1-1-2017

Hybrid Stochastic Systems: Numerical Methods, Limit Results, And Controls

Tuan A. Hoang
Wayne State University,

HYBRID STOCHASTIC SYSTEMS: NUMERICAL METHODS, LIMIT RESULTS, AND CONTROLS

by

TUAN ANH HOANG

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

2017

MAJOR: MATHEMATICS

Approved By:

Advisor Date

DEDICATION

To my family

ACKNOWLEDGEMENTS

It is my great pleasure to acknowledge those whose help and support have made this dissertation possible.

First and foremost, I would like to express my deepest gratitude and appreciation toward my adviser, Professor George Yin, who introduced me to the world of stochastic systems and stochastic control. His continued guidance, support, encouragement and patience greatly influenced and shaped my learning and interests in mathematics. I have been extremely fortunate and privileged to be his student.

I'd also like to extend my gratitude toward Professor Zhimin Zhang, Professor Le Yi Wang, and Professor Pei-Yong Wang for serving on my committee and for their feedbacks on my work. I wish to take this opportunity to extend appreciation toward all the wonderful staff and faculty members of the Math department. I'd like to thank Ms. Barbara Malicke, Ms. Tiana Fluker, and Ms. Maria Leigh Vujic whose patience and care kept me on track on many of administrative matters. My appreciation also goes to Ms. Mary Klamo, Ms. Patricia Bonesteel and the academic coordinators I have worked with who helped plant the seed of love for teaching in me.

I am especially grateful for Professor Alexander Korostelev, Professor Rohini Kumar, Professor Robert R. Bruner, and Professor Kazuhiko Shinki, whose friendliness and enthusiasm come through both in and out of the classroom. Their passionate lectures and discussions broaden and solidify my knowledge in Probability, Statistics, Stochastic Processes Theory, and mathematics in general.

I'd also like to thank Dr. Son Nguyen and Dr. Dung Nguyen for all your cooperation during the course my graduate study. I'd also like to thank Dr. Nguyen Lam, Dr. Ky Tran, Dr. Nghia Tran, and Dang Nguyen for their helpful comments, suggestions, and enthusiasm throughout numerous useful discussions.

Thank you to all my friends who have made the good times so much better and the tough times so much more bearable.

Finally, I am forever grateful for my family for always being supportive and for their confidence in me. I am indebted to all of them for their unconditional love, patience and care. And to my wife, Thu Nguyen, I'd like to especially thank you for your unwavering love and support; I am grateful to have to have you in my life and by my side through thick and thin and I look forward to many more journeys with you in the future.

Thank you!

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

This dissertation is concerned with the so-called stochastic hybrid systems, which are featured by the coexistence of continuous dynamics and discrete events and their interactions. Such systems have drawn much needed attentions in recent years. One of the main reasons is that such systems can be used to better reflect the reality for a wide range of applications in networked systems, communication systems, economic systems, cyber-physical systems, and biological and ecological systems, among others.

Our main interest in this dissertation is centered around one class of such hybrid systems, known as switching diffusions; see [34, 52] and references there in. In such a system, in addition to the driving force of a Brownian motion as in a stochastic system represented by a stochastic differential equation (SDE), there is an additional continuous-time switching process that models the environmental changes or other random factors due to random events not represented in the usual stochastic differential equations. For example, in a financial market model, the switching process (e.g., a Markov chain) depicts such changes as market switches from a bull market to a bear market. In a cyber-physical system of a platoon of un-manned vehicles, the switching process represents the random communication capacity changes because of the interference. People have realized that such switching processes are much more realistic than the fixed configuration counterparts. Because their prevalence, stochastic hybrid systems have been studied extensively. To further our understanding and to treat such systems effectively, this dissertation is devoted to switching diffusions from several angles. In what follows, we give the organization of the dissertation.

In Chapter 2, we develops numerical schemes for stochastic differential equations with

Markovian switching (Markovian switching SDEs). By Markovian switching we meant that the switching process in fact, is a continuous-time Markov chain independent of the driving Brownian motion. Inspired by the well-known Milstein algorithms for solutions of stochastic differential equations, our effort is devoted to designing approximation algorithms with faster convergence rates than the commonly used Euler-Maruyama procedures. Compared to the diffusion case, the presence of the random switching component makes the design of the algorithms and the analysis much more complex. By utilizing a special form of Itô's formula for switching SDEs and special structural of the jumps of the switching component we derived a new scheme to simulate switching SDEs and develop a new approach to establish the convergence of the proposed algorithm. In contrast to the existing literature of numerical solutions for stochastic differential equations and Markovian switching stochastic differential equations, a new approach incorporating martingale methods, quadratic variations, and Markvian stopping times is developed. Detailed and delicate analysis is carried out. Under suitable conditions which are natural extensions of the classical ones, the convergence of the algorithms is established. The rate of convergence is also ascertained. In addition, numerical examples are provided to show the agreement with the theoretical convergence order. The content of this chapter is based on the work [37].

In Chapter 3, we study a limit theorem for general stochastic differential equations with Markovian regime switching. To begin, assume that we have a sequence of stochastic regime switching systems where the discrete switching processes are independent of the continuous state of the systems. The continuous-state component of these systems are governed by stochastic differential equations where the time t and the driving processes $B(\cdot)$ are replaced by $A^n(\cdot)$ and $M^n(\cdot)$, which are non-negative continuous increasing processes and square integrable martingales, respectively. We try to establish the convergence of the sequence of systems to the one described by a state independent regime-switching diffusion process when the two sequence of processes $\{A^n(\cdot)\}$ and $\{M^n(\cdot)\}$ converge to the usual time t and the Brownian motion $B(\cdot)$ in suitable sense. Compared to the corresponding problem of usual SDEs, the presence of the random switching component in our model makes the analysis much more complex. Our model also incorporates very general driving processes and thus makes it different from the existing literature for Markovian regime-switching SDEs. Under suitable conditions, the desired limit theorem is established. Though our motivation stems partially from many approximation schemes for regime-switching SDEs where each sequence of simulations resulted in a sequence of approximation processes of the above mentioned form, our result goes far beyond this situation. The result, besides of the purely theoretical interest, may provide a sort of general theorem to establish the convergence in some other situations as well. The results of this chapter are taken form the work [10].

Chapter 4 is concerned with controlled hybrid systems that are good approximations to controlled switching diffusion processes. The rational is as follows. Although Brownian motion based models are good approximation to the real models, and are easily dealt with in terms of analysis. In real applications, the noise is often non-Markovian and the so-called "white noise" is only an idealization and simplification. The best that one may hope is an approximation of the Brownian motion. Therefore, in lieu of a Brownian motion noise, we use a wide-band noise formulation, which facilitates the treatment of non-Markovian models. The wide-band noise is one whose spectrum has band width wide enough. We work with a basic stationary mixing type process. On top of this wide-band noise process, we allow the system to be subject to random discrete event influence. The discrete event process is a continuous-time Markov chain with a finite state space. Although the state space is finite, we assume that the state space is rather large and the Markov chain is irreducible. We are interested in optimal and equilibrium controls of such systems. Due to the complexity and non-Markovian of the system, obtaining the desired controls is extremely difficult, if not impossible, we therefore contend ourselves with getting the nearly optimal and nearly equilibrium controls. Using a two-time-scale formulation and assuming the Markov chain also subjects to fast variations, combine with weak convergence and singular perturbation test function method we first proved that the when controlled by nearly optimal and equilibrium controls, the state and the corresponding costs of the original systems would "converge" to those of controlled diffusions systems. Using the limit controlled dynamic system as a guidance, we construct controls for the original problem and show that the controls so constructed are near optimal and nearly equilibrium.

The dissertation is concluded with Chapter 5, where we summarize the central themes of the dissertation, provide further discussions and remarks. We also present some directions for future work.

