January 2022

Sequences Of Random Matrices Modulated By A Discrete-Time Markov Chain

Huy Nguyen
Wayne State University

SEQUENCES OF RANDOM MATRICES MODULATED BY A DISCRETE-TIME MARKOV CHAIN

by

HUY 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

2022

MAJOR: MATHEMATICS

Approved By:

____

Advisor Date

____

Advisor

____

____

____

DEDICATION

To My Family

ACKNOWLEDGEMENTS

First and foremost, I would sincerely like to express my gratitude to my advisor, Professor George Yin, for his continued guidance, encouragement, patience, understanding, and support that he offered me during the past five years. Without his generous help and infinite patience, I am afraid that I would never complete the dissertation. Personally speaking, he is a great man who is always humble. Professionally, he works hard, and gives me much to aspire. Being his student has been a great privilege in my life.

I would like to thank Professor Pei-Yong Wang. Thank you for serving as my PhD co-advisor after Professor George Yin moved to the University of Connecticut in the Fall of 2020. His help encouraged me me to stay at Wayne State University to complete the Ph.D. program. Additionally, he taught me several courses that enabled me to complete my research. I truly enjoy his teaching.

I would like to thank Professor Kazuhiko Shinki for serving as a member on my dissertation committee. He taught me several statistics courses which enlarged views about the applications of mathematics to other fields.

I would like to thank Professor Tao Huang and Professor Le Yi Wang for serving as members on my dissertation committee. I thank him for his valuable time and kind instruction in my PhD process.

I would like to show my appreciation to other Professors in the Wayne State Department of Mathematics who taught me and helped me fulfill my program requirements. Specifically, I'd like to acknowledge Prof. Ualbay Umirbayev, Prof. William Cohn, Prof. Fatih Celiker, Prof. Bertram Schreiber, Prof. Robert Bruner, Prof. Luca Candelori, Prof.

Xiaoli Kong, Prof. Shereen Schultz, and Prof. Christopher Leirstein.

I owe my deepest gratitude to Wayne State University, especially the Department of Mathematics, where I have studied and worked for five years. During the graduate study, I received much help and support from the administrative staff. Specifically, I would like to thank Ms. Barbara Malicke, Dr. Tiana Bosley, Ms. Maria Vujic, Ms. Joanne Lewan, Ms. Carla Sylvester, and Ms. Terri Renaud for their kind help. Also, I would like to show my appreciation to Prof. Hengguang Li, Chair of Mathematics Department; Prof. Daniel Isaksen, Director of Graduate program, and many other people who are dedicated to the development and strength of the Mathematics Department.

I would like to express my sincere appreciation to my friends in Vietnam, as well as in America, Mr. Yatin Patel, Dr. Son Nguyen, Prof. Thanh Le, Prof. Quang Nguyen, Mr. Randall White and his wife Mrs. Faith White, as well as Colin and Becky Mansker. I am especially thankful for their very kind support and encouragement during these past years.

Finally, I would like to thank my family for the selfless support. My parents and parents-in-law, my wife and my children have always understood and supported me during the challenges and successes of this PhD process. I am thankful for that. Thank you, and I love you all.

TABLE OF CONTENTS

Dedication ii

Acknowledgements iii

Chapter 1 Introduction 1

1.1 Recent Progress 1

1.2 Markov Modulated Sequences 3

1.3 Outline 4

Chapter 2 Stochastic Differential Equations and Markov Chains 6

2.1 Stochastic process 6

2.2 Itô Integrals 8

2.3 Itô Formula and the Martingale Representation Theorem 10

2.3.1 The Martingale Representation Theorem 12

2.4 Stochastic Differential Equations 13

2.5 Discrete-Time Markov Chains 14

2.5.1 Asymptotic Expansions 19

Chapter 3 Problem Formulation 25

3.1 Problem Formulation and Conditions 25

3.1.1 Decomposition and Subspaces 25

3.1.2 Aggregation and Interpolated Process 27

Chapter 4 Asymptotic Properties 29

Chapter 5 Specialization and Extension 43

5.1 Modulating Markov Chain with P Being Irreducible 43

5.2 Modulating Markov Chain Including Transient States 44

Chapter 6 Further Remarks and Ramification 49
6.1 A Remark on Non-Zero Drift 49
6.2 Ramification 50
6.3 Additional Remarks 55
Appendix A 57
References 60
Abstract 64
Autobiographical Statement 66

CHAPTER 1 INTRODUCTION

This dissertation is devoted to studying matrix-valued random sequences modulated by a discrete-time Markov chain. The main effort is to obtain asymptotic properties of such random processes through appropriate scaling and interpolation. In the introduction, we first review some recent progress in hybrid systems. Then we give motivation of the study of Markov modulated sequences followed by an outline of the dissertation.

1.1 Recent Progress

In recent year, much effort has been devoted to studying hybrid systems. By hybrid systems, we mean systems that include both continuous dynamics and discrete event. One class of such hybrid systems is the so-called switching diffusions in which the underlying systems have a continuous component represented by a diffusion process and a discrete component that is a switching process. The two component process may be written as $(X(t), \alpha(t))$ with $X(t)$ and $\alpha(t)$ representing the continuous and discrete states, respectively. The switching component takes values in a finite state space, say $\mathcal{M}$. For each of the discrete states $\iota \in \mathcal{M}$, $X(t)$ is a solution of a stochastic differential equation with suitable drift and diffusion coefficients depending on ι. The continuous state and discrete events coexist and interact resulting in certain properties that are not seen in either diffusions or switching processes alone. For example, putting two linear differential equations that are both stable through a switching process may result in an unstable system. Likewise, solutions of two equations that are both ergodic may result in a system that is not ergodic. Because of the distinct features, much effort has been devoted to the study of such processes. For an account of switching diffusions when the switching process is a continuous-time Markov chain independent of the Brownian motion, we refer the work of Mao and Yuan [9], when the switching process is a continuous-state-dependent process, see Yin and Zhu [28]. For applications of switched dynamic systems, we mention the works for control systems [2, 10, 15, 16], manufacturing and production planning [17, 18], communication and computer networks [21], hybrid filtering [22], discrete optimization [23], mathematical finance [24], and many references therein.

Parallel to the development of continuous-time systems, effort has also been placed on treating discrete-time systems. The main motivation is reduction of complexity. Based on such motivation, in the work [12, 13], the authors launched a study on discrete-time version of "hybrid switching diffusions". They considered multiple number of vector-valued sequences of random processes modulated by a discrete-time Markov chain and revealed the asymptotic properties of the two-component sequences by using martingale problem formulation and strong approximation using a probabilistic approach. One of main ideas that we used is decomposition and aggregation, which was considered in the work of [1]. In that reference, Courtois illustrated that one can decompose a state space into subspaces leading to a so-called nearly completely decomposable model (see also [19]). In this dissertation, we continue our quest in this direction with the emphasis on a Markov modulated sequences. Nevertheless, the primary sequences we consider are sequences of double arrays or sequences of matrices. The problem becomes more complex. Our aim is to analyze the asymptotic properties of the sequence.

1.2 Markov Modulated Sequences

Before proceeding to our study, we address a couple of questions first. These questions indicates the main features of the random sequences that we are interested in and the motivation for the study.

What are the main features of Markov-modulated sequences? Concentrating on discrete-time models, let us begin with a discrete-time Markov chain α_k with a finite state space $\mathcal{M} = \{1, \ldots, m_0\}$, and suppose that there are m_0 sequences of random variables $\{X_k(\gamma)\}_{\gamma \in \mathcal{M}}$ (in this work, these random variables are matrix-valued), where in the notation $X_k(\gamma)$, k denotes the discrete time and γ is the index the γth sequence. At any given time j, the Markov chain takes a value $\gamma \in \mathcal{M}$, the chain will sojourn in γ for some time. During this time, the sequence $\{X_j(\gamma)\}$ is activated. This sequence will remain active until the Markov chain switches to another state $\gamma_1 \in \mathcal{M}$ at some later time say $k > j$. Then the γ_1th sequence $\{X_k(\gamma_1)\}$ is activated, which will be active until a later time when the Markov chain switching to another state γ_2, and so on. We call $\{X_k(\gamma)\}$ (for $\gamma \in \mathcal{M}$) the primary sequence. The combined sequence of interest in our study can be described by $\{X_k(\alpha_k)\}$. We can also easily write $X_k(\alpha_k) = \sum_{\gamma \in \mathcal{M}} X_k(\gamma) I_{\{\alpha_k = \gamma\}}$, where I_A is the usual indicator function of the set A. The sample paths of α_k are constructed in accordance with the transition probability matrix P.

Why should we be interested in studying Markov modulated sequences? One of the primary motivations for our studies is the treatment of networked systems. In the new era, many physical, biological, and social systems are rather complex. The usual method of modeling is often inadequate. Thus various networked systems come into being. Such systems usually have complex structures and with different components and subsystems that are inter-connected and interacted. To model the inter-connections and the change of the inter-connections, we allow the configuration of the overall system to be time-varying. One way of modeling is that on top of the different components or subsystems, to assume that there is a switching device. The different components (subsystems) can be thought of as satellites around an exchange center. The exchange center dictates which sequence to be active next. The use of a discrete-time Markov chain enables the formulation of the configuration change to be done at random time. Such modeling point has become increasingly more popular in various applications; see for example, consensus formation of multi-agent systems [25], chemostat models and wastewater treatment [11], stochastic ecosystems [20], and discrete optimization [23]. Because of their importance, the study on such systems and related issues has gained resurgent interests and drawn the attention from many researchers.

1.3 Outline

With the motivation given above, this dissertation concentrates on studying matrix-valued sequence. Limit results of double array processes were contained in almost of probability textbooks. The classical work dealt with law of large numbers and related limit theories. In this dissertation, the primary sequences we consider are ϕ-mixing random matrix-valued processes with appropriate mixing rates. Our main effort is on getting the desired asymptotic results. In a certain sense, this work is inspired by [12, 13], but the primary sequence becomes matrix-valued. Similar to the aforementioned references, we assume that the modulating sequence is a discrete-time Markov chain with transition probability matrix displaying certain two-time-scale properties. Technically, comparing to [12, 13], in lieu of using perturbed test function methods, we use direct averaging methods to treat the underlying problems. Our main study is based on the use of functionals, which map the matrix-valued processes to real-valued processes. Then we proceed to study the asymptotic behavior. In addition to such an approach, near the end of the dissertation, we demonstrate how matrix-valued processes may be dealt with directly. In addition, slightly different from the rest of dissertation, the main effort in that chapter is to obtain certain exponential type of bounds of tail probabilities, which is interesting in its own right.

The rest of the dissertation is arranged as follows. In chapter 2, definitions and related properties of Wiener process, stochastic differential equations, Itô's Lemma, Martingale problem formulation and Markov chains are recalled. Chapter 3 gives the precise formulation of the problem. Chapter 4 presents the main asymptotic results. Section 5 concentrates on specializations and extensions. Chapter 6 issues some final remarks to conclude the dissertation. Finally, an appendix is placed at the end of the dissertation, which collects some results on two-time-scale Markov chains.

CHAPTER 2 STOCHASTIC DIFFERENTIAL EQUATIONS AND MARKOV CHAINS

This chapter presents some preliminaries and background materials.

2.1 Stochastic process

Definition 2.1.1. *(see [29], p.10) A stochastic process is a parametrized collection of random variables* $\{X_t\}_{t\in T}$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *and taking values in* $\mathbb{R}^n$.

The parameter space T may be a subset of $\mathbb{R}$ (an interval $[a, b]$, the non-negative integers, but usually the half-line $[0, \infty)$). For each t fixed, the X_t is a random variable $\omega \to X_t(\omega)$; $\omega \in \Omega$. On the other hand, by fixing $\omega \in \Omega$ the function $t \to X_t(\omega)$, $t \in T$ is a path of X_t. For convenience, the notation $X(t, \omega)$ sometimes is used instead of X_t. It means the process X_t can be regarded as a function of two variables $(t, \omega) \to X(t, \omega)$ from $T \times \Omega$ into $\mathbb{R}^n$. The (finite-dimensional) distributions of the process $X = \{X_t\}_{t\in T}$ are the measures $\mu_{t_1,\dots,t_k}$ defined on $\mathbb{R}^{nk}, k = 1, 2, \dots,$ by $\mu_{t_1,\dots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k]; t_i \in T$, where $F_1, \dots, F_k$ denote Borel sets in $\mathbb{R}^n$. Conversely, given a family $\{\nu_{t_1,\dots,t_k}; k \in \mathbb{N}, t_i \in T\}$ of probability measure on $\mathbb{R}^{nk}$ it is able to construct a stochastic process $Y = \{Y_t\}_{t\in T}$ having $\nu_{t_1,\dots,t_k}$ as its finite-dimensional distributions.

Theorem 2.1.2 (Kolmogorov's extension theorem)**.** *(see [29], p. 11) For all* $t_1, \dots, t_k \in T, k \in \mathbb{N}$*, let* $\nu_{t_1,\dots,t_k}$ *be probability measures on* $\mathbb{R}^{nk}$ *s.t.*

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)})$$

for all permutations σ on $\{1, 2, \ldots, k\}$ and

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n)$$

for all $m \in \mathbb{N}$. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_t\}$ on $\Omega, X_t : \Omega \to \mathbb{R}^n$, s.t.

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k],$$

for all $t_i \in T, k \in \mathbb{N}$ and all Borel sets F_i.

Fix $x \in \mathbb{R}^n$ and define $p(t, x, y) = (2\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2t}\right)$ for $y \in \mathbb{R}^n, t > 0$. For any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$, define a measure $\nu_{t_1,\cdots,t_k}$ on $\mathbb{R}^{nk}$ by

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \int_{F_1 \times \cdots \times F_k} p(t_1, x, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k$$

By Kolmogorov's theorem there exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t\}_{t \geq 0}$ on Ω such that the finite-dimensional distributions of B_t are given by

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) = \int_{F_1 \times \cdots F_k} p(t_1, x, x_1) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \ldots dx_k.$$

Definition 2.1.3. *(see [29], p. 12) Such a process is called a (version of) Brownian motion starting at x (observe that $P^x(B_0 = x) = 1$).*

Definition 2.1.4. *(see [29], p. 14) Suppose that $\{X_t\}$ and $\{Y_t\}$ are stochastic processes on $(\Omega, \mathcal{F}, P)$. Then we say that $\{X_t\}$ is a version of (or a modification of) $\{Y_t\}$ if $P(\{\omega; X_t(\omega) =$*

$Y_t(\omega)\}) = 1$ *for all* t.

Theorem 2.1.5 (Kolmogorov's continuity theorem)**.** *(see [29], p. 14) Suppose that the process* $X = \{X_t\}_{t\geq 0}$ *satisfies the following condition: For all* $T > 0$ *there exists positive constants* α, β, D *such that* $E[|X_t - X_s|^\alpha] \leq D \cdot |t-s|^{1+\beta}$*;* $0 \leq s, t \leq T$*. Then there exists a continuous version of* X*.*

2.2 Itô Integrals

Definition 2.2.1. *(see [29], p. 25) Let* $B_t(\omega)$ *be n-dimensional Brownian motion. A* σ*-algebra* $\mathcal{F}_t = \mathcal{F}_t^{(n)}$ *can be defined as the one generated by the random variables* $\{B_i(s)\}_{1\leq i\leq n, 0\leq s\leq t}$*. In other words,* $\mathcal{F}_t$ *is the smallest* σ*-algebra containing all sets of the form* $\{\omega; B_{t_1}(\omega) \in F_1, \cdots, B_{t_k}(\omega) \in F_k\}$*, where* $t_j \leq t$ *and* $F_j \subset \mathbb{R}^n$ *are Borel sets,* $j \leq k = 1, 2, \ldots$*.*

Definition 2.2.2. *(see [29], p. 25) Let* $\{\mathcal{N}_t\}_{t\geq 0}$ *be an increasing family of* σ*-algebras of subsets of* Ω*. A process* $g(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ *is called* $\mathcal{N}_t$*-adapted if for each* $t \geq 0$ *the function* $\omega \to g(t,\omega)$ *is* $\mathcal{N}_t$*-measurable.*

Definition 2.2.3. *(see [29], p. 25) Let* $\mathcal{V} = \mathcal{V}(S,T)$ *be the class of functions* $f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$ *such that*

(i) $(t,\omega) \to f(t,\omega)$ *is* $\mathcal{B} \times \mathcal{F}$*-measurable, where* $\mathcal{B}$ *denotes the Borel* σ*-algebra on* $[0,\infty)$*.*

(ii) $f(t,\omega)$ *is* $\mathcal{F}_t$*-adpated.*

(iii) $E[\int_S^T f^2(t,\omega)dt] < \infty$*.*

Definition 2.2.4 (The Itô integral)**.** *(see [29], p. 29) Let* $f \in \mathcal{V}(S,T)$*. Then the Itô integral of* f *(from* S *to* T*) is defined by* $\int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dB_t(\omega)$ *(limit in* $L^2(P)$*) where* $\{\phi_n\}$ *is a sequence of elementary functions such that* $E\left[\int_S^T (f(t,\omega) - \phi_n(t,\omega))^2 dt\right] \to$

0 as $n \to \infty$.

