\psi_{ij}(p-k) + O\big(\varepsilon + \lambda^k\big). \tag{2.13}$$

Similarly, for $i, j \in \mathcal{M}$, $i \ne j$,

$$\begin{aligned}
&E\big[I(\alpha^\varepsilon_k = i) - \nu_i\big]^2 = \nu_i(1-\nu_i) + O\big(\varepsilon + \lambda^k\big),\\
&E\big[I(\alpha^\varepsilon_k = i) - \nu_i\big]\big[I(\alpha^\varepsilon_k = j) - \nu_j\big] = -\nu_i\nu_j + O\big(\varepsilon + \lambda^k\big).
\end{aligned} \tag{2.14}$$

(iii) Note that $\Psi(k)$ and $\widehat{\Psi}(k)$ above decay exponentially. Because P is irreducible and aperiodic, P has an eigenvalue 1 with multiplicity 1 and all other eigenvalues are inside the unit circle. Thus the λ in the assertion is related to the largest norm of the non-unity eigenvalue. The fact of $\lambda < 1$ yields the geometric or exponential decay.

2.2 Mixing Sequences

In our study, we will work with mixing processes. For two sub-σ-fields $\mathcal{A}, \mathcal{B}$ of $\mathcal{F}$ denote $\phi(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}, P(A)>0} \big|P(B|A) - P(B)\big|$. Recall that a sequence $(X_k : k \in \mathbb{Z})$ is ϕ-mixing (or uniform mixing) if $\phi(n) \to 0$ as $n \to \infty$ where the uniform mixing measure function $\phi(n)$ is defined by

$$\phi(n) = \sup_{k\in\mathbb{Z}} \phi\big(\sigma(\ldots, X_{k-1}, X_k), \sigma(X_{k+n}, X_{k+n+1}, \ldots)\big).$$

The term uniform mixing is taken from [17], [6] and [7], and the mixing rate is modeled after [7, Proposition 2.6].

Remark 2.5. For convenience, we present three mixing inequalities, which will be used frequently in what follows.

Suppose that $(X_k : k \in \mathbb{Z})$ is a ϕ-mixing sequence with mixing measure $\phi(n)$, $X \in \sigma(\ldots, X_{k-1}, X_k)$ and $Y \in \sigma(X_{k+n}, X_{k+n+1}, \ldots)$ such that $\|X\|_p$ and $\|Y\|_q < \infty$ with $p, q \ge 1$, $1/p + 1/q = 1$, where $\|\cdot\|_p$ and $\|\cdot\|_q$ are the usual l_p and l_q-norms respectively (e.g., $\|X\|_p = \big(E|X|^p\big)^{1/p}$). Then the following inequalities hold:

$$\big|EXY - EXEY\big| \le 2\phi(n)^{1/p}\|X\|_p\|Y\|_q, \tag{2.15}$$

$$\big\|E(Y|\sigma(\ldots, X_{k-1}, X_k)) - EY\big\|_p \le 2\phi(n)^{1/q}\|Y\|_p. \tag{2.16}$$

Inequality (2.15) is given in [17, Lemma 1.2.8, p.11] and inequality (2.16) is a special case of [7, Proposition 2.6, p.349]. For convenience, we also present here another inequality, which is a consequence of (2.16), and the Liapunov inequality

$$E\Big|E(Y|\sigma(\ldots, X_{k-1}, X_k)) - EY\Big| \leq 2\phi(n)^{1/q}\|Y\|_p. \tag{2.17}$$

3 Weak Convergence

3.1 Formulation

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Recall that we will use C to denote a generic positive constant with the convention $CC = C$ and $C + C = C$ used.

Let $\varepsilon > 0$ and α_k^ε be a time-homogeneous Markov chain on $(\Omega, \mathcal{F}, P)$ with state space $\mathcal{M}$ containing m_0 states and transition matrix $P_\varepsilon = P + \varepsilon Q$, where $P = (p^{ij})$ is a transition probability matrix and $Q = (q^{ij})$ is a generator of a continuous-time Markov chain (i.e., $p^{ij} \geq 0$ and $\sum_{j=1}^{m_0} p^{ij} = 1$; $q^{ij} \geq 0$ for $i \neq j$ and $\sum_{j=1}^{m_0} q^{ij} = 0$ for each i). Assume that the state space $\mathcal{M}$ can be written as

$$\begin{aligned} \mathcal{M} &= \{s_{11}, \ldots, s_{1,m_1}\} \cup \{s_{21}, \ldots, s_{2,m_2}\} \cup \cdots \cup \{s_{l_0 1}, \ldots, s_{l_0, m_{l_0}}\} \\ &= \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_{l_0}, \end{aligned} \tag{3.1}$$

with $m_0 = m_1 + m_2 + \cdots + m_{l_0}$ and $P = \text{diag}[P^1, P^2, \ldots, P^{l_0}]$, where P^i, $i \leq l_0$, are also transition matrices themselves. The subspace $\mathcal{M}_i$ for each $i = 1, 2, \ldots, l_0$, consists of recurrent states belonging to the ith ergodic class. We also assume that

(A1) For $i \leq l_0$, P^i is irreducible and aperiodic.

Let p_k^ε be the probability vector $p_k^\varepsilon = (P(\alpha_k^\varepsilon = s_{ij})) \in \mathbb{R}^{1 \times m_0}$, and ν^i the stationary distribution corresponding to the transition matrix P_i. Assume that $p_0^\varepsilon = (P(\alpha_0^\varepsilon = s_{ij})) = p_0$ and define an aggregated process $\overline{\alpha}_k^\varepsilon$ of α_k^ε by

$$\overline{\alpha}_k^\varepsilon = i \text{ if } \alpha_k^\varepsilon \in \mathcal{M}_i \quad \text{for} \quad i = 1, \ldots, l_0, \;\; \overline{\alpha}^\varepsilon(t) = \overline{\alpha}_k^\varepsilon \quad \text{for} \quad t \in [\varepsilon k, \varepsilon(k+1)).$$

In this section, we setup the problem in a simplified form, namely, within each $\mathcal{M}_i$ for $i = 1, \ldots, l_0$, there corresponds to only one sequence $\{X(k, i)\}$. This will facilitate the analysis in the next section.

For each $i \le l_0$, let $\{X(k, i)\}$ be a wide-sense (or covariance) stationary sequence of $\mathbb{R}^d$-valued random variables on $(\Omega, \mathcal{F}, P)$ with $X(k, i) = (X_1(k, i), \ldots, X_d(k, i)) \in \mathbb{R}^d$, and $\{(X(k, 1), \ldots, X(k, l_0)) : k \in \mathbb{Z}\}$ is an $\mathbb{R}^{l_0 \times d}$-valued wide-sense stationary sequence. We assume the following conditions hold.

(A2) The sequence $\{(X(k, 1), \ldots, X(k, l_0)) : k \in \mathbb{Z}\}$ is independent of $\{\alpha_k^\varepsilon\}$, and is ϕ-mixing with mixing measure denoted by $\phi(\cdot)$ such that

$$EX(k, i) = 0, \quad E|X(k, i)|^{2(1+\delta)} \le C, \quad \forall\, k \ge 1;\ i = 1, \ldots, l_0, \tag{3.2}$$

for some $\delta > 0$ and $C > 0$ not depend on k, i, and

$$\sum_{n=0}^{\infty} \phi(n)^{\frac{\delta}{1+\delta}} < \infty. \tag{3.3}$$

To proceed, denote $\mathcal{F}_k^X = \sigma\{X(l, i) :\ l \le k,\ i = 1, \ldots, l_0\}$, $\mathcal{F}_k^{\alpha^\varepsilon} = \sigma\{\alpha_l^\varepsilon :\ l \le k\}$. Define

$$z_k^\varepsilon = \sqrt{\varepsilon} \sum_{l=0}^{k-1} X(l, \bar{\alpha}_l^\varepsilon) = \sqrt{\varepsilon} \sum_{l=0}^{k-1} \sum_{i=1}^{l_0} X(l, i) I(\overline{\alpha}_l^\varepsilon = i), \tag{3.4}$$

$$z^\varepsilon(t) = \sqrt{\varepsilon} \sum_{j=0}^{\lfloor t/\varepsilon \rfloor - 1} X(j, \bar{\alpha}_j^\varepsilon), \tag{3.5}$$

where $I(A)$ is the usual indicator function for the event A, and $\lfloor t/\varepsilon \rfloor$ denotes the integer part of the real number t/ε. We are interested in the weak convergence of the process $z^\varepsilon(t)$. It will be shown in the next section that $(z^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to a switching diffusion process $(z(\cdot), \overline{\alpha}(\cdot))$, which is the unique solution of the martingale problem associated with the following operator

$$\mathcal{L}f(x,i) = \frac{1}{2}\sum_{j_1=1}^{d}\sum_{j_2=1}^{d} a^{j_1 j_2}(i)\frac{\partial^2 f(x,i)}{\partial x^{j_1}\partial x^{j_2}} + \overline{Q}f(x,\cdot)(i), \tag{3.6}$$

where

$$A(i) = (a^{j_1 j_2}(i)) = EX(0,i)X'(0,i) + \sum_{k=1}^{\infty}\Big[EX(k,i)X'(0,i) + EX(0,i)X'(k,i)\Big], \tag{3.7}$$

for $i = 1, 2, \ldots, l_0$ and the matrix $\overline{Q}$ is given in (2.3).

Remark 3.1. From (3.2), Cauchy-Schwartz and Liapunov inequalities, there exists a constant C that does not depend on k, l, j, i such that

$$\|X_j(k,i)X_j(l,i)\|_{1+\delta}, \|X_j(k,i)\|_{2(1+\delta)}, \|X_j(k,i)\|_{\frac{2(1+\delta)}{1+2\delta}} \le C. \tag{3.8}$$

Next, we will state a proposition that is needed in our proof. Its proof can be found in [15] (see also [16, Chapter 7]).

Proposition 3.2. *Let $\{x_k^\varepsilon\}$ be a d-dimensional stochastic process in discrete time and $x^\varepsilon(\cdot)$ be its piecewise constant interpolation on the interval $[\varepsilon k, \varepsilon k + \varepsilon)$. Suppose that*

(a) *$(x^\varepsilon(\cdot))$ is tight in $D([0,T),\mathbb{R}^d)$ and $x^\varepsilon(0) \Rightarrow x_0$.*

(b) *The martingale problem with operator $\mathcal{L}$ has a unique solution $x(\cdot)$ in $D([0,T),\mathbb{R}^d)$ for each initial condition.*

(c) *For each* $g(\cdot) \in C_0^2$, *there exists a sequence* $(g^\varepsilon(\cdot))$ *such that*

(c1) $g^\varepsilon(\cdot)$ *is a constant on each interval* $[\varepsilon k, \varepsilon k + \varepsilon)$, *which is measurable (at* εk*) with respect to* $\sigma(x_j^\varepsilon : j \le k)$,

(c2) $$\sup_{0\le k\le T/\varepsilon,\varepsilon} E|g^\varepsilon(\varepsilon k)| + \sup_{0\le k\le T/\varepsilon,\varepsilon} \frac{1}{\varepsilon} E\Big| E(g^\varepsilon(\varepsilon k + \varepsilon)|x_1^\varepsilon, \dots, x_k^\varepsilon) - g^\varepsilon(\varepsilon k)\Big| < \infty,$$

and as $\varepsilon \to 0$ *with* $\varepsilon k \to t$,

(c3) $E|g^\varepsilon(\varepsilon k) - g(x^\varepsilon(\varepsilon k))| \to 0$,

(c4) $$E\Big|\frac{E(g^\varepsilon(\varepsilon k + \varepsilon)|x_1^\varepsilon, \dots, x_k^\varepsilon) - g^\varepsilon(\varepsilon k)}{\varepsilon} - \mathcal{L}g(x^\varepsilon(\varepsilon k))\Big| \to 0.$$

Then $x^\varepsilon(\cdot)$ *converges weakly to* $x(\cdot)$, *the unique solution to the martingale problem with operator* $\mathcal{L}$ *and initial condition* x_0.

Remark 3.3. If $\mathcal{G}_k^\varepsilon$ is a σ-field such that $\sigma(x_1^\varepsilon, \dots, x_k^\varepsilon) \subset \mathcal{G}_k^\varepsilon$ for $k = 1, 2, \dots; \varepsilon > 0$ then

$$|E(g^\varepsilon(\varepsilon k + \varepsilon)|x_1^\varepsilon, \dots, x_k^\varepsilon) - g^\varepsilon(\varepsilon k)| \le E[|E(g^\varepsilon(\varepsilon k + \varepsilon) - g^\varepsilon(\varepsilon k)|\mathcal{G}_k^\varepsilon)||x_1^\varepsilon, \dots, x_k^\varepsilon].$$

Thus,

(c2') $$\sup_{0\le k\le T/\varepsilon,\ \varepsilon} E|g^\varepsilon(\varepsilon k)| + \sup_{0\le k\le T/\varepsilon,\ \varepsilon} \frac{1}{\varepsilon} E\Big|E(g^\varepsilon(\varepsilon k + \varepsilon) \mid \mathcal{G}_k^\varepsilon) - g^\varepsilon(\varepsilon k)\Big| < \infty$$ implies (c2);

(c4') $$E\Big|\frac{E(g^\varepsilon(\varepsilon k + \varepsilon)|\mathcal{G}_k^\varepsilon) - g^\varepsilon(\varepsilon k)}{\varepsilon} - \mathcal{L}g(x^\varepsilon(\varepsilon k))\Big| \to 0$$ implies (c4).

3.2 Weak Convergence

This section presents the main result of this chapter. To obtain the desired weak convergence, we use martingale problem formulations. It requires the verification of tightness of the underlying sequence in an appropriate function space, which is given in Proposition 3.4. Then in the second step, we show that the martingale problem associated with a limit operator has a unique solution, which is stated in Proposition 3.9. The third part of the proof is to characterize the limit process in Theorem 3.10. In the process of obtaining the desired result, a number of technical complements are formulated as lemmas and propositions. They are interesting in their own right. We divide this section into several subsections in accordance with the aforementioned tasks.

3.2.1 Tightness of $(z^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$

We aim to obtain the tightness of $(z^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ here. The main result of this section is the following proposition.

Proposition 3.4. *The process $(z^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ is tight.*

We shall prove this proposition by means of establishing a series of lemmas. Owing to Proposition 2.1, it suffices to work with $z^\varepsilon(\cdot)$. Recall that

$$\mathcal{F}_k^X = \sigma\{X(l,i) : l \le k,\ i = 1, \ldots, l_0\}, \quad \mathcal{F}_k^{\alpha^\varepsilon} = \sigma\{\alpha_l^\varepsilon : l \le k\}.$$

Let

$$\mathcal{F}_t^\varepsilon = \sigma(z^\varepsilon(s) : 0 \le s \le t), \quad \mathcal{G}_k^\varepsilon = \mathcal{F}_k^X \vee \mathcal{F}_k^{\alpha^\varepsilon}.$$

Then it follows that $\mathcal{F}_t^\varepsilon \subset \mathcal{G}_{\lfloor t/\varepsilon \rfloor}^\varepsilon = \mathcal{F}_{\lfloor t/\varepsilon \rfloor}^X \vee \mathcal{F}_{\lfloor t/\varepsilon \rfloor}^{\alpha^\varepsilon}$. To obtain the tightness we need to verify for each $T > 0$ and $t \le T$,

$$\lim_{h\to 0} \limsup_{\varepsilon\to 0} E \sup_{0\le s\le h} E(|z^\varepsilon(t+s) - z^\varepsilon(t)|^2 | \mathcal{F}_t^\varepsilon) = 0, \tag{3.9}$$

and

$$\lim_{K\to\infty} \limsup_{\varepsilon\to 0} P\Big(\sup_{0\le t\le T} |z^\varepsilon(t)| \ge K\Big) = 0, \quad \text{for each } \ T > 0, \tag{3.10}$$

respectively (see [15, Theorem 3, p. 47]). We proceed to prove these in the rest of this section.

Lemma 3.5. *For each $T > 0$ and any $0 < t \le T$, (3.9) holds.*

Proof. By the Cauchy-Schwartz inequality,

$$\begin{aligned} |z^\varepsilon(t+s) - z^\varepsilon(t)|^2 &= \sum_{j=1}^d |z_j^\varepsilon(t+s) - z_j^\varepsilon(t)|^2 \\ &= \sum_{j=1}^d \Big| \sqrt{\varepsilon} \sum_{k=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor - 1} \sum_{i=1}^{l_0} X_j(k,i) I(\overline{\alpha}_k^\varepsilon = i) \Big|^2 \\ &\le l_0 \sum_{j=1}^d \sum_{i=1}^{l_0} \Big| \sqrt{\varepsilon} \sum_{k=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor - 1} X_j(k,i) I(\overline{\alpha}_k^\varepsilon = i) \Big|^2. \end{aligned} \tag{3.11}$$

By the independence of $\{X(k,i)\}$ and $\{\alpha_k^\varepsilon\}$,

$$\begin{aligned}
&E\Big(\Big|\sqrt{\varepsilon}\sum_{k=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+s)/\varepsilon\rfloor-1} X_j(k,i)I(\overline{\alpha}_k^\varepsilon=i)\Big|^2\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\vee\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)\\
&=\varepsilon\sum_{k=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+s)/\varepsilon\rfloor-1} E\Big(X_j^2(k,i)I(\overline{\alpha}_k^\varepsilon=i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\vee\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)\\
&\quad+2\varepsilon\sum_{\lfloor t/\varepsilon\rfloor\le k<l<\lfloor (t+s)/\varepsilon\rfloor} E\Big(I(\overline{\alpha}_k^\varepsilon=i,\overline{\alpha}_l^\varepsilon=i)X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\vee\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)\\
&=\varepsilon\sum_{k=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+s)/\varepsilon\rfloor-1} P\Big(\overline{\alpha}_k^\varepsilon=i\Big|\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)E\Big(X_j^2(k,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)\\
&\quad+2\varepsilon\sum_{\lfloor t/\varepsilon\rfloor\le k<l<\lfloor (t+s)/\varepsilon\rfloor} P\Big(\overline{\alpha}_k^\varepsilon=i,\overline{\alpha}_l^\varepsilon=i\Big|\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)E\Big(X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)\\
&\le\varepsilon\sum_{k=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+h)/\varepsilon\rfloor-1} E\Big(X_j^2(k,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)+2\varepsilon\sum_{\lfloor t/\varepsilon\rfloor\le k<l<\lfloor (t+h)/\varepsilon\rfloor}\Big|E\Big(X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)\Big|\\
&:=\gamma_{\varepsilon,j,i}(h).
\end{aligned}\tag{3.12}$$

Since the inequality (3.12) holds for any s with $0\le s\le h$ and $\mathcal{F}_t^\varepsilon\subset\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\vee\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}$ it follows from (3.11) that

$$\begin{aligned}
\sup_{0\le s\le h} E\big(|z^\varepsilon(t+s)-z^\varepsilon(t)|^2\big|\mathcal{F}_t^\varepsilon\big) &= \sup_{0\le s\le h} E\Big[E\Big(|z^\varepsilon(t+s)-z^\varepsilon(t)|^2\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\vee\mathcal{F}^{\alpha^\varepsilon}_{\lfloor t/\varepsilon\rfloor}\Big)\Big|\mathcal{F}_t^\varepsilon\Big]\\
&\le E\big(\gamma_\varepsilon(h)\big|\mathcal{F}_t^\varepsilon\big),
\end{aligned}\tag{3.13}$$

where $\gamma_\varepsilon(h)=l_0\sum_{j=1}^d\sum_{i=1}^{l_0}\gamma_{\varepsilon,j,i}(h)$.

On the other hand,

$$E\gamma_{\varepsilon,j,i}(h)=\varepsilon\sum_{k=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+h)/\varepsilon\rfloor-1} EX_j^2(k,i)+2\varepsilon\sum_{\lfloor t/\varepsilon\rfloor\le k<l<\lfloor (t+h)/\varepsilon\rfloor} E\Big|E\Big(X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)\Big|.\tag{3.14}$$

Recall that $EX_j(k,i)=EX_j(l,i)=0$. Thus, by the triangle inequality, mixing inequalities (2.17) with $p=1+\delta$ and $q=\frac{1+\delta}{\delta}$, and (2.15) with $p=\frac{2(1+\delta)}{1+2\delta}$ and

$q = 2(1+\delta)$,

$$\begin{aligned}
&E\Big|E\Big(X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big)\Big| \\
&\quad\le E\Big|E\Big(X_j(k,i)X_j(l,i)\Big|\mathcal{F}^X_{\lfloor t/\varepsilon\rfloor}\Big) - EX_j(k,i)X_j(l,i)\Big| \\
&\qquad\qquad + E\Big|EX_j(k,i)X_j(l,i) - EX_j(k,i)EX_j(l,i)\Big| \\
&\quad\le 2\phi\Big(k-\lfloor\frac{t}{\varepsilon}\rfloor\Big)^{\frac{\delta}{1+\delta}}\|X_j(k,i)X_j(l,i)\|_{1+\delta} \\
&\qquad\qquad + 2\phi(l-k)^{\frac{1+2\delta}{2(1+\delta)}}\|X_j(k,i)\|_{2(1+\delta)}\|X_j(k,i)\|_{\frac{2(1+\delta)}{1+2\delta}} \\
&\quad\le C\Big[\phi\Big(k-\Big\lfloor\frac{t}{\varepsilon}\Big\rfloor\Big)^{\frac{\delta}{1+\delta}} + \phi(l-k)^{\frac{\delta}{1+\delta}}\Big].
\end{aligned} \tag{3.15}$$

In the last inequality, we have used (3.8) and the facts that $\phi(k) \le 1$ for all $k \ge 1$ and $\frac{1+2\delta}{2(1+\delta)} > \frac{\delta}{1+\delta}$.

Note that $EX_j^2(k,i) \le C$. Thus, by (3.14) and (3.15),

$$\begin{aligned}
E\gamma_{\varepsilon,j,i}(h) &\le \varepsilon C\Big\lfloor\frac{h}{\varepsilon}\Big\rfloor + \varepsilon C \sum_{\lfloor t/\varepsilon\rfloor \le k<l<\lfloor (t+h)/\varepsilon\rfloor} \Big[\phi\Big(k-\Big\lfloor\frac{t}{\varepsilon}\Big\rfloor\Big)^{\frac{\delta}{1+\delta}} + \phi(l-k)^{\frac{\delta}{1+\delta}}\Big] \\
&\le Ch + \varepsilon C \sum_{0\le k<l\le\lfloor h/\varepsilon\rfloor} \Big[\phi(k)^{\frac{\delta}{1+\delta}} + \phi(l-k)^{\frac{\delta}{1+\delta}}\Big] \\
&= Ch + 2\varepsilon C \sum_{k=1}^{\lfloor h/\varepsilon\rfloor} \Big(\Big\lfloor\frac{h}{\varepsilon}\Big\rfloor - k\Big)\phi(k)^{\frac{\delta}{1+\delta}} \le Ch + 2Ch\sum_{k=1}^{\infty}\phi(k)^{\frac{\delta}{1+\delta}} \le Ch,
\end{aligned} \tag{3.16}$$

where C is a constant not depending on ε, j, i. We have used (3.3) to obtain the last inequality. By (3.13) and the definition of $\gamma_\varepsilon(h)$,

$$\begin{aligned}
&\lim_{h\to 0}\limsup_{\varepsilon\to 0} E\sup_{0\le s\le h} E\Big(|z^\varepsilon(t+s)-z^\varepsilon(t)|^2\Big|\mathcal{F}^\varepsilon_t\Big) \\
&\quad\le \lim_{h\to 0}\limsup_{\varepsilon\to 0} E\gamma_\varepsilon(h) = \lim_{h\to 0}\limsup_{\varepsilon\to 0} l_0\sum_{j=1}^{d}\sum_{i=1}^{l_0} E\gamma_{\varepsilon,j,i}(h) \\
&\quad\le \lim_{h\to 0}\limsup_{\varepsilon\to 0} l_0\sum_{j=1}^{d}\sum_{i=1}^{l_0} Ch = 0.
\end{aligned} \tag{3.17}$$

This proves the lemma. □

To proceed, we verify (3.10). In dealing with dynamic systems, one often uses a truncation device to verify (3.10). That is one works with a truncated process and obtain its tightness and weakly limit and then let the truncation bounds grow to conclude the weak convergence of the untruncated sequence. Here we handle the sequence directly without using truncation. The verification of (3.10) is provided in the next three lemmas.

Lemma 3.6. *Under* (A1) *and* (A2),

$$\lim_{K\to\infty}\limsup_{\varepsilon\to 0}\max_{S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\Big|\sum_{k\in S}X(k,\overline{\alpha}^{\varepsilon}_k)\Big|\geq\frac{K}{\sqrt{\varepsilon}}\Big)=0. \tag{3.18}$$

Proof. By the Markov inequality, for each $i=1,\dots,l_0$, and $S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}$,

$$\begin{aligned}
&P\Big(\Big|\sum_{k\in S}X(k,i)I(\overline{\alpha}^{\varepsilon}_k=i)\Big|\geq\frac{K}{l_0\sqrt{\varepsilon}}\Big)\\
&\leq\frac{\varepsilon l_0^2}{K^2}E\Big|\sum_{k\in S}X(k,i)I(\overline{\alpha}^{\varepsilon}_k=i)\Big|^2\\
&\leq\frac{\varepsilon l_0^2}{K^2}\sum_{j=1}^{d}\Big[\sum_{k\in S}EX_j^2(k,i)+2\sum_{k,l\in S,\ k<l}|E(X_j(k,i)X_j(l,i))|\Big].
\end{aligned} \tag{3.19}$$

Note that $EX_j(k,i)=EX_j(l,i)=0$, so by (2.15) with $p=\frac{2(1+\delta)}{1+2\delta}$ and $q=2(1+\delta)$, we have

$$|E(X_j(k,i)X_j(l,i))|\leq 2\phi(l-k)^{\frac{1+2\delta}{2(1+\delta)}}\|X_j(k,i)\|_{2(1+\delta)}\|X_j(l,i)\|_{\frac{2(1+\delta)}{1+2\delta}}\leq C\phi(l-k)^{\frac{1+2\delta}{2(1+\delta)}}, \tag{3.20}$$

where we have used (3.8) in the last inequality. Since $EX_j^2(k,i)\leq C$ by assumption (A2), it follows from (3.19) and (3.20) that

$$\begin{aligned}
P\Big(\Big|\sum_{k\in S}X(k,i)I(\overline{\alpha}^{\varepsilon}_k=i)\Big|\geq\frac{K}{l_0\sqrt{\varepsilon}}\Big)&\leq\frac{\varepsilon d l_0^2}{K^2}\Big[C|S|+2C\sum_{k,l\in S\ k<l}\phi(l-k)^{\frac{1+2\delta}{2(1+\delta)}}\Big]\\
&\leq\frac{\varepsilon d l_0^2 C|S|}{K^2}\Big[1+\sum_{n=1}^{\infty}\phi(n)^{\frac{1+2\delta}{2(1+\delta)}}\Big]\\
&\leq\frac{\varepsilon d l_0^2 C|S|}{K^2}\Big[1+\sum_{n=1}^{\infty}\phi(n)^{\frac{\delta}{1+\delta}}\Big]\leq\frac{\varepsilon d l_0^2 C|S|}{K^2},
\end{aligned} \tag{3.21}$$

where $|S|$ denotes the cardinality of the set S and (3.3) is used to get the last inequality. Therefore,

$$P\Big(\Big|\sum_{k\in S} X(k,\overline{\alpha}_k^\varepsilon)\Big| \geq \frac{K}{\sqrt{\varepsilon}}\Big) \leq \sum_{i=1}^{l_0} P\Big(\Big|\sum_{k\in S} X(k,i)I(\overline{\alpha}_k^\varepsilon = i)\Big| \geq \frac{K}{l_0\sqrt{\varepsilon}}\Big) \leq \frac{\varepsilon d l_0^3 C|S|}{K^2}. \tag{3.22}$$

In view of (3.22) and the fact that $|S| \leq \frac{T}{\varepsilon}$ for all $S \subset \{0, 1, \ldots, \lfloor T/\varepsilon \rfloor - 1\}$,

$$\begin{aligned} &\lim_{K\to\infty} \limsup_{\varepsilon\to 0} \max_{S\subset\{0,1,\ldots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\Big|\sum_{k\in S} X(k,\overline{\alpha}_k^\varepsilon)\Big| \geq \frac{K}{\sqrt{\varepsilon}}\Big) \\ &\leq \lim_{K\to\infty} \limsup_{\varepsilon\to 0} \max_{S\subset\{0,1,\ldots,\lfloor T/\varepsilon\rfloor-1\}} \frac{\varepsilon d l_0^3 C|S|}{K^2} = 0. \end{aligned} \tag{3.23}$$

The lemma is thus proved. □

To proceed, we need the following Lemma, whose proof can be found in [17] Lemma 2.2.7.