CHAPTER 2 MILSTEIN-TYPE PROCEDURES FOR NUMERICAL SOLUTIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS

2.1 Introduction

In this chapter, we develop numerical algorithms for stochastic differential equations with Markovian switching (in short Markovian switching SDEs). We aim to designing numerical schemes of Milstein type, proving the convergence, and obtaining convergence rate that is better than the commonly used Euler-Maruyama procedures. Our effort is largely motivated by the pressing need of treating hybrid stochastic models involving continuous dynamics and discrete events represented by stochastic differential equations modulated by Markov chains. Recently, much effort has been devoted to the study of switching diffusions [34, 40, 47, 52]. Random switching models have been used in applications as option pricing, jump linear systems in automatic control, hierarchical decision making in production planning [44], estimation in hybrid systems [54], stock liquidation [55], and competitive Lotka-Volterra models in random environments [2], among others.

Because such systems are often highly nonlinear together with the coupling due to the random switching, closed-form solutions are virtually impossible to obtain. Thus significant effort has been devoted to designing feasible and efficient numerical solutions. Nevertheless, to the best of our knowledge, most of the work on numerical methods for Markovian switching diffusion to date has been focusing on Euler-Maruyama schemes; see [12, 40, 47, 48, 53], where different algorithms have been considered under various conditions and convergence modes. In spite of its simplicity, the convergence rate of Euler-Maruyama method is at most of order 1/2.

In numerical methods for stochastic differential equations, a scheme due to Milstein came into being; see [22,36]. The main idea is to use Itô formula for the drift and diffusion coefficients to get a better approximation in each step of the algorithm. The convergence rate was proved to be of order 1. One question naturally arises. For Markovian switching diffusions, can we design Milstein-type schemes? If we can, will such procedures still provide faster (order 1) convergence? There has been no decisive answer to this question to date. In fact, the study on the corresponding numerical algorithms of Milstein type have been scarce or virtually none for switching diffusions. Since the random switching and the discrete and continuous states are tangled together, the analysis is very difficult. Our aim in this chapter is devoted to improving the rates of convergence of numerical solutions for Markovian switching diffusions. To obtain a better convergence rate, we construct a Milstein-type scheme for diffusions with Markovian switching. In contrast to the case of diffusions, the appearance of the discrete component in the regime-switching diffusion leads to some additional terms represented by double stochastic integrals driven by both Brownian motions and discontinuous martingales due to the switching process. This requires special handling of these terms and makes the analysis more complicated. To overcome the difficulties, we use a crucial observation that multiple jump changes (more than two jumps) of the Markov chain in each small interval is of high order in reference to the size of the interval, which enables us to simplify the calculations.

In this chapter,

(1) we design and construct a Milstein-type procedure for numerical solutions of stochastic differential equations with Markovian switching;

(2) we establish the convergence of the algorithm;

(3) we demonstrate order 1 convergence rate and confirm that as in the case of diffusions, the faster convergence of the Milstein-type procedures is preserved;

(4) we numerically verify the convergence rate by providing numerical experimental results.

The rest of the chapter is arranged as follows. Section 2.2 begins with the formulation and preliminaries. Section 2.3 presents the numerical algorithms and states the main result. Section 2.4 is devoted to the study of the convergence of the numerical algorithms. First, we obtain an estimate relating the total number of jumps of the switching process on small intervals and the bound of the moments of the numerical solutions. Next, we use the Itô formula to present the difference between the exact and numerical solutions. The error bounds are then estimated to prove the main result. The performance of the numerical schemes is illustrated by several examples in Section 2.5. Finally, the chapter is concluded with Section 2.6 giving some concluding remarks.

2.2 Formulation and Preliminaries

This section provides the set up of our problem and gives the assumptions and notation as well as some preliminary results regarding the Markovian switching diffusions. Throughout this chapter, we use the same notion $|\cdot|$ to denote the different norms in $\mathbb{R}^d$, $\mathbb{R}^{d\times m}$, or $\mathbb{R}^{d\times d}$ for some fixed positive integers d and m. In this chapter, vectors are column vectors unless specified otherwise, and $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathbb{R}^d$. For $z\in\mathbb{R}^{l_1\times l_2}$, we use z' to denote its transpose. We will use C to denote a generic constant whose value may change from appearance to appearance in this chapter and use $\mathbb{1}$ to denote the usual zero-one indicator function. For $T>0$ and positive integers k and l, we use $C^k(\mathbb{R})$ to denote the set of real-valued functions that are k-times continuously differentiable and use $C^{k,l}([0,T]\times\mathbb{R}^d)$ to denote the set of real-valued functions that are k-times continuously differentiable with respect to the first variable and l-times continuously differentiable with respect to the second variable.

Stochastic differential equations with Markovian switching. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a complete probability space. Let $B(\cdot)=\big(B_1(\cdot),B_2(\cdot),\ldots,B_m(\cdot)\big)'$ be a standard Brownian motion in $\mathbb{R}^m$, and $\{\alpha(t),t\geq 0\}$ be a Markov chain that takes values in the finite set $\mathcal{M}=\{1,2,\ldots,m_0\}$. The dynamic behavior of the Markov chain $\alpha(t)$ is specified by the generator $Q=(q_{i_0j_0}:i_0,j_0\in\mathcal{M})$ satisfying: $q_{i_0j_0}\geq 0$ for $i_0\neq j_0\in\mathcal{M}$ and $q_{i_0i_0}=-\sum_{j_0\neq i_0}q_{i_0j_0}$ for each $i_0\in\mathcal{M}$. We study the numerical approximation to the following stochastic differential equation with Markovian switching

$$dX(t)=b(t,X(t),\alpha(t))dt+\sigma(t,X(t),\alpha(t))dB(t),\quad X(0)=X_0, \tag{2.1}$$

where $b(\cdot,\cdot,\cdot):\mathbb{R}\times\mathbb{R}^d\times\mathcal{M}\to\mathbb{R}^d$, $\sigma(\cdot,\cdot,\cdot):\mathbb{R}\times\mathbb{R}^d\times\mathcal{M}\to\mathbb{R}^{d\times m}$ are vector-valued functions satisfying suitable conditions that will be specified later and $0\leq t\leq T$, a finite time horizon. The initial condition X_0 is a $\mathbb{R}^d$-valued random variable. We assume that X_0, $B(\cdot)$, and $\alpha(\cdot)$ are independent. Thus, the transition probability of Markov chain $\alpha(t)$ satisfies the following equation

$$P\big(\alpha(t+\Delta t)=j_0\big|\alpha(t)=i_0,\alpha(s),X(s),0\leq s\leq t\big)=q_{i_0j_0}\Delta t+o(\Delta t)\quad i_0,j_0\in\mathcal{M},i_0\neq j_0. \tag{2.2}$$

We write

$$X(t) = (X_1(t), X_2(t), \dots, X_d(t))'$$

$$b(t, x, i_0) = (b_1(t, x, i_0), b_2(t, x, i_0), \dots, b_d(t, x, i_0))'$$

and

$$\sigma(t, x, i_0) = (\sigma_1(t, x, i_0), \sigma_2(t, x, i_0), \dots, \sigma_m(t, x, i_0)) \in \mathbb{R}^{d\times m}$$

for $(t, x_0, i_0) \in [0, T] \times \mathbb{R}^d \times \mathcal{M}$, and $\sigma_l = (\sigma_{1,l}, \sigma_{2,l}, \dots, \sigma_{d,l})' \in \mathbb{R}^d$ for $1 \le l \le m$. Denote $\mathcal{F}_t^B = \sigma\{X_0, B_l(s),\ 0 \le s \le t,\ 1 \le l \le m\}$, $\mathcal{F}_t^\alpha = \sigma\{\alpha(s),\ 0 \le s \le t\}$ and $\mathcal{F}_t = \mathcal{F}_t^B \vee \mathcal{F}_t^\alpha$ for $t \ge 0$. Assume that X_0, $b(\cdot,\cdot,\cdot)$, and $\sigma(\cdot,\cdot,\cdot)$ satisfy the following conditions.

Assumption (A). There is a constant C such that for $x, y \in \mathbb{R}^d$, $t \in [0, T]$ and $i_0 \in \mathcal{M}$, $\mathbb{E}|X_0|^2 < C$, and

$$|b(t, x, i_0) - b(t, y, i_0)| + \big|\sigma(t, x, i_0) - \sigma(t, y, i_0)\big| \le C|x - y|,$$

$$|b(t, x, i_0)| + |\sigma(t, x, i_0)| \le C(1 + |x|)$$

Assumption (A) requires the initial data having finite second moment together with the usual Lipschitz continuity and linear growth condition. It follows from Theorem 3.3.13 in [34] that under Assumption (A), equation (2.1) has a unique global solution. In addition, we have the following result regarding the moments of $X(t)$. The proof is similar to those of Theorem 3.3.23 and Theorem 3.3.24 in [34] and is therefore omitted.