Corollary 2.2.5 (The Itô isometry). *(see [29], p. 29)*

$$E\left[\left(\int_S^T f(t,\omega)dB_t\right)^2\right] = E\left[\int_S^T f^2(t,\omega)dt\right] \quad \textit{for all } f \in \mathcal{V}(S,T).$$

Corollary 2.2.6. *(see [29], p. 29) If $f(t,\omega) \in \mathcal{V}(S,T)$ and $f_n(t,\omega) \in \mathcal{V}(S,T)$ for $n = 1, 2, \ldots$ and $E[\int_S^T (f_n(t,\omega) - f(t,\omega))^2 dt] \to 0$ as $n \to \infty$, then $\int_S^T f_n(t,\omega)dB_t(w) \to \int_S^T f(t,\omega)dB_t(\omega)$ in $L^2(P)$ as $n \to \infty$.*

Definition 2.2.7. *(see [29], p. 31) A filtration (on $(\Omega, \mathcal{F})$ is a family $\mathcal{M} = \{\mathcal{M}_t\}_{t\geq 0}$ of σ-algebras $\mathcal{M}_t \subset \mathcal{F}$ such that $\mathcal{M}_s \subset \mathcal{M}_t$ for $0 \leq s < t$. An n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is called a martingale with respect to a filtration $\{\mathcal{M}_t\}_{t\geq 0}$ (and with respect to P) if i) M_t is $\mathcal{M}_t$-measurable for all t, ii) $E[|M_t|] < \infty$ for all t, and iii) $E[M_s|\mathcal{M}_t] = M_t$ for all $s \geq t$.*

Theorem 2.2.8 (Doob's martingale inequality). *(see [29], p. 31) If M_t is martingale such that $t \to M_t(\omega)$ is continuous a.s., then for all $p \geq 1, T \geq 0$ and all $\lambda > 0$*

$$P[\sup_{0\leq t\leq T} |M_t| \geq \lambda] \leq \frac{1}{\lambda^p} E[|M_T|^p].$$

Theorem 2.2.9. *(see [29], p. 32) Let $f \in \mathcal{V}(0,T)$. Then there exists a t-continuous version of $\int_0^t f(s,\omega)dB_s(\omega)$; $0 \leq t \leq T$, i.e. there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, P)$ such that $P[J_t = \int_0^t f dB] = 1$ for all t, $0 \leq t \leq T$.*

Corollary 2.2.10. *(see [29], p. 33) Let $f(t,\omega) \in \mathcal{V}(0,T)$ for all T. Then $M_t(\omega) = \int_0^t f(s,\omega)dB_s$*

is a martingale w.r.t. $\mathcal{F}_t$ and

$$P[\sup_{0\le t\le T}|M_t|\ge\lambda]\le\frac{1}{\lambda^2}\cdot E\left[\int_0^T f(s,\omega)^2ds\right];\quad \lambda,T>0.$$

2.3 Itô Formula and the Martingale Representation Theorem

Definition 2.3.1 (1-dimensional Itô processes)**.** *(see [29], p. 44) Let B_t be a 1-dimensional Brownian motion on $(\Omega,\mathcal{F},P)$. A (1-dimensional) Itô process (or stochastic integral) is a stochastic process X_t on $(\Omega,\mathcal{F},P)$ of the form*

$$X_t=X_0+\int_0^t u(s,\omega)ds+\int_0^t v(s,\omega)dB_s,$$

where $v\in\mathcal{W}_{\mathcal{H}}$, so that $P\left[\int_0^t v^2(s,\omega)ds<\infty \text{ for all } t\ge 0\right]=1$, and u is $\mathcal{H}_t$-adapted i.e. $P\left[\int_0^t |u(s,\omega)|ds<\infty \text{ for all } t\ge 0\right]=1$.

Theorem 2.3.2 (The 1-dimensional Itô formula)**.** *(see [29], p. 44) Let X_t be an Itô process given by $dX_t=udt+vdB_t$, and let $g(t,x)\in C^2([0,\infty)\times\mathbb{R})$. Then $Y_t=g(t,X_t)$ is again an Itô process, and*

$$dY_t=\left(\frac{\partial g}{\partial t}(t,X_t)+u\frac{\partial g}{\partial x}(t,X_t)+\frac{v^2}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)\right)dt+v\frac{\partial g}{\partial x}(t,X_t)dB_t$$

Theorem 2.3.3 (Integration by parts)**.** *(see [29], p. 46) Suppose $f(s,\omega)$ is continuous and of bounded variation with respect to $s\in[0,t]$. Then $\int_0^t f(s)dB_s=f(t)B_t-\int_0^t B_s df_s$.*

Definition 2.3.4 (n-dimensional Itô process)**.** *(see [29], p. 48) Let $B(t,\omega)=(B_1(t,\omega),\cdots,B_m(t,\omega))$*

denote m-dimensional Brownian motion. Suppose that $v_{ij} \in \mathcal{W}_{\mathcal{H}}$ so that

$$P\left[\int_0^t v_{ij}^2(s,\omega)ds < \infty \text{ for all } t \geq 0\right] = 1$$

and u_i is $\mathcal{H}_t$-adapted i.e.

$$P\left[\int_0^t |u_i(s,\omega)|ds < \infty \text{ for all } t \geq 0\right] = 1$$

for all $1 \leq i \leq n, 1 \leq j \leq m$, then a n Itô processes can be defined as the following

$$\begin{cases} dX_1 & = u_1 dt + v_{11} dB_1 + \cdots + v_{1m} dB_m \\ \vdots & \vdots \qquad\qquad \vdots \\ dX_n & = u_n dt + v_{n1} dB_1 + \cdots + v_{nm} dB_m \end{cases}$$

Or, in matrix notation simply

$$dX(t) = udt + vdB(t),$$

where

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}, u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nm} \end{pmatrix}, dB(t) = \begin{pmatrix} dB_1(t) \\ \vdots \\ dB_m(t) \end{pmatrix}$$

Theorem 2.3.5 (The general Itô formula)**.** *(see [29], p. 48) Let $dX(t) = udt + vdB(t)$ be an n-dimensional Itô process. Let $g(t,x) = (g_1(t,x), \cdots, g_p(t,x))$ be a C^2 map from $[0,\infty) \times \mathbb{R}^n$*

into $\mathbb{R}^p$. Then the process $Y(t,w) = g(t, X(t))$ is again an Itô process, whose component number k, Y_k, is given by

$$dY_k = \frac{\partial g_k}{dt}(t,X)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)dX_i + \frac{1}{2}\sum_{ij} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X)dX_i dX_j$$

where $dB_i dB_j = \delta_{ij}dt, dB_i dt = dtdB_i = 0$.

2.3.1 The Martingale Representation Theorem

Lemma 2.3.6. *(see [29], p. 50) Fix $T > 0$. The set of random variables*

$$\{\phi(B_{t_1}, \ldots, B_{t_n}); t_i \in [0,T], \phi \in C_0^\infty(R^n), n = 1, 2, \ldots\}$$

is dense in $L^2(\mathcal{F}_T, P)$.

Lemma 2.3.7. *(see [29], p. 50) The linear span of random variables of the type*

$$\exp\left\{\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)dt\right\}; h \in L^2[0,T] \text{ (deterministic)}$$

is dense in $L^2(\mathcal{F}_T, P)$.

Theorem 2.3.8 (The Itô representation theorem)**.** *(see [29], p. 51) Let $F \in L^2(\mathcal{F}_T^{(n)}, P)$. Then there exists a unique stochastic process $f(t,\omega) \in \mathcal{V}^n(0,T)$ such that $F(\omega) = E[F] + \int_0^T f(t,\omega)dB(t)$.*

Theorem 2.3.9 (The martingale representation theorem)**.** *(see [29], p. 52) Let $B(t) = (B_1(t), \ldots, B_n(t))$ be n-dimensional. Suppose M_t is an $\mathcal{F}_t^{(n)}$-martingale (w.r.t. P) and that $M_t \in L^2(P)$ for all $t \geq 0$. Then there exists a unique stochastic process $g(s,\omega)$ such that $g \in \mathcal{V}^{(n)}(0,t)$ for all $t \geq 0$ and $M_t(\omega) = E[M_0] + \int_0^t g(s,\omega)dB(s)$ a.s., for all $t \geq 0$.*

2.4 Stochastic Differential Equations

Definition 2.4.1. *(see [7], p. 10) Suppose that the $\mathbb{R}^r$-valued random function $b(\cdot)$ is $\mathcal{F}_t$-adapted and satisfies $\int_0^T |b(u)|du < \infty$ w.p. 1. Let the $r \times r$ matrix-valued random function $\sigma(\cdot)$ be $\mathcal{F}_t$-adapted and satisfy $\int_0^T |\sigma(u)|^2 du < \infty$ w.p. 1. A process $x(\cdot)$ defined as*

$$x(t) = x(0) + \int_0^t b(s, x(s))ds + \int_0^t \sigma(s, x(s))dw(s),$$

is called a diffusion. Rewrite it in differential form

$$dx(t) = b(t, x(t))dt + \sigma(t, x(t))dw(t).$$

Theorem 2.4.2 (Existence and uniqueness theorem for stochastic differential equations)**.** *(see [29], p. 68) Let $T > 0$ and $b(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable functions satisfying $|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|)$; $x \in \mathbb{R}^n$, $t \in [0,T]$ for some constant C, (where $|\sigma|^2 = \sum |\sigma_{ij}|^2$) and such that $|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|$; $x, y \in \mathbb{R}^n$, $t \in [0,T]$ for some constant D. Let Z be a random variable which is independent of the σ-algebra $\mathcal{F}_\infty^{(m)}$ generated by $B_s(\cdot)$, $s \geq 0$ and such that $E[|Z|^2] < \infty$. Then the stochastic differential equation $dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$, $0 \leq t \leq T$, $X_0 = Z$ has a unique t-continuous solution $X_t(\omega)$ with the property that $X_t(\omega)$ is adapted to the filtration $\mathcal{F}_t^Z$ generated by Z and $B_s(\cdot)$; $s \leq t$ and $E\left[\int_0^T |X_t|^2 dt\right] < \infty$.*

Definition 2.4.3. *(see [7], p. 10) Let $C^{2,1}$ denote the space of function on $R^r \times [0,T]$ whose mixed partial derivatives up to order 2 in x and order 1 in t are continuous. Let $f(\cdot,\cdot) \in C^{2,1}$.*

Now denote $\mathcal{L}$ as the differential operator which is defined by

$$\mathcal{L}f(x,t) = \sum_i b_i(t,x) f_{x_i}(x,t) + \frac{1}{2}\sum_{i,j} a_{ij}(t,x) f_{x_i x_j}(x,t),$$

where $a(t,x) = [a_{ij}(t,x)] \equiv \sigma(t,x)\sigma'(t,x)$.

2.5 Discrete-Time Markov Chains

Suppose that α_k is a stochastic process taking values in $\mathcal{M}$, which is at most countable (i.e., it is either finite $\mathcal{M} = \{1, 2, \ldots, m_0\}$ or countable $\mathcal{M} = \{1, 2, \ldots\}$).

Definition 2.5.1. *(see [26], p. 26) The process α_k is a Markov chain if*

$$\begin{aligned} p^{ij}_{k,k+1} &= P(\alpha_{k+1} = j | \alpha_k = i) \\ &= P(\alpha_{k+1} = j | \alpha_0 = i_0, \ldots, \alpha_{k-1} = i_{k-1}, \alpha_k = i), \end{aligned}$$

for any $i_0, \ldots, i_{k-1}, i, j \in \mathcal{M}$

Given i, j, if $p^{ij}_{k,k+1}$ is independent of time k, then α_k has stationary transition probabilities. In this case, $P^{(n)} = (P)^n$, where $P = (p^{ij})$ is the transition matrix, and $P^{(n)} = (p^{ij,(n)})$, with $p^{ij,(n)} = P(x_n = j | x_0 = i)$ is the n-step transition matrix.

Definition 2.5.2. *(see [26], p. 27) For a Markov chain α_k, state j is said to be accessible from state i if $p^{ij,(k)} = P(\alpha_k = j | \alpha_0 = i) > 0$ for some $k > 0$. Two states i and j, accessible from each other, are said to communicate. A Markov chain is irreducible if all states communicate with each other. For $i \in \mathcal{M}$, let $d(i)$ denote the period of state i, i.e., the greatest common divisor of all $k \geq 1$ such that $P(\alpha_{k+n} = i | \alpha_n = i) > 0$ (define $d(i) = 0$ if $P(\alpha_{k+n} = i | \alpha_n = i) = 0$ for all k). A Markov chain is called aperiodic if each sate has period one.*

Definition 2.5.3. *(see [26], p. 27) Let P be a transition matrix for a finite-state Markov chain. A row vector $\pi = (\pi^1, \dots, \pi^{m_0})$ with each $\pi^i \geq 0$ is called a stationary distribution of α_k if it is the unique solution to the system of equations*

$$\pi P = \pi$$

$$\sum_i \pi^i = 1.$$

Theorem 2.5.4. *(see [26], p. 27) Let $P = (p^{ij})$ be the transition matrix of an irreducible aperiodic finite-state Markov chain. Then there exist constants $0 < \lambda < 1$ and $c_0 > 0$ such that $|(P)^k - \overline{P}| \leq c_0 \lambda^k$ for $k = 1, 2, \dots,$ where $\overline{P} = \mathbb{1}_{m_0}\pi$, $\mathbb{1}_{m_0} = (1, \dots, 1)' \in \mathbb{R}^{m_0 \times 1}$, and $\pi = (\pi^1, \dots, \pi^{m_0})$ is the stationary distribution of α_k. This implies, in particular, $\lim_{k\to\infty} P^k = \mathbb{1}_{m_0}\pi$.*

Definition 2.5.5. *(see [26], p. 30) A sequence $\{f(i) : i \in \mathcal{M}\}$ is P-harmonic or right-regular if a) $f(\cdot)$ is a real-valued function such that $f(i) \geq 0$ for each $i \in \mathcal{M}$, and b) $f(i) = \sum_{j\in\mathcal{M}} p^{ij} f(j)$ for each $i \in \mathcal{M}$.*

Definition 2.5.6. *(see [26], p. 31) A jump process is a right-continuous stochastic process with piecewise-constant sample paths. Let $\alpha(\cdot) = \{\alpha(t) : t \geq 0\}$ be a jump process defined on $\Omega, \mathcal{F}, P)$ taking values in $\mathcal{M}$. Then $\{\alpha(t) : t \geq 0\}$ is a Markov chain with state space $\mathcal{M}$, if $P(\alpha(t) = i|\alpha(r) : r \leq s) = P(\alpha(t) = i|\alpha(s))$, for all $0 \leq s \leq t$ and $i \in \mathcal{M}$, with $\mathcal{M}$ being either finite or countable.*

Definition 2.5.7 (q-Property)**.** *(see [26], p. 32) A matrix-valued function $Q(t) = (q^{ij}(t))$, for $t \geq 0$, satisfies the q-Property, if*

(a) $q^{ij}(t)$ is Borel measurable for all $i, j \in \mathcal{M}$ and $t \geq 0$;

(b) $q^{ij}(t)$ *is uniformly bounded. That is, there exists a constant* K *such that* $|q^{ij}(t)| \leq K$*, for all* $i, j \in \mathcal{M}$ *and* $t \geq 0$*;*

(c) $q^{ij}(t) \geq 0$ *for* $j \neq i$ *and* $q^{ii} = -\sum_{j \neq i} q^{ij}(t), t \geq 0$*.*

For any real-valued function f on $\mathcal{M}$ and $i \in \mathcal{M}$, write $Q(t)f(\cdot)(i) = \sum_{j \in \mathcal{M}} q^{ij}(t)f(j) = \sum_{j \neq i} q^{ij}(t)(f(j) - f(i))$.