Lemma 3.7. *Let $\{Y_k, k \geq 1\}$ be a ϕ-mixing sequence and η a real number with $0 < \eta < 1$. Suppose that there exists an integer p, $1 \leq p \leq n$, a number $A > 0$ such that*

$$\phi_Y(p) + \max_{p\leq i\leq n} P\Big(|Z_n - Z_i| \geq A\Big) \leq \eta, \tag{3.24}$$

where $Z_n = Y_1 + Y_2 + \cdots + Y_n$ and $\phi_Y(\cdot)$ is the mixing measure function of the sequence $\{Y_k\}$. Then, for any $a \geq 0$ and $b \geq 0$ we have

$$P\Big(\max_{1\leq i\leq n} |Z_i| \geq a + A + b\Big) \leq \frac{1}{1-\eta}\Big[P\Big(|Z_n| \geq a\Big) + P\Big(\max_{1\leq i\leq n} |Y_i| \geq \frac{b}{p-1}\Big)\Big]. \tag{3.25}$$

Lemma 3.8. *Under* (A1) *and* (A2),

$$\lim_{K\to\infty} \limsup_{\varepsilon\to 0} P\Big(\max_{0\leq l\leq \lfloor T/\varepsilon\rfloor-1} \Big|\sum_{k=0}^{l} X(k,\overline{\alpha}_k^\varepsilon)\Big| \geq \frac{K}{\sqrt{\varepsilon}}\Big) = 0. \tag{3.26}$$

Proof. In order to prove (3.26), it suffices to show that for each $\delta > 0$ there exist $K_0 = K(\delta)$ and $\varepsilon_0 = \varepsilon(\delta)$ such that

$$P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k=0}^{l} X(k,\overline{\alpha}_k^\varepsilon)\Big| \ge \frac{K_0}{\sqrt{\varepsilon}}\Big) < \delta, \quad \forall\, \varepsilon < \varepsilon_0. \tag{3.27}$$

Fix $i \in \{1,\dots,l_0\}$. For each $S \subset \{0,1,\dots,\lfloor T/\varepsilon\rfloor - 1\}$, denote

$$\Omega_S^i = \Big\{\text{For } 0 \le k \le \lfloor \tfrac{T}{\varepsilon}\rfloor - 1,\ I(\overline{\alpha}_k^\varepsilon = i) = 1 \text{ if and only if } k \in S\Big\}.$$

It is clear that if S_1 and S_2 are two different subsets of $\{0,1,\dots,\lfloor T/\varepsilon\rfloor - 1\}$ then $\Omega_{S_1}^i$ and $\Omega_{S_2}^i$ are disjoint. Moreover, $\Omega = \cup_{S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}}\Omega_S^i$. Therefore,

$$\begin{aligned}
&P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k=0}^{l} X(k,i)I(\overline{\alpha}_k^\varepsilon = i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big)\\
&= \sum_{S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k=0}^{l} X(k,i)I(\overline{\alpha}_k^\varepsilon = i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big|\Omega_S^i\Big)P(\Omega_S^i)\\
&= \sum_{S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big|\Omega_S^i\Big)P(\Omega_S^i)\\
&= \sum_{S\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big)P(\Omega_S^i).
\end{aligned} \tag{3.28}$$

We have used the independence of $\{X(k,i) : k \ge 0\}$ and $\{\alpha_k^\varepsilon\}$ in the last equation.

To proceed, we fix $S \subset \{0,1,\dots,\lfloor T/\varepsilon\rfloor - 1\}$. By (3.18), for each $i = 1,\dots,l_0$ and $0 < \delta_1 < \frac{1}{2}$ there exist $K_1 = K(\delta_1)$ and $\varepsilon_1 = \varepsilon(\delta_1)$ such that

$$\max_{\widetilde{S}\subset\{0,1,\dots,\lfloor T/\varepsilon\rfloor-1\}} P\Big(\Big|\sum_{k\in\widetilde{S}} X(k,i)\Big| \ge \frac{K_1}{3l_0\sqrt{\varepsilon}}\Big) < \frac{\delta_1}{2}, \quad \forall\, \varepsilon < \varepsilon_1. \tag{3.29}$$

Thus, for $\widetilde{S} \subset \{0,1,\dots,\lfloor T/\varepsilon\rfloor - 1\}$ and $\varepsilon < \varepsilon_1$,

$$P\Big(\Big|\sum_{k\in\widetilde{S}} X(k,i)\Big| \ge \frac{K_1}{3l_0\sqrt{\varepsilon}}\Big) < \frac{\delta_1}{2}. \tag{3.30}$$

Put $N = |S|$ and choose an integer p such that $\phi(p) < \frac{\delta_1}{2}$.

Case 1: $N \le p$. Then we have

$$\begin{aligned} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) &\le P\Big(\sum_{k\in S}|X(k,i)| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) \\ &\le \frac{l_0^2\varepsilon}{K^2}E\Big(\sum_{k\in S}|X(k,i)|\Big)^2 \\ &\le \frac{l_0^2\varepsilon}{K^2}N\sum_{k\in S}E|X(k,i)|^2 \\ &\le \frac{Cl_0^2N^2\varepsilon}{K^2} \le \frac{Cl_0^2p^2\varepsilon}{K^2} \le 2\Big(\frac{\delta_1}{2}+\frac{C}{K^2}\Big). \end{aligned} \tag{3.31}$$

Case 2: $N > p$. We consider the random vectors $X(k,i)$ for $k \in S$ and $k \ge \lfloor T/\varepsilon \rfloor$ and arrange them in increasing order of k. Denote this sequence by $Y_1, Y_2, \ldots, Y_N, \ldots$ It is clear that the mixing measure of the sequence (say $\phi_Y(\cdot)$) is smaller than that of the sequence $\{X(k,i) : k \ge 1\}$. That is, $\phi_Y(n) \le \phi(n)$ for each positive integer n. Taking this fact into account, by the choice of p and (3.30), we get

$$\phi_Y(p) + \max_{1\le i\le N} P\Big(|Z_n - Z_i| \ge \frac{K_1}{3l_0\sqrt{\varepsilon}}\Big) < \delta_1,$$

where Z_n denotes $Y_1 + Y_2 + \cdots + Y_n$. This implies the condition (3.24) of Lemma 3.7 with $A = \frac{K_1}{3l_0\sqrt{\varepsilon}}$ and $\eta = \delta_1$. Hence, by Lemma 3.7 with $A = a = b = \frac{K}{3l_0\sqrt{\varepsilon}}$, $K \ge K_1$, we obtain

$$\begin{aligned} &P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) = P\Big(\max_{0\le l\le N}\Big|\sum_{k=1}^{l} Y_k\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) \\ &\le \frac{1}{1-\delta_1}\Big[P\Big(\Big|\sum_{k=1}^{N} Y_k\Big| \ge \frac{K}{3l_0\sqrt{\varepsilon}}\Big) + P\Big(\max_{0\le k\le N}|Y_k| \ge \frac{K}{3l_0(p-1)\sqrt{\varepsilon}}\Big)\Big] \\ &\le \frac{1}{1-\delta_1}\Big[P\Big(\Big|\sum_{k\in S} X(k,i)\Big| \ge \frac{K}{3l_0\sqrt{\varepsilon}}\Big) + P\Big(\max_{0\le k\le \lfloor T/\varepsilon\rfloor -1}|X(k,i)| \ge \frac{K}{3l_0(p-1)\sqrt{\varepsilon}}\Big)\Big]. \end{aligned} \tag{3.32}$$

By (3.32), (3.30), and

$$\begin{aligned} P\Big(\max_{0\le k\le \lfloor T/\varepsilon\rfloor -1}|X(k,i)| \ge \frac{K}{3l_0(p-1)\sqrt{\varepsilon}}\Big) &\le \sum_{0\le k\le \lfloor T/\varepsilon\rfloor -1} P\Big(|X(k,i)| \ge \frac{K}{3l_0(p-1)\sqrt{\varepsilon}}\Big) \\ &\le \frac{9l_0^2(p-1)^2\varepsilon}{K^2}\sum_{0\le k\le \lfloor T/\varepsilon\rfloor -1} EX^2(k,i) \\ &\le \frac{C}{K^2}, \end{aligned}$$

we have

$$\begin{aligned} P\Big(\max\nolimits_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum\nolimits_{k\in S,\ k\le l} X(k,i)\Big| \ge \tfrac{K}{l_0\sqrt{\varepsilon}}\Big) &\le \frac{1}{1-\delta_1}\Big(\frac{\delta_1}{2}+\frac{C}{K^2}\Big) \\ &\le 2\Big(\frac{\delta_1}{2}+\frac{C}{K^2}\Big). \end{aligned} \tag{3.33}$$

Therefore, from (3.31) and (3.33), for $K > K_1$, $\varepsilon < \varepsilon_1$ and any set $S \subset \{0,1,\ldots,\lfloor T/\varepsilon\rfloor - 1\}$,

$$P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) \le 2\Big(\frac{\delta_1}{2}+\frac{C}{K^2}\Big). \tag{3.34}$$

By choosing $\delta_1 = \frac{\delta}{2l_0}$, $\varepsilon_\delta = \varepsilon_0$ and $K_0 > \max\{K_1, 2\sqrt{\frac{Cl_0}{\delta}}\}$ where C is the constant in (3.34) we have $2(\frac{\delta_1}{2}+\frac{C}{K_0^2}) < \frac{\delta}{l_0}$. Hence, from (3.28) and (3.34), for $K > K_0, \varepsilon < \varepsilon_0$,

$$\begin{aligned} &P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k=0}^{l} X(k,\overline{\alpha}_k^\varepsilon)\Big| \ge \frac{K}{\sqrt{\varepsilon}}\Big) \\ &\le \sum_{i=1}^{l_0} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k=0}^{l} X(k,i)I(\overline{\alpha}_k^\varepsilon = i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big) \\ &\le \sum_{i=1}^{l_0}\sum_{S\subset\{0,1,\ldots,\lfloor T/\varepsilon\rfloor -1\}} P\Big(\max_{0\le l\le \lfloor T/\varepsilon\rfloor -1}\Big|\sum_{k\in S,\ k\le l} X(k,i)\Big| \ge \frac{K}{l_0\sqrt{\varepsilon}}\Big)P(\Omega_S^i) \\ &\le \sum_{i=1}^{l_0}\sum_{S\subset\{0,1,\ldots,\lfloor T/\varepsilon\rfloor -1\}} \Big(\frac{\delta_1}{2}+\frac{C}{K^2}\Big)P(\Omega_S^i) < \frac{\delta}{l_0}\sum_{i=1}^{l_0}\sum_{S\subset\{0,1,\ldots,\lfloor T/\varepsilon\rfloor -1\}} P(\Omega_S^i) = \delta. \end{aligned} \tag{3.35}$$

This gives (3.27) and the Lemma is proved. □

Consequently, Lemma 3.8 yields (3.10). Combining this with Lemma 3.5, we obtain Proposition 3.4.

3.2.2 Uniqueness of Solution to the Martingale Problem

We state the following result of this section.

Proposition 3.9. *The martingale problem associated with the operator* $\mathcal{L}$ *defined by* (3.6) *has a unique solution.*

Proof. By virtue of Lemma 14.8 of [28], it suffices to verify the uniqueness in distribution of a solution $(z(t), \overline{\alpha}(t))$ of the martingale problem associated with the operator A for each $t \in [0, T]$. Consider the characteristic function $\widetilde{\varphi}(x, l) = \exp\{\iota(x\lambda + sl)\}$, for each positive integer l, $x \in \mathbb{R}^{1\times d}, \lambda \in \mathbb{R}^{d\times 1}, s \in \mathbb{R}$, and $\iota^2 = -1$. Note that $x\lambda$ above is just the usual inner product. Define $\varphi^{i_1 i_2}(t) = E[I(\overline{\alpha}(t) = i_1)\widetilde{\varphi}(z(t), i_2)]$ for $1 \le i_1, i_2 \le l_0$. Since $(z(t), \overline{\alpha}(t))$ is a solution of the martingale problem associated with the operator $\mathcal{L}$,

$$\varphi^{i_1 i_2}(t) - \varphi^{i_1 i_2}(0) - \int_0^t \Big\{ \sum_{j,j_0=1}^{d} a^{jj_0}(i_1)(-\lambda_j \lambda_{j_0})\varphi^{i_1 i_2}(u) + \sum_{i_3=1}^{l_0} \bar{q}^{i_3 i_1}\varphi^{i_3 i_2}(u)\Big\} du = 0, \tag{3.36}$$

where $\varphi^{i_1 i_2}(0) = EI(\overline{\alpha}(0) = i_1)\widetilde{\varphi}(0, i_2)$. Let $\varphi(t) = (\varphi^{i_1 i_2}(t), i_1, i_2 = 1, \ldots, l_0)$. Then (3.36) becomes $\varphi(t) = \varphi(0) + \int_0^t \varphi(u)G(u)du$, where $\varphi(0) = (\varphi^{i_1 i_2}(0))$ and G is a matrix-valued function defined by the integrand of (3.36). The equation for $\varphi(t)$ is a linear ordinary differential equation, so it has a unique solution. Thus, $\varphi(t)$ is uniquely determined. As a result, $E\exp\{\iota(z(t)\lambda + \overline{\alpha}(t)s)\} = \sum_{i=1}^{l_0} E\Big(I(\overline{\alpha}(t) = i)\exp\{\iota(z(t)\lambda + is)\}\Big)$ is uniquely determined for all $(\lambda, s) \in \mathbb{R}^{d\times 1} \times \mathbb{R}$. Therefore the distribution of $(z(t), \overline{\alpha}(t))$ is uniquely determined by the well-known uniqueness and inversion formula for characteristic functions. $\square$

3.2.3 Characterization of the Limit

This subsection is devoted to characterization of the limit. We first state the result, and then divide the task of proof into sub-tasks.

Theorem 3.10. *Assume* (A1) *and* (A2). *Then* $(z^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ *converges weakly to* $(z(\cdot), \overline{\alpha}(\cdot))$ *such that the limit is the solution of the martingale problem with operator given by* (3.6).

Proof. We use Proposition 3.2 with $x_k^\varepsilon = (z_k^\varepsilon, \overline{\alpha}_k^\varepsilon)$. For an appropriate function $g(\cdot)$, define the operator $\mathcal{L}^\varepsilon$ by

$$\mathcal{L}^\varepsilon g(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) = \frac{1}{\varepsilon} E_k^\varepsilon [g(z_{k+1}^\varepsilon, \overline{\alpha}_{k+1}^\varepsilon) - g(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon)], \tag{3.37}$$

where E_k^ε denotes the conditional expectation with respect to $\mathcal{G}_k^\varepsilon = \mathcal{F}_k^X \vee \mathcal{F}_k^{\alpha^\varepsilon}$. We will construct a perturbed test function f^ε and show that all conditions in Proposition 3.2 are satisfied. Along this line, we also obtain the representation of the limit operator and the limit covariance matrix. Hence the desired weak convergence follows.

For each $i = 1, \ldots, l_0$, let $f(\cdot, i)$ be any real-valued function with bounded derivatives up to the second order such that the second derivatives are Lipschitz continuous. Define

$$\bar{f}(x, \alpha) = \sum_{i=1}^{l_0} f(x, i) I_{\{\alpha \in \mathcal{M}_i\}}, \quad x \in \mathbb{R}^d, \alpha \in \mathcal{M}. \tag{3.38}$$

Definition (3.38) allows us to replace $f(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon)$ by $\overline{f}(z_k^\varepsilon, \alpha_k^\varepsilon)$. Denote

$$\begin{aligned}
&\hat{\nu} = \text{diag}(\nu^1, \ldots, \nu^{l_0}) \in \mathbb{R}^{l_0 \times m_0},\\
&\overline{\chi}_k^\varepsilon = (I_{\{\overline{\alpha}_k^\varepsilon = 1\}}, \ldots, I_{\{\overline{\alpha}_k^\varepsilon = l_0\}}) \in \mathbb{R}^{1\times l_0},\ \chi_k^\varepsilon = (I_{\{\alpha_k^\varepsilon = s_{ij}\}}) \in \mathbb{R}^{1\times m_0},\\
&\overline{F}(x) = \begin{pmatrix} f(x,1)\mathbb{1}_{m_1} \\ \ldots \\ f(x,l_0)\mathbb{1}_{m_{l_0}} \end{pmatrix} \in \mathbb{R}^{m_0\times 1},\ F(x) = \begin{pmatrix} f(x,1) \\ \ldots \\ f(x,l_0) \end{pmatrix} \in \mathbb{R}^{l_0 \times 1}.
\end{aligned} \tag{3.39}$$

Note that $(P-I)\bar{F}(x) = 0$. Next, we compute $\varepsilon\mathcal{L}^\varepsilon \overline{f}(z_k^\varepsilon, \alpha_k^\varepsilon)$. By (3.37),

$$\begin{aligned}
\varepsilon\mathcal{L}^\varepsilon \overline{f}(z_k^\varepsilon, \alpha_k^\varepsilon) &= E_k^\varepsilon \bar{f}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \overline{f}(z_k^\varepsilon, \alpha_k^\varepsilon)\\
&= E_k^\varepsilon[\overline{f}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \bar{f}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon)] + E_k^\varepsilon[\overline{f}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon) - \bar{f}(z_k^\varepsilon, \alpha_k^\varepsilon)].
\end{aligned} \tag{3.40}$$

By using the Taylor expansion, the second term in the above can be written as

$$E_k^\varepsilon[\overline{f}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon) - \bar{f}(z_k^\varepsilon, \alpha_k^\varepsilon)] = \sqrt{\varepsilon}\bar{f}_z(z_k^\varepsilon, \alpha_k^\varepsilon)X(k, \overline{\alpha}_k^\varepsilon) + \frac{\varepsilon}{2}X'(k, \overline{\alpha}_k^\varepsilon)\bar{f}_{zz}(z_k^\varepsilon, \alpha_k^\varepsilon)X(k, \overline{\alpha}_k^\varepsilon) + e_k^{\varepsilon,1} \tag{3.41}$$

where $\sup_{0<k\le T/\varepsilon} E|e_k^{\varepsilon,1}| = o(\varepsilon)$. In order to estimate the first term in the last equation in (3.40), we have

$$\begin{aligned}
&E_k^\varepsilon(\overline{f}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \overline{f}(z_k^\varepsilon, \alpha_k^\varepsilon))\\
&\quad= \sum_{i_1=1}^{l_0}\sum_{j_1=1}^{m_{i_1}} E_k^\varepsilon\Big[\sum_{i_2=1}^{l_0}\sum_{j_2=1}^{m_{i_2}} \overline{f}(z_{k+1}^\varepsilon, s_{i_2 j_2})P(\alpha_{k+1}^\varepsilon = s_{i_2 j_2}|\alpha_k^\varepsilon = s_{i_1 j_1}) - \overline{f}(z_{k+1}^\varepsilon, s_{i_1 j_1})\Big]\\
&\qquad\qquad \times I(\alpha_k^\varepsilon = s_{i_1 j_1})\\
&\quad= \chi_k^\varepsilon(P_\varepsilon - I)E_k^\varepsilon\overline{F}(z_{k+1}^\varepsilon) = \chi_k^\varepsilon(P - I + \varepsilon Q)E_k^\varepsilon\overline{F}(z_{k+1}^\varepsilon)\\
&\quad= \varepsilon\chi_k^\varepsilon Q E_k^\varepsilon\overline{F}(z_{k+1}^\varepsilon) = \varepsilon\chi_k^\varepsilon Q E_k^\varepsilon[\overline{F}(z_k^\varepsilon) + O(\sqrt{\varepsilon})]\\
&\quad= \varepsilon\chi_k^\varepsilon Q\overline{F}(z_k^\varepsilon) + e_k^{\varepsilon,2} = \varepsilon Q\overline{f}(z_k^\varepsilon, \cdot)(\alpha_k^\varepsilon) + e_k^{\varepsilon,2},
\end{aligned} \tag{3.42}$$

where $\sup_{0<k\le T/\varepsilon} E|e_k^{\varepsilon,2}| = o(\varepsilon)$. The combination of (3.40), (3.41), and (3.42) yields

$$\begin{aligned}\varepsilon\mathcal{L}^\varepsilon\overline{f}(z_k^\varepsilon,\alpha_k^\varepsilon) = \varepsilon Q\bar{f}(z_k^\varepsilon,\cdot)(\alpha_k^\varepsilon) + \frac{\varepsilon}{2}X'(k,\overline{\alpha}_k^\varepsilon)\overline{f}_{zz}(z_k^\varepsilon,\alpha_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon) \\ + \sqrt{\varepsilon}\bar{f}_z(z_k^\varepsilon,\alpha_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon) + e_k^{\varepsilon,1} + e_k^{\varepsilon,2}.\end{aligned} \tag{3.43}$$

Denote $\widetilde{E}_k(\cdot) = E(\cdot|\mathcal{F}_k^{\alpha^\varepsilon})$ and put

$$\begin{aligned}
f_1^\varepsilon(z,i,\varepsilon k) &= \sqrt{\varepsilon}f_z(z,i)\Big(X(k,i) + \sum_{l=k+1}^{T/\varepsilon} E_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)\Big),\\
f_2^\varepsilon(z,i,\varepsilon k) &= \varepsilon\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\Big[E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z,i)X(p,\overline{\alpha}_p^\varepsilon) - \widetilde{E}_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z,i)X(p,\overline{\alpha}_p^\varepsilon)\Big],\\
f_3^\varepsilon(z,i,\varepsilon k) &= \frac{\varepsilon}{2}\sum_{l=k}^{T/\varepsilon}\Big[E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z,i)X(l,\overline{\alpha}_l^\varepsilon) - \widetilde{E}_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z,i)X(l,\overline{\alpha}_l^\varepsilon)\Big],\\
f_4^\varepsilon(z,i,\varepsilon k) &= \varepsilon\sum_{l=k}^{T/\varepsilon} E_k^\varepsilon(\chi_l^\varepsilon - \bar{\chi}_l^\varepsilon\hat{\nu})Q\widetilde{\mathbb{1}}F(z_l^\varepsilon).
\end{aligned} \tag{3.44}$$

Then we have the following proposition, whose proof is long and technical. In order not to interrupt the flow of presentation, the proof is relegated in an appendix and placed at the end of the paper. We will use $\mathrm{tr}(A)$ to denote the trace of A.