Lemma 2.1. *Under Assumption* (A) *the following inequalities hold true with probability one*

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X(t)|^2 \Big| \mathcal{F}_T^\alpha\Big] \le C \quad \textit{and} \quad \mathbb{E}\Big[\sup_{t\in[s,s+h]} \big|X(t) - X(s)\big|^2 \Big| \mathcal{F}_T^\alpha\Big] \le Ch,$$

where the constant C depends only on T and $\mathbb{E}|X_0|^2$.

Martingale associated to the Markov chain and Itô formula. For each pair (i_0, j_0) in $\mathcal{M} \times \mathcal{M}$, $i_0 \neq j_0$, and $t \geq 0$, we define

$$[M_{i_0 j_0}](t) = \sum_{0 \leq s \leq t} \mathbb{1}(\alpha(s-) = i_0)\mathbb{1}(\alpha(s) = j_0), \qquad \langle M_{i_0 j_0} \rangle(t) = \int_0^t q_{i_0 j_0} \mathbb{1}(\alpha(s-) = i_0) ds. \tag{2.3}$$

Then it follows from Lemma IV.21.12 in [43] that the process $M_{i_0 j_0}(t)$, $0 \leq t \leq T$, defined by

$$M_{i_0 j_0}(t) = [M_{i_0 j_0}](t) - \langle M_{i_0 j_0} \rangle(t) \tag{2.4}$$

is a purely discontinuous and square integrable martingale with respect to $\mathcal{F}_t^\alpha$, which is null at the origin. The processes $[M_{i_0 j_0}](t)$ and $\langle M_{i_0 j_0} \rangle(t)$ are the optional and predictable quadratic variations, respectively. For convenience, we denote $M_{i_0 i_0}(t) = 0$ for $i_0 \in \mathcal{M}$ and $0 \leq t \leq T$. We have the following orthogonality relations from the definition of optional quadratic covariations (see [33], Section 1.8):

$$[B_{l_1}, B_{l_2}] = 0 \text{ for } 1 \leq l_1, l_2 \leq m, l_1 \neq l_2,$$

$$[M_{i_0 j_0}, B_l] = 0 \text{ for } i_0, j_0 \in \mathcal{M}, 1 \leq l \leq m,$$

$$[M_{i_0 j_0}, M_{i_1 j_1}] = 0 \text{ for } i_0, j_0, i_1, j_1 \in \mathcal{M}, (i_0, j_0) \neq (i_1, j_1).$$

Let $\mathcal{L}$ denote the generator of system (2.1). For a function $f(\cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^d \times \mathcal{M} \to \mathbb{R}$ such that for each $i_0 \in \mathcal{M}$, $f(\cdot, \cdot, i_0) \in C^{1,2}([0, T] \times \mathbb{R}^d)$,

$$\mathcal{L} f(t, x, i_0) = \frac{\partial}{\partial t} f(t, x, i_0) + \mathcal{L}_{i_0} f(t, x, i_0) + Q f(t, x, \cdot)(i_0) \tag{2.5}$$

for all $(t, x, i_0) \in [0, T] \times \mathbb{R}^d \times \mathcal{M}$, where

$$\mathcal{L}_{i_0} f(t, x, i_0) = b'(t, x, i_0)\nabla_x f(t, x, i_0) + \frac{1}{2}\mathrm{tr}\big(\nabla^2_{xx} f(t, x, i_0) A(t, x, i_0)\big),$$

$$Qf(t, x, \cdot)(i_0) = \sum_{j_0 \in \mathcal{M}} q_{i_0 j_0}\big(f(t, x, j_0) - f(t, x, i_0)\big).$$

Here, ∇_x and ∇^2_{xx} denotes the gradient and Hessian matrix with respect to x, respectively, and $A(t, x, i_0) = \sigma(t, x, i_0)\sigma'(t, x, i_0) \in \mathbb{R}^{d\times d}$. We will use the following form of Itô's lemma to find the stochastic expansion of the solution to (2.1) in Section 2.3. A proof of it will be given in the Appendix.

Lemma 2.2. *For a function $f(\cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^d \times \mathcal{M} \to \mathbb{R}$ such that for each $i_0 \in \mathcal{M}$, $f(\cdot, \cdot, i_0) \in C^{1,2}([0, T] \times \mathbb{R}^d)$ we have*

$$\begin{aligned} f(t, X(t), \alpha(t)) = f(s, x(s), \alpha(s)) &+ \int_s^t \mathcal{L} f(u, X(u), \alpha(u))du \\ &+ \sum_{l=1}^{m} \int_s^t \big\langle \nabla_x f(u, X(u), \alpha(u)), \sigma_l(u, X(u), \alpha(u))\big\rangle dB_l(u) \\ &+ \sum_{i_0 \neq j_0} \int_s^t \big(f(u, X(u), j_0) - f(u, X(u), i_0)\big) dM_{i_0 j_0}(u), \quad 0 \le s \le t \le T. \end{aligned} \tag{2.6}$$

2.3 Numerical Methods

In this section we provide the details of our numerical scheme with constant step size denoted by h. For $n = 0, 1, 2, \ldots$, we use $t_n = nh$ to denote the time mesh points and $\Delta_n = (t_n, t_{n+1}]$ to denote the time intervals in the scheme.

Construction of the Markov chain $\alpha(t)$. For generation of a continuous-time Markov chain, with a given generator $Q = (q_{i_0 j_0}) \in \mathbb{R}^{m_0 \times m_0}$, we quote the method of constructing the Markov chain from [51], Section 2.4. To construct the sample paths of $\alpha(t)$ requires determining its sojourn time at each state and its subsequent moves. The chain sojourns in any given state i_0, $i_0 \in \mathcal{M}$, for a random length of time, η_{i_0}, which has an exponential distribution with parameter $-q_{i_0 i_0}$. Subsequently, the process will enter another state. Each state j_0 (with $j_0 \in \mathcal{M},\ j_0 \neq i_0$) has a probability $q_{i_0 j_0}/(-q_{i_0 i_0})$ of being the chain's next residence. The post-jump location is determined by a discrete random variable Z_{i_0} taking values in $\{1, 2, \ldots, i_0 - 1, i_0 + 1, \ldots, m_0\}$. Its value is specified by

$$Z_{i_0} = \begin{cases} 1, & \text{if } \ U \leq q_{i_0 1}/(-q_{i_0 i_0}), \\ 2, & \text{if } \ q_{i_0 1}/(-q_{i_0 i_0}) < U \leq (q_{i_0 1} + q_{i_0 2})/(-q_{i_0 i_0}), \\ \vdots & \vdots \\ m_0, & \text{if } \ \sum_{j_0 \neq i_0, j_0 < m_0} q_{i_0 j_0}/(-q_{i_0 i_0}) \leq U, \end{cases} \tag{2.7}$$

where U is a random variable uniformly distributed in $(0, 1)$. Thus, the sample path of $\alpha(t)$ is constructed by sampling from exponential and $U(0, 1)$ random variables alternately. With the $\alpha(t)$ generated above, for $n = 0, 1, \ldots$, set $\alpha_n = \alpha_n^h = \alpha(t_n)$ which is the h-skeleton of the Markov chain.

Approximation of the jump times. To develop approximation derived from Milstein's approach for diffusions so as to obtain a better convergence rate than that of Euler-Maruyama method, we need to use higher order Taylor expansion and add additional correction terms. However, different from the diffusion case, the appearance of the discrete component in Markovian switching diffusions makes the calculation of the terms (represented by stochastic integrals driven by the discontinuous martingales associated with the Markov chain) more complicated. To treat these terms, we analyze and approximate the jump times of the Markov chain on each small interval Δ_n.