Definition 2.5.8 (Generator)**.** *(see [26], p. 32) A matrix* $Q(t), t \geq 0$*, is an infinitesimal generator of* $\alpha(\cdot)$ *if it satisfies the* q*-Property, and for any bounded real-valued function* f *defined on* $\mathcal{M}$

$$f(\alpha(t)) - \int_0^t Q(\zeta)f(\cdot)(\alpha(\zeta))d\zeta$$

is a martingale.

Lemma 2.5.9. *(see [26], p. 33) Let* $\mathcal{M} = \{1, \ldots, m_0\}$*. Then* $\alpha(t) \in \mathcal{M}$*,* $t \geq 0$*, is a Markov chain generated by* $Q(t)$ *iff*

$$\left(I_{\{\alpha(t)=1\}}, \ldots, I_{\{\alpha(t)=m_0\}}\right) - \int_0^t \left(I_{\{\alpha(\zeta)=1\}}, \ldots, I_{\{\alpha(\zeta)=m_0\}}\right) Q(\zeta)d\zeta$$

is a martingale.

Let $0 = \tau_0 < \tau_1 < \cdots < \tau_l < \cdots$ be a sequence of jump times of $\alpha(\cdot)$ such that the random variables $\tau_1, \tau_2 - \tau_1, \ldots, \tau_{k+1} - \tau_k, \ldots$ are independent. Let $\alpha(0) = i \in \mathcal{M}$. Then $\alpha(t) = i$ on the interval $[\tau_0, \tau_1)$. The first jump time τ_1 has the probability distribution

$$P(\tau_1 \in B) = \int_B \exp\left\{\int_0^t q^{ii}(s)ds\right\}(-q^{ii}(t))dt,$$

where $B \subset [0, \infty)$ is a Borel set. The post-jump location of $\alpha(t) = j, j \neq i$, is given by $P(\alpha(\tau_1) = j|\tau_1) = \frac{q^{ij}(\tau_1)}{-q^{ii}(\tau_1)}$.

In general, $\alpha(t) = \alpha(\tau_l)$ on the interval $[\tau_l, \tau_{l+1})$. The jump time τ_{l+1} has the conditional probability distribution

$$P(\tau_{l+1}-\tau_l \in B_l|\tau_1, \dots, \tau_l, \alpha(\tau_1), \dots, \alpha(\tau_l)) = \int_{B_l} \exp\left\{\int_{\tau_l}^{t+\tau_l} q^{\alpha(\tau_l)\alpha(\tau_l)}(s)ds\right\}(-q^{\alpha(\tau_l)\alpha(\tau_l)}(t+\tau_l))dt.$$

The post-jump location of $\alpha(t) = j, j \neq \alpha(\tau_l)$ is given by

$$P(\alpha(\tau_{l+1}) = j|\tau_1, \dots, \tau_l, \tau_{l+1}, \alpha(\tau_1), \dots, \alpha(\tau_l)) = \frac{q^{\alpha(\tau_l)j}(\tau_{l+1})}{-q^{\alpha(\tau_l)\alpha(\tau_l)}(\tau_{l+1})}.$$

Theorem 2.5.10. *(see [26], p. 34) Suppose that the matrix $Q(t)$ satisfies the q-Property for $t \geq 0$. Then the following statements hold.*

(a) The process $\alpha(\cdot)$ constructed above is a Markov chain.

(b) The process $f(\alpha(t)) - \int_0^t Q(\zeta)f(\cdot)(\alpha(\zeta))d\zeta$ is a martingale for any uniformly bounded function $f(\cdot)$ on $\mathcal{M}$. Thus $Q(t)$ is indeed the generator of $\alpha(\cdot)$.

(c) The transition matrix $P(t, s)$ satisfies the forward differential equation

$$\frac{dP(t,s)}{dt} = P(t,s)Q(t),\ t \geq s,$$
$$P(s,s) = I,$$

where I is the identity matrix.

(d) *Assume further that $Q(t)$ is continuous in t. Then $P(t,s)$ also satisfies the backward differential equation*

$$\frac{dP(t,s)}{ds} = Q(s)P(t,s), t \geq s,$$
$$P(s,s) = I.$$

Definition 2.5.11 (Irreducibility)**.** *(see [26], p. 34)*

(a) *A generator $Q(t)$ is said to be weakly irreducible if, for each fixed $t \geq 0$, the system of equations*

$$\nu(t)Q(t) = 0,$$
$$\sum_{i=1}^{m_0} \nu^i(t) = 1$$

has a unique solution $\nu(t) =)\nu^1(t), \ldots, \nu^{m_0}(t))$ and $\nu(t) \geq 0$.

(b) *A generator $Q(t)$ is said to be irreducible, if for each fixed $t \geq 0$ the systems of equations has a unique solution $\nu(t)$ and $\nu(t) \geq 0$.*

Definition 2.5.12 (Quasi-Stationary Distribution)**.** *(see [26], p. 35) For $t \geq 0$, $\nu(t)$ is termed a quasi-stationary distribution if it is the unique solution and satisfies $\nu(t) \geq 0$.*

Definition 2.5.13. *(see [26], p. 35) A generator $Q(t)$ is said to be weakly irreducible if, for* *each fixed $t \geq 0$, the system of equations*

$$f(t)Q(t) = 0,$$
$$\sum_{i=1}^{m_0} f^i(t) = 0$$

has only the trivial solution.

2.5.1 Asymptotic Expansions

For a small parameter $\varepsilon > 0$, let α_k^ε be a discrete-time Markov chain depending on ε and having finite state space $\mathcal{M} = \{1, \dots, m_0\}$ and transition matrix $P_k^\varepsilon = P_k + \varepsilon Q_k$, where for each k, P_k is a transition probability matrix and $Q_k = (q_k^{ij})$ is a generator of a continuous-time Markov chain.

Definition 2.5.14. *(see [26], p. 44) Denote p_k^ε be the probability vector*

$$p_k^\varepsilon = (P(\alpha_k^\varepsilon = 1), \dots, P(\alpha_k^\varepsilon = m_0)) \in \mathbb{R}^{1\times m_0}.$$

Assuming that the initial probability p_0^ε is independent of ε, i.e. $p_0^\varepsilon = p_0 = (p_0^1, \dots, p_0^{m_0})$ such that $p_0^i \geq 0$ for $i = 1, \dots, m_0$ and $p_0 \mathbb{1}_{m_0} = \sum_{i=1}^{m_0} p_0^i = 1$, then p_k^ε is a solution of the vector-valued difference equation

$$p_{k+1}^\varepsilon = p_k^\varepsilon P_k^\varepsilon, k = 0, 1, \dots, \lfloor T/\varepsilon \rfloor,$$
$$p_0^\varepsilon = p_0.$$

Theorem 2.5.15. *(see [26], p. 45) Suppose that $P^\varepsilon = P + \varepsilon Q$ and that P is irreducible. For* *an integer* $n_0 > 0$*, for some* $T > 0$*, and for any* $0 \leq k \leq \lfloor T/\varepsilon \rfloor$*, as* $\varepsilon \to 0$ *and* $k \to \infty$*,* $p_k^\varepsilon \to \nu$*, where* ν *is the stationary distribution corresponding to* P*. Moreover, there exist two sequences* $\{\varphi_i(t)\}_{i=0}^{n_0}, 0 \leq t \leq T$*, and* $\{\psi_i(k)\}_{i=0}^{n_0}$ *such that* $|\psi_i(k)| \leq K\lambda_0^k$ *for some* $0 < \lambda_0 < 1$*, that* $\varphi(\cdot)$ *for* $i = 0, \dots, n_0$ *are sufficiently smooth, and that*

$$\sup_{0 \leq k \leq \lfloor T/\varepsilon \rfloor} \left| p_k^\varepsilon - \sum_{i=0}^{n_0} \varepsilon^i \varphi_i(\varepsilon k) - \sum_{i=0}^{n_0} \varepsilon^i \psi_i(k) \right| = \mathcal{O}(\varepsilon^{n_0+1}).$$

Suppose that $T > 0$ and $\varepsilon > 0$ is a small parameter, and that for $0 \leq k \leq \lfloor T/\varepsilon \rfloor$, α_k^ε is a discrete-time Markov chain with state space $\mathcal{M} = \{1, \dots, m_0\}$ and transition matrix $P^\varepsilon = P + \varepsilon Q$, where P is a transition matrix of a discrete-time Markov chain and Q is a generator of a continuous-time Markov chain. Suppose P is given by $P = \operatorname{diag}(P^1, \dots, P^{l_0}) = \begin{pmatrix} P^1 & & \\ & \ddots & \\ & & P^{l_0} \end{pmatrix}$, where $P^i \in \mathbb{R}^{m_i \times m_i}$ are transition matrices and $\sum_{i=1}^{l_0} m_i = m_0$, or $P = \begin{pmatrix} P^1 & & & & \\ & P^2 & & & \\ & & \ddots & & \\ & & & P^{l_0} & \\ P^{*,1} & P^{*,2} & \cdots & P^{*,l_0} & P^* \end{pmatrix}$, where P^i are transition matrices for each $i \leq l_0$ and $(P^{*,1}, \dots, P^{*,l_0}, P^*)$ corresponds to the transient states. Then the state space $\mathcal{M}$ can be rewritten as

$$\begin{aligned} \mathcal{M} &= \{s_{11}, \dots, s_{1m_1}\} \cup \{s_{21}, \dots, s_{2m_2}\} \cup \ldots \cup \{s_{l_01}, \dots, s_{l_0m_{l_0}}\} \\ &= \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_{l_0}, \end{aligned}$$

with $m_0 = m_1+m_2+\cdots+m_{l_0}$. The subspace $\mathcal{M}_i$, for each $i = 1, \ldots, l_0$, consists of recurrent states belonging to the ith ergodic class. Furthermore, suppose that P^ε, P, and P^i for $i \le l_0$ are one-step transition probability matrices such that for each $i \le l_0$, P^i is irreducible and aperiodic.

Theorem 2.5.16. *(see [26], p. 77) Under conditions of being irreducible and aperiodic, the following assertions hold:*

(a) For the probability distribution vector $p_k^\varepsilon = (P(\alpha_k^\varepsilon = s_{ij})) \in \mathbb{R}^{1\times m_0}$, we have

$$p_k^\varepsilon = \theta(\varepsilon k)\mathrm{diag}(\nu^1, \ldots, \nu^{l_0}) + \mathcal{O}(\varepsilon + \lambda^k)$$

for some λ with $0 < \lambda < 1$, where ν^i is the stationary distribution corresponding to the transition matrix P_i, and $\theta(t) = (\theta^1(t), \ldots, \theta^{l_0}(t)) \in \mathbb{R}^{1\times l_0}$ satisfies

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

where

$$\overline{Q} = \mathrm{diag}(\nu^1, \ldots, \nu^{l_0})Q\widetilde{\mathbb{1}}\,\widetilde{\mathbb{1}} = \mathrm{diag}(\mathbb{1}_{m_1}, \ldots, \mathbb{1}_{m_{l_0}}).$$

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

$$(P^\varepsilon)^k = \Phi(t) + \varepsilon\widehat{\Phi}(t) + \Psi(k) + \varepsilon\widehat{\Psi}(k) + \mathcal{O}(\varepsilon^2)$$

where

$$\Psi(t) = \widetilde{\mathbb{1}}\Theta(t)\mathrm{diag}(\nu^1, \ldots, \nu^{l_0})\,\frac{d\Theta(t)}{dt} = \Theta(t)\overline{Q}\,\Theta(0) = I.$$

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

Define continuous-time interpolations by $\alpha^\varepsilon(t)=\alpha^\varepsilon_k$, and $\overline{\alpha}^\varepsilon(t)=\overline{\alpha}^\varepsilon_k$ for $t\in[\varepsilon k,\varepsilon(k+1))$, and denote by $D([0,T];\mathcal{M})$ the space of functions that are defined on $[0,T]$ taking values in $\mathcal{M}$ and that are right continuous and have left limits endowed with the Skorohod topology.

Theorem 2.5.17. *(see [26], p. 78) As* $\varepsilon\to 0$, $\overline{\alpha}^\varepsilon(\cdot)$ *converges weakly to* $\overline{\alpha}(\cdot)$*, a Markov chain generated by* $\overline{Q}$.

For $k=0,\ldots,T/\varepsilon$, $i=1,\ldots,l_0$ and $j=1,\ldots,m_i$, define sequences of occupation measures by

$$\pi^{\varepsilon,ij}_k=\varepsilon\sum_{l=0}^{k-1}\left(I_{\{\alpha^\varepsilon_l=s_{ij}\}}-\nu^{ij}I_{\{\overline{\alpha}^\varepsilon_l=i\}}\right)$$

$$\pi^\varepsilon_k=(\pi^{\varepsilon,ij}_k)\in\mathbb{R}^{1\times m_0}.$$

Define continuous-time interpolations

$$\pi^{\varepsilon,ij}(t)=\pi^{\varepsilon,ij}_k \text{ for t } \in[k\varepsilon,(k+1)\varepsilon),$$

$$\pi^\varepsilon(t)=(\pi^{ij}(t))\in\mathbb{R}^{1\times m_0}.$$

Theorem 2.5.18. *(see [26], p. 79) For* $i=1,\ldots,l_0$, $j=1,\ldots,m_i$,

$$\sup_{0\le k\le T/\varepsilon}E|\pi^{\varepsilon,ij}_k|^2=\mathcal{O}(\varepsilon) \text{ and } \sup_{t\in[0,T]}E|\pi^{\varepsilon,ij}(t)|^2=\mathcal{O}(\varepsilon).$$

For each $i = 1, \ldots, l_0$, $j = 1, \ldots, m_i$, and each $0 < k \le T/\varepsilon$, define sequences of normalized occupation measures

$$n_k^{\varepsilon,ij} = \sqrt{\varepsilon}\sum_{l=0}^{k-1} w^{ij}(\alpha_l^\varepsilon) = \frac{1}{\sqrt{\varepsilon}}\pi_k^{\varepsilon,ij},$$

$$n_k^\varepsilon = (n_k^{\varepsilon,ij} \in \mathbb{R}^{1\times m_0}.$$

where $w^{ij}(\alpha) = I_{\{\alpha=s_{ij}\}} - \nu^{ij} I_{\{\alpha\in\mathcal{M}_i\}}$.

Lemma 2.5.19. *(see [26], p. 80) Let $\mathcal{F}_t^\varepsilon = \sigma\{\alpha^\varepsilon(s) : s \le t\}$. Then*

(a) $\sup_{0\le t\le t+s\le T} E[(n^\varepsilon(t+s) - n^\varepsilon(t))|\mathcal{F}_t^\varepsilon] = \mathcal{O}(\sqrt{\varepsilon})$*;*

(b) For each $i = 1, \ldots, l_0$ and $j = 1, \ldots, m_i$ and for any $0 < s \le \delta$,

$$E[(n^{\varepsilon,ij}(t+s) - n^{\varepsilon,ij}(t))^2|\mathcal{F}_t^\varepsilon] = \mathcal{O}(\delta).$$

Lemma 2.5.20. *(see [26], p. 80) $\{n^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot)\}$ is tight in $D([0,T]; \mathbb{R}^{m_0} \times \mathcal{M})$.*

Define operator $\mathcal{L}$ as the following $\mathcal{L}f(x,i) = \frac{1}{2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i} a^{j_1 j_2}(i)\frac{\partial^2 f(x,i)}{\partial x^{ij_1}\partial x^{ij_2}} + \overline{Q}f(x,\cdot)(i)$, for $i = 1, \ldots, l_0$, where $A(i) = (a^{j_1 J_2}(i))$ is symmetric and nonnegative definite.