Proposition 3.11. *For $f_i^\varepsilon(\cdot,\cdot,\cdot)$, $i=1,2,3,4$ defined above, we have*

$$\sup_{0\le k\le T/\varepsilon} E|f_i^\varepsilon(z_k^\varepsilon,\alpha_k^\varepsilon,\varepsilon k)| \to 0 \quad \textit{as} \quad \varepsilon\to 0 \quad \textit{and} \tag{3.45}$$

$$\begin{aligned}
\varepsilon\mathcal{L}^\varepsilon f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) &= \varepsilon\sum_{l=k+1}^{T/\varepsilon} E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon) f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon) - \sqrt{\varepsilon} f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon) + e_k^{\varepsilon,3},\\
\varepsilon\mathcal{L}^\varepsilon f_2^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) &= -\varepsilon\sum_{l=k+1}^{T/\varepsilon} \mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)E_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big]\\
&\qquad + \varepsilon\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\sum_{l=k+1}^{T/\varepsilon}\widetilde{E}_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big] + e_k^{\varepsilon,4},\\
\varepsilon\mathcal{L}^\varepsilon f_3^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) &= -\frac{\varepsilon}{2}X'(k,\overline{\alpha}_k^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)\\
&\qquad + \frac{\varepsilon}{2}\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\widetilde{E}_k^\varepsilon X(k,\overline{\alpha}_k^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big] + e_k^{\varepsilon,5},\\
\varepsilon\mathcal{L}^\varepsilon f_4^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) &= -\varepsilon(\chi_k^\varepsilon - \overline{\chi}_k^\varepsilon\hat{\nu})Q\bar{F}(z_k^\varepsilon) + e_k^{\varepsilon,6},
\end{aligned} \tag{3.46}$$

where

$$\sup_{0\le k\le T/\varepsilon} E|e_k^{\varepsilon,i}| = o(\varepsilon) \quad \textit{as} \quad \varepsilon\to 0 \quad \textit{for} \quad i=3,4,5,6. \tag{3.47}$$

To proceed, define

$$f^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) = \bar{f}(z_k^\varepsilon,\alpha_k^\varepsilon) + \sum_{i=1}^{4} f_i^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k). \tag{3.48}$$

Then (3.45) gives

$$E|f^\varepsilon(z_k^\varepsilon,\alpha_k^\varepsilon,\varepsilon k) - f(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)| = E|f^\varepsilon(z_k^\varepsilon,\alpha_k^\varepsilon,\varepsilon k) - \bar{f}(z_k^\varepsilon,\alpha_k^\varepsilon)| \to 0 \tag{3.49}$$

as $\varepsilon\to 0$. In addition, according to (3.43) and (3.46), we obtain

$$\begin{aligned}
\mathcal{L}^\varepsilon f^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) &= \mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\sum_{l=k+1}^{T/\varepsilon}\widetilde{E}_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big]\\
&\qquad + \frac{1}{2}\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\widetilde{E}_k^\varepsilon X(k,\overline{\alpha}_k^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big] + \bar{Q}f(z_k^\varepsilon,\cdot)(\overline{\alpha}_k^\varepsilon) + \varepsilon^{-1}e_k^\varepsilon,
\end{aligned} \tag{3.50}$$

where

$$e_k^\varepsilon = \sum_{i=1}^{6} e_k^i \quad \text{and} \quad \sup_{0\le k\le T/\varepsilon} E|e_k^\varepsilon| = o(\varepsilon) \text{ as } \varepsilon\to 0. \tag{3.51}$$

Next, we show that

$$\lim_{\varepsilon\to 0} E|\mathcal{L}^\varepsilon f^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k) - \mathcal{L} f(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon)| = 0. \tag{3.52}$$

We have

$$\begin{aligned} &\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) \sum_{l=k+1}^{T/\varepsilon} \widetilde{E}_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big] \\ &= \sum_{j,j_0=1}^{d} \sum_{i_1,i_2=1}^{l_0} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \sum_{l=k+1}^{T/\varepsilon} \frac{\partial^2 f(z_k^\varepsilon, i_1)}{\partial z_j \partial z_{j_0}} P(\alpha_l^\varepsilon = s_{i_2 j_2} | \alpha_k^\varepsilon = s_{i_1 j_1}) \\ &\qquad\qquad \times I(\alpha_k^\varepsilon = s_{i_1 j_1}) E X_{j_0}(l, i_2) X_j(k, i_1). \end{aligned} \tag{3.53}$$

Note that $|EX_{j_0}(l,i_2)X_j(k,i_1)| \le C\phi(l-k)^{\frac{1+2\delta}{2(1+\delta)}} \le C\phi(l-k)^{\frac{\delta}{1+\delta}}$ by (3.20), so the boundedness of $f_{zz}(\cdot)$ and (3.3) implies that

$$\sum_{j,j_0=1}^{d} \sum_{i_1,i_2=1}^{l_0} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \sum_{l=k+1}^{\infty} \Big| \frac{\partial^2 f(z_k^\varepsilon, i_1)}{\partial z_j \partial z_{j_0}} E X_{j_0}(l, i_2) X_j(k, i_1) \Big| < \infty.$$

Taking this into account and recall from (2.7) that $\lim_{\varepsilon\to 0} P(\alpha_l^\varepsilon = s_{i_2 j_2} | \alpha_k^\varepsilon = s_{i_1 j_1}) = 0$ for $1 \le i_1 \ne i_2 \le l_0$, $k < l$, it follows that

$$\begin{aligned} \lim_{\varepsilon\to 0} E\Big| \sum_{j,j_0=1}^{d} \sum_{1\le i_1 \ne i_2 \le l_0} \sum_{j_1=1}^{m_{i_1}} \sum_{j_2=1}^{m_{i_2}} \sum_{l=k+1}^{T/\varepsilon} \frac{\partial^2 f(z_k^\varepsilon, i_1)}{\partial z_j \partial z_{j_0}} P(\alpha_l^\varepsilon = s_{i_2 j_2} | \alpha_k^\varepsilon = s_{i_1 j_1}) I(\alpha_k^\varepsilon = s_{i_1 j_1}) \\ \times E X_{j_0}(l, i_2) X_j(k, i_1) \Big| = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} &\lim_{\varepsilon\to 0} E\Big| \mathrm{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) \sum_{l=k+1}^{T/\varepsilon} \widetilde{E}_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big] \\ &\quad - \sum_{j,j_0=1}^{d} \sum_{i=1}^{l_0} \sum_{j_1,j_2=1}^{m_i} \sum_{l=k+1}^{T/\varepsilon} \frac{\partial^2 f(z_k^\varepsilon, i)}{\partial z_j \partial z_{j_0}} P(\alpha_l^\varepsilon = s_{i j_2} | \alpha_k^\varepsilon = s_{i j_1}) \\ &\qquad\qquad \times I(\alpha_k^\varepsilon = s_{i j_1}) E X_{j_0}(l, i) X_j(k, i) \Big| = 0. \end{aligned} \tag{3.54}$$

Again, because of (2.7), $\lim_{\varepsilon\to 0}\sum_{j_2=1}^{m_i} P(\alpha_l^\varepsilon = s_{ij_2}|\alpha_k^\varepsilon = s_{ij_1}) = 1$. Thus,

$$\lim_{\varepsilon\to 0} E\Big|\sum_{j,j_0=1}^{d}\sum_{i=1}^{l_0}\sum_{j_1,j_2=1}^{m_i}\sum_{l=k+1}^{T/\varepsilon}\frac{\partial^2 f(z_k^\varepsilon,i)}{\partial z_j\partial z_{j_0}}P(\alpha_l^\varepsilon = s_{ij_2}|\alpha_k^\varepsilon = s_{ij_1})I(\alpha_k^\varepsilon = s_{ij_1})EX_{j_0}(l,i)X_j(k,i) - \sum_{j,j_0=1}^{d}\sum_{i=1}^{l_0}\sum_{l=k+1}^{\infty}\frac{\partial^2 f(z_k^\varepsilon,i)}{\partial z_j\partial z_{j_0}}I(\overline{\alpha}_k^\varepsilon = i)EX_{j_0}(l,i)X_j(k,i)\Big| = 0. \tag{3.55}$$

By (3.54), (3.55), and the stationarity,

$$\begin{aligned}&\lim_{\varepsilon\to 0} E\Big|\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\sum_{l=k+1}^{T/\varepsilon}\widetilde{E}_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big]\\ &\qquad - \sum_{j,j_0=1}^{d}\sum_{i=1}^{l_0}\sum_{l=1}^{\infty}\frac{\partial^2 f(z_k^\varepsilon,i)}{\partial z_j\partial z_{j_0}}I(\overline{\alpha}_k^\varepsilon = i)EX_{j_0}(l,i)X_j(0,i)\Big|\\ &= \lim_{\varepsilon\to 0} E\Big|\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\sum_{l=k+1}^{T/\varepsilon}\widetilde{E}_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big]\\ &\qquad - \sum_{i=1}^{l_0}\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,i)\sum_{l=1}^{\infty}EX(l,i)X'(0,i)\Big]I(\overline{\alpha}_k^\varepsilon = i)\Big| = 0.\end{aligned} \tag{3.56}$$

Similarly,

$$\lim_{\varepsilon\to 0} E\Big|\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\widetilde{E}_k^\varepsilon X(k,\overline{\alpha}_k^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big] - \sum_{i=1}^{l_0}\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,i)EX(0,i)X'(0,i)\Big]I(\overline{\alpha}_k^\varepsilon = i)\Big| = 0. \tag{3.57}$$

Therefore, (3.52) follows from (3.6), (3.7), (3.50), (3.55), (3.57), and the fact $EX(l,i)X'(0,i) = EX(0,i)X'(l,i)$.

Next, by the mixing inequality (3.3) and the moment condition (3.2), the same argument as above yields

$$\sup_{0\le k\le T/\varepsilon,\varepsilon} E|\mathcal{L}^\varepsilon f^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)| < \infty. \tag{3.58}$$

Hence, by (3.49), (3.52), and (3.58), conditions $(c)(1)$, $(c)(2')$, $(c)(3)$, and $(c)(4')$ in Proposition 3.2 are satisfied. On the other hand, by virtue of Propositions 3.4 and

3.9, the conditions (a) and (b) are fulfilled. So the proof of the theorem follows by Proposition 3.2 and the Remark 3.3. $\square$

Remark 3.12. As given in Remark 2.2 (iii), $\overline{\alpha}^{\varepsilon}(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$, a Markov chain generated by $\overline{Q}$. Define a stochastic process

$$\bar{z}(t) = \int_0^t \sigma(\overline{\alpha}(s))dw(s), \tag{3.59}$$

where $w(\cdot)$ is a standard Brownian motion and $\sigma(i)\sigma'(i) = A(i)$ with $A(\cdot)$ is given in (3.7). Then for each $i = 1\ldots, l_0$, for any $f(\cdot, i)$ that is a real-valued function with bounded derivatives up to second order and with Lipschitz continuous second derivatives, $f(\overline{z}(t), \overline{\alpha}(t)) - \int_0^t \mathcal{L}f(\bar{z}(s), \overline{\alpha}(s))ds$ is a martingale. Therefore, $(\overline{z}(\cdot), \overline{\alpha}(\cdot))$ is a solution of the martingale problem associated with operator $\mathcal{L}$. In view of Proposition 3.9, the uniqueness of the martingale problem with operator $\mathcal{L}$ implies that $(\overline{z}(\cdot), \overline{\alpha}(\cdot))$ has same distribution as that of $(z(\cdot), \overline{\alpha}(\cdot))$.

3.3 Ramifications

In this section, we obtain further results and ramifications as a consequence of the previous sections. These results are in the light of reduction of computational complexity. It indicates that we can aggregate the Markovian states in an appropriate way so that the aggregated process is much easier to deal with. For the original sequence, we have to deal with $|\mathcal{M}|$ sequences $\{X(l, s_{ij})\}$, whereas in the aggregated process, we need only examine $|\overline{\mathcal{M}}|$ sequences.

Label the state space $\mathcal{M}$ as $\mathcal{M} = \{s_{11}, \ldots, s_{1m_1}\} \cup \cdots \cup \{s_{l_0 1}, \ldots, s_{l_0 m_{l_0}}\}$. Let

$\{(X(k,\alpha), \alpha \in \mathcal{M}) : k \in \mathbb{Z}\}$ be a wide-sense stationary sequence of $\mathbb{R}^{|\mathcal{M}|\times d}$-valued random variables on $(\Omega, \mathcal{F}, P)$. Thus, compared to Section 2, we have a total of $|\mathcal{M}|$ sequences to deal with. We replace (A2) by the following condition.

(A2') The sequence $\{(X(k,\alpha), \alpha \in \mathcal{M}) : k \in \mathbb{Z}\}$ is independent of the Markov process $\{\alpha_k^\varepsilon\}$, and is ϕ-mixing with mixing measure denoted by $\phi(\cdot)$. Moreover, assume that there exists $\delta > 0$ and a constant C that does not depend on k and α such that

$$EX(k,\alpha) = 0, \quad E|X(k,\alpha)|^{2(1+\delta)} \le C, \quad \forall\, k \ge 1;\ \alpha \in \mathcal{M}, \tag{3.60}$$

and (3.3) holds.

Aggregating the Markov states in each $\mathcal{M}_i$ into one state leads to the definition of the following centered and scaled sequences associated with the aggregated Markov states:

$$\widehat{z}_k^\varepsilon = \sqrt{\varepsilon} \sum_{l=0}^{k-1} \sum_{i=1}^{l_0} \sum_{j=1}^{m_i} X(l, s_{ij})[I(\alpha_l^\varepsilon = s_{ij}) - \nu^{ij} I(\overline{\alpha}_l^\varepsilon = i)], \quad \widehat{z}^\varepsilon(t) = \widehat{z}_k^\varepsilon,\ t \in [\varepsilon k, \varepsilon k + \varepsilon). \tag{3.61}$$

Using the techniques presented in the last section, we can establish the following results. The detailed proof is omitted.

Theorem 3.13. *Assume* (A1) *and* (A2'). *The process* $(\widehat{z}^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ *converges weakly to* $(\widehat{z}(\cdot), \overline{\alpha}(\cdot))$ *such that the limit is the solution of the martingale problem with operator given by*

$$\mathcal{L} f(x,i) = \frac{1}{2} \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \widehat{a}^{j_1 j_2}(i) \frac{\partial^2 f(x,i)}{\partial x^{j_1} \partial x^{j_2}} + \overline{Q} f(x,\cdot)(i), \quad i = 1, 2, \ldots, l_0, \tag{3.62}$$

where $\widehat{A}(i) = (\widehat{a}^{j_1 j_2}(i))$ *and*

$$\begin{aligned}\widehat{A}(i) = \sum_{j=1}^{m_i} \Big\{ & \nu^{ij}(1-\nu^{ij}) EX(0,s_{ij})X'(0,s_{ij}) \\ & + \sum_{k=1}^{\infty} \nu^{ij}\psi^{ij,ij}(k)\big[EX(0,s_{ij})X'(k,s_{ij}) + EX(k,s_{ij})X'(0,s_{ij})\big]\Big\} \\ + \sum_{1\le j_1<j_2\le m_i} & \Big\{ \sum_{k=1}^{\infty}\Big[\nu^{ij_1}\psi^{ij_1,ij_2}(k)\big[EX(0,s_{ij_1})X'(k,s_{ij_2}) + EX(k,s_{ij_2})X'(0,s_{ij_1})\big] \\ & + \nu^{ij_2}\psi^{ij_2,ij_1}(k)\big[EX(0,s_{ij_2})X'(k,s_{ij_1}) + EX(k,s_{ij_1})X'(0,s_{ij_2})\big]\Big] \\ & - \nu^{ij_1}\nu^{ij_2}\big[EX(0,s_{ij_1})X'(0,s_{ij_2}) + EX(0,s_{ij_2})X'(0,s_{ij_1})\big]\Big\}.\end{aligned} \tag{3.63}$$

Remark 3.14. As a special case, we consider a Markov chain α_k^ε with transition probability matrix given by $P_\varepsilon = P + \varepsilon Q$, where P is irreducible and Q is a generator of a continuous-time Markov chain. That is, the states of the Markov chain belong to one weakly irreducible class. Assume that $\mathcal{M} = \{1, 2, \dots, m\}$ and for each $k \ge 0$, $i \in \mathcal{M}$, $X(k,i) \in \mathbb{R}^d$ and is wide-sense stationary mixing. This is a consequence of the main result. Define

$$\widehat{Z}_k^\varepsilon = \sqrt{\varepsilon}\sum_{l=0}^{k-1}\sum_{i=1}^{m} X(l,i)[I(\alpha_l^\varepsilon = i) - \nu_i], \quad \widehat{Z}^\varepsilon(t) = \widehat{Z}_k^\varepsilon, \ \ t \in [\varepsilon k, \varepsilon k + \varepsilon), \tag{3.64}$$

where $\nu = (\nu_1, \dots, \nu_m)$ is the stationary distribution associated with the transition matrix P. Then it can be shown that under conditions (A1) and (A2') with the modification mentioned above, $\widehat{Z}^\varepsilon(\cdot)$ converges weakly to $\widehat{Z}(\cdot)$, a d-dimensional Brownian motion with mean zero and covariance Σt where

$$\begin{aligned}\Sigma = \sum_{i=1}^{m} \Big[\nu_i(1-\nu_i)EX(0,i)X'(0,i) + \sum_{k=1}^{\infty} \nu_i \psi^{ii}(k)\big[EX(0,i)X'(k,i) + EX(k,i)X'(0,i)\big]\Big] \\ + \sum_{1\le i<j\le m} \Bigg\{ \sum_{k=1}^{\infty} \Big[\nu_i \psi^{ij}(k)\big[EX(0,i)X'(k,j) + EX(k,j)X'(0,i)\big] \\ + \nu_j \psi^{ji}(k)\big[EX(0,j)X'(k,i) + EX(k,i)X'(0,j)\big]\Big] \\ - \nu_i\nu_j\big[EX(0,i)X'(0,j) + EX(0,j)X'(0,i)\big]\Bigg\}. \end{aligned} \tag{3.65}$$

In addition to the process $(\widehat{z}^\varepsilon(t), \overline{\alpha}^\varepsilon(t))$, we may define

$$\begin{aligned}\widetilde{z}_k^\varepsilon &= \sqrt{\varepsilon}\sum_{l=0}^{k-1}\sum_{i=1}^{l_0}\sum_{j=1}^{m_i} X(l,s_{ij})I(\alpha_l^\varepsilon = s_{ij}) = \sqrt{\varepsilon}\sum_{l=0}^{k-1} X(l,\alpha_l^\varepsilon),\ \widetilde{z}^\varepsilon(t) = \widetilde{z}_k^\varepsilon, \quad t\in[\varepsilon k, \varepsilon k+\varepsilon), \\ \overline{z}_k^\varepsilon &= \sqrt{\varepsilon}\sum_{l=0}^{k-1}\sum_{i=1}^{l_0}\sum_{j=1}^{m_i} X(k,s_{ij})\nu^{ij} I(\overline{\alpha}_l^\varepsilon = i),\ \overline{z}^\varepsilon(t) = \overline{z}_k^\varepsilon, \quad t\in[\varepsilon k, \varepsilon k+\varepsilon), \end{aligned} \tag{3.66}$$

where $\lfloor t/\varepsilon \rfloor$ denotes the integer part of the real number t/ε.

Remark 3.15. Under the conditions of Theorem 3.13, we establish the following results.

(i) $(\widetilde{z}^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(\widetilde{z}(\cdot), \overline{\alpha}(\cdot))$ such that the limit is the solution of the martingale problem with operator given by

$$\mathcal{L}f(x,i) = \frac{1}{2}\sum_{j_1=1}^{d}\sum_{j_2=1}^{d} \widetilde{a}^{j_1 j_2}(i)\frac{\partial^2 f(x,i)}{\partial x^{j_1}\partial x^{j_2}} + \overline{Q}f(x,\cdot)(i), \quad i = 1,2,\ldots,l_0 \tag{3.67}$$

where

$$
\begin{aligned}
\widetilde{A}(i) &= \left(\widetilde{a}^{j_1 j_2}(i)\right) \\
&= \sum_{j=1}^{m_i} \nu^{ij} EX(0,s_{ij})X'(0,s_{ij}) + \sum_{j_1,j_2=1}^{m_i} \sum_{k=1}^{\infty} \nu^{ij_1}\left(\nu^{ij_2} + \psi^{ij_1,ij_2}(k)\right) \\
&\qquad \times \Big[EX(0,s_{ij_1})X'(k,s_{ij_2}) + EX(k,s_{ij_2})X'(0,s_{ij_1})\Big].
\end{aligned}
\tag{3.68}
$$

(ii) $(\overline{z}^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(\overline{z}(\cdot), \overline{\alpha}(\cdot))$ such that the limit is the solution of the martingale problem with operator given by

$$
\mathcal{L}f(x,i) = \frac{1}{2}\sum_{j_1=1}^{d}\sum_{j_2=1}^{d} \overline{a}^{j_1 j_2}(i)\frac{\partial^2 f(x,i)}{\partial x^{j_1}\partial x^{j_2}} + \overline{Q}f(x,\cdot)(i), \quad i = 1,2,\ldots,l_0 \tag{3.69}
$$

where

$$
\overline{A}(i) = (\overline{a}^{j_1 j_2}(i)) = E\overline{X}(0,i)\overline{X}'(0,i) + \sum_{k=1}^{\infty}\Big[E\overline{X}(k,i)\overline{X}'(0,i) + E\overline{X}(0,i)\bar{X}'(k,i)\Big], \tag{3.70}
$$

$\overline{X}(k,i) = \sum_{j=1}^{m_i} X(k,s_{ij})\nu^{ij}$ and the matrix $\overline{Q}$ is given in (2.3).

These results illustrate the aggregation and associated limit results from a slightly different angle.

Example 3.16. a, Let $\varepsilon > 0$ and α_k^ε be a time-homogeneous Markov chain with the state space $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 = \{s_{11}, s_{12}\} \cup \{s_{21}, s_{22}\}$ and transition matrix $P_\varepsilon = P + \varepsilon Q$ with

$$
P = \mathrm{diag}[P^1, P^2] = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{7} & \frac{6}{7} \\ 0 & 0 & \frac{2}{7} & \frac{5}{7} \end{pmatrix}, \quad Q = \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix}.
$$

Then $\nu^1 = (\frac{1}{3}, \frac{2}{3})$, $\nu^2 = (\frac{1}{4}, \frac{3}{4})$ and

$$(P^1)^k = \begin{pmatrix} \frac{1}{3} + \frac{2}{3\cdot 4^k} & \frac{2}{3} - \frac{2}{3\cdot 4^k} \\ \frac{1}{3} - \frac{1}{3\cdot 4^k} & \frac{2}{3} + \frac{1}{3\cdot 4^k} \end{pmatrix}, \quad (P^2)^k = \begin{pmatrix} \frac{1}{4} + \frac{3}{4\cdot(-7)^k} & \frac{3}{4} - \frac{3}{4\cdot(-7)^k} \\ \frac{1}{4} - \frac{1}{4\cdot(-7)^k} & \frac{3}{4} + \frac{1}{4\cdot(-7)^k} \end{pmatrix}.$$

By (2.3) and (2.6),

$$\overline{Q} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}, \quad \Psi(k) = \begin{pmatrix} \frac{2}{3\cdot 4^k} & \frac{-2}{3\cdot 4^k} & 0 & 0 \\ \frac{-1}{3\cdot 4^k} & \frac{1}{3\cdot 4^k} & 0 & 0 \\ 0 & 0 & \frac{3}{4\cdot(-7)^k} & \frac{-3}{4\cdot(-7)^k} \\ 0 & 0 & \frac{-1}{4\cdot(-7)^k} & \frac{1}{4\cdot(-7)^k} \end{pmatrix}.$$

Let $\{(X(k,1), X(k,2)) : k \geq 0\}$ and $\{(X(k, s_{11}), X(k, s_{12}), X(k, s_{21}), X(k, s_{22})) : k \geq 0\}$ be two sequences of m-dependent, wide-sense stationary random variables in $\mathbb{R}^d$ satisfying the conditions (A2) and (A2') respectively. Denote

$$z^\varepsilon(t) = \sqrt{\varepsilon} \sum_{j=0}^{\lfloor t/\varepsilon \rfloor - 1} X(j, \bar{\alpha}_j^\varepsilon), \quad \hat{z}^\varepsilon(t) = \sqrt{\varepsilon} \sum_{l=0}^{\lfloor t/\varepsilon \rfloor - 1} \sum_{i=1}^{2} \sum_{j=1}^{2} X(l, s_{ij})[I(\alpha_l^\varepsilon = s_{ij}) - \nu^{ij} I(\bar{\alpha}_l^\varepsilon = i)].$$

Then, by Theorems 3.10 and 3.13, $(z^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ and $(\hat{z}^\varepsilon(\cdot), \bar{\alpha}^\varepsilon(\cdot))$ respectively converge weakly to $(z(\cdot), \bar{\alpha}(\cdot))$ and $(\hat{z}(\cdot), \bar{\alpha}(\cdot))$ such that the limits are the solutions of the martingale problems with operators respectively given by

$$\mathcal{L} f(x,i) = \frac{1}{2} \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} a^{j_1 j_2}(i) \frac{\partial^2 f(x,i)}{\partial x^{j_1} \partial x^{j_2}} + \overline{Q} f(x, \cdot)(i),$$
$$\widehat{\mathcal{L}} f(x,i) = \frac{1}{2} \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \widehat{a}^{j_1 j_2}(i) \frac{\partial^2 f(x,i)}{\partial x^{j_1} \partial x^{j_2}} + \overline{Q} f(x, \cdot)(i),$$

where $A(i) = (a^{j_1 j_2}(i)) = EX(0,i)X'(0,i) + \sum_{i=1}^{\infty} \left[EX(k,i)X'(0,i) + EX(0,i)X'(k,i) \right]$

and $\widehat{A}(i) = (\widehat{a}^{j_1 j_2}(i))$, $i = 1, 2$ with

$$\begin{aligned}
\widehat{A}(1) = \frac{2}{9} E\Big[& X(0, s_{11})X'(0, s_{11}) + X(0, s_{12})X'(0, s_{12}) \\
& -X(0, s_{11})X'(0, s_{12}) - X(0, s_{12})X'(0, s_{11})\Big] \\
+ \sum_{k=1}^{m} \frac{2}{9 \cdot 4^k} E\Big[& X(0, s_{11})X'(k, s_{11}) + X(k, s_{11})X'(0, s_{11}) \\
& +X(0, s_{12})X'(k, s_{12}) + X(k, s_{12})X'(0, s_{12}) - X(0, s_{11})X'(k, s_{12}) \\
& -X(k, s_{12})X'(0, s_{11}) - X(0, s_{12})X'(k, s_{11}) - X(k, s_{11})X'(0, s_{12})\Big], \\
\widehat{A}(2) = \frac{3}{16} E\Big[& X(0, s_{21})X'(0, s_{21}) + X(0, s_{22})X'(0, s_{22}) \\
& -X(0, s_{21})X'(0, s_{22}) - X(0, s_{22})X'(0, s_{21})\Big] \\
+ \sum_{k=1}^{m} \frac{3}{16 \cdot (-7)^k} E\Big[& X(0, s_{21})X'(k, s_{21}) + X(k, s_{21})X'(0, s_{21}) \\
& +X(0, s_{22})X'(k, s_{22}) + X(k, s_{22})X'(0, s_{22}) - X(0, s_{21})X'(k, s_{22}) \\
& -X(k, s_{22})X'(0, s_{21}) - X(0, s_{22})X'(k, s_{21}) - X(k, s_{21})X'(0, s_{22})\Big].
\end{aligned}$$

b, Let α_k^ε be a time-homogeneous Markov chain with the state space $\mathcal{M} = \{1, 2\}$ and transition matrix $P_\varepsilon = P + \varepsilon Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} + \varepsilon \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix}$. Then $\nu = (\frac{1}{3}, \frac{2}{3})$. By (2.6), $\Psi(k) = \begin{pmatrix} \frac{2}{3 \cdot 4^k} & \frac{-2}{3 \cdot 4^k} \\ \frac{-1}{3 \cdot 4^k} & \frac{1}{3 \cdot 4^k} \end{pmatrix}$. Let $\{(X(k, 1), X(k, 2)) : k \geq 0\}$ be a sequence of m-dependent, wide-sense stationary random variables in $\mathbb{R}^d$ satisfying conditions (A2). Denote $\widehat{Z}^\varepsilon(t) = \sqrt{\varepsilon} \sum_{l=0}^{\lfloor t/\varepsilon \rfloor - 1} \sum_{i=1}^{2} X(l, i)[I(\alpha_l^\varepsilon = i) - \nu_i]$. By Remark 4.2, $\widehat{Z}^\varepsilon(\cdot)$ converges weakly to $\widehat{Z}(\cdot)$, a d-dimensional Brownian motion with mean zero and covariance Σt where

$$\begin{aligned}
\Sigma = & \frac{2}{9} E\big[X(0, 1)X'(0, 1) + X(0, 2)X'(0, 2) - X(0, 1)X'(0, 2) - X(0, 2)X'(0, 1)\big] \\
& + \sum_{k=1}^{m} \frac{2}{9 \cdot 4^k} E\big[X(0, 1)X'(k, 1) + X(k, 1)X'(0, 1) + X(0, 2)X'(k, 2) + X(k, 2)X'(0, 2) \\
& \qquad -X(0, 1)X'(k, 2) - X(k, 2)X'(0, 1) - X(0, 2)X'(k, 1) - X(k, 1)X'(0, 2)\big].
\end{aligned}$$

3.4 Proof of Proposition 3.11

The proof is divided into several steps. Each step is formulated as a claim.