We observe that the stochastic integrals driven by the discontinuous martingales associated with the Markov chain disappear on Δ_n if there is no jump within the interval (see Lemma 2.7). In addition, two or more jumps take places within this interval is only of order $O(h^2)$, which is negligible. Therefore, we only need to consider the case that there is exactly one jump occurs in each interval Δ_n. In what follow, we will define the function $\varsigma(\cdot)$ representing the jump time in Δ_n when only one jump occurs. We will ignore the case of having none or more than one jump in that interval by putting $\varsigma(n) = t_{n+1}$.

From the above construction of the Markov chain, we can compute the time of the k-th jump denoted by τ_k for $k \geq 0$. For convenience, define $\tau_0 = 0$. We denote the sojourn time of the Markov chain at the state previous to the one of the $(k+1)$-th jump by η_k, i.e., $\eta_k = \tau_{k+1} - \tau_k$. For $k \geq 0$ denote $n_k = n_k^h = \left\lceil \frac{\tau_k}{h} \right\rceil - 1$, where $\lceil x \rceil$ denotes the least integer greater than or equals x (i.e., $\lceil x \rceil - 1 < x \leq \lceil x \rceil$). It follows that $n_k h < \tau_k \leq (n_k + 1)h$. Next, we define the function $\varsigma(\cdot) : \mathbb{N} \cup \{0\} \to \mathbb{R}_+$ as follow:

- If $n_{k-1} < n = n_k < n_{k+1}$ for some $k \geq 1$ we define $\varsigma(n) = \varsigma^h(n) = \tau_k$.

- If $n_{k-1} < n < n_k$ or $n_{k-1} < n = n_k = n_{k+1} = \cdots = n_{k+p} < n_{k+p+1}$ for some $k, p \geq 1$ we define $\varsigma(n) = \varsigma^h(n) = t_{n+1}$.

According to the definition of $\varsigma(\cdot)$, for any $n \geq 0$, $\varsigma(n) = t_{n+1}$ if there is none or there are more than one jump occurring in Δ_n, and $t_n < \varsigma(n) < t_{n+1}$ if there is only one jump occurring in Δ_n. In the latter case, the Markov chain jumps from the state $\alpha(t_n)$ to the state $\alpha(t_{n+1})$ at the time $\varsigma(n)$.

Numerical scheme. For $f : \mathbb{R} \times \mathbb{R}^d \times \mathcal{M} \to \mathbb{R}$ such that $f(\cdot, \cdot, i_0) \in C^{1,2}([0, T] \times \mathbb{R}^d)$ for each $i_0 \in \mathcal{M}$, denote

$$L^l f(t,x,i_0) = \sum_{k=1}^{d} \sigma_{k,l}(t,x,i_0) \frac{\partial f}{\partial x_k}(t,x,i_0) = \big\langle \nabla_x f(t,x,i_0), \sigma_l(t,x,i_0) \big\rangle, \quad l = 1,2,\ldots,m,$$

and

$$L^0 f(t,x,i_0) = \frac{\partial f}{\partial t} f(t,x,i_0) + \sum_{k=1}^{d} \left[b_k(t,x,i_0) - \frac{1}{2} \sum_{l=1}^{m} L^l \sigma_{k,l}(t,x,i_0) \right] \frac{\partial f}{\partial x_k} f(t,x,i_0).$$

It follows that

$$\mathcal{L} f(t,x,i_0) = L^0 f(t,x,i_0) + \frac{1}{2} \sum_{l=1}^{m} L^l L^l f(t,x,i_0) + Qf(t,x,\cdot)(i_0). \tag{2.8}$$

To approximate the solution to (2.1) we propose the following algorithm

$$\begin{aligned} Y_{n+1}^h = Y_n^h + hb(t_n, Y_n^h, \alpha_n) + \sum_{l=1}^{m} \sigma_l(t_n, Y_n^h, \alpha_n) \Delta_n B_l + \sum_{l_1,l_2=1}^{m} L^{l_2} \sigma_{l_1}(t_n, Y_n^h, \alpha_n) I_{l_1,l_2}(n) \\ + \sum_{l=1}^{m} \big[\sigma_l(t_n, Y_n^h, \alpha_{n+1}) - \sigma_l(t_n, Y_n^h, \alpha_n)\big] \big[B_l(t_{n+1}) - B_l(\varsigma(n))\big], \end{aligned} \tag{2.9}$$

for $h > 0$, $n = 0, 1, ...$, where, for $l, l_1, l_2 = 1, 2, \ldots, m$ and $n = 0, 1, \ldots$,

$$\Delta_n B_l = B_l\big(t_{n+1}\big) - B_l(t_n), \quad I_{l_1,l_2}(n) = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dB_{l_2}(s_2) dB_{l_1}(s_1). \tag{2.10}$$

A detailed derivation of the above scheme is given in Section 2.4.

Remark 2.3. Similar to the Milstein scheme for stochastic differential equations without switching, in algorithm (2.9), we implicitly assume that the terms $I_{l_1,l_2}(n)$ with $1 \le n \le T/h$, can be simulated. As shown in [22], Chapter 10, these multiple stochastic integrals cannot be easily expressed in terms of $\Delta_n B_{l_1}$ and $\Delta_n B_{l_2}$, the increments of the Brownian motions. In many important practical problems, the diffusion coefficients have special properties that allow the Milstein scheme to be simplified avoiding the use of double stochastic integrals involving different components of the Brownian motions. For instance, with additive noise

$\sigma_{k,l}(t,x,i_0) = \sigma_{k,l}(t,i_0)$, or linear noise where $\sigma_{k,l}(t,x,i_0) = \sigma_{k,l}(t,i_0)x_k$, $l = 1,\ldots,m$, $k = 1,\ldots,d$ and $(t,x,i_0) \in [0,T] \times \mathbb{R}^d \times \mathcal{M}$, the double stochastic integrals can be simplified. Another special case is that of diagonal noise, where $d = m$ and each component X_k of the process X is disturbed only by the corresponding Brownian motion B_k and the diagonal diffusion coefficient $\sigma_{k,k}(t,x,i_0)$ depends only on (t,x_k,i_0). A more general, but important special case is that of commutative noise in which the diffusion matrix satisfies the commutativity condition $L^{l_1}\sigma_{k,l_2}(t,x,i_0) = L^{l_2}\sigma_{k,l_1}(t,x,i_0)$ for all $l_1,l_2 = 1,\ldots,m$, $k = 1,\ldots,d$ and $(t,x,i_0) \in [0,T] \times \mathbb{R}^d \times \mathcal{M}$.

We assume the following assumptions throughout this work. This set of assumptions is a natural extension of those in Theorem 10.3.5 in [22] for the case SDE with Markovian switching. In addition to Assumption (A), we assume that the initial approximation is close to X_0 in the second moments. This is hardly a restriction. In reality, we often take $Y_0^h = X_0$ and even X_0 is often not random. We use the current condition to accommodate more complex cases. Moreover, we assume the growth condition for the partial derivatives, and the local Hölder condition for the coefficients and appropriate derivatives.

Assumption (B). Assumption (A) holds. In addition, there is a constant C such that for $0 \le l_0 \le m$, $1 \le l, l_1 \le m$, $x, y \in \mathbb{R}^d$, $0 \le s,t \le T$ and each $i \in \mathcal{M}$, $[\mathbb{E}|X_0 - Y_0^h|^2]^{1/2} \le Ch$, $|L^{l_1}\sigma_l(t,x,i_0) - L^{l_1}\sigma_l(t,y,i_0)| \le C|x-y|$, and

$$|L^{l_0}b(t,x,i_0)| + |L^{l_0}\sigma_l(t,y,i_0)| + |L^{l_1}L^{l_1}b(t,x,i_0)| + |L^{l_1}L^{l_1}\sigma_l(t,x,i_0)| \le C(1+|x|),$$

$$|b(s,x,i_0) - b(t,x,i_0)| + |\sigma_l(s,x,i_0) - \sigma_l(t,x,i_0)| + |L^{l_1}\sigma_l(s,x,i_0) - L^{l_1}\sigma_l(t,x,i_0)| \le C(1+|x|)|s-t|^{\frac{1}{2}}.$$

Denote $X_n^h = X(t_n)$ for $n = 0,1,\ldots,T/h$ where $X(t)$ is the solution to (2.1) and T/h is understood to be $\lfloor T/h \rfloor$, the integer part of T/h. In what follows, for notational simplicity, we suppress the notation $\lfloor\cdot\rfloor$. We are now in a position to state our main theorem.