Theorem 2.5.21. *(see [26], p. 81) $(n^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(n(\cdot), \overline{\alpha}(\cdot))$ such that the limit is the solution of the martingale problem with operator $\mathcal{L}$.*

Suppose that P^ε, P and P^i for $i \le l_0$ are one-step transition probability matrices such that for each $i \le l_0$, P^i is irreducible and aperiodic. All the eigenvalues of P^* are inside the unit circle. Then we have

Theorem 2.5.22. *(see [26], p. 83)*

(a) *The probability vector* $p_k^\varepsilon = \theta(\varepsilon k)\mathrm{diag}(\nu^1, \ldots, \nu^{l_0}, 0_{m_*}) + \mathcal{O}(\varepsilon + \lambda^k)$*, where* $0_{m_*} \in \mathbb{R}^{1\times m_*}$ *and* $\theta(t) = (\theta^1(t), \ldots, \theta^{l_0}(t)) \in \mathbb{R}^{1\times l_0}$ *satisfies* $\frac{d\theta(t)}{dt} = \theta(t)\overline{Q}_*$*,* $\theta^i(0) = p^i(0)\mathbb{1}_{m_i} - p^*(0)a^i$*. The transition matrix satisfies* $P^\varepsilon(\varepsilon k_0, \varepsilon k) = \Phi(t_0, t) + \Psi(k_0, k) + \varepsilon\widehat{\Phi}(t_0, t) + \varepsilon\widehat{\Psi}(k_0, k) + \mathcal{O}(\varepsilon^2)$*, for some* λ *with* $0 < \lambda < 1$*, where* $\Phi(t_0, t) = \widetilde{\mathbb{1}}_*\Theta_*(t_0, t)\nu_*$ *with* $\Theta_*(t_0, t) = \mathrm{diag}(\Theta(t_0, t)I_{m_*\times m_*})$ *whre* $\Theta(t_0, t) = (\theta^{ij}(t_0, t))$ *satisfies the differential equation* $\frac{\partial\Theta(t_0,t)}{\partial t} = \Theta(t_0, t)\overline{Q}_*(t)$*,* $\Theta(t_0, t_0) = I$*.*

(b) *For each* $j = 1, \ldots, m_i$*,* $\sup_{t\in[0,T]} E|\pi^{\varepsilon,ij}(t)|^2 = \begin{cases} \mathcal{O}(\varepsilon), & \text{for } i = 1, \ldots, l_0, \\ \mathcal{O}(\varepsilon^2), & \text{for } i = *; \end{cases}$

(c) $\overline{\alpha}^\varepsilon(\cdot)$ *converges weakly to* $\overline{\alpha}(\cdot)$*, a Markov chain generated by* $\overline{Q}_*$*;*

CHAPTER 3 PROBLEM FORMULATION

3.1 Problem Formulation and Conditions

Consider a sequence of double arrays or a sequence of matrices modulated by a discrete-time Markov chain α_k with a finite state space $\mathcal{M}$. We denote the sequence by $\{X_k(\imath,\jmath,\alpha_k)\}$. This work aims to investigate the asymptotic properties of $\{X_k(\imath,\jmath,\alpha_k)\}$. Note that $k = 0,1,2,\dots$ denotes the discrete time, and $(\imath,\jmath)$ denotes the indices of the matrices with $\imath,\jmath = 1,\dots,d$. That is, for each k and each $\gamma \in \mathcal{M}$, $X_k(\imath,\jmath,\gamma) \in \mathbb{R}^{d\times d}$ is a $d \times d$ matrix. In fact, for each $\gamma \in \mathcal{M}$,

$$X_k(\gamma) = \begin{pmatrix} X_k(1,1,\gamma) & X_k(1,2,\gamma) & \dots & X_k(1,d,\gamma) \\ X_k(2,1,\gamma) & X_k(2,2,\gamma) & \dots & X_k(2,d,\gamma) \\ \dots & \dots & \dots & \dots \\ X_k(d,1,\gamma) & X_k(d,2,\gamma) & \dots & X_k(d,d,\gamma) \end{pmatrix} \in \mathbb{R}^{d\times d}. \tag{3.1}$$

For simplicity, in what follows, we will suppress the $\imath,\jmath$ dependence and write it in a more compact notation as $X_k(\gamma)$.

Remark 3.1.1. In lieu of a square matrix $X_k(\gamma) \in \mathbb{R}^{d\times d}$, we can consider $X_k(\gamma) \in \mathbb{R}^{d_1\times d_2}$ with $d_1 \neq d_2$. All subsequent development goes through with no essential difficulties. Only the notation is somewhat different. In fact, we give a remark on how to treat sequences of non-square matrix-valued random elements at the last chapter of this dissertation.

3.1.1 Decomposition and Subspaces

Suppose that the state space of the Markov chain is

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cdots \cup \mathcal{M}_{l_0}. \tag{3.2}$$

We assume that the Markov chain is nearly completely decomposable in the sense [1], the subspace of $\mathcal{M}_\iota$ is given by

$$\mathcal{M}_\iota = \{s_{\iota 1}, \dots, s_{\iota m_\iota}\}, \tag{3.3}$$

with the cardinality of $\mathcal{M}_\iota$ given by

$$|\mathcal{M}_\iota| = m_\iota \text{ for each } \iota = 1, \dots, l_0,$$

so that

$$m_0 = m_1 + m_2 + \cdots + m_{l_0}. \tag{3.4}$$

Thus, the cardinality of $\mathcal{M}$ is $|\mathcal{M}| = m_0$. Note that in the above, we used $|\cdot|$ to denote the cardinality of a set. In what follows, we also use $|\cdot|$ to denote the norm of a vector (absolute value of a scalar in particular). They should be clear from the context.

Corresponding to $\mathcal{M}$, the transition matrix of the Markov chain is given by $P^\varepsilon = P + \varepsilon Q$, where $\varepsilon > 0$ is a small parameter. Suppose

$$P = \begin{pmatrix} P^1 & & & \\ & P^2 & & \\ & & \ddots & \\ & & & P^{l_0} \end{pmatrix}, \tag{3.5}$$

where P^ι for $\iota \le l_0$ are themselves probability transition matrices; Q is a generator of a continuous-time Markov chain. That is, $q_{\gamma\gamma_1} \ge 0$ and for each $\gamma \in \mathcal{M}$, $\sum_{\gamma_1 \in \mathcal{M}} q_{\gamma\gamma_1} = 0$. To highlight the ε-dependence of the transition probabilities, we denote the Markov chain by

α_k^ε. For this Markov chain, we assume the following conditions hold.

(A1) For each $\iota = 1, \ldots, l_0$, P^ι is irreducible and aperiodic.

Remark 3.1.2. A consequence of (A1) is the following fact. Consider a Markov chain with transition matrix P^ι. Then this Markov chain is ergodic. That is, there is a stationary distribution ν^ι such that

$$\begin{cases} \nu^\iota P^\iota & = \nu^\iota \\ \nu^\iota \mathbb{1}_{m_i} & = 1, \end{cases} \tag{3.6}$$

where $\mathbb{1}_{m_\iota} = (1, \ldots, 1)'$ is an m_ι-dimensional column vector with all components being 1. In fact, ν^ι is the unique solution of the system (3.6). It is easily seen that (3.6) can also be written as

$$\nu^\iota (P^\iota - I_{m_\iota}) = 0,$$

where I_{m_ι} denotes the m_ι-dimensional identity matrix. Moreover, the n-step transition probability matrix $(P^\iota)^n$ converges to $\mathbb{1}_{m_\iota}\nu^\iota$ exponentially fast in that there is a $K > 0$ and a λ satisfying $0 < \lambda < 1$ such that

$$(P^\iota)^n \to \mathbb{1}_{m_\iota}\nu^\iota \text{ as } n \to \infty \text{ and } |(P^\iota)^n - \mathbb{1}_{m_\iota}\nu^\iota| \le K\lambda^n,$$

and that $\mathbb{1}_{m_\iota}\nu^\iota$ has identical rows.

3.1.2 Aggregation and Interpolated Process

To proceed, we define an aggregated state space $\overline{\mathcal{M}}$ as

$$\overline{\mathcal{M}} = \{1, \ldots, l_0\}, \tag{3.7}$$

and an aggregation process

$$\overline{\alpha}_k^\varepsilon = \iota \text{ if } \ \alpha_k^\varepsilon \in \mathcal{M}_\iota. \tag{3.8}$$

Define also an interpolated process $\overline{\alpha}^\varepsilon(\cdot)$ as

$$\overline{\alpha}^\varepsilon(t) = \overline{\alpha}_k^\varepsilon \text{ for } \ t \in \big[\varepsilon k, \varepsilon(k+1)\big). \tag{3.9}$$

A number of preliminary results concerning the processes α_k^ε and $\overline{\alpha}_k^\varepsilon$ are relegated to an appendix. These results, in particular, Lemma 6.3.1 and Remark 6.3.2 in the appendix will be used throughout the rest of the dissertation.

Next, we specify the primary sequence $\{X_k(\iota)\}$. For each $\iota \in \overline{\mathcal{M}}$, $\{X_k(\iota)\}$ is independent of $\{\alpha_k^\varepsilon\}$ and is a ϕ-mixing sequence with mixing measure $\widetilde{\phi}(\cdot)$. More details of the mixing sequence will be specified later. To study $X_k(\iota) \in \mathbb{R}^{d\times d}$, we consider a functional $\widetilde{f}(\cdot)$: $\mathbb{R}^{d\times d} \to \mathbb{R}$. For each $\iota \in \overline{\mathcal{M}}$, and write $f_k(\iota) = \widetilde{f}(X_k(\iota))$. There are many such functionals of interests. For example, $f_k(\iota) = \text{tr}(X_k(\iota))$, or $f_k(\iota) = |X_k(\iota)|$, where $|\cdot|$ is the induced norm of $X_k(\iota)$, etc. In the above, we used the idea of mapping a sequence of $d \times d$ matrices into a sequence of real numbers so that the asymptotic properties of $\{X_k(\iota)\}$ can be explained by $\{f_k(\iota)\}$. Note that

$$\begin{aligned} &f_k(\iota) = \widetilde{f}(X_k(\iota)) \ \text{ for each } \ \iota \in \overline{\mathcal{M}} \ \text{ so} \\ &f_k(\overline{\alpha}_k^\varepsilon) = \widetilde{f}(X_k(\overline{\alpha}_k^\varepsilon)). \end{aligned} \tag{3.10}$$

Our study is through proper centered, scaled, and interpolated sequences.

CHAPTER 4 ASYMPTOTIC PROPERTIES

Denote the σ-algebras by $\mathcal{F}_k^X = \sigma\{X_j(\iota) : j \le k, \iota = 1, \ldots, l_0\}$, $\mathcal{F}_k^{\alpha^\varepsilon} = \sigma\{\alpha_j^\varepsilon : j \le k\}$, and $\mathcal{F}_k = \sigma\{X_j(\iota), \alpha_j^\varepsilon : j \le k, \iota = 1, \ldots, l_0\}$. We are interested in a centered and scaled sequence. To proceed, we specify $\widetilde{f}(\cdot)$ in what follows.

Define

$$\overline{f}(\iota) = E\widetilde{f}(X_k(\iota)). \tag{4.1}$$

We proceed to define a centered and rescaled sequence y_k^ε as

$$y_k^\varepsilon = \sqrt{\varepsilon}\sum_{j=0}^{k-1}\sum_{\iota\in\overline{\mathcal{M}}}[\widetilde{f}(X_j(\iota)) - \overline{f}(\iota)]I_{\{\overline{\alpha}_j^\varepsilon=\iota\}}. \tag{4.2}$$

Note that $\overline{f}(\overline{\alpha}_j^\varepsilon) = E[\widetilde{f}(X_j(\overline{\alpha}_j^\varepsilon))|\mathcal{F}_j^{\alpha^\varepsilon}]$. Thus y_k^ε may also be written as

$$y_k^\varepsilon = \sqrt{\varepsilon}\sum_{j=0}^{k-1}\{\widetilde{f}(X_j(\overline{\alpha}_j^\varepsilon)) - E[\widetilde{f}(X_j(\overline{\alpha}_j^\varepsilon))|\mathcal{F}_j^{\alpha^\varepsilon}]\}. \tag{4.3}$$

(A2) The functional $\widetilde{f}(x)$ is continuous in its argument and there is a $c_0 \ge 1$ such that $|\widetilde{f}(x)| \le K(1 + |x|^{c_0})$.

(A3) For each $\iota \in \overline{\mathcal{M}}$, the sequence $\{X_k(\iota)\}$ satisfies

$$\begin{aligned} &EX_k(\iota) = 0 \\ &E|X_k(\iota)|^{2(c_0+\Delta)} < \infty \end{aligned} \tag{4.4}$$

for some $\Delta > 0$ and the mixing measure satisfies

$$\sum_{k=0}^{\infty} \widetilde{\phi}^{1/2}(k) < \infty. \tag{4.5}$$

Remark 4.0.1. We note the following.

- Condition (A2) indicates that $\widetilde{f}(\cdot)$ is of polynomial growth in its argument and is a continuous functional. The intuition of this conditions stems from the examples mentioned before.
- Equation (4.3) gives us the equivalence of the definition of y_k^ε using conditional expectation. However, for our analysis to follow, it is more convenient to use (4.2).
- Because $\{X_k(\iota)\}$ is stationary mixing, $\{\widetilde{f}(X_k(\iota))\}$ is also a mixing sequence.