Step 1: $\sup_{0\le k\le T/\varepsilon} E|f_i^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)|\to 0$ for $i=1,\dots,4$.

(1) $\sup_{0\le k\le T/\varepsilon} E|f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)|\to 0$. By Cauchy-Schwartz inequality and the boundedness of the first derivative of f,

$$\begin{aligned} E|f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)| &\le \sqrt{\varepsilon}E\Big[|f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)|\Big|\sum_{l=k}^{T/\varepsilon}E_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)\Big|\Big] \\ &\le \sqrt{\varepsilon}CE\Big|\sum_{l=k}^{T/\varepsilon}E_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)\Big| \le \sqrt{\varepsilon}C\sum_{j=1}^{d}E\Big|\sum_{l=k}^{T/\varepsilon}E_k^\varepsilon X_j(l,\overline{\alpha}_l^\varepsilon)\Big| \\ &\le \sqrt{\varepsilon}C\sum_{j=1}^{d}\sum_{l=k}^{T/\varepsilon}E|E_k^\varepsilon X_j(l,\overline{\alpha}_l^\varepsilon)|. \end{aligned} \tag{3.71}$$

By the independence of $\{X_j(k,i)\}$ and $\{\alpha_k^\varepsilon\}$,

$$\begin{aligned} E|E_k^\varepsilon X_j(l,\overline{\alpha}_l^\varepsilon)| &= E\Big|\sum_{i=1}^{l_0}E_k^\varepsilon(X_j(l,i)I(\overline{\alpha}_l^\varepsilon=i))\Big| \\ &= E\Big|\sum_{i=1}^{l_0}E\Big(X_j(l,i)I(\overline{\alpha}_l^\varepsilon=i)\Big|\mathcal{F}_k^X\vee\mathcal{F}_k^{\alpha^\varepsilon}\Big)\Big| \\ &= E\Big|\sum_{i=1}^{l_0}E\big(X_j(l,i)\big|\mathcal{F}_k^X\big)E\big(I(\overline{\alpha}_l^\varepsilon=i)\big|\mathcal{F}_k^{\alpha^\varepsilon}\big)\Big| \\ &\le \sum_{i=1}^{l_0}E\big|E\big(X_j(l,i)\big|\mathcal{F}_k^X\big)\big|E\big|E\big(I(\overline{\alpha}_l^\varepsilon=i)\big|\mathcal{F}_k^{\alpha^\varepsilon}\big)\big| \\ &\le \sum_{i=1}^{l_0}E\big|E\big(X_j(l,i)\big|\mathcal{F}_k^X\big)\big|. \end{aligned} \tag{3.72}$$

Note that $EX_j(l,i)=0$, so by the inequality (2.17) with $p=1+\delta$ and $q=\frac{1+\delta}{\delta}$,

$$E\big|E\big(X_j(l,i)\big|\mathcal{F}_k^X\big)\big| \le 2\phi(l-k)^{\frac{\delta}{1+\delta}}\|X_j(l,i)\|_{1+\delta} \le C\phi(l-k)^{\frac{\delta}{1+\delta}}. \tag{3.73}$$

Next, from (3.71), (3.72), and (3.73), we have

$$E|f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)| \le \sqrt{\varepsilon}C\sum_{j=1}^{d}\sum_{l=k}^{T/\varepsilon}\sum_{i=1}^{l_0}\phi(l-k)^{\frac{\delta}{1+\delta}} \le \sqrt{\varepsilon}C\sum_{n=0}^{\infty}\phi(n)^{\frac{\delta}{1+\delta}} = C\sqrt{\varepsilon}. \tag{3.74}$$

The last identity follows from the assumptions (A2). Since the inequality (3.74) holds for all k with $0 \le k \le T/\varepsilon$, we get $\sup_{0\le k\le T/\varepsilon} E|f_1^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k)| \to 0$ as desired.

(2) $\sup_{0\le k\le T/\varepsilon} E|f_2^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k)| \to 0$ as $\varepsilon \to 0$.

For $j, j_0 = 1, \ldots, d$ we have

$$\begin{aligned}
&\Big|\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\Big[E_k^\varepsilon X_j(l,\overline{\alpha}_l^\varepsilon)X_{j_0}(p,\overline{\alpha}_p^\varepsilon) - \widetilde{E}_k^\varepsilon X_j(l,\overline{\alpha}_l^\varepsilon)X_{j_0}(p,\overline{\alpha}_p^\varepsilon)\Big]\Big| \\
&= \Big|\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}\Big[E\Big(X_j(l,i_1)X_{j_0}(p,i_2)I(\overline{\alpha}_l^\varepsilon=i_1)I(\overline{\alpha}_p^\varepsilon=i_2)\Big|\mathcal{F}_k^X\vee\mathcal{F}_k^{\alpha^\varepsilon}\Big) \\
&\qquad\qquad -E\Big(X_j(l,i_1)X_{j_0}(p,i_2)I(\overline{\alpha}_l^\varepsilon=i_1)I(\overline{\alpha}_p^\varepsilon=i_2)\Big|\mathcal{F}_k^{\alpha^\varepsilon}\Big)\Big]\Big| \\
&= \Big|\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}\Big[E\Big(X_j(l,i_1)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - EX_j(l,i_1)X_{j_0}(p,i_2)\Big] \\
&\qquad\qquad \times E\Big(I(\overline{\alpha}_l^\varepsilon=i_1)I(\overline{\alpha}_p^\varepsilon=i_2)\Big|\mathcal{F}_k^{\alpha^\varepsilon}\Big)\Big| \\
&\le \sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}\Big|E\Big(X_j(l,i_1)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - EX_j(l,i_1)X_{j_0}(p,i_2)\Big|.
\end{aligned} \tag{3.75}$$

Since $l > p \ge k$, by the mixing inequality (2.17) with $p = 1+\delta$ and $q = \frac{1+\delta}{\delta}$,

$$\begin{aligned}
&E\Big|E\Big(X_j(l,i_1)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - EX_j(l,i_1)X_{j_0}(p,i_2)\Big| \\
&= E\Big|E\Big(E(X_j(l,i_1)|\mathcal{F}_p^X)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - E\Big(E(X_j(l,i_1)|\mathcal{F}_p^X)X_{j_0}(p,i_2)\Big)\Big| \\
&\le 2\phi(p-k)^{\frac{\delta}{1+\delta}}\|E(X_j(l,i_1)|\mathcal{F}_p^X)X_{j_0}(p,i_2)\|_{1+\delta}.
\end{aligned} \tag{3.76}$$

Next, by the mixing inequality (2.16) with $p = 2(1+\delta)$, $q = \frac{2(1+\delta)}{1+2\delta}$,

$$\begin{aligned}
\|E(X_j(l,i_1)|\mathcal{F}_p^X) - EX_j(l,i_1)\|_{2(1+\delta)} &\le 2\phi(l-p)^{\frac{1+2\delta}{2(1+\delta)}}\|X_j(l,i_1)\|_{2(1+\delta)} \\
&\le 2\phi(l-p)^{\frac{\delta}{1+\delta}}\|X_j(l,i_1)\|_{2(1+\delta)}.
\end{aligned} \tag{3.77}$$

We have used the fact $\frac{\delta}{1+\delta} < \frac{1+2\delta}{2(1+\delta)}$ in the last inequality. Note that $EX_j(l,i_1) = 0$, so by (3.76), Cauchy-Schwartz inequality and (3.77) we obtain

$$\begin{aligned} &E\Big|E\Big(X_j(l,i_1)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - EX_j(l,i_1)X_{j_0}(p,i_2)\Big| \\ &\le 2\phi(p-k)^{\frac{\delta}{1+\delta}}\|E(X_j(l,i_1)|\mathcal{F}_p^X)\|_{2(1+\delta)}\|X_{j_0}(p,i_2)\|_{2(1+\delta)} \\ &\le 4\phi(p-k)^{\frac{\delta}{1+\delta}}\phi(l-p)^{\frac{\delta}{1+\delta}}\|X_j(l,i_1)\|_{2(1+\delta)}\|X_{j_0}(p,i_2)\|_{2(1+\delta)} \\ &\le C\phi(p-k)^{\frac{\delta}{1+\delta}}\phi(l-p)^{\frac{\delta}{1+\delta}}. \end{aligned} \tag{3.78}$$

The constant C in the last inequality does not depend on l, p, i_1, i_2, j, j_0 because of (3.2). Since $\sum_{k=0}^{\infty}\phi(k)^{\frac{\delta}{1+\delta}} < \infty$, it follows from (3.75) and (3.78) that

$$\begin{aligned} &E\Big|\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\Big[E_k^{\varepsilon}X_j(l,\overline{\alpha}_l^{\varepsilon})X_{j_0}(p,\overline{\alpha}_p^{\varepsilon}) - \widetilde{E}_k^{\varepsilon}X_j(l,\overline{\alpha}_l^{\varepsilon})X_{j_0}(p,\overline{\alpha}_p^{\varepsilon})\Big]\Big| \\ &\le \sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}E\Big|E\Big(X_j(l,i_1)X_{j_0}(p,i_2)\Big|\mathcal{F}_k^X\Big) - EX_j(l,i_1)X_{j_0}(p,i_2)\Big| \\ &\le C\sum_{p=k}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}\phi(p-k)^{\frac{\delta}{1+\delta}}\phi(l-p)^{\frac{\delta}{1+\delta}} \\ &\le Cl_0^2\Big(\sum_{l=0}^{\infty}\phi(l)^{\frac{\delta}{1+\delta}}\Big)\Big(\sum_{p=1}^{\infty}\phi(p)^{\frac{\delta}{1+\delta}}\Big) \le C. \end{aligned} \tag{3.79}$$

By the boundedness of $f_{zz}(\cdot,\cdot)$, (3.79) implies

$$\sup_{0\le k\le T/\varepsilon} E|f_2^{\varepsilon}(z_k^{\varepsilon},\overline{\alpha}_k^{\varepsilon},\varepsilon k)| = O(\varepsilon) \to 0 \text{ as } \varepsilon \to 0.$$

(3) $\sup_{0\le k\le T/\varepsilon} E|f_3^{\varepsilon}(z_k^{\varepsilon},\overline{\alpha}_k^{\varepsilon},\varepsilon k)| \to 0$ as $\varepsilon \to 0$. This can be done by using the argument of Step 1 (2).

(4) $\sup_{0\le k\le T/\varepsilon} E|f_4^{\varepsilon}(z_k^{\varepsilon},\overline{\alpha}_k^{\varepsilon},\varepsilon k)| \to 0$ as $\varepsilon \to 0$. The assertion is directly implied by the boundedness of f and the virtue of (2.8).

Step 2: Claim: (3.46) and (3.47) hold. This step is divided into four sub-steps as follow.

(1) Claim:

$$\varepsilon\mathcal{L}^\varepsilon f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)=\varepsilon\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)-\sqrt{\varepsilon}f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)+e_k^{\varepsilon,3} \tag{3.80}$$

where $e_k^{\varepsilon,3}$ satisfies (3.47) with $i=3$. We have

$$\begin{aligned}\varepsilon\mathcal{L}^\varepsilon f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)&=E_k^\varepsilon f_1^\varepsilon(z_{k+1}^\varepsilon,\overline{\alpha}_{k+1}^\varepsilon,\varepsilon k+\varepsilon)-f_1^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k)\\&=\sqrt{\varepsilon}\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_{k+1}^\varepsilon)-f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\\&\quad+\sqrt{\varepsilon}\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)-f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\\&\quad-\sqrt{\varepsilon}f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon).\end{aligned} \tag{3.81}$$

To proceed, we evaluate first two terms in the last equation of (3.81). First, since z_{k+1}^ε is $\mathcal{F}_k^\varepsilon$-measurable, we have

$$\begin{aligned}&E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_{k+1}^\varepsilon)-f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\\&=E_k^\varepsilon\Big[\Big(\overline{f}_z(z_{k+1}^\varepsilon,\alpha_{k+1}^\varepsilon)-\overline{f}_z(z_{k+1}^\varepsilon,\alpha_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\\&=\sum_{i_3=1}^{l_0}\sum_{j_3=1}^{m_{i_3}}\sum_{i_2=1}^{l_0}\sum_{j_2=1}^{m_{i_2}}\sum_{i_1=1}^{l_0}\sum_{j_1=1}^{m_{i_1}}E_k^\varepsilon\Big[I(\alpha_k^\varepsilon=s_{i_1j_1},\alpha_{k+1}^\varepsilon=s_{i_2j_2},\alpha_l^\varepsilon=s_{i_3j_3})\\&\qquad\times\Big(\overline{f}_z(z_{k+1}^\varepsilon,s_{i_2j_2})-\overline{f}_z(z_{k+1}^\varepsilon,s_{i_1j_1})\Big)X(l,i_3)\Big]\\&=\sum_{i_3=1}^{l_0}\sum_{j_3=1}^{m_{i_3}}\sum_{i_2=1}^{l_0}\sum_{j_2=1}^{m_{i_2}}\sum_{i_1=1}^{l_0}\sum_{j_1=1}^{m_{i_1}}P\big(\alpha_l^\varepsilon=s_{i_3j_3}\big|\alpha_{k+1}^\varepsilon=s_{i_2j_2}\big)P\big(\alpha_{k+1}^\varepsilon=s_{i_2j_2}\big|\alpha_k^\varepsilon=s_{i_1j_1}\big)\\&\qquad\times I(\alpha_k^\varepsilon=s_{i_1j_1})\Big(\overline{f}_z(z_{k+1}^\varepsilon,s_{i_2j_2})-\bar{f}_z(z_{k+1}^\varepsilon,s_{i_1j_1})\Big)E\big(X(l,i_3)\big|\mathcal{F}_k^X\big).\end{aligned} \tag{3.82}$$

Observe that if $i_1=i_2$ then $\bar{f}_z(z_{k+1}^\varepsilon,s_{i_2j_2})-\overline{f}_z(z_{k+1}^\varepsilon,s_{i_1j_1})=0$. In case $i_1\neq i_2$, by noting that $P_\varepsilon=P+\varepsilon Q$, we get $P\big(\alpha_{k+1}^\varepsilon=s_{i_2j_2}\big|\alpha_k^\varepsilon=s_{i_1j_1}\big)\le C\varepsilon$, where the constant C could be chosen as the maximum of the absolute values of all entries of Q. Taking this and the boundedness of $\overline{f}_z(\cdot,\cdot)$ into account, it follows from (3.82) that

$$\Big|E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_{k+1}^\varepsilon)-f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\Big| \le C\varepsilon\sum_{i_3=1}^{l_0}\sum_{j=1}^{d}\big|E\big(X_j(l,i_3)\big|\mathcal{F}_k^X\big)\big|. \tag{3.83}$$

Thus, by inequality (3.73), we have

$$\begin{aligned}
&\sqrt{\varepsilon}E\Big|\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_{k+1}^\varepsilon)-f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big]\Big| \\
&\quad\le \varepsilon\sqrt{\varepsilon}C\sum_{l=k+1}^{T/\varepsilon}\sum_{i=1}^{l_0}\sum_{j=1}^{d}E\big|E\big(X_j(l,i)\big|\mathcal{F}_k^X\big)\big| \\
&\quad\le \varepsilon\sqrt{\varepsilon}Cl_0d\sum_{l=k+1}^{T/\varepsilon}\phi(l-k)^{\frac{\delta}{1+\delta}} \le \varepsilon\sqrt{\varepsilon}Cl_0d\sum_{n=1}^{\infty}\phi(n)^{\frac{\delta}{1+\delta}} \le \varepsilon\sqrt{\varepsilon}C.
\end{aligned} \tag{3.84}$$

Next, note that all norms in $\mathbb{R}^d$ are equivalent, so, since the second derivatives of f are bounded and Lipschitz continuous, for $z, z' \in \mathbb{R}^d$ and $i = 1, 2, \ldots, l_0$, $|f_{zz}(z,i) - f_{zz}(z',i)|_\infty \le \min\{C, C|z-z'|_1\}$. Here, for a matrix $A = (a_{ij})$, $|\cdot|_\infty$ is taken to be $|A|_\infty = \max_{i,j}|a_{ij}|$ and $|z|_1$ is the usual 1-norm, $|z|_1 = \sum_{j=1}^d |z_j|$. Noting that $z_{k+1}^\varepsilon - z_k^\varepsilon = \sqrt{\varepsilon}X(k,\overline{\alpha}_k^\varepsilon)$, by a Taylor expansion,

$$\begin{aligned}
&\Big|\sqrt{\varepsilon}\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)-f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big] - \varepsilon\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)\Big| \\
&= \Big|\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon)-f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)-\sqrt{\varepsilon}X'(k,\overline{\alpha}_k^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big]\sqrt{\varepsilon}\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon X(l,\overline{\alpha}_l^\varepsilon)\Big| \\
&\le \varepsilon C\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}|X(k,i_1)|_1\min\{1,\sqrt{\varepsilon}|X(k,i_1)|_1\}\sum_{l=k+1}^{T/\varepsilon}\big|E\big(X(l,i_2)\big|\mathcal{F}_k^X\big)\big|_1 \\
&\le \varepsilon^{5/4}C\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}|X(k,i_1)|_1\sum_{l=k+1}^{T/\varepsilon}\big|E\big(X(l,i_2)\big|\mathcal{F}_k^X\big)\big|_1 \\
&\qquad+\varepsilon C\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}|X(k,i_1)|_1 I\big(|X(k,i_1)|_1>\varepsilon^{-1/4}\big)\sum_{l=k+1}^{T/\varepsilon}\big|E\big(X(l,i_2)\big|\mathcal{F}_k^X\big)\big|_1.
\end{aligned} \tag{3.85}$$

By Hölder inequality, for $j, j_0 = 1, \ldots, d$ and $i_1, i_2 = 1, \ldots, l_0$,

$$\begin{aligned} E\big|X_j(k,i_1)E\big(X_{j_0}(l,i_2)\big|\mathcal{F}_k^X\big)\big| &\le \|X_j(k,i_1)\|_{\frac{2(1+\delta)}{1+2\delta}}\big\|E\big(X_{j_0}(l,i_2)\big|\mathcal{F}_k^X\big)\big\|_{2(1+\delta)} \\ &\le C\phi(l-k)^{\frac{\delta}{1+\delta}}\|X_j(k,i_1)\|_{\frac{2(1+\delta)}{1+2\delta}}\|X_{j_0}(l,i_2)\|_{2(1+\delta)} \\ &\le C\phi(l-k)^{\frac{\delta}{1+\delta}}. \end{aligned} \tag{3.86}$$

We have used the inequality (3.77) in the second inequality together with the fact that $\|X_j(k,i_1)\|_{\frac{2(1+\delta)}{1+2\delta}}$ and $\|X_{j_0}(l,i_2)\|_{2(1+\delta)}$ are bounded in the last one. Therefore,

$$\begin{aligned} &\varepsilon^{5/4}CE\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}|X(k,i_1)|_1\sum_{l=k+1}^{T/\varepsilon}\big|E\big(X(l,i_2)\big|\mathcal{F}_k^X\big)\big|_1 \\ &\le \varepsilon^{5/4}Cl_0^2d^2\sum_{n=1}^{\infty}\phi(n)^{\frac{\delta}{1+\delta}} \le C\varepsilon^{5/4}. \end{aligned} \tag{3.87}$$

Similarly, by Hölder inequality, for $j, j_0 = 1, \ldots, d$ and $i_1, i_2 = 1, \ldots, l_0$,

$$\begin{aligned} &E\Big|X_j(k,i_1)I\big(|X(k,i_1)|_1 > \varepsilon^{-1/4}\big)E\big(X_{j_0}(l,i_2)\big|\mathcal{F}_k^X\big)\Big| \\ &\le \|X_j(k,i_1)\|_{2(1+\delta)}\big\|I\big(|X(k,i_1)|_1 > \varepsilon^{-1/4}\big)\big\|_{\frac{1+\delta}{\delta}}\big\|E\big(X_{j_2}(l,i_2)\big|\mathcal{F}_k^X\big)\big\|_{2(1+\delta)} \\ &\le C\phi(l-k)^{\frac{\delta}{1+\delta}}P\Big(|X(k,i_1)|_1 > \varepsilon^{\frac{-1}{4}}\Big)^{\frac{\delta}{1+\delta}} \\ &\le C\phi(l-k)^{\frac{\delta}{1+\delta}}\varepsilon^{\frac{\delta}{4(1+\delta)}}E\sum_{j_1=1}^{d}|X_{j_1}(k,i_1)| \le C\phi(l-k)^{\frac{\delta}{1+\delta}}\varepsilon^{\frac{\delta}{4(1+\delta)}}. \end{aligned} \tag{3.88}$$

We have used the Chebyshev's inequality in the third line above. Therefore,

$$\begin{aligned} &CE\sum_{i_1=1}^{l_0}\sum_{i_2=1}^{l_0}|X(k,i_1)|_1 I\big(|X(k,i_1)|_1 > \varepsilon^{-1/4}\big)\sum_{l=k+1}^{T/\varepsilon}\big|E\big(X(l,i_2)\big|\mathcal{F}_k^X\big)\big|_1 \\ &\le C\varepsilon^{1+\frac{\delta}{4(1+\delta)}}\sum_{n=1}^{\infty}\phi(n)^{\frac{\delta}{1+\delta}} \le C\varepsilon^{1+\frac{\delta}{4(1+\delta)}}. \end{aligned} \tag{3.89}$$

By (3.85), (3.87), and (3.89), we have

$$\begin{aligned} &E\Big|\sqrt{\varepsilon}\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon\Big[\Big(f_z(z_{k+1}^\varepsilon,\overline{\alpha}_k^\varepsilon) - f_z(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\Big)X(l,\overline{\alpha}_l^\varepsilon)\Big] \\ &\qquad -\varepsilon\sum_{l=k+1}^{T/\varepsilon}E_k^\varepsilon X'(l,\overline{\alpha}_l^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)\Big| \\ &\le C\big(\varepsilon^{\frac{5}{4}} + \varepsilon^{1+\frac{\delta}{4(1+\delta)}}\big). \end{aligned} \tag{3.90}$$

Thus (3.80) and (3.47) for $i = 3$ follows from (3.81), (3.84) and (3.90).

(2) Claim:

$$\begin{aligned}\varepsilon\mathcal{L}^\varepsilon f_2^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k) = -\varepsilon \sum_{l=k+1}^{T/\varepsilon} \operatorname{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) E_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big] \\ +\varepsilon \operatorname{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) \sum_{l=k+1}^{T/\varepsilon} \widetilde{E}_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big] + e_k^{\varepsilon,4},\end{aligned} \tag{3.91}$$

where $e_k^{\varepsilon,4}$ satisfies (3.47) with $i = 4$.