Theorem 2.4. *Assume that Assumption* (B) *holds. Then there exists a constant C independent of h such that* $\mathbb{E}\Big[\sup_{0\le n\le T/h}\big|X_n^h - Y_n^h\big|^2\Big] \le Ch^2$.

2.4 Proof of Main Result

In this section, we provide the proof of our main result after establishing a number of preliminary lemmas. We first provide a bound on the probability that the Markov chain α has more than N jumps and prove the boundedness of the second moment of the approximate solution Y_n^h. These results are repeatedly used in the subsequent proofs. We next give a detailed derivation of the proposed numerical scheme (2.9) and then proceed to give estimates on various error terms and conclude the section with the proof of the main theorem.

2.4.1 Total Number of Jumps of the Markov Chain and the Boundedness of the Second Moments

For a fixed number h, $0 < h < 1$, and $n = 0, 1, \ldots$, denote by N_n the total number of jumps of the chain in the interval Δ_n with the sequence of jump times $t_n = \tau_0^n < \tau_1^n < \tau_2^n < \ldots < \tau_{N_n}^n < t_{n+1}$. We now provide a bound on the probability that the Markov chain $\alpha(\cdot)$ has more than N jumps on a time interval of length h.

Lemma 2.5. *The following inequality holds true.*

$$P\big(N_n \ge N\big) \le q^N h^N, \quad N \ge 1 \tag{2.11}$$

where $q = \max\{-q_{j_0j_0} : j_0 \in \mathcal{M}\}$ *and* $n = 0, 1, \ldots$ *As a consequence, if* $h < 1/(2q)$ *there is a constant C independent of n such that*

$$\mathbb{E}N_n \le Ch. \tag{2.12}$$

Proof. Denote $\eta_p^n = \tau_{p+1}^n - \tau_p^n$ for $0 \le p \le N_n$. Then on the set $\{N_n \ge 1\}$, $\eta_0^n, \eta_1^n, \ldots, \eta_{N_n-1}^n$ are the times between the consecutive jumps and are conditionally independent random variables. In addition, if $N_n \ge 1$ and the chain jumps from state i_{p-1} to state i_p at the time τ_p^n for $1 \le p \le N_n$ then η_p^n has the exponential distribution with parameter $-q_{i_p i_p}$. Since $q = \max\{-q_{j_0 j_0} : j_0 \in \mathcal{M}\}$, by the strong Markov property of $\alpha(t)$, we have

$$\begin{aligned} \mathbb{P}\big(N_n \ge N\big) &\le \mathbb{P}\Big(\sum_{p=0}^{N-1} \eta_p^n < h\Big) \le \prod_{p=0}^{N-1} \mathbb{P}\big(\eta_p^n < h\big) \\ &= \prod_{p=0}^{N-1} \Big(1 - e^{q_{i_p i_p} h}\Big) \le \prod_{p=0}^{N-1} \big(-q_{i_p i_p} h\big) \le q^N h^N, \end{aligned} \tag{2.13}$$

for $N \ge 1$ and $n \ge 0$. Therefore, (2.11) follows. Inequality (2.12) is a consequence of (2.11) and the identity $\mathbb{E}N_n = \sum_{N=1}^{\infty} \mathbb{P}(N_n \ge N)$. □

Under Assumption (B), we obtain the boundedness of the second moments of Y_n^h.

Lemma 2.6. *Under Assumption* (B)*, there exists a constant C independent of h such that*

$$\sup_{0 \le n \le T/h} \mathbb{E}|Y_n^h|^2 \le C. \tag{2.14}$$

Proof. For algorithm (2.9), we have

$$Y_{k,n+1}^h = Y_{k,n}^h + \Big\{ h b_k(t_n, Y_n^h, \alpha_n) + \sum_{l=1}^{m} \sigma_{k,l}(t_n, Y_n^h, \alpha_n)\Delta_n B_l \tag{2.15}$$

$$+ \sum_{l_1,l_2=1}^{m} L^{l_2}\sigma_{k,l_1}(t_n, Y_n^h, \alpha_n) I_{l_1,l_2}(n)$$

$$+ \sum_{l=1}^{m} \big[\sigma_{k,l}(t_n, Y_n^h, \alpha_{n+1}) - \sigma_{k,l}(t_n, Y_n^h, \alpha_n)\big]\big[B_l(t_{n+1}) - B_l(\varsigma(n))\big]\Big\}. \tag{2.16}$$

Thus, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
|Y^h_{k,n+1}|^2 \le |Y^h_{k,n}|^2 + 2Y^h_{k,n}\Big\{ & hb(t_n, Y^h_n, \alpha_n) + \sum_{l=1}^m \sigma_l(t_n, Y^h_n, \alpha_n)\Delta_n B_l \\
& + \sum_{l_1,l_2=1}^m L^{l_2}\sigma_{l_1}(t_n, Y^h_n, \alpha_n) I_{l_1,l_2}(n) \\
& + \sum_{l=1}^m \big[\sigma_l(t_n, Y^h_n, \alpha_{n+1}) - \sigma_l(t_n, Y^h_n, \alpha_n)\big]\big[B_l(t_{n+1}) - B_l(\varsigma(n))\big]\Big\} \\
& + \Big(m^2 + 2m + 1\Big)\Big\{h^2|b(t_n, Y^h_n, \alpha_n)|^2 + \sum_{l=1}^m |\sigma_l(t_n, Y^h_n, \alpha_n)|^2 |\Delta_n B_l|^2 \\
& + \sum_{l_1,l_2=1}^m |L^{l_2}\sigma_{l_1}(t_n, Y^h_n, \alpha_n)|^2 |I_{l_1,l_2}(n)|^2 \\
& + \sum_{l=1}^m \big|\sigma_l(t_n, Y^h_n, \alpha_{n+1}) - \sigma_l(t_n, Y^h_n, \alpha_n)\big|^2 \big[B_l(t_{n+1}) - B_l(\varsigma(n))\big]^2\Big\}. \qquad (2.17)
\end{aligned}
$$

Again, by virtue of algorithm (2.9), Y^h_n is independent of $\Delta_n B_l$, $I_{l_1,l_2}(n)$ and $B_l(t_{n+1}) - B_l(\varsigma(n))$ for $1 \le l, l_1, l_2 \le m$. Note that $\mathbb{E}\big[\Delta_n B_l\big] = \mathbb{E}[B_l(t_{n+1}) - B_l(\varsigma(n))] = \mathbb{E}\big[I_{l_1,l_2}(n)\big] = 0$. In addition, it follows from (2.10) that $\mathbb{E}|\Delta_n B_l|^2 \le h$, $\mathbb{E}|B_l(t_{n+1}) - B_l(\varsigma(n))|^2 \le h$, $\mathbb{E}|I_{l_1,l_2}(n)|^2 = \frac{h^2}{2}$. Thus, using these facts and Assumption (B) and taking the expectations on both sides of (2.17) yields

$$
\begin{aligned}
\mathbb{E}|Y^h_{n+1}|^2 \le E|Y^h_n|^2 + hC\mathbb{E}\big[|Y^h_n|\big(1 + |Y^h_n|\big)\big] + C\mathbb{E}\Big\{ & h^2\big(1 + |Y^h_n|^2\big) + \sum_{l=1}^m \big(1 + |Y^h_n|^2\big)|\Delta_n B_l|^2 \\
+ \sum_{l_1,l_2=1}^m \big(1 + |Y^h_n|^2\big)|I_{l_1,l_2}(n)|^2 & + \sum_{l=1}^m \big(1 + |Y^h_n|^2\big)\big[B_l(t_{n+1}) - B_l(\varsigma(n))\big]^2\Big\} \\
\le \mathbb{E}|Y^h_n|^2 + Ch\big(1 + \mathbb{E}|Y^h_n|^2\big) \le \mathbb{E}|Y^h_0|^2 & + Ch\sum_{i=0}^n \big(1 + \mathbb{E}|Y^h_i|^2\big). \qquad (2.18)
\end{aligned}
$$