Now we are in a position to study the asymptotic properties of the sequence $\{y_k^\varepsilon\}$. The study is through the examination of a continuous-time interpolation. In fact, we define $y^\varepsilon(t) = y_k^\varepsilon$ for $t \in [\varepsilon k, \varepsilon k + \varepsilon)$. We aim to obtain a weak convergence result for the two component sequence $\{y^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot)\}$. Our effort will be devoted to showing that $(y^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(y(\cdot), \overline{\alpha}(\cdot))$, a switching diffusion limit. To this end, consider the following operator for a suitable function $h(\cdot, \cdot)$ that is smooth with respect to the first variable y,

$$\mathcal{L}h(y, \iota) = \frac{1}{2} a(\iota) \frac{d^2 h(y, \iota)}{dy^2} + \overline{Q}h(y, \cdot)(\iota), \ \iota \in \overline{\mathcal{M}}, \tag{4.6}$$

where

$$\overline{Q} = \begin{pmatrix} \nu^1 & & \\ & \ddots & \\ & & \nu^{l_0} \end{pmatrix} Q \begin{pmatrix} 1\!\!1_{m_1} & & \\ & \ddots & \\ & & 1\!\!1_{m_{l_0}} \end{pmatrix}, \tag{4.7}$$
$$1\!\!1_{m_i} = (1, \dots, 1)' \in \mathbb{R}^{m_i \times 1},$$

$$\overline{Q}h(y, \cdot)(\iota) = \sum_{\ell=1}^{l_0} \overline{q}_{\iota\ell} h(y, \ell), \tag{4.8}$$

and $a(\iota)$ is defined as

$$a(\iota) = E[\widehat{f}_0(\iota)]^2 + 2\sum_{j=1}^{\infty} E\widehat{f}_0(\iota)[\widehat{f}_j(\iota)]. \tag{4.9}$$

Lemma 4.0.2. *Let $(y(\cdot), \overline{\alpha}(\cdot))$ be a solution of the martingale problem associated with the operator defined in* (4.6). *Then the solution of the martingale problem is unique.*

Proof. The proof is similar to [26, Lemma 14.8], thus we will only provide a short outline here. It suffices to verify the uniqueness in distribution of a solution $(y(\cdot), \overline{\alpha}(\cdot))$ of the martingale problem associated with the operator $\mathcal{L}$ for each $0 \le t \le T$. To this end, consider the characteristic function $\widetilde{\rho}(y, l) = E\exp(\imath(y\lambda + sl))$, where $\imath^2 = -1$, for each positive integer l, each $y, \lambda, s \in \mathbb{R}$. Define $\rho^{\iota\iota_1}(t) = E[I_{\{\overline{\alpha}(t)=\iota\}}\widetilde{\rho}(y(t), \iota_1)]$ for $\iota, \iota_1 \in \overline{\mathcal{M}}$. Because $(y(t), \overline{\alpha}(t))$ is a solution of the martingale problem associated with the operator $\mathcal{L}$, $\rho^{\iota\iota_1}(t)$ is the solution of a linear ordinary differential equation. The details are similar to [26, Proof of Lemma 4.9], which is omitted. As a result, it has a unique solution. Thus, $(y(t), \overline{\alpha}(t))$ is uniquely determined. So the desired uniqueness follows. □

Because of the polynomial growth of $\{\widetilde{f}(X_k(\iota))\}$, we need to work with a truncated version of the sequence. For fixed but otherwise arbitrary positive real number N, we define

$$y_k^{\varepsilon,N} = \sqrt{\varepsilon} \sum_{j=0}^{k-1} \sum_{\iota \in \overline{\mathcal{M}}} [\widetilde{f}^N(X_j^N(\iota)) - \overline{f}^N(\iota)] I_{\{\overline{\alpha}_j^\varepsilon = \iota\}}, \tag{4.10}$$

where

$$\begin{aligned} &\widetilde{f}^N(X_j^N(\iota)) = \widetilde{f}(X_j^N(\iota)) q^N(\widetilde{f}^N(X_j^N(\iota))), \\ &\overline{f}^N(\iota) = E\widetilde{f}^N(X_j^N(\iota)), \\ &q^N(y) \text{ is a smooth function with } q^N(y) = \begin{cases} 1 & \text{if } y \in \widetilde{S}_N, \\ 0 & \text{if } y \in \mathbb{R} - \widetilde{S}_{N+1}, \end{cases} \end{aligned} \tag{4.11}$$

where $\widetilde{S}_n = \{y \in \mathbb{R} : |y| < n\}$ denotes the interval containing the origin of radius n in $\mathbb{R}$. Because of the possible unboundedness, we need to overcome the difficulty by use of the truncation device. That is, we first obtain the desired properties for

$$y^{\varepsilon,N}(t) = y_k^{\varepsilon,N} \text{ for } t \in [\varepsilon k, \varepsilon k + \varepsilon),$$

the interpolation of the truncated process $y_k^{\varepsilon,N}$. Note that

$$y^{\varepsilon,N}(t) = \sqrt{\varepsilon} \sum_{j=0}^{\lfloor t/\varepsilon \rfloor - 1} \sum_{\iota \in \overline{\mathcal{M}}} [\widetilde{f}^N(X_j^N(\iota)) - \overline{f}^N(\iota)] I_{\{\overline{\alpha}_j^\varepsilon = \iota\}}. \tag{4.12}$$

For notational convenience, we define

$$\begin{aligned} \widehat{f}_j(\iota) &= [\widetilde{f}(X_j(\iota)) - \overline{f}(\iota)] \\ \widehat{f}_j^N(\iota) &= [\widetilde{f}^N(X_j^N(\iota)) - \overline{f}^N(\iota)], \end{aligned} \tag{4.13}$$

respectively.

Lemma 4.0.3. *Assume that* (A1)-(A3). *Then* $\{y^{\varepsilon,N}(\cdot)\}$ *is tight in* $D[0,T]$*, the space of functions that are right continuous with left limits endowed with the Skorohod topology.*

Proof. Denote the conditional expectation with respect to $\mathcal{F}_t^\varepsilon$ by E_t^ε. We use K to denote a generic positive constant whose value may different for different appearances. For any $\delta > 0$, t, and $s > 0$ satisfying $s \leq \delta$, consider

$$\begin{aligned}
&E_t^\varepsilon |y^{\varepsilon,N}(t+s) - y^{\varepsilon,N}(t)|^2 \\
&= \varepsilon E_t^\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor -1} \sum_{k=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor -1} \sum_{\iota \in \overline{\mathcal{M}}} [\widetilde{f}^N(X_k^N(\iota)) - \overline{f}^N(\iota)][\widetilde{f}^N(X_j^N(\iota)) - \overline{f}^N(\iota)] I_{\{\overline{\alpha}_k^\varepsilon = \iota\}} \\
&\leq K\varepsilon E_t^\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor -1} \sum_{k \geq j} \sum_{\iota \in \overline{\mathcal{M}}} [\widetilde{f}^N(X_k^N(\iota)) - \overline{f}^N(\iota)][\widetilde{f}^N(X_j^N(\iota)) - \overline{f}^N(\iota)] I_{\{\overline{\alpha}_k^\varepsilon = \iota\}} \\
&= K\varepsilon E_t^\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor -1} \sum_{k \geq j} \sum_{\iota_1, \iota \in \overline{\mathcal{M}}} E_j \widehat{f}_k^N(\iota) \widehat{f}_j^N(\iota) E_j I_{\{\overline{\alpha}_k^\varepsilon = \iota\}}
\end{aligned} \tag{4.14}$$

Because $\overline{\alpha}_k^\varepsilon$ is an aggregation of α_k^ε, it may not be a Markov chain, but the value is bounded. It follows that for $j < k$,

$$E_j I_{\{\overline{\alpha}_k^\varepsilon = \iota\}} = P(\overline{\alpha}_k^e = \iota | \mathcal{F}_j^\varepsilon) \leq 1.$$

Because $\{X_k(\iota)\}$ is a mixing sequence, the truncated sequence $\{\widehat{f}^N(j,\iota)\}$ is a bounded mixing sequence. Because $\widehat{f}^N(j,\iota_2)$ is $\mathcal{F}_j^\varepsilon$-measurable, [7, Lemma 4, p.82] implies that

$$E_j \widehat{f}_k^N(\iota) \widehat{f}_j^N(\iota) = \widehat{f}_j^N(\iota) E_j \widehat{f}_k^N(\iota) \leq \widetilde{\phi}(k-j) \widehat{f}_j^N(\iota). \tag{4.15}$$

Note that $\sum_{k \geq j} \widetilde{\phi}(k-j) < \infty$, that $E_t^\varepsilon \widehat{f}_j^N(\iota_2)$ is a bounded random variable that is $\mathcal{F}_t^\varepsilon$-

measurable, and that

$$\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor - 1} K \leq K\varepsilon(t+s-t)/\varepsilon = Ks \leq K\delta.$$

We have that there is a random variable $\gamma^\varepsilon(t)$ that is $\mathcal{F}_t^\varepsilon$-measurable such that

$$\begin{aligned} &\varepsilon E_t^\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor - 1} \sum_{k \geq j} \widetilde{\phi}(k-j) \widehat{f}_j^N(\iota) \\ &\leq K\varepsilon E_t^\varepsilon \sum_{j=\lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor - 1} \widehat{f}_j^N(\iota) \\ &\leq K\gamma^\varepsilon(\delta). \end{aligned} \tag{4.16}$$

Combining (4.14)-(4.16), we have

$$E_t^\varepsilon |y^{\varepsilon,N}(t+s) - y^{\varepsilon,N}(t)|^2 \leq K\gamma^\varepsilon(\delta)$$

such that

$$\lim_{\delta \to 0} \limsup_{\varepsilon \to 0} E\gamma^\varepsilon(t) = 0.$$

Thus, [7, Theorem 3, p. 47] yields that $\{y^{\varepsilon,N}(\cdot)\}$ is tight. $\square$

Next, we obtain another lemma.

Lemma 4.0.4. *Under assumption* (A1)*,* $\{\overline{\alpha}^\varepsilon(\cdot)\}$ *is tight in* $D[0,T]$.

The proof is given in [26, Theorem 4.3]. We omit the details. The next result is a direct consequence of Lemma 4.0.3 and Lemma 4.0.4.

Corollary 4.0.5. *Under* (A1)-(A3)*,* $\{y^{\varepsilon,N}(\cdot), \overline{\alpha}^\varepsilon(\cdot)\}$ *is tight.*

Because $\{y^{\varepsilon,N}(\cdot), \overline{\alpha}^\varepsilon(\cdot)\}$ is tight, by Prohorov's theorem [8, Chapter 7], the pair of sequences is sequentially compact. Select a weakly convergent subsequence with limit denoted by $(y^N(\cdot), \overline{\alpha}(\cdot))$. Without loss of generality, we still index the sequence by ε. We proceed to characterize the limit process. We shall show that the limit is the solution of the martingale problem with operator

$$\mathcal{L}^N h(y,\iota) = \frac{1}{2} a^N(\iota) \frac{d^2 h(y,\iota)}{dy^2} + \overline{Q} h(y,\cdot)(\iota), \ \iota \in \overline{\mathcal{M}}, \tag{4.17}$$

where $a^N(\iota)$ is defined as

$$a^N(\iota) = E[\widehat{f}^N(0,\iota)]^2 + 2\sum_{j=1}^{\infty} E\widehat{f}^N(0,\iota)\widehat{f}^N(j,\iota), \tag{4.18}$$

with $\widehat{f}^N(i,\iota)$ defined in (4.13). In the above, again, we are using $g^N(y) = g(y)q_N(y)$ with the $q_N(y)$ being the truncation function.

To proceed, it is sometimes convenient to write $y_k^{\varepsilon,N}$ in recursive form. It is easily verified that

$$y_{k+1}^{\varepsilon,N} = y_k^{\varepsilon,N} + \sqrt{\varepsilon} \sum_{\iota \in \overline{\mathcal{M}}} \widehat{f}^N(k,\iota) I_{\{\overline{\alpha}_k^{\varepsilon} = \iota\}}. \tag{4.19}$$

For each $\iota \in \overline{\mathcal{M}}$, let $h(\cdot,\iota) \in C_0^2$ (where C_0^2 denotes the class of functions that are C^2 with compact support). Let $\overline{h}(y,\alpha)$ be defined as

$$\overline{h}(y,\alpha) = \begin{cases} h(y,1), & \text{if } \alpha \in \mathcal{M}_1, \\ \ldots\ \ldots & \\ h(y,l_0), & \text{if } \alpha \in \mathcal{M}_{l_0}. \end{cases} \tag{4.20}$$

Using the structure of $\overline{h}(y, \alpha)$, it is readily verified that

$$\begin{aligned} & h(y^{\varepsilon,N}(t+s), \overline{\alpha}^{\varepsilon}(t+s)) - h(y^{\varepsilon,N}(t), \overline{\alpha}^{\varepsilon}(t)) \\ & \quad = \overline{h}(y^{\varepsilon,N}(t+s), \alpha^{\varepsilon}(t+s)) - \overline{h}(y^{\varepsilon,N}(t), \alpha^{\varepsilon}(t)). \end{aligned} \tag{4.21}$$

Choose a sequence $\{n_\varepsilon\}$ such that $n_\varepsilon \to \infty$ as $\varepsilon \to 0$, but $\delta_\varepsilon = \varepsilon n_\varepsilon \to 0$. Then

$$\begin{aligned} & \overline{h}(y^{\varepsilon,N}(t+s), \alpha^{\varepsilon}(t+s)) - \overline{h}(y^{\varepsilon,N}(t), \alpha^{\varepsilon}(t)) \\ & \quad = \sum_{ln_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon + n_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon})] \\ & \quad = \sum_{ln_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon})] \\ & \qquad + \sum_{ln_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon + n_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon})]. \end{aligned} \tag{4.22}$$

Note that because of the continuity of $h(\cdot)$ and hence the continuity of $\overline{h}(\cdot)$, we can replace

$$\sum_{ln_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon + n_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon + n_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon})]$$

with

$$\sum_{ln_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon + n_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^{\varepsilon}_{ln_\varepsilon})]$$

since $\varepsilon n_\varepsilon \to 0$. Furthermore, we can write the above sum as

$$\begin{aligned}
&\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} [\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^\varepsilon_{ln_\varepsilon+n_\varepsilon}) - \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, \alpha^\varepsilon_{ln_\varepsilon})] \\
&\quad = \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \sum_{1\le j\le m_i, 1\le i\le l_0} \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, s_{ij}) I_{\{\alpha^\varepsilon_k=s_{ij}\}} \\
&\quad = \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \sum_{1\le j\le m_i, 1\le i\le l_0} \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, s_{ij}) [I_{\{\alpha^\varepsilon_k=s_{ij}\}} - \nu^i_j I_{\{\alpha^\varepsilon_k\in\mathcal{M}_i\}}] \\
&\qquad + \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \sum_{1\le j\le m_i, 1\le i\le l_0} \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, s_{ij}) \nu^i_j I_{\{\alpha^\varepsilon_k\in\mathcal{M}_i\}}
\end{aligned} \tag{4.23}$$

First, we state a lemma, whose proof is essentially in [26].

Lemma 4.0.6. *Under condition* (A1)*,*

$$E^\varepsilon_m \Big\{ \sum_{k=m}^{n-1} [I_{\{\alpha^\varepsilon_k=s_{ij}\}} - \nu^i_j I_{\{\alpha^\varepsilon_k\in\mathcal{M}_i\}}] \Big\}^2 = \mathcal{O}(\varepsilon).$$

With Lemma 4.0.6 at our hand, for any positive integer κ, any $t_m \le t$ with $m \le \kappa$, any $t, s > 0$, and any bounded and continuous function $g(\cdot, i)$ (for each $i \in \overline{\mathcal{M}}$), it is readily seen that

$$\begin{aligned}
&Eg(\overline{h}(y^{\varepsilon,N}(t_m), \alpha^\varepsilon(t_m) : m \le \kappa) \\
&\quad \times \Big[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \sum_{1\le j\le m_i, 1\le i\le l_0} \overline{h}(y^{\varepsilon,N}_{ln_\varepsilon}, s_{ij}) [I_{\{\alpha^\varepsilon_k=s_{ij}\}} - \nu^i_j I_{\{\alpha^\varepsilon_k\in\mathcal{M}_i\}}] \Big] \\
&\qquad\qquad \to 0 \text{ as } \varepsilon \to 0.
\end{aligned} \tag{4.24}$$

Moreover, we can further show that

$$\begin{aligned}
&Eg(\overline{h}(y^{\varepsilon,N}(t_m),\alpha^{\varepsilon}(t_m):m\le\kappa)\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{1\le j\le m_i,1\le i\le l_0}\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon},s_{ij})\nu^i_j I_{\{\alpha^\varepsilon_k\in\mathcal{M}_i\}}\\
&\to Eh(y^N(t_m),\overline{\alpha}(t_m):m\le\kappa)\int_t^{t+s}\overline{Q}h(y^N(u),\iota)I_{\{\overline{\alpha}(u)=\iota\}}du\\
&=Eh(y^N(t_m),\overline{\alpha}(t_m):m\le\kappa)\int_t^{t+s}\overline{Q}h(y^N(u),\overline{\alpha}(u))du.
\end{aligned}\tag{4.25}$$

In the above, we have used $h(y,\overline{\alpha}^\varepsilon(u))=\overline{h}(y,\alpha^\varepsilon(u))$, the weak convergence of $(y^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ to $(y^N(\cdot),\overline{\alpha}(\cdot))$, and the Skorohod representation.