We have

$$\begin{aligned}&\varepsilon\mathcal{L}^\varepsilon f_2^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k) \\ &= E_k^\varepsilon f_2^\varepsilon(z_{k+1}^\varepsilon, \overline{\alpha}_{k+1}^\varepsilon, \varepsilon k + \varepsilon) - f_2^\varepsilon(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon, \varepsilon k) \\ &= \varepsilon E_k^\varepsilon \operatorname{tr}\Big[\big(\bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon)\big) \sum_{p=k+1}^{T/\varepsilon} \sum_{l=p+1}^{T/\varepsilon} \big[E_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(p, \overline{\alpha}_p^\varepsilon) \\ &\qquad - \widetilde{E}_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(p, \overline{\alpha}_p^\varepsilon)\big]\Big] \\ &\quad + \varepsilon \operatorname{tr}\Big[\big(\bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon) - \bar{f}_{zz}(z_k^\varepsilon, \alpha_k^\varepsilon)\big) \sum_{p=k+1}^{T/\varepsilon} \sum_{l=p+1}^{T/\varepsilon} \big[E_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(p, \overline{\alpha}_p^\varepsilon) \\ &\qquad - \widetilde{E}_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(p, \overline{\alpha}_p^\varepsilon)\big]\Big] \\ &\quad - \varepsilon \sum_{l=k+1}^{T/\varepsilon} \operatorname{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) E_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big] \\ &\quad + \varepsilon \operatorname{tr}\Big[f_{zz}(z_k^\varepsilon, \overline{\alpha}_k^\varepsilon) \sum_{l=k+1}^{T/\varepsilon} \widetilde{E}_k^\varepsilon X(l, \overline{\alpha}_l^\varepsilon) X'(k, \overline{\alpha}_k^\varepsilon)\Big].\end{aligned} \tag{3.92}$$

To proceed, we evaluate the first two terms in the last equation of (3.92). Similar to

(3.82), for $k < p < l \le T/\varepsilon$, we have

$$\varepsilon E_k^\varepsilon \mathrm{tr}\Big[\big(\bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon)\big)\big[E_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon) - \widetilde{E}_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon)\big]\Big]$$
$$= \sum_{i_4=1}^{l_0}\sum_{j_4=1}^{m_{i_4}}\sum_{i_3=1}^{l_0}\sum_{j_3=1}^{m_{i_3}}\sum_{i_2=1}^{l_0}\sum_{j_2=1}^{m_{i_2}}\sum_{i_1=1}^{l_0}\sum_{j_1=1}^{m_{i_1}} \varepsilon I(\alpha_k^\varepsilon = s_{i_1j_1})$$
$$\times P(\alpha_{k+1}^\varepsilon = s_{i_2j_2}|\alpha_k^\varepsilon = s_{i_1j_1})P(\alpha_p^\varepsilon = s_{i_3j_3}|\alpha_{k+1}^\varepsilon = s_{i_2j_2})P(\alpha_l^\varepsilon = s_{i_4j_4}|\alpha_p^\varepsilon = s_{i_3j_3})$$
$$\times \mathrm{tr}\Big[\big(\bar{f}_{zz}(z_{k+1}^\varepsilon, s_{i_2j_2}) - \bar{f}_{zz}(z_{k+1}^\varepsilon, s_{i_1j_1})\big)\big[E\big(X(l,i_4)X'(p,i_3)\big|\mathcal{F}_k^X\big) - EX(l,i_4)X'(p,i_3)\big]\Big]. \tag{3.93}$$

The same argument as what follows (3.82) gives

$$\varepsilon\Big|E_k^\varepsilon \mathrm{tr}\Big[\big(\overline{f}_{zz}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon)\big)\big[E_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon) - \widetilde{E}_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon)\big]\Big]\Big|$$
$$\le \varepsilon^2 C \sum_{i_3,i_4=1}^{l_0}\sum_{j,j_0=1}^{d} \big|E\big(X_j(l,i_4)X_{j_0}(p,i_3)\big|\mathcal{F}_k^X\big) - EX_j(l,i_4)X_{j_0}(p,i_3)\big|. \tag{3.94}$$

Similar to (3.79) in Step 1 (2), we get

$$\sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon} E\big|E\big(X_j(l,i_4)X_{j_0}(p,i_3)\big|\mathcal{F}_k^X\big) - EX_j(l,i_4)X_{j_0}(p,i_3)\big| \le C. \tag{3.95}$$

Therefore, by (3.94) and (3.95),

$$E\Big|\varepsilon E_k^\varepsilon \mathrm{tr}\Big[\big(\overline{f}_{zz}(z_{k+1}^\varepsilon, \alpha_{k+1}^\varepsilon) - \bar{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon)\big)\sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\big[E_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon)$$
$$- \widetilde{E}_{k+1}^\varepsilon X(l, \overline{\alpha}_l^\varepsilon)X'(p, \overline{\alpha}_p^\varepsilon)\big]\Big]\Big| \le C\varepsilon^2 d^2 l_0^2 = C\varepsilon^2. \tag{3.96}$$

Next, note that by boundedness and Lipschitz condition of $f_{zz}(\cdot)$, $|\overline{f}_{zz}(z_{k+1}^\varepsilon, \alpha_k^\varepsilon) - \bar{f}_{zz}(z_k^\varepsilon, \alpha_k^\varepsilon)|_\infty \le C\min\{1, \sqrt{\varepsilon}|X(k, \alpha_k^\varepsilon)|_1\}$, where $|A|_\infty = \max_{i,j}|a_{ij}|$ for $A = (a_{ij})$ and

$|z|_1 = |z_1| + |z_2| + \cdots + |z_d|$ for $z \in \mathbb{R}^d$. Thus,

$$\begin{aligned}
&E\Big|\varepsilon \mathrm{tr}\Big[\big(\bar{f}_{zz}(z^\varepsilon_{k+1},\alpha^\varepsilon_k) - \bar{f}_{zz}(z^\varepsilon_k,\alpha^\varepsilon_k)\big) \\
&\qquad\qquad \times \sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\big[E^\varepsilon_k X(l,\overline{\alpha}^\varepsilon_l)X'(p,\overline{\alpha}^\varepsilon_p) - \widetilde{E}^\varepsilon_{k+1}X(l,\overline{\alpha}^\varepsilon_l)X'(p,\overline{\alpha}^\varepsilon_p)\big]\Big]\Big| \\
&\le \varepsilon C \sum_{j,j_0=1}^{d}\sum_{i_1,i_2,i_3=1}^{l_0}\sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon} E\Big\{\min\{1,\sqrt{\varepsilon}|X(k,i_1)|_1\} \\
&\qquad\qquad \times\Big|\big[E\big(X_j(l,i_3)X_{j_0}(p,i_2)\big|\mathcal{F}^X_k\big) - EX_j(l,i_3)X_{j_0}(p,i_2)\big]\Big|\Big\} \\
&\le \varepsilon C \sum_{j,j_0=1}^{d}\sum_{i_1,i_2,i_3=1}^{l_0}\sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon} \big\|\min\{1,\sqrt{\varepsilon}|X(k,i_1)|_1\}\big\|_{\frac{1+\delta}{\delta}} \\
&\qquad\qquad \times\Big\|\big[E\big(X_j(l,i_3)X_{j_0}(p,i_2)\big|\mathcal{F}^X_k\big) - EX_j(l,i_3)X_{j_0}(p,i_2)\big]\Big\|_{1+\delta}.
\end{aligned} \tag{3.97}$$

We have just used Hölder inequality in the last inequality with $p = \frac{1+\delta}{\delta}, q = 1+\delta$. By using (2.16) with $p = 1+\delta$ and $q = \frac{1+\delta}{\delta}$, instead of (2.17) in (3.76), similar argument to (3.77), (3.78) and (3.79) yields

$$\sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\Big\|\big[E\big(X_j(l,i_3)X_{j_0}(p,i_2)\big|\mathcal{F}^X_k\big) - EX_j(l,i_3)X_{j_0}(p,i_2)\big]\Big\|_{1+\delta} \le C \tag{3.98}$$

where C does not depend on ε and k. Since $\big\|\min\{1,\sqrt{\varepsilon}|X(k,i_1)|_1\}\big\|_{\frac{1+\delta}{\delta}} \to 0$ as $\varepsilon \to 0$, it follows from (3.97) and (3.98) that

$$\begin{aligned}
&\sup_{0\le k<T/\varepsilon} E\Big|\varepsilon \mathrm{tr}\Big[\big(\overline{f}_{zz}(z^\varepsilon_{k+1},\alpha^\varepsilon_k) - \bar{f}_{zz}(z^\varepsilon_k,\alpha^\varepsilon_k)\big) \\
&\qquad \times \sum_{p=k+1}^{T/\varepsilon}\sum_{l=p+1}^{T/\varepsilon}\big[E^\varepsilon_k X(l,\overline{\alpha}^\varepsilon_l)X'(p,\overline{\alpha}^\varepsilon_p) - \widetilde{E}^\varepsilon_{k+1}X(l,\overline{\alpha}^\varepsilon_l)X'(p,\overline{\alpha}^\varepsilon_p)\big]\Big]\Big| = o(\varepsilon).
\end{aligned} \tag{3.99}$$

Thus, (3.91) and (3.47) with $i = 4$ are implied by (3.92), (3.96), and (3.99).

The following statement is obtained by the same argument as in Step 2 (2).

(3) The following assertion holds:

$$\begin{aligned}\varepsilon\mathcal{L}^\varepsilon f_3^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) = -\frac{\varepsilon}{2}X'(k,\overline{\alpha}_k^\varepsilon)f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)X(k,\overline{\alpha}_k^\varepsilon)\\ +\frac{\varepsilon}{2}\mathrm{tr}\Big[f_{zz}(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon)\widetilde{E}_k^\varepsilon X(k,\overline{\alpha}_k^\varepsilon)X'(k,\overline{\alpha}_k^\varepsilon)\Big]+e_k^{\varepsilon,5}\end{aligned} \tag{3.100}$$

where $e_k^{\varepsilon,5}$ satisfies (3.47) with $i=5$.

Finally, by direct computation we obtain the following result.

(4) $\varepsilon\mathcal{L}^\varepsilon f_4^\varepsilon(z_k^\varepsilon,\overline{\alpha}_k^\varepsilon,\varepsilon k) = -\varepsilon(\chi_k^\varepsilon - \overline{\chi}_k^\varepsilon\hat{\nu})Q\overline{F}(z_k^\varepsilon)+e_k^{\varepsilon,6}$ where $e_k^{\varepsilon,6}$ satisfies (3.47) with $i=6$. This concludes the proof of the proposition. □

4 Strong Approximation

4.1 Formulation

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space. We may assume without loss of generality that the probability space accommodates all the random variables and processes of our interest. Throughout this chapter, we use C to denote a generic positive constant with the convention $CC = C$ and $C + C = C$ used.

Let $\varepsilon > 0$ and α_k^ε be a time-homogeneous Markov chain on $(\Omega, \mathcal{F}, P)$ with state space $\mathcal{M} = \{1, 2, \ldots, m\}$ and transition matrix

$$P^\varepsilon = P + \varepsilon Q, \tag{4.1}$$

where $P = (p^{ij})$ is a transition probability matrix and $Q = (q^{ij})$ is a generator of a continuous-time Markov chain (i.e., $p^{ij} \geq 0$ and $\sum_{j=1}^m p^{ij} = 1$; $q^{ij} \geq 0$ for $i \neq j$ and $\sum_{j=1}^m q^{ij} = 0$ for each i). Suppose that P is irreducible and aperiodic with the stationary distribution denoted by $\nu = (\nu_1, \nu_2, \ldots, \nu_m) \in \mathbb{R}^{1\times m}$. Denote by p_k^ε the probability vector $p_k^\varepsilon = (P(\alpha_k^\varepsilon = 1), \cdots, P(\alpha_k^\varepsilon = m)) \in \mathbb{R}^{1\times m}$. Assume that the initial probability p_0^ε is independent of ε, i.e., $p_0^\varepsilon = p_0 = (p_0^1, p_0^2, \ldots, p_0^m)$.

For each $i \in \mathcal{M}$, let $\{X(k, i)\}$ be a wide-sense stationary sequence of real-valued random variables on $(\Omega, \mathcal{F}, P)$ such that $\{(X(k, 1), X(k, 2), \ldots, X(k, m)) : k \in \mathbb{Z}\}$ is an $\mathbb{R}^m$-valued wide-sense stationary sequence. We assume that the sequence $\{(X(k, 1), X(k, 2), \ldots, X(k, m)) : k \in \mathbb{Z}\}$ is independent of the Markov process $\{\alpha_k^\varepsilon\}$ and is ϕ-mixing.

Denote

$$\overline{X}(k) = \sum_{i=1}^{m} \nu_i X(k,i), \tag{4.2}$$

where for each $i \in \mathcal{M}$, $\{X(k,i)\}$ is independent of α_k^ε. That is, $\overline{X}(k)$ can be viewed as an average of $\{X(k,i)$: $i \in \mathcal{M}\}$ with respect to the stationary measure ν. It can be seen (see Remark 4.2) that $\{X(k,\alpha_k^\varepsilon)\}$ is ϕ-mixing. Thus it is ergodic, and as a result,

$$\varepsilon \sum_{k=0}^{\lfloor t/\varepsilon \rfloor -1} [X(k,\alpha_k^\varepsilon) - \overline{X}(k)] = \varepsilon \sum_{k=0}^{\lfloor t/\varepsilon \rfloor -1} \sum_{i=1}^{m} [I_{\{\alpha_k^\varepsilon = i\}} - \nu_i] X(k,i) \to 0 \quad \text{as} \quad \varepsilon \to 0$$

in probability and also with probability one, where $\lfloor z \rfloor$ denotes the integer part of the real number z. Such a result is of interest to many applications in discrete optimization, manufacturing, and wireless communication; see [26, 28] and references therein. The practical implication is that we can "replace" the complex stochastic process by its limit or average in an appropriate sense. How close is this approximation? With more effort, we can further show that

$$X^\varepsilon(t) = \sqrt{\varepsilon} \sum_{k=0}^{\lfloor t/\varepsilon \rfloor -1} [X(k,\alpha_k^\varepsilon) - \overline{X}(k)] \text{ converges weakly to a Brownian motion } B(\cdot), \tag{4.3}$$

with appropriate covariance as $\varepsilon \to 0$. What can we say about the rate of convergence of the process $X^\varepsilon(\cdot)$? What is the almost sure behavior of the underlying process? These questions are our focus in this chapter. We are interested in the almost sure behavior of the sequence $X^\varepsilon(t)$ defined in (4.3). We aim to find the strong approximation of (4.3).

Denote

$$\mathcal{F}_n^X = \sigma\{(X(k,1), X(k,2), \ldots, X(k,m)) : \ k \le n\},$$
$$\phi_X(n) = \sup_{N\in\mathbb{Z}} \phi\Big(\sigma(X(k,i) : k \le N, i \in \mathcal{M}), \sigma(X(l,i) : l \ge N+n, i \in \mathcal{M})\Big).$$

We pose the following conditions.

(A)
- P is irreducible and aperiodic.
- $\{\alpha_k^\varepsilon\}$ is independent of $\{X(k,i) : k \in \mathbb{Z}, i \in \mathcal{M}\}$.
- $\{(X(k,1), X(k,2) \ldots, X(k,m)) : k \in \mathbb{Z}\}$ is an $\mathbb{R}^m$-valued wide-sense stationary, ϕ-mixing sequence with mean 0 and mixing measure given by $\phi_X(n) < C/n^{\frac{4}{3}(1+\beta)}$ for some positive constants C and β.
- $\sup_{k,i} E|X(k,i)|^4 < \infty$.

Remark 4.1. The proof of our main result is based on the mixing property of the sequences $\{\alpha_k^\varepsilon\}$. Under the conditions of Lemma 2.3, the finite state space, and the transition probability (4.1), for sufficiently small ε, the $\{\alpha_k^\varepsilon\}$ are ϕ-mixing with exponential mixing rates. In fact, by virtue of [5, Equation (2.2), p.173] and the ergodicity of P, there exists a number n_0 such that all entries of P^{n_0} are positive. Thus, by the continuity with respect to ε, for $\varepsilon > 0$ small enough, all entries of $(P^\varepsilon)^{n_0}$ are bounded below by a positive number not depending on ε. Denote the bound by q. This implies that P^ε is ergodic with the unique ergodic distribution $\nu^\varepsilon = (\nu_1^\varepsilon, \ldots, \nu_m^\varepsilon)$. By using the result in [5, equation (2.2), p.173], $|p^{\varepsilon,ij}(n) - \nu_j^\varepsilon| \le (1 - mq)^{\frac{n}{n_0}-1}$, $\forall i, j = 1, \ldots, m$, where $p^{\varepsilon,ij}(n) = P(\alpha_n^\varepsilon = j | \alpha_0^\varepsilon = i)$. From this, we can show that $\{\alpha_k^\varepsilon\}$ is mixing with exponential rate. Moreover, if $\phi_{\alpha^\varepsilon}(n)$ is the mixing measure of α_k^ε then there is a positive number $\lambda_0 < 1$ such that $\phi_{\alpha^\varepsilon}(n) < \lambda_0^n$.

Remark 4.2. By [6, Theorem 1, p.4] and the independence of $\{X(k,i)\}$ and $\{\alpha_k^\varepsilon\}$, the sequence $\{[I_{\{\alpha_k^\varepsilon=i\}} - \nu_i]X(k,i)\}$ is ϕ-mixing with the mixing measure $\phi^\varepsilon(n) \leq \phi_X(n) + \phi_{\alpha^\varepsilon}(n)$. In addition, by Remark 4.1 and condition (A), there is a constant C independent of ε such that $\phi^\varepsilon(n) \leq C/n^{\frac{4}{3}(1+\beta)}$. Therefore, without lost of generality, we can suppress the superscript ε in the mixing function $\phi^\varepsilon(n)$.

4.2 Strong Approximation

This section is devoted to obtaining strong approximation results. We use the idea of a step-by-step approximation, which is inspired by the approach used in [27]. Nevertheless, the actual techniques are quite different since continuous-time Markov chains are considered in [27], whereas in our case, discrete-time sequences are treated. Moreover, in addition to the modulating Markov chain, there are a number of random sequences $\{X(k,i) : i \in \mathcal{M}\}$ as well. To obtain the desired result, we use a blocking technique, which is originally appeared in [23]. This approach enables us to effectively "partition" the sequences. Recall the definition of $\Psi(k) = (\psi_{ij}(k)) \in \mathbb{R}^{m\times m}$ given by (2.11). We are in a position to present the main result.

4.2.1 Main Results

Theorem 4.3. *Assume that condition* (A) *holds. Then there exist a constant* $\theta > 0$ *and a* (*possibly non-standard*) *Brownian motion* $\widetilde{W}(t)$ *with* $E\widetilde{W}(t) = 0$ *and* $E[\widetilde{W}(t)]^2 =$

$\sigma^2 t$ *where*

$$\sigma^2 = \sum_{i=1}^{m} \Big[EX(0,i)^2 \nu_i (1-\nu_i) + 2 \sum_{k=1}^{\infty} EX(0,i)X(k,i)\nu_i \psi_{ii}(k) \Big]$$
$$+ \sum_{1 \le i < j \le m} \Big\{ 2 \sum_{k=1}^{\infty} \Big[\nu_i \psi_{ij}(k) EX(0,i)X(k,j) + \nu_j \psi_{ji}(k) EX(k,i)X(0,j) \Big]$$
$$-2\nu_i \nu_j EX(0,i)X(0,j) \Big\} \tag{4.4}$$

such that

$$\sup_{0 \le t \le T} \Big| X^{\varepsilon}(t) - \widetilde{W}(t) \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.5}$$

Remark 4.4. (i) Equation (4.5) is understood to be in the sense that

$$\lim_{\varepsilon \to 0_+} \frac{\sup_{0 \le t \le T} \Big| X^{\varepsilon}(t) - \widetilde{W}(t) \Big|}{\varepsilon^{\theta}} = 0 \ \text{ a.s.}$$

It will be seen in the proof that we can select any positive number θ such that $0 < \theta < \min\{1/8, \beta/4\}$, where β is given in condition (A). For instance, we can choose $0 < \theta < 1/8$ if $\beta = 1/2$. In this case, the condition for mixing rate is $\phi_X(n) < Cn^{-2}$ for some constant C. Then

$$\sup_{0 \le t \le T} |X^{\varepsilon}(t) - \widetilde{W}(t)| = o(\varepsilon^{1/8}) \ \text{ a.s.}$$

Due to the modulating Markov chains and the mixing processes used, the rate is slower than the classical rate for a single i.i.d. sequence with zero mean and bounded fourth moments which is $o(\varepsilon^{1/4})$ obtained directly by using the usual Skorohod embedding method. This is expected because of a family of random sequences is considered and they are correlated by the Markov chain. This also hints the rate of the Markov modulated sequence to be a product of rate of convergence of the scaled occupation meausre for the Markov chain α_k^ε (see [28, Chapter 4]) and that of the sequence $\{X(k,i)\}$ for a fixed $i \in \mathcal{M}$.

(ii) If one deals with a single mixing process, and if one is only interested in getting error bounds for a fixed t, then much sharper results are possible. We refer the reader to [17] for a discussion of the related results. The difficulty of our problem is: The sequences under consideration, in particular, the mixing rates depend not only on n but also on ε. Thus effectively, we have to deal with "double arrays" rather than a "single" sequence. This makes the estimates much more difficult resulting in lower rate of convergence compared to the classical results for a single sequence.

Proof of Theorem 4.3. To facilitate the presentation, the proof is divided into four steps.

Step 1. Approximate $X^\varepsilon(t)$ by a martingale $M_\varepsilon(t)$ defined in (4.6). The main result of this step is given in Proposition 4.5.

Choose $l > 1$, which will be used in the subsequent development. For each ε we divide the series $\sum_j \sum_{i=1}^m [I(\alpha_j^\varepsilon = i) - \nu_i] X(j,i)$ into several blocks with the size of each block being approximately $\varepsilon^{-1/l}$. Define

$$Z_{\varepsilon,n} = \sqrt{\varepsilon} \sum_{j=\lfloor (n-1)/\varepsilon^{1/l} \rfloor}^{\lfloor n/\varepsilon^{1/l} \rfloor - 1} \sum_{i=1}^m [I(\alpha_j^\varepsilon = i) - \nu_i] X(j,i), \quad n \geq 1,$$

$$\widetilde{\mathcal{F}}_n^\varepsilon = \mathcal{F}_{\lfloor n/\varepsilon^{1/l} \rfloor - 1}^{\alpha^\varepsilon} \vee \mathcal{F}_{\lfloor n/\varepsilon^{1/l} \rfloor - 1}^{X}, \quad n \geq 1.$$

It is clear that $Z_{\varepsilon,n}$ is $\widetilde{\mathcal{F}}_n^\varepsilon$-measurable. Define $Y_{\varepsilon,1} = Z_{\varepsilon,1}$ and

$$Y_{\varepsilon,n} = \sum_{j=n}^{2n-1} \left[E(Z_{\varepsilon,j} | \widetilde{\mathcal{F}}_n^\varepsilon) - E(Z_{\varepsilon,j} | \widetilde{\mathcal{F}}_{n-1}^\varepsilon) \right], \quad n \geq 2.$$

Then $Y_{\varepsilon,n}$ is also $\widetilde{\mathcal{F}}^{\varepsilon}_n$-measurable. Moreover, $E(Y_{\varepsilon,n}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1}) = 0$ for all $n \geq 2$. Define

$$M_{\varepsilon,n} = \textstyle\sum_{j=1}^{n} Y_{\varepsilon,j}, \ n \geq 1, \qquad M_\varepsilon(t) = M_{\varepsilon,\lfloor t/\varepsilon^{(l-1)/l}\rfloor} \quad 0 \leq t \leq T. \tag{4.6}$$

We can show that $(M_{\varepsilon,n}, \widetilde{\mathcal{F}}^{\varepsilon}_n)$ is a martingale. To proceed, approximate $X^\varepsilon(t)$ by $M_\varepsilon(t)$. The result on uniform approximation is presented next.

Proposition 4.5. *If* $\theta < \frac{1}{4} - \frac{1}{4l}$, *then*

$$\sup_{0\leq t\leq T} \Big| M_\varepsilon(t) - X^\varepsilon(t) \Big| = o(\varepsilon^\theta) \quad a.s. \tag{4.7}$$

Proof of Proposition 4.5. For each $n \geq 2$,

$$M_{\varepsilon,n} = \sum_{j=1}^{n} Y_{\varepsilon,j} = \sum_{j=1}^{n} Z_{\varepsilon,j} + \sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_n) - \sum_{j=1}^{n-1} E(Z_{\varepsilon,2j} + Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}^{\varepsilon}_j).$$

Therefore,

$$\begin{aligned} M_{\varepsilon,\lfloor t/\varepsilon^{(l-1)/l}\rfloor} = X_{\varepsilon,\left\lfloor \frac{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}{\varepsilon^{1/l}} \right\rfloor} &+ \sum_{j=\lfloor t/\varepsilon^{(l-1)/l}\rfloor+1}^{2\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}\Big) \\ &- \sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j} + Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}^{\varepsilon}_j\Big). \end{aligned} \tag{4.8}$$

This yields that

$$\begin{aligned} \sup_{0\leq t\leq T} |M_\varepsilon(t) - X^\varepsilon(t)| &= \sup_{0\leq t\leq T} \Big| M_{\varepsilon,\lfloor t/\varepsilon^{(l-1)/l}\rfloor} - X_{\varepsilon,\lfloor t/\varepsilon\rfloor} \Big| \\ &\leq \sup_{0\leq t\leq T} \Big| X_{\varepsilon,\left\lfloor \frac{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}{\varepsilon^{1/l}} \right\rfloor} - X_{\varepsilon,\lfloor t/\varepsilon\rfloor} \Big| + \sup_{0\leq t\leq T} \Big| \sum_{j=\lfloor t/\varepsilon^{(l-1)/l}\rfloor+1}^{2\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}\Big) \Big| \\ &\quad + \sup_{0\leq t\leq T} \Big| \sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j} + Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}^{\varepsilon}_j\Big) \Big|. \end{aligned} \tag{4.9}$$

To estimate the left-hand side, by virtue of (4.9), it suffices to examine each term on the right-hand side. The estimates of these terms are presented in Proposition 4.6. The result is stated next, and its proof is relegated to Section 4.3 to maintain the continuity of the flow of presentation.

Proposition 4.6. *There exists a constant C independent of ε such that*

(i)

$$P\Big(\sup_{0\le t\le T}\Big|\sum_{j=\lfloor t/\varepsilon^{(l-1)/l}\rfloor+1}^{2\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}\Big)\Big|\ge\varepsilon^{\theta}\Big)\le C\varepsilon^{1+\frac{1}{l}-4\theta}, \tag{4.10}$$

(ii)

$$P\Big(\sup_{0\le t\le T}\Big|\sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j}+Z_{\varepsilon,2j+1}\Big|\widetilde{\mathcal{F}}^{\varepsilon}_{j}\Big)\Big|\ge\varepsilon^{\theta}\Big)\le C\varepsilon^{\frac{1}{2}-\theta}, \tag{4.11}$$

(iii)

$$P\Big(\max_{0\le t\le T}\Big|X_{\varepsilon,\left\lfloor\frac{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}{\varepsilon^{1/l}}\right\rfloor}-X_{\varepsilon,\lfloor t/\varepsilon\rfloor}\Big|\ge\varepsilon^{\theta}\Big)\le C\varepsilon^{1-\frac{1}{l}-4\theta}. \tag{4.12}$$

Now we can complete the proof of Proposition 4.5. It follows from (4.9), (4.10), (4.11), and (4.12) that

$$\begin{aligned}
&P\Big(\sup_{0\le t\le T}\Big|M_{\varepsilon,\lfloor t/\varepsilon^{(l-1)/l}\rfloor}-X_{\varepsilon,\lfloor t/\varepsilon\rfloor}\Big|\ge\varepsilon^{\theta}\Big)\\
&\le P\Big(\sup_{0\le t\le T}\Big|X_{\varepsilon,\left\lfloor\frac{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}{\varepsilon^{1/l}}\right\rfloor}-X_{\varepsilon,\lfloor t/\varepsilon\rfloor}\Big|\ge\frac{\varepsilon^{\theta}}{3}\Big)\\
&\quad+P\Big(\sup_{0\le t\le T}\Big|\sum_{j=\lfloor t/\varepsilon^{(l-1)/l}\rfloor+1}^{2\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}\Big)\Big|\ge\frac{\varepsilon^{\theta}}{3}\Big)\\
&\quad+P\Big(\sup_{0\le t\le T}\Big|\sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j}+Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}^{\varepsilon}_{j}\Big)\Big|\ge\frac{\varepsilon^{\theta}}{3}\Big)\\
&\le C\Big(\varepsilon^{1-\frac{1}{l}-4\theta}+\varepsilon^{1+\frac{1}{l}-4\theta}+\varepsilon^{\frac{1}{2}-\theta}\Big)\le C\varepsilon^{1-\frac{1}{l}-4\theta}.
\end{aligned} \tag{4.13}$$

If $\theta<\frac{1}{4}-\frac{1}{4l}$, then $\widetilde{\theta}=1-\frac{1}{l}-4\theta>0$, and

$$P\Big(\sup_{0\le t\le T}\Big|M_{\varepsilon,\lfloor t/\varepsilon^{(l-1)/l}\rfloor}-X_{\varepsilon,\lfloor t/\varepsilon\rfloor}\Big|\ge\varepsilon^{\theta}\Big)\le C\varepsilon^{\widetilde{\theta}}. \tag{4.14}$$

Let $\varepsilon_n = n^{-2/\widetilde{\theta}}$. Then from (4.14),

$$\sum_{n=1}^{\infty} P\Big(\sup_{0\le t\le T}\Big|M_{\varepsilon_n,\lfloor t/\varepsilon_n^{(l-1)/l}\rfloor} - X_{\varepsilon_n,\lfloor t/\varepsilon_n\rfloor}\Big| \ge \varepsilon_n^{\theta}\Big) < \infty.$$

The Borel-Cantelli lemma implies that

$$\sup_{0\le t\le T}\Big|M_{\varepsilon_n,\lfloor t/\varepsilon_n^{(l-1)/l}\rfloor} - X_{\varepsilon_n,\lfloor t/\varepsilon_n\rfloor}\Big| \le O(\varepsilon_n^{\theta}) \quad \text{a.s.}$$

According to the choice of ε_n and (4.8), (4.7) follows. $\square$

In view of Proposition 4.5, to prove (4.5), it suffices to show that there exists a standard Brownian motion $W(t)$ such that

$$\sup_{0\le t\le T}\Big|M_{\varepsilon}(t) - W(\sigma t)\Big| = o(\varepsilon^{\theta}) \quad \text{a.s.} \tag{4.15}$$

for some $\theta > 0$ with σ defined in (4.4).