By the Gronwall inequality, (2.18) yields $\sup_{0\le n\le T/h} \mathbb{E}|Y^h_n|^2 \le \big(CT + \mathbb{E}|Y_0|^2\big)e^{CT}$. $\square$

2.4.2 Derivation of the Scheme

We are now in a position to provide a detailed derivation of the proposed scheme (2.9). For $n \ge 0$, denote $X^h_n = X(t_n) = (X^h_{1,n}, X^h_{2,n}, \ldots, X^h_{d,n})'$. Applying the Itô formula (2.6) to

$f = b_k$ and $f = \sigma_{k,l}$ for $k = 1, 2, \dots, d$ and $l = 1, 2, \dots, m$, we have the following equation for the k-th component of $X(t_{n+1})$

$$
\begin{aligned}
X^h_{k,n+1} = X^h_{k,n} + \int_{t_n}^{t_{n+1}} \Big[& b_k\big(t_n, X^h_n, \alpha_n\big) + \int_{t_n}^{s} \mathcal{L} b_k(u, X(u), \alpha(u)) du \\
& + \sum_{l=1}^{m} \int_{t_n}^{s} L^l b_k(u, X(u), \alpha(u)) dB_l(u) \\
& + \sum_{i_0 \neq j_0} \int_{t_n}^{s} \big(b_k(u, X(u), j_0) - b_k(u, X(u), i_0) \big) dM_{i_0 j_0}(u) \Big] ds \\
& + \sum_{l_1=1}^{m} \int_{t_n}^{t_{n+1}} \Big[\sigma_{k,l_1}\big(t_n, X^h_n, \alpha_n\big) + \int_{t_n}^{s} \mathcal{L} \sigma_{k,l_1}(u, X(u), \alpha(u)) du \\
& + \sum_{l_2=1}^{m} \int_{t_n}^{s} L^{l_2} \sigma_{k,l_1}(u, X(u), \alpha(u)) dB_{l_2}(u) \\
& + \sum_{i_0 \neq j_0} \int_{t_n}^{s} \big(\sigma_{k,l_1}(u, X(u), j_0) - \sigma_{k,l_1}(u, X(u), i_0) \big) dM_{i_0 j_0}(u) \Big] dB_{l_1}(s).
\end{aligned}
\tag{2.19}
$$

Rearranging terms in the last equation and making use of the identity $M_{i_0 j_0} = [M_{i_0 j_0}] - \langle M_{i_0 j_0} \rangle$ and the notation $I_{l_1, l_2}(n)$, we obtain

$$
\begin{aligned}
X^h_{k,n+1} = {} & X^h_{k,n} + h b_k\big(t_n, X^h_n, \alpha_n\big) + \sum_{l=1}^{m} \sigma_{k,l}\big(t_n, X^h_n, \alpha_n\big) \Delta_n B_l \\
& + \sum_{l=1}^{m} \sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0) \big) d[M_{i_0 j_0}](u) dB_l(s) \\
& + \sum_{l_1, l_2 = 1}^{m} L^{l_2} \sigma_{k,l_1}(t_n, X^h_n, \alpha_n) I_{l_1, l_2}(n) + \sum_{j=1}^{6} r_{k,n,j},
\end{aligned}
\tag{2.20}
$$

for $k = 1, 2, \ldots, d$ and $0 \le n \le T/h$, where

$$r_{k,n,1} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathcal{L} b_k(u, X(u), \alpha(u)) du ds, \tag{2.21}$$

$$r_{k,n,2} = \sum_{l=1}^{m} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} L^l b_k(u, X(u), \alpha(u)) dB_l(u) ds, \tag{2.22}$$

$$r_{k,n,3} = \sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(b_k(u, X(u), j_0) - b_k(u, X(u), i_0)\big) dM_{i_0 j_0}(u) ds, \tag{2.23}$$

$$r_{k,n,4} = \sum_{l=1}^{m} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \mathcal{L} \sigma_{k,l}(u, X(u), \alpha(u)) du dB_l(s), \tag{2.24}$$

$$r_{k,n,5} = \sum_{l=1}^{m} \sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), i_0) - \sigma_{k,l}(u, X(u), j_0)\big) d\big\langle M_{i_0 j_0} \big\rangle (u) dB_l(s), \tag{2.25}$$

$$r_{k,n,6} = \sum_{l_1, l_2 = 1}^{m} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \Big[L^{l_2} \sigma_{k,l_1}(u, X(u), \alpha(u)) - L^{l_2} \sigma_{k,l_1}(t_n, X_n^h, \alpha_n)\Big] dB_{l_2}(u) dB_{l_1}(s). \tag{2.26}$$

Guided by the Milstein scheme in the case of diffusion without switching, we should include the first four terms on the right hand side of (2.20) into our numerical scheme. However, different from the traditional Milstein schemes for stochastic differential equations, we need to include the fifth term on the right hand side of (2.20), which involves the double stochastic integrals with respect to the optional quadratic variation processes $[M_{i_0 j_0}]$ and the Brownian motions. An explanation for the above choice is that, based on the definition of $[M_{i_0 j_0}]$, the total contribution of the fifth term after all iterations in the scheme is $O(h^{1/2})$. The following lemma gives a more convenient representation for this double integral term. Its proof is postponed to the appendix.

Lemma 2.7. *If $N_n \geq 1$, for $k = 1, 2, \ldots, d$ and $l = 1, 2, \ldots, m$ we have*

$$\sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\big) d[M_{i_0 j_0}](u) dB_l(s)$$
$$= \sum_{i=1}^{N_n} \Big[\sigma_{k,l}(\tau_i^n, X(\tau_i^n), \alpha(\tau_i^n)) - \sigma_{k,l}(\tau_i^n, X(\tau_i^n), \alpha(\tau_{i-1}^n))\Big] \Big(B_l(t_{n+1}) - B_l(\tau_i^n)\Big). \tag{2.27}$$

If $N_n = 0$ the left-hand side equals 0.

Since we have $\alpha(\tau_1^n) = \alpha_{n+1}$ and $\alpha(\tau_0^n) = \alpha_n$ on the set $\{N_n = 1\}$, it follows from the above Lemma that

$$\begin{aligned}
&\sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\big) d[M_{i_0 j_0}](u) dB_l(s) \\
&= \mathbb{1}\big(N_n = 1\big) \Big[\sigma_{k,l}(\tau_1^n, X(\tau_1^n), \alpha_{n+1}) - \sigma_{k,l}(\tau_1^n, X(\tau_1^n), \alpha_n)\Big] \Big(B_l(t_{n+1}) - B_l(\tau_1^n)\Big) \\
&\quad + \mathbb{1}\big(N_n \geq 2\big) \sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\big) d[M_{i_0 j_0}](u) dB_l(s).
\end{aligned} \tag{2.28}$$

Denote

$$\begin{aligned}
r_{k,n,7} = \sum_{l=1}^{m} \mathbb{1}\big(N_n = 1\big) \Big[&\Big(\sigma_{k,l}(\tau_1^n, X(\tau_1^n), \alpha_{n+1}) - \sigma_{k,l}(t_n, X_n^h, \alpha_{n+1})\Big) \\
&- \Big(\sigma_{k,l}(\tau_1^n, X(\tau_1^n), \alpha_n) - \sigma_{k,l}(t_n, X_n^h, \alpha_n)\Big)\Big] \Big(B_l(t_{n+1}) - B_l(\tau_1^n)\Big),
\end{aligned} \tag{2.29}$$

$$r_{k,n,8} = \sum_{l=1}^{m} \mathbb{1}\big(N_n \geq 2\big) \sum_{i_0 \neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\big) d[M_{i_0 j_0}](u) dB_l(s). \tag{2.30}$$