Next we work on the terms of the third line of (4.22). We have

$$\begin{aligned}
&\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon}[\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon+n_\varepsilon},\alpha^\varepsilon_{ln_\varepsilon})-\overline{h}(y^{\varepsilon,N}_{ln_\varepsilon},\alpha^\varepsilon_{ln_\varepsilon})]\\
&\quad=\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon}h_y(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^\varepsilon_{ln_\varepsilon})[y^{\varepsilon,N}_{ln_\varepsilon+n_\varepsilon}-y^{\varepsilon,N}_{ln_\varepsilon}]\\
&\qquad+\frac{1}{2}h_{yy}(y^{\varepsilon,N,+}_{ln_\varepsilon},\overline{\alpha}^\varepsilon_{ln_\varepsilon})[y^{\varepsilon,N}_{ln_\varepsilon+n_\varepsilon}-y^{\varepsilon,N}_{ln_\varepsilon}]^2\\
&\quad=\sqrt{\varepsilon}\sum_{\iota\in\overline{\mathcal{M}}}\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}h_y(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^\varepsilon_{ln_\varepsilon})\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\widehat{f}^N_j(\iota)I_{\{\overline{\alpha}^\varepsilon_j=\iota\}}\\
&\qquad+\varepsilon\frac{1}{2}\sum_{\iota\in\overline{\mathcal{M}}}\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}h_{yy}(y^{\varepsilon,N,+}_{ln_\varepsilon},\overline{\alpha}^\varepsilon_{ln_\varepsilon})\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\widehat{f}^N_j(\iota)I_{\{\overline{\alpha}^\varepsilon_j=\iota\}}\Big]^2,
\end{aligned}\tag{4.26}$$

where h_y and h_{yy} denote the first and the second derivatives of h with respect to y, respectively, $y^{\varepsilon,N,+}_{ln_\varepsilon}$ denotes the quantity on the line segment joining $y^{\varepsilon,N}_{ln_\varepsilon}$ and $y^{\varepsilon,N}_{ln_\varepsilon+n_\varepsilon}$.

We next average out the terms in the last two lines of (4.26). For any positive integer κ, any $t_m\le t$ with $m\le\kappa$, any $t,s>0$, and any bounded and continuous function $g(\cdot,i)$

(for each $i \in \overline{\mathcal{M}}$), because $h_y(y^{\varepsilon,N}_{ln_\varepsilon}, \overline{\alpha}^\varepsilon_{ln_\varepsilon})$ is $\mathcal{F}_{ln_\varepsilon}$-measurable, we have

$$\begin{aligned}
&\sqrt{\varepsilon} E g(y^{\varepsilon,N}(t_m), \overline{\alpha}^\varepsilon(t_m) : m \le \kappa, t_m \le t) \sum_{\iota \in \overline{\mathcal{M}}} \sum_{ln_\varepsilon = \lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor} h_y(y^{\varepsilon,N}_{ln_\varepsilon}, \overline{\alpha}^\varepsilon_{ln_\varepsilon}) \sum_{j=ln_\varepsilon}^{ln_\varepsilon + n_\varepsilon - 1} \widehat{f}^N_j(\iota) I_{\{\overline{\alpha}^\varepsilon_j = \iota\}} \\
&= \sqrt{\varepsilon} E g(y^{\varepsilon,N}(t_m), \overline{\alpha}^\varepsilon(t_m) : m \le \kappa, t_m \le t) \sum_{\iota \in \overline{\mathcal{M}}} \sum_{ln_\varepsilon = \lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor} h_y(y^{\varepsilon,N}_{ln_\varepsilon}, \overline{\alpha}^\varepsilon_{ln_\varepsilon}) \\
&\qquad\qquad \times E_{ln_\varepsilon} \sum_{j=ln_\varepsilon}^{ln_\varepsilon + n_\varepsilon - 1} \widehat{f}^N_j(\iota) I_{\{\overline{\alpha}^\varepsilon_j = \iota\}} \\
&\to 0 \ \text{ as } \ \varepsilon \to 0,
\end{aligned} \tag{4.27}$$

by the mixing property of the sequence $\widehat{f}^N_j$.

Next for the last term, we note that the limit of

$$\begin{aligned}
&\frac{\varepsilon}{2} E g(y^{\varepsilon,N}(t_m), \overline{\alpha}^\varepsilon(t_m) : m \le \kappa, t_m \le t) \\
&\times \sum_{\iota \in \overline{\mathcal{M}}} \sum_{ln_\varepsilon = \lfloor t/\varepsilon \rfloor}^{\lfloor (t+s)/\varepsilon \rfloor} h_{yy}(y^{\varepsilon,N,+}_{ln_\varepsilon}, \overline{\alpha}^\varepsilon_{ln_\varepsilon}) \Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon + n_\varepsilon - 1} \widehat{f}^N_j(\iota) I_{\{\overline{\alpha}^\varepsilon_j = \iota\}} \Big]^2
\end{aligned}$$

is the same as that of

$$
\begin{aligned}
&\frac{\varepsilon}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^{\varepsilon}(t_m):m\le\kappa,t_m\le t)\\
&\times\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^{\varepsilon}_{ln_\varepsilon})\sum_{\iota\in\overline{\mathcal{M}}}\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\widehat{f}^N_j(\iota)I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big]^2\\
&\quad=\frac{1}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^{\varepsilon}(t_m):m\le\kappa,t_m\le t)\\
&\qquad\times\sum_{\iota\in\overline{\mathcal{M}}}\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}\delta_\varepsilon h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^{\varepsilon}_{ln_\varepsilon})\frac{1}{n_\varepsilon}\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\widehat{f}^N_j(\iota)I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big]^2\\
&\quad=\frac{1}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^{\varepsilon}(t_m):m\le\kappa,t_m\le t)\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}\delta_\varepsilon h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^{\varepsilon}_{ln_\varepsilon})\\
&\qquad\times\sum_{\iota\in\overline{\mathcal{M}}}\frac{1}{n_\varepsilon}\Big\{\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}E_{ln_\varepsilon}\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)E_{ln_\varepsilon}I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\\
&\qquad\qquad-\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}E\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)E_{ln_\varepsilon}I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big]\\
&\qquad\qquad+\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}E\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)E_{ln_\varepsilon}I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big\}.
\end{aligned}
\tag{4.28}
$$

Moveover, the limit of

$$
\begin{aligned}
&\frac{1}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^{\varepsilon}(t_m):m\le\kappa,t_m\le t)\\
&\times\sum_{\iota\in\overline{\mathcal{M}}}\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}\delta_\varepsilon h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^{\varepsilon}_{ln_\varepsilon})\frac{1}{n_\varepsilon}\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}[E_{ln_\varepsilon}\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\\
&\qquad\qquad-E\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)][E_{ln_\varepsilon}I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big]
\end{aligned}
$$

is the same as that of

$$
\begin{aligned}
&\frac{1}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^{\varepsilon}(t_m):m\le\kappa,t_m\le t)\\
&\times\sum_{\iota\in\overline{\mathcal{M}}}\sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor(t+s)/\varepsilon\rfloor}\delta_\varepsilon h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^{\varepsilon}_{ln_\varepsilon})\frac{1}{n_\varepsilon}\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k>j}^{ln_\varepsilon+n_\varepsilon-1}[E_{ln_\varepsilon}\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)-E\widehat{f}^N_j(\iota)\widehat{f}^N_k(\iota)]\\
&\qquad\qquad\times E_{ln_\varepsilon}I_{\{\overline{\alpha}^{\varepsilon}_j=\iota\}}\Big].
\end{aligned}
\tag{4.29}
$$

Because $\{\widehat{f}_j(\iota)\}$ is a stationary mixing sequence by (A3), $\{\widehat{f}_j^N(\iota)\}$ is a bounded stationary mixing sequence. With a slight abuse of notation, we still denote the mixing measure by $\widetilde{\phi}(\cdot)$. Then by virtue of [7, Lemma 4, p.82],

$$|E_{ln_\varepsilon}\widehat{f}_j^N(\iota)\widehat{f}_k^N(\iota) - E\widehat{f}_j^N(\iota)\widehat{f}_k^N(\iota)| \le K\widetilde{\phi}^{1/2}(k-j)\widehat{\phi}^{1/2}(j-ln_\varepsilon).$$

Moreover, in view of (A3),

$$\sum_\iota \widetilde{\phi}^{1/2}(\iota) < \infty.$$

It then yields that for some $K > 0$,

$$\begin{aligned}
&\frac{1}{2}Eg(y^{\varepsilon,N}(t_m),\overline{\alpha}^\varepsilon(t_m) : m \le \kappa, t_m \le t) \\
&\quad \times \sum_{\iota\in\overline{\mathcal{M}}} \sum_{ln_\varepsilon=\lfloor t/\varepsilon\rfloor}^{\lfloor (t+s)/\varepsilon\rfloor} \delta_\varepsilon h_{yy}(y^{\varepsilon,N}_{ln_\varepsilon},\overline{\alpha}^\varepsilon_{ln_\varepsilon})\frac{1}{n_\varepsilon}\Big[\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k>j}^{ln_\varepsilon+n_\varepsilon-1} [E_{ln_\varepsilon}\widehat{f}_j^N(\iota)\widehat{f}_k^N(\iota) - E\widehat{f}_j^N(\iota)\widehat{f}_k^N(\iota)] \\
&\qquad\qquad \times E_{ln_\varepsilon} I_{\{\overline{\alpha}_j^\varepsilon=\iota\}}\Big] \\
&\quad \to 0 \text{ as } \varepsilon \to 0.
\end{aligned} \tag{4.30}$$

Next, the bounded stationary mixing implies the ergodicity. As a result, in view of (A

$$\begin{aligned}
&\frac{1}{n_\varepsilon}\sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1}\sum_{k=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} E\widehat{f}_j^N(\iota)\widehat{f}_k^N(\iota) \\
&\quad \to E[\widehat{f}^N(\iota)]^2 + 2\sum_{j=1}^\infty E[\widehat{f}_j^N(\iota)\widehat{f}_0^N(\iota)] \text{ as } \varepsilon \to 0 \\
&\quad = a^N(\iota);
\end{aligned} \tag{4.31}$$

see (4.18). That is, $a^N(\iota)$ is similar to $a(\iota)$, except a truncation device is used. Note that $E_{ln_\varepsilon} I_{\{\overline{\alpha}_j^\varepsilon=\iota\}} = P(\overline{\alpha}_j^\varepsilon = \iota|\mathcal{F}^\varepsilon_{ln_\varepsilon})$ and $(y^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(y^N(\cdot),\overline{\alpha}(\cdot))$. Combining the results obtained thus far, we obtain that $(y^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(y^N(\cdot), \overline{\alpha}(\cdot))$ as $\varepsilon \to 0$, where $(y^N(\cdot), \overline{\alpha}(\cdot))$ is the solution of the martingale problem with operator $\mathcal{L}^N$, where $\mathcal{L}^N$ is the operator defined in (4.17). Summarizing what have been obtained, we have the following theorem.

Theorem 4.0.7. *Under assumptions* (A1)–(A3)*, as* $\varepsilon \to 0$*,* $(y^{\varepsilon,N}(\cdot), \overline{\alpha}^{\varepsilon}(\cdot))$ *converges weakly to* $(y^N(\cdot), \overline{\alpha}(\cdot))$*, which is the solution of the martingale problem with operator* $\mathcal{L}^N$*.*

Next, we show that the untruncated sequence $(y^{\varepsilon}(\cdot), \overline{\alpha}^{\varepsilon}(\cdot))$ is also convergent. This step is done by sending $N \to \infty$.

Theorem 4.0.8. *Under the conditions of Theorem* 4.0.7*, as* $\varepsilon \to 0$*,* $(y^{\varepsilon}(\cdot), \overline{\alpha}^{\varepsilon}(\cdot))$ *converges weakly to* $(y(\cdot), \overline{\alpha}(\cdot))$ *such that the limit is the solution of the martingale problem with operator* $\mathcal{L}$*.*

The idea of proof follows from that of Corollary [7, p.46]. So we will be very brief. The essential idea is on the use of the uniqueness of the martingale problem. Denote by $P(\cdot)$ and $P^N(\cdot)$ the measures induced by $y(\cdot)$ and $y^N(\cdot)$ on the Borel subsets of $D[0,T]$, respectively. The measure P is unique because of the uniqueness of the martingale problem. Thus, for each $T_1 \le T$, $P(\cdot)$ agrees with $P^N(\cdot)$ on all Borel subsets of the set of paths in $D[0,T]$ whose values are in S_N for each $t \le T_1$. However, $P(\sup_{t \le T_1} |y(t)| \le N) \to 1$ as $N \to \infty$. This implies that $y^{\varepsilon,N}(\cdot)$ converges weakly to $y^N(\cdot)$ implies $y^{\varepsilon}(\cdot)$ converges weakly to $y(\cdot)$. Because that the limit is unique, the chosen subsequence is irrelevant.

CHAPTER 5 SPECIALIZATION AND EXTENSION

In this chapter, we look at two cases. In the first case, the modulating Markov chain is given by $P^\varepsilon = P + \varepsilon Q$, where P is irreducible. In the second case, we allow the P to include transient states.

5.1 Modulating Markov Chain with P Being Irreducible

We consider the case that the modulating Markov chain is $P^\varepsilon = P + \varepsilon Q$, where Q is a generator of a continuous-time Markov chain as before, and P itself is a transition matrix that is irreducible and aperiodic. The state space of the Markov chain is $\mathcal{M} = \{1, \ldots, m\}$. It is easily seen that in this case, the stationary distribution associated with the transition matrix P is $\nu = (\nu_1, \ldots, \nu_m) \in \mathbb{R}^{1\times m}$. In this case, it can be proved as in [26], the asymptotic expansions of the transition matrices can be constructed so that

$$(P^\varepsilon)^k = \mathbb{1}\nu + \mathcal{O}(\varepsilon + \lambda^k),$$

where $0 < \lambda < 1$ and $\mathbb{1} = (1, \ldots, 1)' \in \mathbb{R}^{m\times 1}$. As in the previous chapters, we define

$$\begin{aligned}
\widetilde{y}_n \quad &= \sqrt{\varepsilon}\sum_{k=0}^{n-1}(f_k(i) - Ef_k(i))[I_{\{\alpha_k^\varepsilon = i\}} - \nu_i] \\
&= \sqrt{\varepsilon}\sum_{k=0}^{n-1}(\widetilde{f}(X_k(i)) - E\widetilde{f}(X_k(i)))[I_{\{\alpha_k^\varepsilon = i\}} - \nu_i], \\
\widetilde{y}^\varepsilon(t) &= \widetilde{y}_n, \quad \text{for } \ t \in [\varepsilon n, \varepsilon n + \varepsilon),
\end{aligned} \tag{5.1}$$

where $f_k(i) = \widetilde{f}(X_k(i))$ is as defined in (3.10). For simplicity, define

$$\widehat{f}_k(i) = f_k(i) - \overline{f}_k(i) = f_k(i) - Ef_k(i).$$

Using the techniques of previous chapter, we can obtain the following limit theorem. For brevity, we omit the verbatim proof.

Proposition 5.1.1. *Assume that the P is as given in this section, and that* (A2) *and* (A3) *hold. Then $\widetilde{y}^{\varepsilon}(\cdot)$ converges weakly to a real-valued Brownian motion with variance $\widetilde{\sigma}^2 t$, where*

$$\begin{aligned}\widetilde{\sigma}^2 = \sum_{i=1}^{m} \Big[\nu_i(1-\nu_i)E\widehat{f}_0^2(i) + 2\sum_{k=1}^{\infty}\nu_i\psi_{ii}(k)\big[E\widehat{f}_0(i)\widehat{f}_k(i)\big]\Big] \\ + \sum_{1\le i<j\le m}\Bigg\{\sum_{k=1}^{\infty}\Big[2\nu_i\psi_{ij}(k)2E\widehat{f}_0(i)\widehat{f}_k(j) \\ + 2\nu_j\psi_{ji}(k)E\widehat{f}_0(j)\widehat{f}_k(i)\Big] \\ - 2\nu_i\nu_j E\widehat{f}_0(i)\widehat{f}_0(j)\Bigg\},\end{aligned} \tag{5.2}$$

where $\psi_{ij}(k)$ denotes the ijth entry of the matrix $\Psi(k)$ as in the asymptotic expansions (6.16).