Note that $M_\varepsilon(t) = M_\varepsilon(k\varepsilon^{(l-1)/l})$ if $k \le \frac{t}{\varepsilon^{(l-1)/l}} < k+1$. Hence

$$\begin{aligned}\sup_{0\le t\le T}\Big|M_{\varepsilon}(t) - W(\sigma t)\Big| \le &\max_{1\le k\le \lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\Big|M_\varepsilon(k\varepsilon^{(l-1)/l}) - W(\sigma k\varepsilon^{(l-1)/l})\Big|\\ &+ \sup_{0\le k\le \lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \sup_{k\varepsilon^{(l-1)/l}\le t\le (k+1)\varepsilon^{(l-1)/l}}\Big|W(\sigma k\varepsilon^{(l-1)/l}) - W(\sigma t)\Big|.\end{aligned} \tag{4.16}$$

The estimate of the first term on the left-hand side of (4.16) is obtained in step 3, whereas the last term in (4.16) is dealt with in step 4. We next give the formula of σ and prepare for step 3.

Step 2. Preliminary estimates. The following lemma gives the representation of σ. Its proof is deferred until Section 4.

Proposition 4.7. *If* $\bar{\beta} < \widehat{\beta} = \min\{1, \beta\}$ *then there exists a constant* C *such that for each* $n \geq 1$*,*

$$\Big|\frac{1}{n}E\Big[\sum_{k=1}^{n}\sum_{i=1}^{m}[I(\alpha_k^\varepsilon = i) - \nu_i]X(k,i)\Big]^2 - \sigma^2\Big| \leq C(\varepsilon + n^{-\bar{\beta}}), \tag{4.17}$$

where σ^2 *is given by* (4.4).

Recall that $M_\varepsilon(k\varepsilon^{(l-1)/l}) = M_{\varepsilon,k}$ and $(M_{\varepsilon,k}, \widetilde{\mathcal{F}}_k^\varepsilon)$ is a martingale. By virtue of the martingale version of the Skorohod representation theorem (see [11, Theorem A.1, p.269]), there exist nonnegative random variables $\tau_{\varepsilon,k}$ such that

$$\begin{aligned}&\Big\{M_\varepsilon(k\varepsilon^{(l-1)/l}), k = 1, \ldots, \Big\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\Big\rfloor\Big\}\\ &\quad = \Big\{W\Big(\varepsilon^{(l-1)/l}\big(\tau_{\varepsilon,1} + \tau_{\varepsilon,2} + \cdots + \tau_{\varepsilon,k}\big)\Big), k = 1, \cdots, \Big\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\Big\rfloor\Big\} \quad \text{in distribution,}\end{aligned} \tag{4.18}$$

with $W(\cdot)$ being a standard Brownian motion.

Now, let $\mathcal{F}_k^\varepsilon = \sigma(Y_{\varepsilon,1}, Y_{\varepsilon,2}, \ldots, Y_{\varepsilon,k})$. Let $\mathcal{G}_0^\varepsilon$ be the trivial σ-field and let $\mathcal{G}_k^\varepsilon$ be the σ-field generated by $\mathcal{F}_k^\varepsilon$ and $\sigma(W(t) : 0 \leq t \leq \varepsilon^{(l-1)/l}\sum_{i=1}^{k}\tau_{\varepsilon,i})$. Again, from [11, Theorem A.1, p.269]), we have $\tau_{\varepsilon,k}$ is $\mathcal{G}_k^\varepsilon$-measurable. Moreover,

$$E[\varepsilon^{(l-1)/l}\tau_{\varepsilon,k}] = EY_{\varepsilon,k}^2 \ \text{ for } \ k \geq 1, \tag{4.19}$$

and

$$E\Big(\varepsilon^{(l-1)/l}\tau_{\varepsilon,k}\Big|\mathcal{G}_{k-1}^\varepsilon\Big) = E\Big(Y_{\varepsilon,k}^2\Big|\mathcal{F}_{k-1}^\varepsilon\Big), \quad \text{for } k = 1, \ldots, \Big\lfloor\tfrac{T}{\varepsilon^{(l-1)/l}}\Big\rfloor. \tag{4.20}$$

We have the following estimates. The proofs are given in Section 4.3.

Proposition 4.8. (i) *If* $0 < \theta < \frac{1}{2} - \frac{1}{2l}$*, then*

$$\max_{0\leq k\leq\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\Big[\tau_{\varepsilon,j} - E(\tau_{\varepsilon,j}|\mathcal{G}_{j-1}^\varepsilon)\Big]\Big| = o(\varepsilon^\theta) \quad a.s. \tag{4.21}$$

(ii) *If* $0 < \theta < \frac{1}{2} - \frac{1}{2l}$, *then*

$$\max_{0 \le k \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| \sum_{j=1}^{k} \Big[Y_{\varepsilon,j}^2 - E(Y_{\varepsilon,j}^2 | \mathcal{F}_{j-1}^{\varepsilon}) \Big] \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.22}$$

(iii) *If* $0 < \theta < \frac{1}{2l}$, *then*

$$\max_{0 \le k \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| \sum_{j=1}^{k} [Y_{\varepsilon,j}^2 - Z_{\varepsilon,j}^2] \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.23}$$

(iv) *If* $\theta < \frac{1}{2} - \frac{1}{2l}$, *then*

$$\max_{1 \le k \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| \sum_{j=1}^{k} \Big[Z_{\varepsilon,j}^2 - E Z_{\varepsilon,j}^2 \Big] \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.24}$$

(v) *Let* $\widehat{\beta} = \min\{\beta, 1\}$. *If* $\theta < \frac{\widehat{\beta}}{l}$, *then*

$$\max_{1 \le k \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| \sum_{j=1}^{k} E Z_{\varepsilon,j}^2 - k\varepsilon^{(l-1)/l} \sigma^2 \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.25}$$

where σ *is defined by* (4.4).

Step 3. Estimate $|M_\varepsilon(k\varepsilon^{(l-1)/l}) - W(\sigma k \varepsilon^{(l-1)/l})|$. The result is stated in the following proposition.

Proposition 4.9. *For any* $0 < \theta < \min\{\frac{1}{4} - \frac{1}{4l}, \frac{1}{4l}, \frac{\beta}{2l}\}$,

$$\max_{1 \le k \le \lfloor \frac{T}{\varepsilon^{(l-1)/l}} \rfloor} \Big| M_\varepsilon(k\varepsilon^{(l-1)/l}) - W(\sigma k \varepsilon^{(l-1)/l}) \Big| = o(\varepsilon^{\theta}) \quad a.s. \tag{4.26}$$

Proof. From (4.20) and the triangle inequality,

$$\begin{aligned}
&\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\tau_{\varepsilon,j}-k\varepsilon^{(l-1)/l}\sigma^2\right|\\
&\le \max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\Big[\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big]\right|\\
&\quad+\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\sum_{j=1}^{k}\Big[Y^2_{\varepsilon,j}-E(Y^2_{\varepsilon,j}|\mathcal{F}^{\varepsilon}_{j-1})\Big]\right|+\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\sum_{j=1}^{k}[Y^2_{\varepsilon,j}-Z^2_{\varepsilon,j}]\right|\\
&\quad+\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\sum_{j=1}^{k}\Big[Z^2_{\varepsilon,j}-EZ^2_{\varepsilon,j}\Big]\right|+\max_{1\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\sum_{j=1}^{k}EZ^2_{\varepsilon,j}-k\varepsilon^{(l-1)/l}\sigma^2\right|.
\end{aligned} \tag{4.27}$$

By (4.27), (4.21), (4.22), (4.23), (4.24), and (4.25), for $0<\theta<\min\{\frac{1}{2}-\frac{1}{2l},\frac{1}{2l},\frac{\widehat{\beta}}{l}\}$ with $\widehat{\beta}=\min\{\beta,1\}$,

$$\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\tau_{\varepsilon,j}-k\varepsilon^{(l-1)/l}\sigma^2\right|=o(\varepsilon^{\theta})\quad \text{a.s.} \tag{4.28}$$

Thus, it follows from (4.28) and [4, Theorem 1.1.1] that

$$\begin{aligned}
&\sup_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\left|W\Big(\varepsilon^{(l-1)/l}\big(\tau_{\varepsilon,1}+\cdots+\tau_{\varepsilon,k}\big)\Big)-W\Big(k\varepsilon^{(l-1)/l}\sigma^2\Big)\right|\\
&\le \sup_{0\le s\le \sigma^2 T}\sup_{0\le h\le \varepsilon^{\theta}}\Big|W(s+h)-W(s)\Big|=O\Big(\varepsilon^{\theta/2}\log^{1/2}\Big(\frac{1}{\varepsilon}\Big)\Big)\quad \text{a.s.}
\end{aligned} \tag{4.29}$$

Since $\min\{\frac{1}{4}-\frac{1}{4l},\frac{1}{4l},\frac{\beta}{2l}\}=\frac{1}{2}\min\{\frac{1}{2}-\frac{1}{2l},\frac{1}{2l},\frac{\widehat{\beta}}{l}\}$, the proof follows from (4.29) and (4.18). $\square$

Step 4. Estimate $|W(\sigma k\varepsilon^{(l-1)/l})-W(\sigma t)|$.

Proposition 4.10. *For any $\theta<\frac{l-1}{2l}$, we have*

$$\sup_{0\le k\le \lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \sup_{k\varepsilon^{(l-1)/l}\le t\le (k+1)\varepsilon^{(l-1)/l}}\left|W(\sigma k\varepsilon^{(l-1)/l})-W(\sigma t)\right|=o(\varepsilon^{\theta}). \tag{4.30}$$

Proof. Using [4, Theorem 1.1.1],

$$\lim_{\varepsilon\to 0}\frac{\sup\limits_{0\le k\le\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \sup\limits_{k\varepsilon^{(l-1)/l}\le t\le(k+1)\varepsilon^{(l-1)/l}}\left|W(\sigma k\varepsilon^{(l-1)/l})-W(\sigma t)\right|}{\left(2\sigma\varepsilon^{(l-1)/l}\log(\sigma^{-1}\varepsilon^{-(l-1)/l})\right)^{1/2}}=1\quad\text{a.s.}$$

This implies (4.30). □

Completion of the Proof of Theorem 4.3. Put $\theta_0=\min\{\frac{1}{2}-\frac{1}{2l},\frac{1}{2l},\frac{\beta}{l}\}$. By Proposition 4.9, for $\theta<\frac{\theta_0}{2}$,

$$\max_{1\le k\le\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\left|M_\varepsilon(k\varepsilon^{(l-1)/l})-W(\sigma k\varepsilon^{(l-1)/l})\right|=o(\varepsilon^\theta)\quad\text{a.s.}\tag{4.31}$$

On the other hand, by Proposition 4.10, for $\theta<\frac{l-1}{2l}$,

$$\sup_{0\le k\le\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \sup_{k\varepsilon^{(l-1)/l}\le t\le(k+1)\varepsilon^{(l-1)/l}}\left|W(\sigma k\varepsilon^{(l-1)/l})-W(t)\right|=o(\varepsilon^\theta).\tag{4.32}$$

Therefore, by (4.16), for $0<\theta<\min\{\frac{\theta_0}{2},\frac{l-1}{2l}\}$, we have

$$\sup_{0\le t\le T}\left|M_\varepsilon(t)-W(\sigma t)\right|=o(\varepsilon^\theta).\tag{4.33}$$

Since $\frac{\theta_0}{2}\le\frac{1}{4}-\frac{1}{4l}$, by Proposition 4.5, for $\theta<\frac{\theta_0}{2}$

$$\sup_{0\le t\le T}\left|M_\varepsilon(t)-X^\varepsilon(t)\right|=o(\varepsilon^\theta)\quad\text{a.s.}\tag{4.34}$$

By (4.33) and (4.34), for $0<\theta<\min\{\frac{\theta_0}{2},\frac{l-1}{2l}\}$,

$$\sup_{0\le t\le T}\left|X^\varepsilon(t)-W(\sigma t)\right|=o(\varepsilon^\theta).\tag{4.35}$$

Choose $l=2$, then $\frac{\theta_0}{2}=\min\{\frac{1}{8},\frac{\widehat{\beta}}{4}\}=\min\{\frac{1}{8},\frac{\beta}{4}\}<\frac{l-1}{2l}=\frac{1}{4}$. Hence, (4.35) holds true for $0<\theta<\min\{\frac{1}{8},\frac{\beta}{4}\}$. The theorem is thus proved for $\widetilde{W}(t)=W(\sigma t)$. □

4.2.2 Examples

Example 4.11. For the purpose of demonstration, only a very simple example (a one dimensional parameter optimization problem) is considered. Suppose that one is interested in finding the minima of a function $J(\cdot) : \mathbb{R} \mapsto \mathbb{R}$, in which only the noisy corrupted observations or measurements $\nabla J(\Theta_n) + \check{X}_n$ are available for each n, where $\check{X}_n = X(n, \alpha_n^\varepsilon) - \overline{X}(n)$ represents the noise, and $\{X(n, \alpha_n^\varepsilon)\}$ and $\overline{X}(n)$ are as defined in the beginning of Section 2.2. Not only does the measurement noise include the usual noise processes, but also there is a switching process representing the random environment resulting in the regime-switching from one discrete state to another.

To carry out the desired optimization task, we use stochastic approximation methods. This amount to construct a recursive algorithm of the form

$$\Theta_{n+1} = \Theta_n - \varepsilon \nabla J(\Theta_n) + \varepsilon \check{X}_n.$$

Note that for simplicity, we have assumed that $\nabla J(\Theta)$+noise is available. If we can observe only function values with noise, then a noisy finite difference method is needed for the gradient approximation.

Define $\Theta^\varepsilon(t) = \Theta_n$ for $t \in [n\varepsilon, n\varepsilon + \varepsilon)$. Suppose that there is a unique Θ^* (a unique minimizer of $J(\cdot)$) such that $\nabla J(\Theta^*) = 0$. Then with the mixing condition proposed together with the Markov chain α_n, it can be shown that $\Theta^\varepsilon(\cdot + t_\varepsilon) \to \Theta^*$ as $\varepsilon \to 0$ in probability, where $t_\varepsilon \to \infty$ as $\varepsilon \to 0$.

To analyze the rate of convergence, suppose that $\nabla J(\Theta) = H(\Theta - \Theta^*) + O(|\Theta - \Theta^*|^2)$ and $H < 0$. Note a particular case is that J is quadratic in Θ, and ∇J is linear in Θ. Define $u_n = (\Theta_n - \Theta^*)/\sqrt{\varepsilon}$. Then under suitable conditions, it can be shown (see [16, Chapter 10]) that there is an N_ε such that $E|u_n|^2 = O(\varepsilon)$ for $n \geq N_\varepsilon$. Then it can be shown that

$$u_{\lfloor t/\varepsilon \rfloor} = \sqrt{\varepsilon} \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} (1 - \varepsilon H)^{\lfloor t/\varepsilon \rfloor - 1 - k} \check{X}_k + \varepsilon^{3/2} \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} (1 - \varepsilon H)^{\lfloor t/\varepsilon \rfloor - 1 - k} O(|u_k|^2).$$

Recall that $X^\varepsilon(t) = \sqrt{\varepsilon} \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} \check{X}_k$. Then by Theorem 4.3, $\sup_{0 \leq t \leq T} |X^\varepsilon(t) - \widetilde{W}(t)| = o(\varepsilon^\theta)$. Clearly, the dominating part of $u_{\lfloor t/\varepsilon \rfloor}$ is

$$U_{\lfloor t/\varepsilon \rfloor - 1} = \sqrt{\varepsilon} \sum_{k=0}^{\lfloor t/\varepsilon \rfloor} (1 - \varepsilon H)^{\lfloor t/\varepsilon \rfloor - 1 - k} \check{X}_k.$$

Define

$$U^\varepsilon(t) = X^\varepsilon(t) - \varepsilon H \sum_{k=1}^{\lfloor t/\varepsilon \rfloor - 1} (1 - \varepsilon H)^{\lfloor t/\varepsilon \rfloor - 1 - k} X^\varepsilon(\varepsilon k).$$

Roughly, by Theorem 4.3, $X^\varepsilon(\cdot)$ can be replaced by $B(\cdot)$ with an additional error of the order $o(\varepsilon^\theta)$. Thus Theorem 4.3 will help us to obtain further analyze the asymptotics of $U^\varepsilon(\cdot)$.

Example 4.12. As alluded to in the introduction, to reflect the feature of random environment, random processes $X^\varepsilon(t)$ is used frequently. Often, one wishes to find the excursion probability

$$P(\sup_{0 \leq t \leq T} X^\varepsilon(t) \geq a) \quad \text{for some} \quad a > 0. \tag{4.36}$$

Such an estimate is not all simple due to the complex structure of the processes. However, Theorem 4.3 provides us with a viable alternative, namely, to use $X^\varepsilon(t) = \widetilde{W}(t) + o(\varepsilon^\theta)$ a.s. Thus with an error of the order $o(\varepsilon^\theta)$, the calculation of (4.36) reduces to the use of the excursion of a Brownian motion. First,

$$P(\sup_{0\le t\le T}\widetilde{W}(t)\ge a)=\sqrt{\frac{2}{\pi T}}\int_a^\infty \exp(-x^2/2T)dx.$$

The last part of the above equation follows from [13, p. 346]. Then Theorem 4.3 tells use that there is a function $\kappa(\varepsilon)$ satisfying $\kappa(\varepsilon)=o(\varepsilon^\theta)$ such that

$$\begin{aligned}
&P(\sup_{0\le t\le T}X^\varepsilon(t)\ge a)\\
&\quad=P(\sup_{0\le t\le T}\widetilde{W}(t)\ge a-\kappa(\varepsilon))\\
&\quad=\sqrt{\frac{2}{\pi T}}\Big(\int_a^\infty+\int_{a-\kappa(\varepsilon)}^a\Big)\exp(-x^2/2T)dx\\
&\quad=\sqrt{\frac{2}{\pi T}}\int_a^\infty \exp(-x^2/2T)dx+o(\varepsilon^\theta).
\end{aligned}$$

4.3 Proofs of Technical Results

This section is divided into three subsections. Each subsection provides the proof of one proposition. Within a subsection, we organize the results into a number of lemmas if it is needed.

4.3.1 Proof of Proposition 4.6

As a preparation, we first prove some lemmas.

Lemma 4.13. *Let $\{U_n, n\ge 1\}$ be a ϕ-mixing sequence, n, N positive integers, $0\le k_1<k_2<\cdots<k_N$ integers. Denote $S_k(\iota)=\sum_{j=k}^{k+\iota-1}U_j$ for $k,\iota\ge 1$. Assume that there exists a positive number η, $0<\eta<1$, an integer p, $1\le p\le n$ and a number $A>0$ such that*

$$\phi(p)+\max_{1\le j\le N}\max_{p\le\iota\le n}P\Big(\Big|S_{k_j}(n)-S_{k_j}(\iota)\Big|\ge A\Big)\le\eta.$$

Then for any $a, b \geq 0$, we have

$$P\Big(\max_{1\leq j\leq N}\max_{1\leq \iota\leq n}\Big|S_{k_j}(\iota)\Big| \geq a+A+b\Big)$$
$$\leq \frac{1}{1-\eta}\sum_{j=1}^{N} P\Big(\Big|S_{k_j}(n)\Big| \geq a\Big) + \frac{1}{1-\eta}P\Big(\max_{1\leq j\leq N}\max_{0\leq \iota\leq n-1}|U_{k_j+\iota}| \geq \frac{b}{p-1}\Big). \tag{4.37}$$

Proof of Lemma 4.13. For $1 \leq \iota \leq n, 1 \leq j \leq N$, denote $T_j = \max_{1\leq \iota\leq n}\Big|S_{k_j}(\iota)\Big|$, $E_j = \Big\{\max_{1\leq k<j} T_k < a+A+b \leq T_j\Big\}$ and

$$E_j^{\iota} = E_j \cap \Big\{\max_{1\leq k<\iota}\Big|S_{k_j}(k)\Big| < a+A+b \leq \Big|S_{k_j}(\iota)\Big|\Big\}.$$

Then

$$P\Big(\max_{1\leq j\leq N}\max_{1\leq \iota\leq n}\Big|S_{k_j}(\iota)\Big| \geq a+A+b\Big)$$
$$\leq \sum_{j=1}^{N}\Big\{P\Big(E_j\cap\{|S_{k_j}(n)|\geq a\}\Big) + \sum_{\iota=1}^{n-1}P\Big(E_j^{\iota}\cap\{|S_{k_j}(n)-S_{k_j}(\iota)|\geq A+b\}\Big)\Big\}. \tag{4.38}$$

We have

$$\begin{aligned}
&\sum_{\iota=1}^{n-1}P\Big(E_j^{\iota}\cap\{|S_{k_j}(n)-S_{k_j}(\iota)|\geq A+b\}\Big)\\
&\quad\leq \sum_{\iota=1}^{n-p-1}P\Big(E_j^{\iota}\cap\{|S_{k_j}(\iota+p-1)-S_{k_j}(\iota)|\geq b\}\Big)\\
&\qquad+\sum_{\iota=1}^{n-p-1}P\Big(E_j^{\iota}\cap\{|S_{k_j}(n)-S_{k_j}(\iota+p-1)|\geq A\}\Big)\\
&\qquad+\sum_{\iota=n-p}^{n-1}P\Big(E_j^{\iota}\cap\{|S_{k_j}(n)-S_{k_j}(\iota)|\geq A+b\}\Big) \qquad (4.39)\\
&\quad\leq \sum_{\iota=1}^{n-1}P\Big(E_j^{\iota}\cap\Big\{\max_{1\leq j\leq N}\max_{0\leq \iota\leq n-1}|U_{k_j+\iota}|\geq \frac{b}{p-1}\Big\}\Big)\\
&\qquad+\sum_{\iota=1}^{n-p-1}P(E_j^{\iota})\Big(P\big(\big|S_{k_j}(n)-S_{k_j}(\iota+p-1)\big|\geq A\big)+\phi(p)\Big)\\
&\quad\leq P\Big(E_j\cap\{\max_{1\leq j\leq N}\max_{0\leq \iota\leq n-1}|U_{k_j+\iota}|\geq \frac{b}{p-1}\}\Big)+\eta P(E_j).
\end{aligned}$$

Thus,

$$
\begin{aligned}
P\Big(&\max_{1\le j\le N}\max_{1\le\iota\le n}\Big|S_{k_j}(\iota)\Big|\ge a+A+b\Big)\\
&\le\sum_{j=1}^{N}P\Big(|S_{k_j}(n)|\ge a\Big)+P\Big(\max_{1\le j\le N}\max_{0\le\iota\le n-1}|U_{k_j+\iota}|\ge\frac{b}{p-1}\Big)\\
&\quad+\eta P\Big(\max_{1\le j\le N}\max_{1\le\iota\le n}\Big|S_{k_j}(\iota)\Big|\ge a+A+b\Big).
\end{aligned}
\tag{4.40}
$$

This proves the lemma. □

The following Lemma follows directly from [17, Lemma 2.2.5]

Lemma 4.14. *Let $\{U_n, n\ge 1\}$ be a mixing sequence such that $EU_n^{2(1+\delta)}\le C$ and $\phi(n)\le\frac{C}{n}$ for some constant C and $\delta>0$. Denote $S_n=\sum_{j=1}^{n}U_j$, then there is a constant C not depending on n such that $E|S_n|^{2(1+\delta)}\le Cn^{1+\delta}$.*

By assumption (A) and Remark 4.2, the mixing sequence $U_j=\sum_{i=1}^{m}[I(\alpha_j^\varepsilon=i)-\nu_i]X(j,i)$ satisfies all conditions of Lemma 4.14 with $\delta=1$ and C does not depend on ε. Thus, the lemma yields the following result.

Corollary 4.15. *There exists a constant C that does not depend on ε and n such that*

$$
E|Z_{\varepsilon,n}|^4\le C\varepsilon^{2(1-1/l)},\quad E|Z_{\varepsilon,n}|^2\le C\varepsilon^{1-1/l}. \tag{4.41}
$$

Lemma 4.16. *For any $n>0$,*

$$
E\Big|\sum_{j=n+1}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}_n^\varepsilon)\Big|^4\le C\varepsilon^2, \tag{4.42}
$$

where C is a constant independent of ε and n.