We can rewrite (2.20) as

$$
\begin{aligned}
X^h_{k,n+1} &= X^h_{k,n} + hb_k\big(t_n, X^h_n, \alpha_n\big) + \sum_{l=1}^m \sigma_{k,l}\big(t_n, X^h_n, \alpha_n\big)\Delta_n B_l \\
&\quad + \sum_{l=1}^m \mathbb{1}\big(N_n = 1\big)\Big[\sigma_{k,l}(t_n, X^h_n, \alpha_{n+1}) - \sigma_{k,l}(t_n, X^h_n, \alpha_n)\Big]\Big(B_l(t_{n+1}) - B_l(\tau^n_1)\Big) \\
&\quad + \sum_{l_1,l_2=1}^m L^{l_2}\sigma_{k,l_1}(t_n, X^h_n, \alpha_n)I_{l_1,l_2}(n) + \sum_{j=1}^8 r_{k,n,j}.
\end{aligned}
\tag{2.31}
$$

Next, we write $Y^h_n = \big(Y^h_{1,n}, Y^h_{2,n}, \ldots, Y^h_{d,n}\big)'$. Since $\varsigma(n) = \tau^n_1$ on the set $\{N_n = 1\}$ and $\varsigma(n) = t_{n+1}$ on the set $\{N_n \neq 1\}$, we have $B_l(t_{n+1}) - B_l(\varsigma(n)) = \mathbb{1}\big(N_n = 1\big)\big[B_l(t_{n+1}) - B_l(\tau^n_1)\big]$. Thus, in view of (2.31), the consideration regarding the terms involving double integrals with respect to the optional quadratic variation processes $[M_{i_0j_0}]$ and the Brownian motions, and the discussion on approximation of the jump times of the Markov chain α, the component sequences $(Y^h_{k,n}, n \geq 0)$ of the approximate solution should satisfy the following recursive equation

$$
\begin{aligned}
Y^h_{k,n+1} &= Y^h_{k,n} + hb_k(t_n, Y^h_n, \alpha_n) + \sum_{l=1}^m \sigma_{k,l}(t_n, Y^h_n, \alpha_n)\Delta_n B_l + \sum_{l_1,l_2=1}^m L^{l_2}\sigma_{k,l_1}(t_n, Y^h_n, \alpha_n)I_{l_1,l_2}(n) \\
&\quad + \sum_{l=1}^m \mathbb{1}\big(N_n = 1\big)\Big[\sigma_{k,l}(t_n, Y^h_n, \alpha_{n+1}) - \sigma_{k,l}(t_n, Y^h_n, \alpha_n)\Big]\Big(B_l(t_{n+1}) - B_l(\tau^n_1)\Big)
\end{aligned}
\tag{2.32}
$$

for $k = 1, 2, \ldots, d$ and $n \geq 0$. That was how we came up with the proposed numerical scheme (2.9).

We now give an estimate on the difference between the exact solution and the approximate one at each grid point. For $n \geq 0$, $k = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, 8$, denote

$$
R_{k,n,j} = \sum_{i=0}^n r_{k,i,j}. \tag{2.33}
$$

By applying recursively equations (2.31) and (2.32) we obtain

$$
\begin{aligned}
X^h_{k,n+1} - Y^h_{k,n+1} = X^h_{k,0} - Y^h_{k,0} &+ h\sum_{i=0}^{n}\Big(b_k\big(t_i, X_i^h, \alpha_i\big) - b_k\big(t_i, Y_i^h, \alpha_i\big)\Big) \\
&+ \sum_{i=0}^{n}\sum_{l=1}^{m}\Big[\sigma_{k,l}\big(t_i, X_i^h, \alpha_i\big) - \sigma_{k,l}\big(t_i, Y_i^h, \alpha_i\big)\Big]\Delta_i B_l \\
&+ \sum_{i=0}^{n}\sum_{l_1,l_2=1}^{m}\Big[L^{l_2}\sigma_{k,l_1}(t_i, X_i^h, \alpha_i) - L^{l_2}\sigma_{k,l_1}(t_i, Y_i^h, \alpha_i)\Big] I_{l_1,l_2}(i) \\
&+ \sum_{i=0}^{n}\sum_{l=1}^{m}\mathbb{1}(N_i = 1)\Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i)\Big) \\
&- \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i)\Big)\Big]\Big[B_l(t_{i+1}) - B_l(\tau_1^i)\Big] + \sum_{j=1}^{8} R_{k,n,j}.
\end{aligned}
$$

This and the Cauchy-Schwarz inequality imply

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n+1}|X^h_{k,p} - Y^h_{k,p}|^2 \qquad (2.34)\\
&\le C\mathbb{E}|X^h_{k,0} - Y^h_{k,0}|^2 + h^2 CE\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\Big(b_k\big(t_i, X_i^h, \alpha_i\big) - b_k\big(t_i, Y_i^h, \alpha_i\big)\Big)\Big|^2 \\
&+ C\mathbb{E}\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\sum_{l=1}^{m}\Big[\sigma_{k,l}\big(t_i, X_i^h, \alpha_i\big) - \sigma_{k,l}\big(t_i, Y_i^h, \alpha_i\big)\Big]\Delta_i B_l\Big|^2 \\
&+ C\mathbb{E}\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\sum_{l_1,l_2=1}^{m}\Big[L^{l_2}\sigma_{k,l_1}(t_i, X_i^h, \alpha_i) - L^{l_2}\sigma_{k,l_1}(t_i, Y_i^h, \alpha_i)\Big] I_{l_1,l_2}(i)\Big|^2 \\
&+ C\mathbb{E}\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\sum_{l=1}^{m}\mathbb{1}(N_i = 1)\Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i)\Big) \\
&- \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i)\Big)\Big]\Big[B_l(t_{i+1}) - B_l(\tau_1^i)\Big]\Big|^2 \\
&+ \sum_{j=1}^{8} C\mathbb{E}\sup_{0\le p\le n}\big|R_{k,p,j}\big|^2. \qquad (2.35)
\end{aligned}
$$

To establish the convergence of the proposed scheme, we need to study the right-hand side of (2.35). This is done in the remaining part of this section.

2.4.3 Error Bounds

Lemma 2.8. *Assume that Assumption (B) holds. For $1 \le n \le T/h$ and $k = 1, 2, \ldots, d$, we have the following inequality*

$$
\begin{aligned}
& h^2 \mathbb{E} \sup_{0 \le p \le n} \Big| \sum_{i=0}^{p} \Big(b_k\big(t_i, X_i^h, \alpha_i\big) - b_k\big(t_i, Y_i^h, \alpha_i\big) \Big) \Big|^2 \\
& + \mathbb{E} \sup_{0 \le p \le n} \Big| \sum_{i=0}^{p} \sum_{l=1}^{m} \Big[\sigma_{k,l}\big(t_i, X_i^h, \alpha_i\big) - \sigma_{k,l}\big(t_i, Y_i^h, \alpha_i\big) \Big] \Delta_i B_l \Big|^2 \\
& + \mathbb{E} \sup_{0 \le p \le n} \Big| \sum_{i=0}^{p} \sum_{l_1, l_2 = 1}^{m} \Big[L^{l_2} \sigma_{k,l_1}(t_i, X_i^h, \alpha_i) - L^{l_2} \sigma_{k,l_1}(t_i, Y_i^h, \alpha_i) \Big] I_{l_1, l_2}(i) \Big|^2 \\
& + \mathbb{E} \sup_{0 \le p \le n} \Big| \sum_{i=0}^{p} \sum_{l=1}^{m} 1\!\!1(N_i = 1) \Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i) \Big) \\
& \qquad\qquad - \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i) \Big) \Big] \Big[B_l(t_{i+1}) - B_l(\tau_1^i) \Big] \Big|^2 \\
& \le Ch \sum_{p=0}^{n} \mathbb{E} \sup_{0 \le i \le p} \big| X_i^h - Y_i^h \big|^2, \qquad (2.36)
\end{aligned}
$$

where C is a constant independent of h.