5.2 Modulating Markov Chain Including Transient States

In a finite-state Markov chain, there is at least one recurrent state (i.e., not all states are transient). Iosifescu [6, p. 94] has indicated that any transition probability matrix of a finite-state Markov chain with stationary transition probabilities can be put into either the form of (3.5) (i.e., inclusion of all recurrent states) of the form

$$P = \begin{pmatrix} P^1 & & & & \\ & P^2 & & & \\ & & \ddots & & \\ & & & P^{l_0} & \\ P^{*,1} & P^{*,2} & \cdots & P^{*,l_0} & P^* \end{pmatrix}, \tag{5.3}$$

where each P^i is itself a transition matrix within the ιth recurrent class $\mathcal{M}_\iota$ for $\iota \le l_0$. In fact, (5.3) takes care of the situation that there are transient states. The last row of the partitioned matrix in (5.3), namely, $(P^{*,1}, \cdots P^{*,l_0}, P^*)$ in (5.3) is resulted from the inclusion of the transient states. It turns out $P^{*,\iota}$, $\iota = 1, \dots, l_0$, are the transition probabilities from the transient states to the recurrent states in $\mathcal{M}_\iota$, and P^* is the transition probabilities within the transient states. We assume the following conditions hold.

(A1') For the transition matrix P given by (5.3), for each $\iota \le l_0$, P^ι is transition probability matrix that is irreducible and aperiodic. In addition, P^* is a matrix having all of its eigenvalues inside the unit circle.

To proceed, we need some notation. Define $\widetilde{\mathbb{1}}$ and $\widetilde{\mathbb{1}}_*$ as

$$\begin{aligned} \widetilde{\mathbb{1}} &= \operatorname{diag}(\mathbb{1}_{m_1}, \dots, \mathbb{1}_{m_{l_0}}) = \begin{pmatrix} \mathbb{1}_{m_1} & & \\ & \ddots & \\ & & \mathbb{1}_{m_{l_0}} \end{pmatrix} \quad \text{and} \\ \widetilde{\mathbb{1}}_* &= \begin{pmatrix} \widetilde{\mathbb{1}} & 0_{(m_0 - m_*)\times m_*} \\ \Xi_* & 0_{m_* \times m_*} \end{pmatrix}, \end{aligned} \tag{5.4}$$

where $0_{l_1 \times l_2}$ denotes a 0 matrix with dimension $l_1 \times l_2$, and

$$\Xi_* = (\xi^1, \dots, \xi^{l_0}) \in \mathbb{R}^{m_* \times (m_1 + \cdots + m_{l_0})} \tag{5.5}$$

with

$$\xi^\iota = -(P^* - I)^{-1} P^{*,\iota} \mathbb{1}_{m_\iota}, \quad \iota = 1, \dots, l_0. \tag{5.6}$$

Note that $\xi^\iota \ge 0$ for $\iota = 1, \dots, l_0$ (in the sense that all components of ξ^ι are nonnegative).

Moreover, using

$$P^{*}\mathbb{1}_{m_*} + \sum_{\iota=1}^{l_0} P^{*,\iota}\mathbb{1}_{m_\iota} = \mathbb{1}_{m_*},$$

it is easily verified that

$$\sum_{\iota=1}^{l_0} \xi^{\iota} = \mathbb{1}_{m_*}. \tag{5.7}$$

That is, for each ι, ξ^{ι} represents the probability of transition from the transient states to the ιth recurrent class $\mathcal{M}_{\iota}$.

We can still aggregate the states in α_k^{ε}. Nevertheless, we only aggregate the states in each recurrent class, but do not aggregate states in the transient class. Eventually, these transient states will go to the respective recurrent classes. To proceed, partition the matrix Q as

$$Q = \begin{pmatrix} Q^{11} & Q^{12} \\ Q^{21} & Q^{22} \end{pmatrix}, \tag{5.8}$$

where

$$Q^{11} \in \mathbb{R}^{(m_0-m_*)\times(m_0-m_*)},\ Q^{12} \in \mathbb{R}^{(m_0-m_*)\times m_*},$$
$$Q^{21} \in \mathbb{R}^{m_*\times(m_0-m_*)},\ \text{ and } Q^{22} \in \mathbb{R}^{m_*\times m_*},$$

and set

$$\overline{Q}_* = \operatorname{diag}(\nu^1, \dots, \nu^{l_0})(Q^{11}\widetilde{\mathbb{1}} + Q^{12}A_*). \tag{5.9}$$

Denote

$$\nu_* = \operatorname{diag}(\nu^1, \dots, \nu^{l_0}, 0_{m_*\times m_*}).$$

Next, we define an aggregated process

$$\overline{\alpha}_k^\varepsilon = \begin{cases} \iota, \text{ if } \alpha_k^\varepsilon \in \mathcal{M}_\iota, \\ U_{\iota_1}, \text{ if } \alpha_k^\varepsilon = s_{*\iota_1}, \end{cases} \tag{5.10}$$

where U_{ι_1} is given by

$$U_{\iota_1} = I_{\{0 \le U \le \xi^{1,\iota_1}\}} + 2I_{\{\xi^{1,\iota_1} < U \le \xi^{1,\iota_1} + \xi^{2,\iota_1}\}} + \cdots + l_0 I_{\{\xi^{1,\iota_1} + \cdots + \xi^{l_0 - 1,\iota_1} < U \le 1\}},$$

with U being a random variable that is independent of the Markov chain α_k^ε and that is uniformly distributed on the unit interval $[0,1]$. Define also the interpolation of $\overline{\alpha}_k$ by

$$\overline{\alpha}^\varepsilon(t) = \overline{\alpha}_k \text{ for } t \in [\varepsilon k, \varepsilon k + \varepsilon). \tag{5.11}$$

With the preparation above, we can proceed to study the limit of the modulated matrix sequences. We consider $\{X_k(\overline{\alpha}_k^\varepsilon)\}$, but this time the α_k^ε includes the transient states as given in this section. Redefine $\widetilde{f}(X_k(\iota))$ and $f_k(\iota)$ as in (3.10), but $f_k(\overline{\alpha}_k^\varepsilon)$ is define with $\overline{\alpha}_k^\varepsilon$ given by (5.10). Modify the definition of the operator $\mathcal{L}$ with

$$\widehat{\mathcal{L}}h(y,\iota) = \frac{1}{2}a(\iota)\frac{d^2h(y,\iota)}{dy^2} + \overline{Q}_*h(y,\cdot)(\iota), \ \iota \in \overline{\mathcal{M}}, \tag{5.12}$$

where

$$\overline{Q}_*h(y,\cdot)(\iota) = \sum_{\iota_1=1}^{l_0} \overline{q}_{*,\iota\iota_1}h(y,\iota_1), \tag{5.13}$$

where $\overline{Q}_*$ is given by (5.9) and $\overline{q}_{*,\iota\iota_1}$ denotes the $\iota\iota_1$th entry of $\overline{Q}_*$.

Define the centered and normalized sequence $\widehat{y}_k^\varepsilon$ as in (4.2) but with the aggregated process $\overline{\alpha}_k^e$ defined by (5.10). Define also $\widehat{y}^\varepsilon(t) = \widehat{y}_k^\varepsilon$ for $t \in [\varepsilon k, \varepsilon k + \varepsilon)$. Using Lemma 6.3.3 and modifying the proofs in the previous sections, we obtain the following result.

Theorem 5.2.1. *Assume conditions* (A1'), (A2)*, and* (A3) *hold. Then* $(\widehat{y}^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ *converges weakly to* $(\widehat{y}(\cdot), \overline{\alpha}(\cdot))$. *The limit process* $(\widehat{y}(\cdot), \overline{\alpha}(\cdot))$ *is the solution of the martingale problem with operator* $\widehat{\mathcal{L}}$ *defined in* (5.12).

CHAPTER 6 FURTHER REMARKS AND RAMIFICATION

6.1 A Remark on Non-Zero Drift

This dissertation concentrates on the use of a martingale problem formulation. It has been shown that (e.g., in Section 4), the limit $(y(\cdot), \overline{\alpha}(\cdot))$ is a solution of the martingale problem with operator $\mathcal{L}$. Such a result can also be described by means of limit stochastic differential equation. In fact, if we consider the process $y(\cdot)$, we can write it as

$$y(t) = \int_0^t \sqrt{a(\overline{\alpha}(s))} dW(s),$$

where $W(\cdot)$ is a real-valued standard Brownian motion. Similar results hold for the case of Markov modulated sequence including the transient states.

In view of the scaling, we could include a drift term. For example, for the results in Section 4, we may consider another nonlinear function $g(\cdot)$ and denote

$$\begin{aligned} & g_k(\iota) = \widetilde{g}(X_k(\iota)) \ \text{ for each } \ \iota \in \overline{\mathcal{M}} \ \text{ so} \\ & g_k(\overline{\alpha}_k^\varepsilon) = \widetilde{g}(X_k(\overline{\alpha}_k^\varepsilon)). \end{aligned} \tag{6.1}$$

Define a scaled sequence as

$$\begin{aligned} & z_k^\varepsilon = \varepsilon \sum_{j=0}^{k-1} \sum_{\iota \in \overline{\mathcal{M}}} g_j(\iota) I_{\{\overline{\alpha}_j^\varepsilon = \iota\}} + \sqrt{\varepsilon} \sum_{j=0}^{k-1} \sum_{\iota \in \overline{\mathcal{M}}} [\widetilde{f}(X_j(\iota)) - \overline{f}(\iota)] I_{\{\overline{\alpha}_j^\varepsilon = \iota\}} \\ & z^\varepsilon(t) = z_k^\varepsilon \ \text{ for } \ t \in [\varepsilon k, \varepsilon k + \varepsilon). \end{aligned} \tag{6.2}$$

Then using the techniques of Section 4, $(z^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(z(\cdot), \overline{\alpha}(\cdot))$ such that $z(\cdot)$ is a solution of the following stochastic equation

$$z(t) = \int_0^t \overline{g}(\overline{\alpha}(s))ds + \int_0^t \sqrt{a(\overline{\alpha}(s))}dW(s),$$

where

$$\overline{g}(\iota) = E\widetilde{g}(X_k(\iota)) \text{ for } \iota \in \mathcal{M}. \tag{6.3}$$

Alternatively, the above stochastic integral equation may be written as a stochastic differential equation

$$dz(t) = \overline{g}(\overline{\alpha}(t))dt + \sqrt{a(\overline{\alpha}(t))}dW(t).$$

6.2 Ramification

We have studied sequences of matrix-valued random elements by taking a functional to map the matrix-valued processes into real-valued processes. This section takes a different approach by dealing with the matrix-valued processes directly. Specifically, we aim to obtain exponential-type of probability upper bounds. The problem is interesting in their own right. First let us state the following condition.

(A4) There is a Markov chain α_k with a finite state space $\mathcal{M} = \{1, \ldots, m\}$ and transition matrix P such that P is irreducible and aperiodic. For each $i \in \mathcal{M}$, $\{X_k(i)\}$ is a sequence of independent and identically distributed $d_1 \times d_2$-valued normal random variables;

$$X_1(i) \sim N_{d_1,d_2}(M(i), \Psi_1 \otimes \Psi_2). \tag{6.4}$$

The Markov chain α_k and $\{X_k(i)\}$ are independent.

Remark 6.2.1. In this section, we still assume that there is a finite-state Markov chain and that the m sequences $\{X_k(i)\}$ for $i = 1, \ldots, m$ are modulated by α_k. However, we assume the chain is irreducible and has a simpler structure.

Note that (6.4) indicates that $X_1(i) \in \mathbb{R}^{d_1 \times d_2}$ is normally distributed with mean $M(i) \in \mathbb{R}^{d_1 \times d_2}$ and covariance $\Psi_1 \otimes \Psi_2$, where both $\Psi_1 \in \mathbb{R}^{d_1 \times d_1}$ and $\Psi_2 \in \mathbb{R}^{d_2 \times d_2}$ are symmetric positive definite, and $\otimes$ denotes the Kronecker product [3]. Note that in view of our assumption, for different $i \in \mathcal{M}$, only the means of $X_1(i)$ are different. The covariances are the same for all $i \in \mathcal{M}$. This is more or less for convenience in the subsequent development although different covariance matrices can be dealt with.

It is known that

(i) by definition [4, p. 55], $X_1(i) \sim N_{d_1,d_2}(M(i), \Psi_1 \otimes \Psi_2)$ if and only if the vectorization vec$(X_1'(i))$ is a $(d_1 d_2)$-dimensional normal vector satisfying

$$\mathrm{vec}(X_1'(i)) \sim N(\mathrm{vec}(M'(i)), \Psi_1 \otimes \Psi_2);$$

(ii) the density function of $X_1(i)$ is given by ([4, p. 56])

$$\begin{aligned} \widetilde{f}(x) = {} & (2\pi)^{-\frac{1}{2}d_1 d_2}[\det(\Psi_1)]^{-\frac{1}{2}d_2}[\det(\Psi_2)]^{-\frac{1}{2}d_1} \\ & \times \exp\Big\{\mathrm{tr}\big[-\frac{1}{2}\Psi_1^{-1}(x - M(i))\Psi_2^{-1}(x - M(i))'\big]\Big\}; \end{aligned} \tag{6.5}$$

(iii) the moment generating function of $X_1(i)$ is given by

$$\widetilde{G}(z, i) = \exp\Big\{\mathrm{tr}\big(z'M(i) + \frac{1}{2}z'\Psi_1 z \Psi_2\big)\Big\}. \tag{6.6}$$

In addition, the moment generating function of $X_1(i) - M(i)$ is given by

$$G(z) = \exp\Big\{ \mathrm{tr}\big(\frac{1}{2} z' \Psi_1 z \Psi_2\big) \Big\} := e^{\gamma(z)}, \tag{6.7}$$

where $\gamma(z)$ is a positive real-valued function independent of i. We use the $e^{\gamma(z)}$ to simply the notation in the following calculation.

In contrast to the previous sections, we study exponential type of probability bounds for sum of a block of random variables. To begin, for each l, define

$$S_\kappa = \sum_{j=l}^{l+\kappa} \varepsilon_j \left(\sum_{i\in\mathcal{M}} [X_j(i) - M(i)] I_{\{\alpha_j = i\}} \right), \tag{6.8}$$

where $\{\varepsilon_j\}$ is a sequence of positive real numbers satisfying $\varepsilon_j \to 0$ as $j \to \infty$ and $\sum_{j=0}^{\infty} \varepsilon_j = \infty$ and $\sum_{j=0}^{\infty} \varepsilon_j^2 < \infty$. For any $e > 0$, with the motivation of finding the tail probability error bounds of

$$P(\max_{\kappa\le n} |S_\kappa| \ge e) = P\left(\max_{\kappa\le n} \Big| \sum_{j=l}^{l+\kappa} \varepsilon_j [X_j(\alpha_j) - M(\alpha_j)] \Big| \ge e \right), \tag{6.9}$$

we examine $\zeta(e)$ defined as for any $z \in \mathbb{R}^{d_1\times d_2}$ and arbitrary $\beta > 0$,

$$\zeta(e) = P\left(\max_{\kappa\le n} \exp\left\{ \mathrm{tr}\Big(\beta z' \sum_{j=l}^{l+\kappa} \varepsilon_j \big[\sum_{i\in\mathcal{M}} [X_j(i) - M(i)] I_{\{\alpha_j = i\}} \big] \Big) \ge \exp\left(\beta e \frac{z'z}{|z|} \right) \right\} \right), \tag{6.10}$$

but $z'z/|z| = |z|$, the term $\exp(\beta e z'z/|z|)$ above can be replaced by $\exp(\beta e|z|)$.