Proof of Lemma 4.16. By the independence between $\{\alpha_k^\varepsilon\}$ and $\{X(k,i): i=1,\dots,m; k\in\mathbb{Z}\}$, and (2.12), there exists a constant C independent of ε such that for any $j \geq \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor$,

$$E\Big([I(\alpha^\varepsilon_j = i) - \nu_i]X(j,i)\Big|\widetilde{\mathcal{F}}^\varepsilon_n\Big) = E\Big([I(\alpha^\varepsilon_j = i) - \nu_i]\Big|\mathcal{F}^{\alpha^\varepsilon}_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor - 1}\Big)E\Big(X(j,i)\Big|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor - 1}\Big)$$
$$\leq C\Big(\varepsilon + \lambda^{j - \lfloor \frac{k}{\varepsilon^{1/l}} \rfloor + 1}\Big)E\Big(X(j,i)\Big|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor - 1}\Big). \tag{4.43}$$

Therefore, there exists a constant C independent of ε such that

$$\begin{aligned} E\Big|\sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^\varepsilon_n)\Big|^4 &= \varepsilon^2 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m E([I(\alpha^\varepsilon_j = i) - \nu_i]X(j,i)|\widetilde{\mathcal{F}}^\varepsilon_n)\Big|^4 \\ &\leq C\varepsilon^2 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \big(\varepsilon + \lambda^{j - \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor + 1}\big)\big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\big|\Big|^4 \\ &\leq C\varepsilon^6 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\big|\Big|^4 \\ &\quad + C\varepsilon^2 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \lambda^{j - \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor + 1}\big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\big|\Big|^4. \end{aligned} \tag{4.44}$$

We have used the elementary inequality $(a+b)^4 \leq 8(a^4+b^4)$ for $a, b \in \mathbb{R}$ in the second inequality of (4.44). By the Hölder inequality, for $1 \leq n \leq T/\varepsilon^{(l-1)/l}$,

$$\begin{aligned} &\varepsilon^6 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\big|\Big|^4 \\ &\leq \varepsilon^6 \Big(m\Big\lfloor \frac{n-1}{\varepsilon^{1/l}} \Big\rfloor\Big)^3 \Big(\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m E\Big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\Big|^4\Big) \leq C\varepsilon^2. \end{aligned} \tag{4.45}$$

Also by the Hölder inequality,

$$\begin{aligned} &\varepsilon^2 E\Big|\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \lambda^{j - \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor + 1}\big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\big|\Big|^4 \\ &\leq \varepsilon^2 \Big(\sum_{j=\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \lambda^{j - \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor + 1}\Big)^3 \Big(\sum_{j=\lfloor \frac{n}{\varepsilon l 1/l} \rfloor}^{\lfloor \frac{2n-1}{\varepsilon^{1/l}} \rfloor - 1} \sum_{i=1}^m \lambda^{j - \lfloor \frac{n}{\varepsilon^{1/l}} \rfloor + 1} E\Big|E(X(j,i)|\mathcal{F}^X_{\lfloor \frac{n}{\varepsilon^{1/l}} \rfloor})\Big|^4\Big) \leq C\varepsilon^2, \end{aligned} \tag{4.46}$$

where C is a constant independent of ε and n. Thus, (4.42) follows from (4.44), (4.45), and (4.46). $\square$

Remark 4.17. (i) As a direct consequence of (4.42), there is a constant C independent of ε and n such that

$$E\Big|\sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n})\Big|^2 \le C\varepsilon. \tag{4.47}$$

(ii) Similar to the proof of Lemma 4.16, we can show that

$$E\Big|E\Big(Z_{\varepsilon,2n}+Z_{\varepsilon,2n+1}\Big|\widetilde{\mathcal{F}}^{\varepsilon}_{n}\Big)\Big|^2 \le C\sqrt{\varepsilon}\sum_{k=\lfloor(2n-1)/\varepsilon^{1/l}\rfloor}^{\lfloor(2n+1)/\varepsilon^{1/l}\rfloor-1}\Big(\varepsilon+\lambda^{k-\lfloor n/\varepsilon^{1/l}\rfloor+1}\Big). \tag{4.48}$$

Proof of Proposition 4.6. (i) By Chebyshev inequality and (4.42) we have

$$\begin{aligned}
&P\Big(\sup_{0\le t\le T}\Big|\sum_{j=\lfloor t/\varepsilon^{(l-1)/l}\rfloor+1}^{2\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}\Big)\Big| \ge \varepsilon^{\theta}\Big)\\
&= P\Big(\sup_{0\le k\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\sum_{j=k+1}^{2k-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{k}\Big)\Big|^4 \ge \varepsilon^{4\theta}\Big)\\
&\le \varepsilon^{-4\theta}\sum_{k=0}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} E\Big|\sum_{j=k+1}^{2k-1} E\Big(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{k}\Big)\Big|^4\\
&\le C\varepsilon^{1+\frac{1}{l}-4\theta}.
\end{aligned} \tag{4.49}$$

(ii) By the Chebyshev inequality and (4.48),

$$
\begin{aligned}
&P\Big(\sup_{0\le t\le T}\Big|\sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j}+Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}_j^{\varepsilon}\Big)\Big|\ge\varepsilon^{\theta}\Big)\\
&\le \varepsilon^{-\theta}E\Big\{\sup_{0\le t\le T}\Big|\sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big(Z_{\varepsilon,2j}+Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}_j^{\varepsilon}\Big)\Big|\Big\}\\
&\le \varepsilon^{-\theta}\sum_{j=1}^{\lfloor t/\varepsilon^{(l-1)/l}\rfloor-1} E\Big|E\Big(Z_{\varepsilon,2j}+Z_{\varepsilon,2j+1}|\widetilde{\mathcal{F}}_j^{\varepsilon}\Big)\Big|\\
&\le C\varepsilon^{\frac{1}{2}-\theta}\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor-1}\ \sum_{k=\lfloor(2j-1)/\varepsilon^{1/l}\rfloor}^{\lfloor(2j+1)/\varepsilon^{1/l}\rfloor-1}\Big(\varepsilon+\lambda^{k-\lfloor j/\varepsilon^{1/l}\rfloor}\Big)\\
&\le C\varepsilon^{\frac{1}{2}-\theta}\Big(C+\sum_{j=1}^{\infty}\lambda^{(j-1)/\varepsilon^{1/l}}\Big)\\
&\le C\varepsilon^{\frac{1}{2}-\theta}.
\end{aligned}
\tag{4.50}
$$

Hence, (4.11) is proved.

(iii) To proceed, we use Lemma 4.13 to carry out certain estimates. To apply the lemma, denote $U_j=\sqrt{\varepsilon}\sum_{i=1}^{m}[I(\alpha_j^{\varepsilon}=i)-\nu_i]X(j,i)$, $A=\frac{\varepsilon^{\theta}}{3}$, $n=\Big\lfloor\frac{1}{\varepsilon^{1/l}}\Big\rfloor+1$, $N=\Big\lfloor\frac{T}{\varepsilon^{(l-1)/l}}\Big\rfloor$ and $k_{\iota}=\Big\lfloor\frac{\iota}{\varepsilon^{1/l}}\Big\rfloor$ for $\iota=0,\dots,N$. Then $\{U_j\}$ is a ϕ-mixing sequence with the ϕ^{ε} defined in Remark 4.1. Note that $\sup_{j,i}E|X(j,i)|^4<\infty$, so there exists a constant C such that $E|U_j|^4<C\varepsilon^2$ for each $j=1,2,\dots$ By Chebyshev inequality,

$$
P\Big(\max_{0\le j\le\lfloor T/\varepsilon\rfloor}\big|U_j\big|\ge\varepsilon^{\theta}\Big)\le\sum_{j=0}^{\lfloor T/\varepsilon\rfloor}P\Big(\big|U_j\big|\ge\varepsilon^{\theta}\Big)\le\sum_{j=0}^{\lfloor T/\varepsilon\rfloor}\varepsilon^{-4\theta}E\big|U_j\big|^4\le C\varepsilon^{1-4\theta}. \tag{4.51}
$$

For k, $\iota=1,2,\dots$, denote

$$
S_k^{\varepsilon}(\iota)=\sum_{j=k}^{k+\iota-1}U_j=\sqrt{\varepsilon}\sum_{j=k}^{k+\iota-1}\sum_{i=1}^{m}[I(\alpha_j^{\varepsilon}=i)-\nu_i]X(j,i).
$$

By Lemma 4.14, $E|S_{k_j}^{\varepsilon}(n)|^4\le Cn^2\varepsilon^2$. Since $n=\Big\lfloor\frac{1}{\varepsilon^{1/l}}\Big\rfloor+1$, by Chebyshev inequality,

$$
P\Big(\big|S_{k_j}^{\varepsilon}(n)\big|\ge\frac{\varepsilon^{\theta}}{3}\Big)\le 81\varepsilon^{-4\theta}E\big|S_{k_j}^{\varepsilon}(n)\big|^4\le C\varepsilon^{2-4\theta-\frac{2}{l}}.
$$

Similarly, for $\iota = 1, 2, \ldots, n$,

$$P\Big(\big|S^{\varepsilon}_{k_j}(n) - S^{\varepsilon}_{k_j}(\iota)\big| \geq \frac{\varepsilon^{\theta}}{3}\Big) \leq C\varepsilon^{2-4\theta-\frac{2}{l}}.$$

This implies that there exists a constant C independent of ε, θ, l such that

$$\max_{1\leq j\leq N}\max_{1\leq \iota\leq n} P\Big(\big|S^{\varepsilon}_{k_j}(n) - S^{\varepsilon}_{k_j}(\iota)\big| \geq \frac{\varepsilon^{\theta}}{3}\Big) \leq C\varepsilon^{2-4\theta-\frac{2}{l}}.$$

Hence, if $2 - 4\theta - \frac{2}{l} > 0$, we can choose ε small enough so that

$$\max_{1\leq j\leq N}\max_{1\leq \iota\leq n} P\Big(\big|S^{\varepsilon}_{k_j}(n) - S^{\varepsilon}_{k_j}(\iota)\big| \geq \frac{\varepsilon^{\theta}}{3}\Big) \leq \frac{1}{4}.$$

By Remark 4.1, $\phi^{\varepsilon}(p) < 1/4$ for some fixed large integer p independent of ε. Therefore, the condition of Lemma 4.13 is satisfied with p, $\eta = \frac{1}{2}$ and k_{ι}, N, n, A are defined above. Hence, according to Lemma 4.13 with $\eta = \frac{1}{2}$,

$$\begin{aligned}
&P\Big(\max_{1\leq k\leq \lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \max_{1\leq j\leq \lfloor \frac{1}{\varepsilon^{1/l}}\rfloor+1}\Big|S^{\varepsilon}_{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor}(j)\Big| \geq \varepsilon^{\theta}\Big)\\
&\leq 2\sum_{k=1}^{\lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor} P\Big(\Big|S^{\varepsilon}_{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor}\Big(\lfloor \frac{1}{\varepsilon^{1/l}}\rfloor + 1\Big)\Big| \geq \frac{\varepsilon^{\theta}}{3}\Big)\\
&\quad +2P\Big(\max_{1\leq k\leq \lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor}\ \max_{1\leq j\leq \lfloor \frac{1}{\varepsilon^{1/l}}\rfloor+1}\big|U_{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor+j}\big| \geq \frac{\varepsilon^{\theta}}{3(p-1)}\Big)\\
&= 2\sum_{k=1}^{\lfloor \frac{T}{\varepsilon^{(l-1)/l}}\rfloor} P\Big(\Big|S^{\varepsilon}_{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor}\Big(\lfloor \frac{1}{\varepsilon^{1/l}}\rfloor + 1\Big)\Big| \geq \frac{\varepsilon^{\theta}}{3}\Big) + 2P\Big(\max_{1\leq j\leq \lfloor \frac{T}{\varepsilon}\rfloor}\big|U_j\big| \geq \frac{\varepsilon^{\theta}}{3(p-1)}\Big)\\
&\leq 2\Big\lfloor \frac{T}{\varepsilon^{(l-1)/l}}\Big\rfloor C\varepsilon^{2-4\theta-\frac{2}{l}} + 2C\varepsilon^{1-4\theta}\\
&\leq C\varepsilon^{1-4\theta-\frac{1}{l}}.
\end{aligned} \tag{4.52}$$

In (4.52), we have used (4.51) in the third inequality. Therefore,

$$P\Big(\max_{0\leq t\leq T}\Big|X_{\varepsilon,\Big\lfloor \frac{\lfloor t/\varepsilon^{(l-1)/l}\rfloor}{\varepsilon^{1/l}}\Big\rfloor} - X_{\varepsilon,\lfloor t/\varepsilon\rfloor}\Big| \geq \varepsilon^{\theta}\Big) \leq C\varepsilon^{1-4\theta-\frac{1}{l}}.$$

This gives (4.12). $\square$

4.3.2 Proof of Proposition 4.7

To prove Proposition 4.7, we need the following lemma.

Lemma 4.18. *Let* $\{(X(k,1),\dots,X(k,m)) : k \geq 1\}$ *be wide-sense stationary,* ϕ-*mixing sequence in* $\mathbb{R}^m$. *Assume that there exists a constant* C *such that* $\phi(n) \leq Cn^{-\frac{4}{3}(1+\beta)}$ *for all* $n \geq 1$ *and that* $EX(k,i) = 0$, $E|X(k,i)|^4 \leq 1$ $\forall k \geq 1, i = 1,\dots,m$. *Then there exists a constant* C *such that for each* $n \geq 1$, $1 \leq i,j \leq m$,

$$n \sum_{k>n} |EX(1,i)X(k,j)| \leq Cn^{1-\beta}. \tag{4.53}$$

Proof of Lemma 4.18. By means of the mixing inequality (2.15) with $p = \frac{4}{3}$, $q = 4$,

$$\begin{aligned} |EX(1,i)X(k,j)| &= |EX(1,i)X(k,j) - EX(1,i)EX(k,j)| \\ &\leq 2\phi^{3/4}(k-1)\|X(1,i)\|_{4/3}\|X(k,j)\|_4 \\ &\leq Ck^{-(1+\beta)}. \end{aligned} \tag{4.54}$$

It follows that

$$n \sum_{k>n} |EX(1,i)X(k,j)| \leq Cn \sum_{k>n} k^{-(1+\beta)} \leq Cn^{1-\beta}.$$

This implies (4.53). $\square$

Now we are in a position to prove Proposition 4.7.

Proof of Proposition 4.7. We have

$$\begin{aligned}
&\frac{1}{n}E\Big[\sum_{i=1}^{m}\sum_{k=1}^{n}[I(\alpha_k^\varepsilon = i) - \nu_i]X(k,i)\Big]^2 \\
&= \frac{1}{n}\sum_{i=1}^{m}E\Big[\sum_{k=1}^{n}[I(\alpha_k^\varepsilon = i) - \nu_i]X(k,i)\Big]^2 \\
&\quad + \frac{2}{n}\sum_{1\le i<j\le m}E\Big\{\Big[\sum_{k=1}^{n}[I(\alpha_k^\varepsilon = i) - \nu_i]X(k,i)\Big]\Big[\sum_{p=1}^{n}[I(\alpha_p^\varepsilon = j) - \nu_j]X(p,j)\Big]\Big\} \\
&= I(n,\varepsilon) + J(n,\varepsilon).
\end{aligned} \tag{4.55}$$

Step a. Compute $I(n,\varepsilon)$. For each $i \in \mathcal{M}$ denote

$$\begin{aligned}
I_i(n,\varepsilon) &= \frac{1}{n}\sum_{k=1}^{n}E[I(\alpha_k^\varepsilon = i) - \nu_i]^2 EX(k,i)^2 \\
&\quad + \frac{2}{n}\sum_{1\le k<p\le n}^{n}E\Big[[I(\alpha_k^\varepsilon = i) - \nu_i][I(\alpha_p^\varepsilon = i) - \nu_i]\Big]E[X(k,i)X(p,i)] \\
&= I_i^1(n,\varepsilon) + I_i^2(n,\varepsilon).
\end{aligned} \tag{4.56}$$

Then $I(n,\varepsilon) = \sum_{i=1}^{n} I_i(n,\varepsilon)$. Since $EX^2(k,i) = EX^2(0,i)\ \forall\, k \ge 0$, by (2.14),

$$I_i^1(n,\varepsilon) = \nu_i[1-\nu_i]EX(0,i)^2 + O(\varepsilon + \frac{1}{n}). \tag{4.57}$$

By (2.13) and the stationarity of the sequence $\{X(n,i)\}$,

$$I_i^2(n,\varepsilon) = \frac{2}{n}\sum_{1\le k<p\le n}\Big[\psi_{ii}(p-k)\nu_i + O(\varepsilon + \lambda^k)\Big]E[X(0,i)X(p-k,i)]. \tag{4.58}$$

Since $\sum_{k=0}^{\infty}\lambda^k < \infty$ and $EX(0,i)X(q,i) = O(q^{-1-\beta})$ (by (4.54)),

$$\begin{aligned}
\frac{2}{n}\sum_{1\le k<p\le n}O(\varepsilon + \lambda^k)E[X(0,i)X(p-k,i)] &= \frac{2}{n}\sum_{q=1}^{n-1}\Big[E[X(0,i)X(q,i)]\sum_{k=1}^{n-q}O(\varepsilon+\lambda^k)\Big] \\
&= \frac{2}{n}\sum_{q=1}^{n-1}\Big[O(q^{-1-\beta})\Big((n-q)O(\varepsilon) + O(1)\Big)\Big] \\
&= O(\varepsilon + \frac{1}{n}).
\end{aligned} \tag{4.59}$$

We have used the fact that $\sum_{q=1}^{\infty} q^{-1-\beta} < \infty$ in the last identity. Next, for the first term on the right-hand side of (4.58), we have

$$\begin{aligned}
&\frac{2}{n}\sum_{1\le k<p\le n}\psi_{ii}(p-k)\nu_i E[X(0,i)X(p-k,i)] \\
&\quad=\sum_{k=1}^{n-1}\frac{2(n-k-1)}{n}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] \\
&\quad=2\sum_{k=1}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)]-2\sum_{k=n}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] \\
&\qquad-\sum_{k=1}^{n-1}\frac{2(k+1)}{n}\psi_{ii}(k)\nu_i O(k^{-1-\beta}).
\end{aligned} \tag{4.60}$$

Note that $\nu_i < 1$ and $\psi_{ii}(n)$ is uniformly bounded, so by Lemma 4.18,

$$\sum_{k=n}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] = O(n^{-\beta}).$$

On the other hand, by the uniform boundedness of $\psi_{ii}(k)$ again,

$$\sum_{k=1}^{n-1}\frac{2(k+1)}{n}\psi_{ii}(k)\nu_i O(k^{-1-\beta}) = \frac{1}{n}\sum_{k=1}^{n-1}\psi_{ii}(k)\nu_i O(k^{-\beta}) = \begin{cases} O(n^{-1}), & \text{if } \beta > 1 \\ O(n^{-1}\log n), & \text{if } \beta = 1 \\ O(n^{-\beta}), & \text{if } \beta < 1. \end{cases} \tag{4.61}$$

In view of (4.60) and (4.61), for $\bar{\beta} < \min\{1, \beta\}$,

$$\frac{2}{n}\sum_{1\le k<p\le n}\psi_{ii}(p-k)\nu_i E[X(0,i)X(p-k,i)] = 2\sum_{k=1}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] + O(n^{-\bar{\beta}}). \tag{4.62}$$

By (4.58), (4.59), and (4.62),

$$I_i^2(n,\varepsilon) = 2\sum_{k=1}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] + O(\varepsilon + n^{-\bar{\beta}}).$$

This equation, (4.56), (4.57), and $\bar{\beta} < 1$ yield

$$I_i(n,\varepsilon) = \nu_i[1-\nu_i]EX(0,i)^2 + 2\sum_{k=1}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)] + O(\varepsilon + n^{-\bar{\beta}}).$$

Therefore,

$$I(n,\varepsilon) = \sum_{i=1}^{m}\Big[\nu_i[1-\nu_i]EX(0,i)^2 + 2\sum_{k=1}^{\infty}\psi_{ii}(k)\nu_i E[X(0,i)X(k,i)]\Big] + O(\varepsilon + n^{-\bar{\beta}}). \tag{4.63}$$

Step b. Compute $J(n,\varepsilon)$. Denote

$$J_{ij}(n,\varepsilon) = \frac{2}{n}E\Big[\Big(\sum_{k=1}^{n}[I(\alpha_k^\varepsilon = i) - \nu_i]X(k,i)\Big)\Big(\sum_{p=1}^{n}[I(\alpha_p^\varepsilon = j) - \nu_j]X(p,j)\Big)\Big].$$

Then $J(n,\varepsilon) = \sum_{1\le i<j\le m} J_{ij}(n,\varepsilon)$. We can write $J_{ij}(n,\varepsilon) = J_{ij}^1(n,\varepsilon) + J_{ij}^2(n,\varepsilon) + J_{ij}^3(n,\varepsilon)$, where

$$\begin{aligned}
J_{ij}^1(n,\varepsilon) &= \frac{2}{n}\sum_{1\le k<p\le n} E\big[\big(I(\alpha_k^\varepsilon = i)-\nu_i\big)\big(I(\alpha_p^\varepsilon = j)-\nu_j\big)\big]E[X(k,i)X(p,j)],\\
J_{ij}^2(n,\varepsilon) &= \frac{2}{n}\sum_{1\le p<k\le n} E\big[\big(I(\alpha_k^\varepsilon = i)-\nu_i\big)\big(I(\alpha_p^\varepsilon = j)-\nu_j\big)\big]E[X(k,i)X(p,j)],\\
J_{ij}^3(n,\varepsilon) &= \frac{2}{n}\sum_{k=1}^{n} E\big[\big(I(\alpha_k^\varepsilon = i)-\nu_i\big)\big(I(\alpha_k^\varepsilon = j)-\nu_j\big)\big]E[X(k,i)X(k,j)].
\end{aligned}$$

Similar to step a,

$$\begin{aligned}
J_{ij}^1(n,\varepsilon) &= 2\sum_{k=1}^{\infty}\nu_i\psi_{ij}(k)E[X(0,i)X(k,j)] + O(\varepsilon + n^{-\bar{\beta}}),\\
J_{ij}^2(n,\varepsilon) &= 2\sum_{k=1}^{\infty}\nu_j\psi_{ji}(k)E[X(k,i)X(0,j)] + O(\varepsilon + n^{-\bar{\beta}}),\\
J_{ij}^3(n,\varepsilon) &= -2\nu_i\nu_j E[X(0,i)X(0,j)] + O(\varepsilon + n^{-1}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
J(n,\varepsilon) = \sum_{1\le i<j\le m}\Bigg\{ & 2\sum_{k=1}^{\infty}\Big[\nu_i\psi_{ij}(k)EX(0,i)X(k,j) + \nu_j\psi_{ji}(k)EX(k,i)X(0,j)\Big]\\
& -2\nu_i\nu_j EX(0,i)X(0,j)\Bigg\} + O(\varepsilon + n^{-\bar{\beta}}).
\end{aligned} \tag{4.64}$$

From (4.63), (4.64), and (4.55), we obtain (4.17). Note that all the constants involved in $O(\cdot)$ depend only on β and the constant given by (4.53). Thus the proposition is proved. □

4.3.3 Proof of Proposition 4.8

We first establish three lemmas. The first lemma is a consequence of Lemma 4.13; see also [22]. The proof is omitted.

Lemma 4.19. *Let $\{U_k, k \geq 1\}$ be a ϕ-mixing sequence with $0 < \eta < 1$. Suppose that there exist an integer p with $1 \leq p \leq n$ and a number $A > 0$ such that*

$$\phi(p) + \max_{p \leq i \leq n} P\Big(|S_n - S_i| \geq A\Big) \leq \eta. \tag{4.65}$$

Then, for any $a \geq 0$ and $b \geq 0$, we have

$$\begin{aligned} &P\Big(\max_{1 \leq i \leq n} |S_i| \geq a + A + b\Big) \\ &\leq \frac{1}{1-\eta} P\Big(|S_n| \geq a\Big) + \frac{1}{1-\eta} P\Big(\max_{1 \leq i \leq n} |U_i| \geq \frac{b}{p-1}\Big). \end{aligned} \tag{4.66}$$

Lemma 4.20. *There exists a constant C such that for all $\theta > 0$,*

$$P\Big(\max_{0 \leq j \leq \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| Z_{\varepsilon,j}^2 - EZ_{\varepsilon,j}^2 \Big| \geq \varepsilon^\theta\Big) \leq C\varepsilon^{1-\frac{1}{l}-2\theta}. \tag{4.67}$$

Proof of Lemma 4.20. According to (4.41), $EZ_{\varepsilon,j}^4 \leq C\varepsilon^{2-\frac{2}{l}}$. Thus,

$$\begin{aligned} P\Big(\max_{0 \leq j \leq \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \Big| Z_{\varepsilon,j}^2 - EZ_{\varepsilon,j}^2 \Big| \geq \varepsilon^\theta\Big) &\leq \sum_{j=0}^{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} P\Big(\Big| Z_{\varepsilon,j}^2 - EZ_{\varepsilon,j}^2 \Big| \geq \varepsilon^\theta\Big) \\ &\leq \varepsilon^{-2\theta} \sum_{j=0}^{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} E\Big| Z_{\varepsilon,j}^2 - EZ_{\varepsilon,j}^2 \Big|^2 \\ &\leq \varepsilon^{-2\theta} \sum_{j=0}^{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} EZ_{\varepsilon,j}^4 \leq C\varepsilon^{1-\frac{1}{l}-2\theta}. \end{aligned} \tag{4.68}$$

The proof is completed. □

Lemma 4.21. *Let $\{V_n\}$ be a ϕ-mixing sequence satisfying $\phi(n) \le \frac{C}{n}$ and $EV_n^4 \le C^2$ for all $n \ge 1$. Denote*

$$T_m(n) = \sum_{k=m+1}^{m+n} \big(V_k^2 - EV_k^2\big), \quad \widetilde{\tau}(n) = \sup_m \|T_m(n)\|_2,$$

where $\|\cdot\|_2$ denotes the L_2 norm. Then for each $\rho > 0$, there exists a constant $K = K(C,\rho)$ such that $\widetilde{\tau}(n) \le K\sqrt{n}\big(\log 2n\big)^{3+\rho}$.

Proof of Lemma 4.21. The proof here is similar to that of [17, Theorem 9.1.1]. Choose $d = \lfloor 2n \big/ \big(\log 2n\big)^{2+2\rho} \rfloor$. By the triangle inequality,

$$\|T_m(n)\|_2 \le \Big\|T_m\big(\lfloor\frac{n}{2}\rfloor\big) + T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big\|_2 + 2\widetilde{\tau}(d) + 2\widetilde{\tau}(1). \tag{4.69}$$

By the definition of $\widetilde{\tau}(n)$,

$$\Big\|T_m\big(\lfloor\frac{n}{2}\rfloor\big) + T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big\|_2^2 \le 2\widetilde{\tau}^2(\lfloor\frac{n}{2}\rfloor) + 2E\Big[T_m\big(\lfloor\frac{n}{2}\rfloor\big)T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big]. \tag{4.70}$$

Since $ET_m\big(\lfloor\frac{n}{2}\rfloor\big) = ET_{m+\lfloor n/2\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big) = 0$, by the mixing inequality (2.15),

$$\Big|E\Big[T_m\big(\lfloor\frac{n}{2}\rfloor\big)T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big]\Big| \le 2\sqrt{\varphi(d)}\Big\|T_m\big(\lfloor\frac{n}{2}\rfloor\big)\Big\|_2\Big\|T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big\|_2. \tag{4.71}$$

Since

$$\begin{aligned}
\Big\|T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\tfrac{n}{2}\rfloor\big)\Big\|_2 &= \Big\|\sum_{k=m+\lfloor n/2\rfloor+d+1}^{m+2\lfloor n/2\rfloor+d} (V_k^2 - EV_k^2)\Big\|_2 \\
&\le \sum_{k=m+\lfloor n/2\rfloor+d+1}^{m+2\lfloor n/2\rfloor+d} \big\|V_k^2 - EV_k^2\big\|_2 = \sum_{k=m+\lfloor n/2\rfloor+d+1}^{m+2\lfloor n/2\rfloor+d} \Big(EV_k^4 - \big(EV_k^2\big)^2\Big)^{\frac{1}{2}} \\
&\le \sum_{k=m+\lfloor n/2\rfloor+d+1}^{m+2\lfloor n/2\rfloor+d} \Big(EV_k^4\Big)^{\frac{1}{2}} \le C\Big\lfloor\frac{n}{2}\Big\rfloor,
\end{aligned} \tag{4.72}$$

from (4.71) we have

$$\Big|E\Big[T_m\big(\lfloor\frac{n}{2}\rfloor\big)T_{m+\lfloor\frac{n}{2}\rfloor+d}\big(\lfloor\frac{n}{2}\rfloor\big)\Big]\Big| \le 2\sqrt{\varphi(d)}C\Big\lfloor\frac{n}{2}\Big\rfloor\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big). \tag{4.73}$$

By (4.69), (4.70), and (4.73),

$$\begin{aligned}
\|T_m(n)\|_2 &\le 2\widetilde{\tau}(d)+2\widetilde{\tau}(1)+\Big[2\widetilde{\tau}^2\big(\lfloor\frac{n}{2}\rfloor\big)+4\sqrt{\varphi(d)}C\Big\lfloor\frac{n}{2}\Big\rfloor\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)\Big]^{\frac{1}{2}}\\
&\le 2\widetilde{\tau}(d)+2\widetilde{\tau}(1)+\Big[2\widetilde{\tau}^2\big(\lfloor\frac{n}{2}\rfloor\big)+4C\sqrt{\frac{C}{d}}\Big\lfloor\frac{n}{2}\Big\rfloor\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)\Big]^{\frac{1}{2}}\\
&\le 2\widetilde{\tau}(d)+2\widetilde{\tau}(1)+\sqrt{2}\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)+C\sqrt{\frac{2C}{d}\Big\lfloor\frac{n}{2}\Big\rfloor}\\
&\le 2\widetilde{\tau}(d)+2\widetilde{\tau}(1)+\sqrt{2}\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)+\sqrt{\frac{C(\log 2n)^{2+2\rho}}{n}\frac{C}{2}n}\\
&= 2\widetilde{\tau}\Big(\Big\lfloor\frac{2n}{\big(\log 2n\big)^{2+2\rho}}\Big\rfloor\Big)+2\widetilde{\tau}(1)+\sqrt{2}\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)+C^{3/2}\sqrt{n}\frac{(\log 2n)^{1+\rho}}{2}.
\end{aligned}$$

Thus,

$$\|T_m(n)\|_2 \le 2\widetilde{\tau}\Big(\Big\lfloor\frac{2n}{\big(\log 2n\big)^{2+2\rho}}\Big\rfloor\Big)+2\widetilde{\tau}(1)+\sqrt{2}\widetilde{\tau}\big(\lfloor\frac{n}{2}\rfloor\big)+C^{3/2}\sqrt{n}\frac{(\log 2n)^{1+\rho}}{2}. \tag{4.74}$$

By induction, we can show from (4.74) that there exists a constant K such that

$$\widetilde{\tau}(n) \le K\sqrt{n}\big(\log 2n\big)^{3+\rho}.$$

The proof is completed. □

Proof of Proposition 4.8. The proof is divided into several steps. (i) Note that $\{\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1}) : j=1,2,\ldots\}$ is a martingale difference sequence. Thus, by

Burkholder's inequality,

$$
\begin{aligned}
&P\Big(\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\Big[\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big]\Big|\ge\varepsilon^{\theta}\Big)\\
&\le \varepsilon^{2(\frac{l-1}{l}-\theta)}E\Big(\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\sum_{j=1}^{k}\Big[\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big]\Big|\Big)^2\\
&\le C\varepsilon^{2(\frac{l-1}{l}-\theta)}E\Big(\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big[\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big]^2\Big)\\
&\le C\varepsilon^{2(\frac{l-1}{l}-\theta)}2\Big(E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\tau_{\varepsilon,j}\Big|^2+E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big|^2\Big).
\end{aligned}
\tag{4.75}
$$

By Jensen's inequality, $E\Big|\tau_{\varepsilon,j}\Big|^2\ge E\Big|E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big|^2$. Thus, by (4.19) and (4.75),

$$
\begin{aligned}
&P\Big(\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\varepsilon^{(l-1)/l}\sum_{j=1}^{k}\Big[\tau_{\varepsilon,j}-E(\tau_{\varepsilon,j}|\mathcal{G}^{\varepsilon}_{j-1})\Big]\Big|\ge\varepsilon^{\theta}\Big)\\
&\le C\varepsilon^{-2\theta}2^2E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\varepsilon^{(l-1)/l}\tau_{\varepsilon,j}\Big|^2\le 4C\varepsilon^{-2\theta}E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|Y_{\varepsilon,j}\Big|^4.
\end{aligned}
\tag{4.76}
$$

Recall that

$$
Y_{\varepsilon,n}=Z_{\varepsilon,n}+\sum_{j=n+1}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n})-\sum_{j=n}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1}).
$$

So, by the Hölder inequality,

$$
E|Y_{\varepsilon,n}|^4\le C\Big(E|Z_{\varepsilon,n}|^4+E\Big|\sum_{j=n+1}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n})\Big|^4+E\Big|\sum_{j=n}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1})\Big|^4\Big),\tag{4.77}
$$

where $C=3^3$ is independent of ε.

By virtue of (4.42),

$$
E\Big|\sum_{j=n+1}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n})\Big|^4+E\Big|\sum_{j=n}^{2n-1}E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1})\Big|^4\le C\varepsilon^2.
$$

In addition, by (4.41), $E|Z_{\varepsilon,n}|^4\le C\varepsilon^{2(1-\frac{1}{l})}$. Hence, from (4.77) we have $E\Big|Y_{\varepsilon,n}\Big|^4\le C\varepsilon^{2(1-\frac{1}{l})}$, where C is a constant independent of ε. This implies that

$$
C\varepsilon^{-2\theta}E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|Y_{\varepsilon,j}\Big|^4\le C\varepsilon^{-2\theta}E\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\varepsilon^{2(1-\frac{2}{l})}=C\varepsilon^{1-2\theta-1/l}.\tag{4.78}
$$

Noting $l > 3$, if $\theta < \frac{1}{2} - \frac{1}{2l}$, then $1 - 2\theta - \frac{1}{l} > 0$. Thus, similar to (4.14) by Borel-Cantelli lemma, we obtain (4.21).

(ii) (4.22) can be proved by the similar argument to (i).

(iii) Denote $Y^*_{\varepsilon,n} = Y_{\varepsilon,n} - Z_{\varepsilon,n}$. Then

$$Y^*_{\varepsilon,n} = \sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n}) - \sum_{j=n}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1})$$

and

$$\sum_{n=1}^{k}[Y^2_{\varepsilon,n} - Z^2_{\varepsilon,n}] = \sum_{n=1}^{k}[(Z_{\varepsilon,n} + Y^*_{\varepsilon,n})^2 - Z^2_{\varepsilon,n}] = \sum_{n=1}^{k}[(Y^*_{\varepsilon,n})^2 + 2Y^*_{\varepsilon,n}Z_{\varepsilon,n}].$$

Thus,

$$\begin{aligned} &\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{n=1}^{k}[Y^2_{\varepsilon,n} - Z^2_{\varepsilon,n}]\Big| \\ &\le \max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{n=1}^{k}(Y^*_{\varepsilon,n})^2\Big| + 2 \max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{n=1}^{k} Y^*_{\varepsilon,n}Z_{\varepsilon,n}\Big| \\ &\le \sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} (Y^*_{\varepsilon,n})^2 + 2 \sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|Y^*_{\varepsilon,n}Z_{\varepsilon,n}\Big| \end{aligned} \tag{4.79}$$

In view of (4.47),

$$E\Big|\sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n})\Big|^2 \le C\varepsilon, \quad E\Big|\sum_{j=n}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1})\Big|^2 \le C\varepsilon.$$

Therefore,

$$E\big|Y^*_{\varepsilon,n}\big|^2 = E\Big|\sum_{j=n+1}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n}) - \sum_{j=n}^{2n-1} E(Z_{\varepsilon,j}|\widetilde{\mathcal{F}}^{\varepsilon}_{n-1})\Big|^2 \le C\varepsilon,$$

and

$$E \sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \big|Y^*_{\varepsilon,n}\big|^2 \le C\varepsilon^{1/l}. \tag{4.80}$$

Next, by (4.41), $EZ_{\varepsilon,n}^2 \le C\varepsilon^{(l-1)/l}$. Thus, the Cauchy-Schwartz inequality yields

$$\sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} E\big|Y_{\varepsilon,n}^* Z_{\varepsilon,n}\big| \le \sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big(E\big|Y_{\varepsilon,n}^*\big|^2\Big)^{\frac{1}{2}} (EZ_{\varepsilon,n}^2)^{\frac{1}{2}} \le C \sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \varepsilon^{\frac{1}{2}+\frac{l-1}{2l}} = C\varepsilon^{\frac{1}{2l}}. \tag{4.81}$$

Then (4.79), (4.80), (4.81), and the Chebyshev inequality lead to

$$\begin{aligned} &P\Big(\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{n=1}^{k} \big[Y_{\varepsilon,n}^2 - Z_{\varepsilon,n}^2\big]\Big| \ge \varepsilon^{\theta}\Big) \\ &\le \varepsilon^{-\theta} E\Big[\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{n=1}^{k} \big[Y_{\varepsilon,n}^2 - Z_{\varepsilon,n}^2\big]\Big|\Big] \\ &\le \varepsilon^{-\theta} E\Big[\sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \big|Y_{\varepsilon,n}^*\big|^2 + 2\sum_{n=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \big|Y_{\varepsilon,n}^* Z_{\varepsilon,n}\big|\Big] \\ &\le C\varepsilon^{\frac{1}{2l}-\theta}. \end{aligned} \tag{4.82}$$

The bound in (4.82) and the Borel-Cantelli lemma imply that for $\theta < \frac{1}{2l}$,

$$\max_{0\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{j=1}^{k} \big[Y_{\varepsilon,j}^2 - Z_{\varepsilon,j}^2\big]\Big| = o(\varepsilon^{\theta}) \quad \text{a.s.}$$

Thus (4.23) is proved.

(iv) Similar to the proofs of (i)-(iii), to prove (4.24), our main task is to estimate the following probability

$$P\Big(\max_{1\le k\le \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{j=1}^{k} \Big[Z_{\varepsilon,j}^2 - EZ_{\varepsilon,j}^2\Big]\Big| \ge \varepsilon^{\theta}\Big). \tag{4.83}$$

Observe that this probability is the left-hand side of (4.66) with $U_k = U_{\varepsilon,k} = Z_{\varepsilon,k}^2 - EZ_{\varepsilon,k}^2$, $S_k = S_{\varepsilon,k} = \sum_{i=1}^{k} U_{\varepsilon,i}$, $n = \lfloor T/\varepsilon^{(l-1)/l}\rfloor$ and $A = a = b = \varepsilon^{\theta}/3$. By virtue of Lemma 4.19 with $p = 2$ and $\eta = 1/2$, to estimate (4.83) it requires to verify (4.65) and estimate the right-hand side of (4.66).

Since $\{U_{\varepsilon,k} : k \ge 1\}$ is defined based on blocks of the sequence $\{[I(\alpha_j^{\varepsilon} = i) - \nu_i]X(k,i)\}$ with block size approximately $\lfloor 1/\varepsilon^{1/l}\rfloor$, it is also a mixing sequence with the mixing measure ϕ_U^ε smaller than that of the sequence $\{[I(\alpha_j^\varepsilon = i) - \nu_i]X(k,i)\}$. More precisely, $\phi_U^\varepsilon(k) \le \phi^\varepsilon(k(\lfloor \frac{1}{\varepsilon^{1/l}} \rfloor - 1))$. By Remark 4.2, $\phi_U^\varepsilon(2) < 1/4$ for ε small enough.

To complete verifying (4.65), we will use the notations in Lemmas 4.19 and 4.21 to prove that for sufficiently small $\varepsilon > 0$,

$$\max_{2 \le i \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} P\Big(\Big|S_{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} - S_i\Big| \ge \frac{\varepsilon^\theta}{3}\Big) \le \frac{1}{4}. \tag{4.84}$$

Denote $V_k = V_{\varepsilon,k} = \varepsilon^{-\frac{1}{2}(1-\frac{1}{l})} Z_{\varepsilon,k}$, $T_m(i) = \sum_{k=m+1}^{m+i}(V_k^2 - EV_k^2)$, $\widetilde{\tau}(i) = \sup_m \|T_m(i)\|_2$ for $m \ge 0$ and $i, k \ge 1$. By (4.41), $EV_k^4 \le C$ for some constant C independent of ε. On the other hand, the mixing condition in Lemma 4.21 follows by assumption (A), the remark after Lemma 4.19 and the fact that $\{V_k\}$ and $\{U_k\}$ have the same mixing measure. Thus, according to Lemma 4.21, for any $\rho > 0$, there exists a constant $K = K(C,\rho)$ such that

$$\widetilde{\tau}(n) \le K\sqrt{n}(\log 2n)^{3+\rho}. \tag{4.85}$$

By noting that $S_k - S_i = \varepsilon^{1-\frac{1}{l}} T_i(k-i)$, Chebyshev inequality and (4.85) with $n = \lfloor T/\varepsilon^{(l-1)/l} \rfloor$ yield,

$$\begin{aligned}
&\max_{2 \le i \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} P\Big(\Big|S_{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} - S_i\Big| \ge \frac{\varepsilon^\theta}{3}\Big) \\
&\quad \le \max_{2 \le i \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} 9\varepsilon^{-2\theta} E\Big|S_{\lfloor T/\varepsilon^{(l-1)/l} \rfloor} - S_i\Big|^2 \\
&\quad \le 9\varepsilon^{-2\theta}\varepsilon^{2-\frac{2}{l}} \max_{1 \le i \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} E\Big|T_i\Big(\Big\lfloor \frac{T}{\varepsilon^{(l-1)/l}} \Big\rfloor - i\Big)\Big|^2 \\
&\quad \le 9\varepsilon^{2-\frac{2}{l}-2\theta} \max_{1 \le i \le \lfloor T/\varepsilon^{(l-1)/l} \rfloor} \widetilde{\tau}^2\Big(\Big\lfloor \frac{T}{\varepsilon^{(l-1)/l}} \Big\rfloor - i\Big) \\
&\quad \le C\varepsilon^{2-\frac{2}{l}-2\theta} \Big\lfloor \frac{T}{\varepsilon^{(l-1)/l}} \Big\rfloor \Big[\log\Big(2\Big\lfloor \frac{T}{\varepsilon^{(l-1)/l}} \Big\rfloor\Big)\Big]^{6+2\rho} \\
&\quad = C\varepsilon^{1-\frac{1}{l}-2\theta}(\log \varepsilon)^{6+2\rho}.
\end{aligned} \tag{4.86}$$

Thus, (4.84) holds for $1-\frac{1}{l}-2\theta>0$ and sufficiently small positive ε. Since $\phi_U^\varepsilon(2)<\frac{1}{4}$ for ε small enough, (4.84) yields

$$\phi_U^\varepsilon(2)+\max_{2\le i\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor} P\Big(\Big|S_{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}-S_i\Big|\ge\frac{\varepsilon^\theta}{3}\Big)\le\frac{1}{2}, \tag{4.87}$$

i.e., (4.65) holds for $\{U_k^\varepsilon\}$ with $n=\lfloor T/\varepsilon^{(l-1)/l}\rfloor$, $A=a=b=\varepsilon^\theta/3$, $p=2$ and $\eta=1/2$.

Next, in view of Lemma 4.19 with above notations of $\{U_k^\varepsilon\}$, n, A, a, b, p, and η,

$$\begin{aligned} &P\Big(\max_{0\le k\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\sum_{j=1}^{k}\Big[Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big]\Big|\ge\varepsilon^\theta\Big)\\ &\le 2P\Big(\Big|\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big[Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big]\Big|\ge\frac{\varepsilon^\theta}{3}\Big)+2P\Big(\max_{0\le j\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big|\ge\frac{e^\theta}{3}\Big). \end{aligned} \tag{4.88}$$

Similar to (4.86),

$$P\Big(\Big|\sum_{j=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big[Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big]\Big|\ge\frac{\varepsilon^\theta}{3}\Big)\le C\varepsilon^{1-\frac{1}{l}-2\theta}(\log\varepsilon)^{6+2\rho}. \tag{4.89}$$

From (4.67),

$$P\Big(\max_{0\le j\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big|\ge\frac{e^\theta}{3}\Big)\le C\varepsilon^{1-\frac{1}{l}-2\theta}. \tag{4.90}$$

Thus, by (4.88), (4.89), and (4.90), we obtain

$$P\Big(\max_{0\le k\le\lfloor T/\varepsilon^{(l-1)/l}\rfloor}\Big|\sum_{j=1}^{k}\Big[Z_{\varepsilon,j}^2-EZ_{\varepsilon,j}^2\Big]\Big|\ge\varepsilon^\theta\Big)\le C\varepsilon^{1-\frac{1}{l}-2\theta}(\log\varepsilon)^{6+2\rho}. \tag{4.91}$$

By using the Borel-Cantelli lemma as in the proof of Proposition 4.5 we obtain (4.24) for $\theta<\frac{1}{2}-\frac{1}{2l}$.

(v) By virtue of (4.17) with $n=\varepsilon^{1/l}$, for all $\bar\beta<\min\{1,\beta\}$ there exists a constant

C that does not depend on ε and k such that for all $k \geq 1$,

$$\Big|\varepsilon^{1/l} E\Big[\sum_{j=\lfloor \frac{k-1}{\varepsilon^{1/l}}\rfloor}^{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor - 1} \sum_{i=1}^{m} [I(\alpha_j^\varepsilon = i) - \nu_i] X(j,i)\Big]^2 - \sigma^2\Big| \leq C(\varepsilon + \varepsilon^{\bar{\beta}/l}). \tag{4.92}$$

Thus, by (4.92) and the formula of $Z_{\varepsilon,k}$,

$$\begin{aligned}\Big|EZ_{\varepsilon,k}^2 - \varepsilon^{(l-1)/l}\sigma^2\Big| &= \varepsilon^{(l-1)/l}\Big|E\Big[\varepsilon^{1/l} \sum_{j=\lfloor \frac{k-1}{\varepsilon^{1/l}}\rfloor}^{\lfloor \frac{k}{\varepsilon^{1/l}}\rfloor - 1} \sum_{i=1}^{m} [I(\alpha_j^\varepsilon = i) - \nu_i] X(j,i)\Big]^2 - \sigma^2\Big| \\ &\leq C\varepsilon^{(l-1)/l}(\varepsilon + \varepsilon^{\bar{\beta}/l}).\end{aligned} \tag{4.93}$$

Since the constant C is independent of k,

$$\begin{aligned}\max_{1\leq k \leq \lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|\sum_{j=1}^{k} EZ_{\varepsilon,j}^2 - k\varepsilon^{(l-1)/l}\sigma^2\Big| &\leq \sum_{k=1}^{\lfloor T/\varepsilon^{(l-1)/l}\rfloor} \Big|EZ_{\varepsilon,k}^2 - \varepsilon^{(l-1)/l}\sigma^2\Big| \\ &\leq C(\varepsilon + \varepsilon^{\bar{\beta}/l}).\end{aligned} \tag{4.94}$$

This proves (4.25). $\square$

5 Further Remarks

This work has been devoted to limit results of a class of suitably scaled random processes modulated by a Markov chain with finite state space. The original processes are in discrete time. The limit, however, are continuous-time processes. Under simple conditions, it is demonstrated in Chapter 3 that the limits are switching diffusions. The main techniques used are weak convergence methods.

Chapter 4 has focused on strong approximation of a suitably scaled sequence of processes modulated by a Markov chain with the assumption that the Markov chain is ergodic. Corresponding to a weak convergence result of the centered and scaled sequence, it ascertains the rate of convergence by means of strong approximation. It also provides insight for application in networks and systems involving such sequences. Note that in this chapter, $\{X(k,i)\}$ is assumed to be a wide-sense stationary sequence. This condition can be relaxed; non-stationary sequences (e.g., non-stationary mixing sequences) may be treated, but more work is needed in this direction. The crucial point is to have sufficiently fast mixing rate.

For future study, Markov chains including transient states can be considered. For such cases, we will only aggregate states in each recurrent class and leave the transient states alone. Essentially the same techniques enable us to reach similar conclusions. Another worthwhile direction is to examine the convergence rates. Furthermore, one may consider large deviation type estimates and study the associated empirical measure processes, which are motivated by system identification and tracking randomly varying processes under binary-valued and quantized data.

It is conceivable the results obtained here will be useful for carrying out control and optimization tasks for Markov modulated sequences. Future work may also be directed to system identifications when the observation sequence is modeled by Markov modulated processes.

REFERENCES

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

[2] H. Brunnel and B. Kim, *Discrete-Time Models for Communication Systems Including ATM*, Kluwer, Boston, 1993.

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

[4] M. Csörgö and P. Révész, *Strong approximations in probability and statistics*, Probability and Mathematical Statistics. Academic Press Inc. [Harcourt Brace Jovanovich Publisher], New York, NY, 1981.

[5] J.L. Doob, *Stochastics Processes*, J. Wiley, New York, NY, 1953.

[6] P. Doukhan, *Mixing: Properties and Examples*, Lecturer Notes in Statistics Vol.85, Springer, New York, NY, 1994.

[7] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, J. Wiley, New York, NY, 1986.

[8] J.D. Hamilton and R. Susmel, Autoregressive conditional heteroskedasticity and changes in regime, *J. Econometrics*, **64**, 307-333, 1994.

[9] J.P. Hespanha, *Stochastic Hybrid Systems: Application to Communication Networks*, Springer, Berlin, 2004.

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

[11] P. Hall and C.C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, New York, 1980.

[12] F.C. Hoppensteadt and W.L. Miranker, Multitime methods for systems of difference equations, *Studies Appl. Math.* **56** (1977), 273-289.

[13] S. Karlin and H.M. Taylor, *A First Course in Stochastic Processes*, 2nd Edition, Academic Press, New York, NY, 1975.

[14] V. Korolykuk and A. Swishchuk, *Evolution of Systems in Random Media*, CRC Press, Boca Raton, 1995.

[15] H.J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.

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

[17] Z.Y. Liu and C.R. Lu, *Limit Theory for Mixing Dependent Random Variables*, Science Press, Kluwer Academic, New York, Dordrecht, 1996.

[18] X. Mao and C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperical College Press, London, 2006.

[19] W.A. Massey and W. Whitt, Uniform acceleration expansions for Markov chains with time-varying rates, *Ann. Appl. Probab.*, **8** (1998), 1130-1155.

[20] M. Miyazawa and H. Takagi, Advances in discrete time queues, *Queueing Systems: Theory and Applications*, **18** (1994).

[21] G.C. Papanicolaou, D. Stroock, and S.R.S. Varadhan, Martingale approach to some limit theorems, in *Proc. 1976 Duke Univ. Conf. on Turbulence*, Durham, NC, 1976.

[22] M. Peligrad, An invariance principle for ϕ-mixing sequences, *Ann. Probab.*, **13** (1985), 1304-1313.

[23] W. Philipp and W. Stout, Almost sure invariance principles for partial sums of weakly dependent random variables, *Mem. Amer. Math. Soc.*, **2** (1975), issue 2, no. 161.

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

[25] M. Woodward, *Communication and Coputer Networks: Modelling with Discrete-Time Queues*, California IEEE Computer Society Press, Los Alamitos, CA, 1994.

[26] G. Yin, V. Krishnamurthy, and C. Ion, Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization, *SIAM J. Optim.*, **14** (2004), 1187-1215.

[27] G. Yin and H. Zhang, Singularly perturbed Markov chains: Limit results and applications, *Ann. Appl. Probab.*, **17** (2007), 207-229.

[28] G. Yin and Q. Zhang, *Discrete-time Markov Chains: Two-time-scale Methods and Applications*, Springer, New York. 2005.

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

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

ABSTRACT

ASYMPTOTIC PROPERTIES OF MARKOV MODULATED SEQUENCES WITH FAST AND SLOW TIME SCALES

by

SON LUU NGUYEN

December 2010

Advisor: Dr. G. George Yin

Major: Mathematics (Applied)

Degree: Doctor of Philosophy

In this dissertation we investigate asymptotic properties of Markov modulated random processes having two-time scales. The model contains a number of mixing sequences modulated by a randomly switching process that is a discrete-time Markov chain. The motivation of our study stems from applications in manufacturing systems, communication networks, and economic systems, in which regime-switching models are used.

This dissertation focuses on asymptotic properties of the Markov modulated processes under suitable scaling. Our main effort is devoted to obtaining weak convergence and strong approximation results.

AUTOBIOGRAPHICAL STATEMENT

SON LUU NGUYEN

Education

- Ph.D. in Applied Mathematics, December 2010 (expected)
  Wayne State University, Detroit, Michigan
- M.A. in Mathematical Statistics, May, 2010
  Wayne State University, Detroit, Michigan
- B.S. in Mathematics, June 2002
  Hanoi University of Sciences, Vietnam

Awards

1. Outstanding Research Award, Department of Mathematics, Wayne State University, April 2009.

2. Paul Weiss Award in Recognition of Outstanding Achievement in the Ph.D. Program, Department of Mathematics, Wayne State University, April 2009 and April 2010.

3. The Karl W. and Helen L. Folley Endowed Mathematics Scholarship, Department of Mathematics, Wayne State University, April 2008.

4. Graduate Student Professional Travel Award, Department of Mathematics, Wayne State University, May 2007 and July 2008.

5. The Thomas C. Rumble Fellowship, Department of Mathematics, Wayne State University, August 2005 – May 2006.

List of Publications

1. G. Yin, S.L. Nguyen, L.Y. Wang, and C. Xu, Time-inhomogeneous Markov chains and ergodicity arising from nonlinear dynamic systems and optimization. *Progress in nonlinear analysis research*, Nova Sci. Publ., New York, 5–19, 2009.

2. S.L. Nguyen and G. Yin, Asymptotic properties of hybrid random processes modulated by Markov chains, to appear in *Nonlinear Analysis: Theory, Methods and Applications*, **71** (2009), e1638–e1648.

3. S.L. Nguyen and G. Yin, Asymptotic properties of Markov modulated random sequences with fast and slow time scales, to appear in *Stochastics*.

4. S.L. Nguyen and G. Yin, Weak convergence of Markov modulated random sequences, to appear in *Stochastics*.

5. S.L. Nguyen and G. Yin, Almost sure error bounds for numerical solutions of stochastic differential equations, submitted.