Redefine by $\mathcal{F}_k$ the σ-algebra generated by $\{X_j(i), \alpha_j : j < k, i = 1, \ldots, m\}$. Note that for two square matrices A and B, $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$. Using the familiar Markov inequality, we obtain

$$\begin{aligned}\zeta(e) &= \exp(-\beta e|z|)E\exp\Big\{\mathrm{tr}\Big(\beta\sum_{j=l}^{l+\kappa}\varepsilon_j\big[\sum_{i\in\mathcal{M}}z'[X_j(i)-M(i)]I_{\{\alpha_j=i\}}\big]\Big)\Big\}\\ &= \exp(-\beta e|z|)EE_{l+\kappa}\exp\Big\{\Big(\beta\sum_{j=l}^{l+\kappa}\varepsilon_j\mathrm{tr}\big[\sum_{i\in\mathcal{M}}z'[X_j(i)-M(i)]I_{\{\alpha_j=i\}}\big]\Big)\Big\}\\ &= \exp(-\beta e|z|)E\exp\Big\{\beta\sum_{j=l}^{l+\kappa-1}\varepsilon_j\sum_{i\in\mathcal{M}}\mathrm{tr}\big(z'[X_j(i)-M(i)]\big)I_{\{\alpha_j=i\}}\Big\}\\ &\qquad\times E_{l+\kappa}\exp\Big[\beta\varepsilon_{\kappa+l}\sum_{i\in\mathcal{M}}\mathrm{tr}\big(z'[X_{\kappa+l}(i)-M(i)]\big)I_{\{\alpha_{\kappa+l}=i\}}\Big]\\ &= \exp(-\beta e|z|)EE_{l+\kappa-1}\exp\Big\{\beta\sum_{j=l}^{l+\kappa-1}\varepsilon_j\sum_{i\in\mathcal{M}}\mathrm{tr}\big(z'[X_j(i)-M(i)]\big)I_{\{\alpha_j=i\}}\Big\}\\ &\qquad\times\exp(\beta^2\varepsilon_{k+l}^2\gamma(z)).\end{aligned}\tag{6.11}$$

To reach the last line in (6.11), we have used the independence of $X_{\kappa+l}(i)$ and $\alpha_{\kappa+l}$, the i.i.d. assumption on $\{X_k(i)\}$, the moment generating function of the form (6.7) that does not depend on i, and noting the property of conditional probability

$$\sum_{i\in\mathcal{M}}p_{\alpha_{\kappa+l-1},i}=1$$

with $p_{\alpha_{\kappa+l-1},i}$ denoting the conditional probability from the state $\alpha_{\kappa+l-1}$ to i, to obtain

$$\begin{aligned}&E_{l+\kappa}\exp\Big[\beta\varepsilon_{\kappa+l}\sum_{i\in\mathcal{M}}\mathrm{tr}\big(z'[X_{\kappa+l}(i)-M(i)]\big)I_{\{\alpha_{\kappa+l}=i\}}\Big]\\ &\quad=\sum_{i\in\mathcal{M}}\exp(\beta^2\varepsilon_{\kappa+l}^2\gamma(z))p_{\alpha_{\kappa+l-1},i}\\ &\quad=\exp(\beta^2\varepsilon_{\kappa+l}^2\gamma(z)).\end{aligned}$$

Working on (6.11), repeatedly taking conditional expectations $E_{l+\kappa-1}$, $E_{l+\kappa-2}$, $E_{l+\kappa-3},\ldots,$

and so on one at a time, we finally arrive at

$$\zeta(e) = \exp\Big(\big(\beta^2\gamma(z)\sum_{j=l}^{\kappa+l}\varepsilon_j^2\big) - \beta e|z|\Big). \tag{6.12}$$

Treating the index of the exponential as a quadratic function of β, it is readily seen that the minimum of the function is reached at

$$\beta^* = \frac{e|z|}{2\gamma(z)\sum_{j=l}^{\kappa+l}\varepsilon_j^2}.$$

Using β^* in (6.12), we then obtain the desired error bound

$$\zeta(e) \le \exp\left(-\frac{e^2|z|^2}{4\gamma(z)\sum_{j=l}^{\kappa+l}\varepsilon_j^2}\right). \tag{6.13}$$

Remark 6.2.2. Note that for any $l > 0$, it can proved that

$$\sum_{j=l}^{l+\kappa}\varepsilon_j\left(\sum_{i\in\mathcal{M}}[X_j(i) - M(i)]I_{\{\alpha_j=i\}}\right) \to 0$$

as $\kappa \to \infty$, which is essentially a law of large numbers type of result. However, for each $e > 0$, one would not be able to obtain any estimate on the tail probability

$$P\Big(\mathrm{tr}\big(\sum_{j=l}^{l+\kappa}\varepsilon_j z'\big(\sum_{i\in\mathcal{M}}[X_j(i) - M(i)]I_{\{\alpha_j=i\}}\big)\big) \ge |z|e\Big)$$

using either law of large numbers or central limit theorems. Now, our result in (6.13) provided such error bounds. Because of our condition $\sum_j \varepsilon_j^2 < \infty$,

$$\sum_{j=l}^{\kappa+l} \varepsilon_j^2 \to 0 \ \text{ as } l \to \infty.$$

Thus, (6.13) indicates that the tail probability is exponentially small. The bound belongs to large deviations type of estimates. The exponential bounds we obtained are particularly useful for applications using stochastic approximation; see [8].

In [8], one mainly dealt with vector-valued processes. In the above, we indicated how matrix-valued estimates can be handled. For simplicity, here we treated independent and identically distributed random elements that have Gaussian distributions. This condition can be much relaxed with the use of correlated noise. What is crucial is that we need a condition on the bounds of the conditional moment generating function similar to the exact moment generating function; see [8, Section 5.3].

6.3 Additional Remarks

This dissertation has been devoted to Markov modulated random sequences that are matrix valued. Our main effort is on obtaining scaling limit of the underlying sequences. We used a functional mapping the matrix-valued processes to $\mathbb{R}$. An alternative approach is to pile up the columns of the matrices into a big vector. Then one can examine the limit process through vectorization.

We finally close this dissertation by mentioning that a number of topics related to the problems considered in this work are worthwhile to be pursued further. It is possible to consider certain jump processes as a component in the primary sequence leading to a jump diffusion type limit with Markov modulation. In addition, it is also a good effort to replace the modulating Markov chain by a semi-Markov process (see [5] for discussions on semi-Markov processes). Much of these require further analysis and consideration.

APPENDIX: PRELIMINARY RESULTS

This chapter presents some preliminary results to be used in the study of the dissertation, in particular, in the study of scaling limit. It includes two parts. The first part considers Markov chains having only recurrent states, whereas the second part considers Markov chains include also transient states.

Recurrent Classes. We state a lemma whose proof can be found in [26, Theorem 4.1 and Theorem 4.3]. In fact, a full asymptotic expansion (or asymptotic series) was developed in [26], but for us in this dissertation, we only need the following results.

Lemma 6.3.1. *Assume* (A1). *Then the following results hold:*

(i) *Consider the probability vector* $p_k^\varepsilon \in \mathbb{R}^{1\times m_0}$. *Then*

$$p_k^\varepsilon = \theta(\varepsilon k)\begin{pmatrix} \nu^1 & & \\ & \ddots & \\ & & \nu^{l_0} \end{pmatrix} + \mathcal{O}(\varepsilon + \lambda^k) \tag{6.14}$$

where λ *is a constant satisfying* $0 < \lambda < 1$, *and* $\theta(t) = (\theta^i(t) : \iota = 1, \ldots, l_0) \in \mathbb{R}^{1\times l_0}$ *is a solution of the following initial value problem*

$$\begin{aligned} \frac{d\theta(t)}{dt} &= \theta(t)Q \\ \theta(0) &= p_0 \begin{pmatrix} \mathbb{1}_{m_1} & & \\ & \ddots & \\ & & \mathbb{1}_{m_{l_0}} \end{pmatrix}, \end{aligned} \tag{6.15}$$

where $\overline{Q}$ *is given by* (4.7), *and* z' *is the transpose of* z.

(ii) *For some finite $T > 0, k \leq T/\varepsilon$, the* kth *step transition probability matrix*

$$(P^\varepsilon)^k = (P + \varepsilon Q)^k$$

satisfies

$$(P^\varepsilon)^k = \Phi(\varepsilon k) + \varepsilon\Phi_1(\varepsilon k) + \Psi(k) + \varepsilon\Psi_1(k) + \mathcal{O}(\varepsilon^2) \tag{6.16}$$

where

$$\Phi(t) = \begin{pmatrix} \mathbb{1}_{m_1} & & \\ & \ddots & \\ & & \mathbb{1}_{m_{l_0}} \end{pmatrix} \begin{pmatrix} \nu^1 & & \\ & \ddots & \\ & & \nu^{l_0} \end{pmatrix} \tag{6.17}$$

and

$$\begin{cases} \dfrac{d\Theta(t)}{dt} & = \Theta(t)\overline{Q} \\ \Theta(0) & = I \end{cases} \tag{6.18}$$

$\Phi(\varepsilon k)$ *and* $\Phi_1(\varepsilon k)$ *are uniformly bounded in* $[0, T]$*, and* $\Psi(k)$ *and* $\Psi_1(k)$ *decay exponentially in that*

$$|\Psi(k)| \leq K\lambda^k \tag{6.19}$$

$$|\Psi_1(k)| \leq K\lambda^k \tag{6.20}$$

for some $K > 0$ and some $0 < \lambda < 1$.

(iii) *Recall $\overline{\alpha}^\varepsilon(\cdot)$ defined in* (3.9). *Then $\overline{\alpha}^\varepsilon(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$, which is a continuous-time Markov chain with generator $\overline{Q}$ defined in* (4.7).

Remark 6.3.2. Let us make a couple of remarks below.

(i) In view of the asymptotic expansion

$$(P^{\varepsilon})^k = \Phi(\varepsilon k) + \mathcal{O}(\varepsilon + \lambda^k) \tag{6.21}$$

(ii) We can write out $\Psi(k)$. It can be chosen so that $\Phi(0) + \Psi(0) = I_{m_0}$, and

$$\Psi(k) = \Psi(0)(P)^k$$
$$\Psi(k) = \begin{pmatrix} (I_{m_1} - \mathbb{1}_{m_1}\nu^1)(P^1)^k & & \\ & \ddots & \\ & & (I_{m_{l_0}} - \mathbb{1}_{m_{l_0}}\nu^{l_0})(P^{l_0})^k \end{pmatrix},$$

where I_l denotes an $l \times l$ identity matrix. For I_{m_0}, we will also use I if there is no confusion.

Inclusion of Transient States. When the Markov chain α_k^ε includes also transient states as in Section 5.2, we can still carry out the asymptotic expansions as in [26, Chapter 3]. Recall all the notation given in Section 5.2. As for the aggregated process $\overline{\alpha}_k^\varepsilon$, we have the following lemma, whose proof can be found in [26, Chpater 4].

Lemma 6.3.3. *Assume* (A1'). *Then* $\overline{\alpha}^\varepsilon(\cdot)$ *converges weakly to* $\overline{\alpha}(\cdot)$ *such that the limit is a continuous-time Markov chain with generator* $\overline{Q}_*$ *given in* (5.9).

REFERENCES

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

[2] F. Delebecque and J. Quadrat, Optimal control for Markov chains admitting strong and weak interactions, *Automatica* **17** (1981), 281–296.

[3] A. Graham, *Kronecker Product and Matrix Calculus: with Applications*, Ellis Horwood Lt., New York, 1981.

[4] A.K. Gupta and D.K. Nagar, *Matrix Variate Distributions*, Chapman & Hall/CRC, Monographs and Surveys in Pure and Applied Mathematics, Boca Raton, 2000.

[5] V.S. Korolyuk and N. Limnios, Diffusion approximation with equilibrium of evolutionary systems switched by semi-Markov processes, translation in *Ukrainian Math. J.* **57** (2005), 1466–1476.

[6] M. Iosifescu, *Finite Markov Processes and Their Applications*, Wiley, Chichester, 1980.

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

[8] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd Edition, Springer-Verlag, New York, 2003.

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

[10] M. Mariton, *Jump Linear Systems in Automatic Control*, Marcel Dekker, Inc., New York, 1990.

[11] D. Nguyen, N. Nguyen, and G. Yin, General Nonlinear Stochastic Systems Motivated by Chemostat Models: Complete Characterization of Long-Time Behavior, Optimal Controls, and Applications to Wastewater Treatment, *Stochastic Process Appl.*, **130** (2020) 4608–4642.

[12] S.L. Nguyen and G. Yin, Asymptotic properties of Markov modulated random sequences with fast and slow time scales, *Stochastics*, **82** (2010), 445–474.

[13] S.L. Nguyen and G. Yin, Weak convergence of Markov modulated random sequences, *Stochastics*, **82** (2010), 521–552.

[14] R.E. O'Malley, Jr., *Singular Perturbation Methods for Ordinary Differential Equations*, Springer-Verlag, New York, 1991.

[15] Z.G. Pan and T. Başar, H^∞-control of Markovian jump linear systems and solutions to associated piecewise-deterministic differential games, in *New Trends in Dynamic Games and Applications*, G.J. Olsder Ed., 61–94, Birkhäuser, Boston, MA, 1995.

[16] Z.G. Pan and T. Başar, H^∞ control of large scale jump linear systems via averaging and aggregation, in *Proc. 34th IEEE Conf. Decision Control*, 2574-2579, New Orleans, LA, 1995.

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

[18] S.P. Sethi, H. Zhang, and Q. Zhang, *Average-Cost Control of Stochastic Manufacturing Systems*, Springer-Verlag, New York, NY, 2004.

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

[20] K. Tran and G. Yin, Stochastic competitive Lotka-Volterra ecosystems under partial observation: Feedback controls for permanence and extinction, *J. Franklin Inst.*, **351** (2014) 4039—4064.

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

[22] G. Yin and S. Dey, Weak convergence of hybrid filtering problems involving nearly completely decomposable hidden Markov chains, *SIAM J. Control Optim.*, **41** (2003), 1820–1842.

[23] 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.

[24] G. Yin, R.H. Liu, and Q. Zhang, Recursive algorithms for stock Liquidation: A stochastic optimization approach, *SIAM J. Optim.*, **13** (2002), 240–263.

[25] G. Yin, Q. Yuan, L.Y. Wang, Asynchronous stochastic approximation algorithms for networked systems: Regime-switching topologies and multi-scale structure, *SIAM Journal: Multiscale Modeling and Simulation*, **11** (2013), 813–839.

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

[27] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*, Second Edition, Springer, New York, 2013.

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

[29] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, Sixth Edition, Springer, 2003.

ABSTRACT

SEQUENCES OF RANDOM MATRICES MODULATED BY A DISCRETE-TIME MARKOV CHAIN

by

HUY NGUYEN

August 2022

Advisor:	Dr. George Yin
Co-Advisor:	Dr. Pei-Yong Wang
Major:	Mathematics
Degree:	Doctor of Philosophy

In this dissertation, we consider a number of matrix-valued random sequences that are modulated by a discrete-time Markov chain having a finite space. Assuming that the state space of the Markov chain is large, our main effort in this work is devoted to reducing the complexity. To achieve this goal, our formulation uses time-scale separation of the Markov chain. The state-space of the Markov chain is split into subspaces. Next, the states of the Markov chain in each subspace are aggregated into a "super" state. Then we normalize the matrix-valued sequences that are modulated by the two-time-scale Markov chain. Under simple conditions, we derive a scaling limit of the centered and scaled sequence by using a martingale averaging approach. The limit is considered through a functional. It is shown that the scaled and interpolated sequence converges weakly to a switching diffusion. Towards the end of the work, we also indicate how we may handle matrix-valued processes directly. Certain tail probability estimates are obtained.

Keywords. Matrix-valued random sequence, mixing process, Markov chain, two-time-scale

formulation.

AUTOBIOGRAPHICAL STATEMENT

HUY NGUYEN

Education

- Ph.D. in Mathematics (expected), 2022

  Wayne State University, Detroit, Michigan, USA

- M.S. in Mathematics, 2007

  Vinh University, Nghe An, Vietnam

- B.S. in Mathematics, 2004

  Vinh University, Nghe An, Vietnam

Selected List of Publications

1. G. YIN, Z. WEN, H. QIAN, AND H. NGUYEN, *Numerical solutions for optimal control of stochastic Kolmogorov systems*, Journal of Systems Science and Complexity **34** (2021), 1703–1722.

2. N.V. QUANG, N.N. HUY AND L.H. SON, *The degenerate convergence criterion and Feller's weak law of large numbers for double arrays in noncommutative probability*, Statistics and Probability Letters , **83** (2013), 1812 -1818.

3. N.V. QUANG AND N.N. HUY, *Weak law of large numbers for adapted double arrays of random variables*, Journal of the Korean Mathematical Society, **45** (2008), 795 -805.