Proof. To bound the first term in the left-hand side of (2.36), we use the Cauchy-Schwarz inequality and the Lipschitz continuity in Assumption (B) to obtain

$$
h^2 \mathbb{E} \sup_{0 \le p \le n} \Big| \sum_{i=0}^{p} \Big(b_k\big(t_i, X_i^h, \alpha_i\big) - b_k\big(t_i, Y_i^h, \alpha_i\big) \Big) \Big|^2 \qquad (2.37)
$$

$$
\begin{aligned}
& \le n h^2 \sum_{i=0}^{n} \mathbb{E} \Big[b_k\big(t_i, X_i^h, \alpha_i\big) - b_k\big(t_i, Y_i^h, \alpha_i\big) \Big]^2 \\
& \le Ch \sum_{i=0}^{n} \mathbb{E} \big| X_i^h - Y_i^h \big|^2 \le Ch \sum_{p=0}^{n} \mathbb{E} \sup_{0 \le i \le p} \big| X_i^h - Y_i^h \big|^2. \qquad (2.38)
\end{aligned}
$$

Next, we deal with the last term in the left-hand side of (2.36). Denote $\mathcal{G}_p = \mathcal{G}_p^h = \mathcal{F}_{t_{p+1}}^B \vee \mathcal{F}_T^\alpha$

and

$$M_p = \sum_{i=0}^{p} \sum_{l=1}^{m} \mathbb{1}\big(N_i = 1\big) \Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i) \Big) \\ - \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i) \Big) \Big] \Big[B_l(t_{i+1}) - B_l(\tau_1^i) \Big].$$

Since $B_l(\cdot)$ and $\alpha(\cdot)$ are independent $(M_p, \mathcal{G}_p, p \geq 0)$ is a martingale. By Lemma 2.1, Lemma 2.6 and Assumption (B), $(M_p, \mathcal{G}_p, p \geq 0)$ is a square-integrable martingale. In addition, $\tau_1^i \wedge t_{i+1}$ is a stopping time with respect to $\mathcal{F}_t$, $\tau_1^i \wedge t_{i+1} = \tau_1^i$ on the set $\{N_i = 1\}$ and

$$\mathbb{E}\Big\{ \Big[B_l(t_{i+1}) - B_l(\tau_1^i \wedge t_{i+1}) \Big]^2 \Big| \mathcal{F}_{\tau_1^i \wedge t_{i+1}} \vee \mathcal{F}_T^\alpha \Big\} = \mathbb{E}\Big[B_l(t_{i+1}) - B_l(\tau_1^i \wedge t_{i+1}) \Big]^2 \leq Ch. \tag{2.39}$$

Hence, by the Burkholder-Davis-Gundy inequality,

$$\begin{aligned} &\mathbb{E} \sup_{0 \leq p \leq n} \Big| \sum_{i=0}^{p} \sum_{l=1}^{m} \mathbb{1}(N_i = 1) \Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i) \Big) \\ &\quad - \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i) \Big) \Big] \Big[B_l(t_{i+1}) - B_l(\tau_1^i) \Big] \Big|^2 \\ &= \mathbb{E} \sup_{0 \leq p \leq n} |M_p|^2 \\ &\leq C\mathbb{E} \sum_{i=0}^{n} \sum_{l=1}^{m} \Big| \mathbb{1}\big(N_i = 1\big) \Big[\Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, X_i^h, \alpha_i) \Big) \\ &\quad - \Big(\sigma_{k,l}(t_i, Y_i^h, \alpha_{i+1}) - \sigma_{k,l}(t_i, Y_i^h, \alpha_i) \Big) \Big] \Big[B_l(t_{i+1}) - B_l(\tau_1^i) \Big] \Big|^2. \end{aligned}$$

Conditioning on $\mathcal{F}^B_{\tau_1^i \wedge t_{i+1}} \vee \mathcal{F}_T^\alpha$, using (2.39), and noting N_i, X_i^h and Y_i^h being measurable with respect to $\mathcal{F}^B_{\tau_1^i\wedge t_{i+1}} \vee \mathcal{F}^\alpha_T$, and the Lipschitz continuity in Assumption (B), we have

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}|M_p|^2\\
&\quad\le C\mathbb{E}\sum_{i=0}^{n}\sum_{l=1}^{m}\Big|\mathbb{1}\big(N_i=1\big)\Big[\Big(\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})-\sigma_{k,l}(t_i,X_i^h,\alpha_i)\Big)\\
&\qquad-\Big(\sigma_{k,l}(t_i,Y_i^h,\alpha_{i+1})-\sigma_{k,l}(t_i,Y_i^h,\alpha_i)\Big)\Big]\Big|^2\mathbb{E}\Big\{\Big[B_l(t_{i+1})-B_l(\tau_1^i\wedge t_{i+1})\Big]^2\Big|\mathcal{F}_{\tau_1^i\wedge t_{i+1}}\vee\mathcal{F}^\alpha_T\Big\}\\
&\quad\le Ch\sum_{i=0}^{n}\sum_{l=1}^{m}\mathbb{E}\Big[\Big(\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})-\sigma_{k,l}(t_i,Y_i^h,\alpha_{i+1})\Big) \qquad (2.40)\\
&\qquad\qquad-\Big(\sigma_{k,l}(t_i,X_i^h,\alpha_i)-\sigma_{k,l}(t_i,Y_i^h,\alpha_i)\Big)\Big]^2\\
&\quad\le Ch\sum_{p=0}^{n}\mathbb{E}\sup_{0\le i\le p}\big|X_i^h-Y_i^h\big|^2. \qquad (2.41)
\end{aligned}
$$

Similarly, we can use the Burkholder-Davis-Gundy inequality and the following inequality

$$
\mathbb{E}|\Delta_i B_l|^2\le Ch,\quad \mathbb{E}|I_{l_1,l_2}(i)|^2\le Ch\qquad 0\le i\le T/h;\;\; l,l_1,l_2=1,2,\dots,m
$$

instead of (2.39) to bound the remaining terms in the left-hand side of (2.36) and get

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\sum_{l=1}^{m}\Big[\sigma_{k,l}\big(t_i,X_i^h,\alpha_i\big)-\sigma_{k,l}\big(t_i,Y_i^h,\alpha_i\big)\Big]\Delta_iB_l\Big|^2\\
&\quad+\mathbb{E}\sup_{0\le p\le n}\Big|\sum_{i=0}^{p}\sum_{l_1,l_2=1}^{m}\Big[L^{l_2}\sigma_{k,l_1}(t_i,X_i^h,\alpha_i)-L^{l_2}\sigma_{k,l_1}(t_i,Y_i^h,\alpha_i)\Big]I_{l_1,l_2}(i)\Big|^2\\
&\qquad\le Ch\sum_{p=0}^{n}\mathbb{E}\sup_{0\le i\le p}\big|X_i^h-Y_i^h\big|^2. \qquad (2.42)
\end{aligned}
$$

By combining (2.38), (2.41), and (2.42), the proof is complete. □

Lemma 2.9. *Assume that Assumption (B) is satisfied. Then, for $k=1,2,\dots,d$, there is a constant C independent of h such that $\mathbb{E}\sup_{0\le p\le T/h}\big|R_{k,p,7}\big|^2\le Ch^2$.*

Proof. Let $\mathcal{G}_{p,7}=\mathcal{G}^h_{p,7}=\mathcal{F}_{\tau_1^{p+1}\wedge t_{p+2}}\vee\mathcal{F}^\alpha_T$ for $p=1,2,\dots$. It is clear from (2.29) and (2.33) that

$$R_{k,p,7} = \sum_{i=0}^{p} \sum_{l=1}^{m} \mathbb{1}(N_i = 1)\Big[\Big(\sigma_{k,l}(\tau_1^i, X(\tau_1^i), \alpha_{i+1}) - \sigma_{k,l}(\tau_1^i, X(\tau_1^i), \alpha_i)\Big)$$
$$- \Big(\sigma_{k,l}(t_i, X_i^h, \alpha_{i+1})) - \sigma_{k,l}(t_i, X_i^h, \alpha_i)\Big)\Big]\Big[B_l(t_{i+1}) - B_l(\tau_1^i)\Big]$$

is $\mathcal{G}_{p,7}^h$-measurable. Note that $\tau_1^{p+1} = \tau_1^{p+1} \wedge t_{p+2}$ on the set $\{N_{p+1} = 1\}$. In addition, $\tau_1^{p+1} \wedge t_{p+2}$ is a stopping time with respect to $\mathcal{F}_t$ and $\mathbb{E}\big[B_l(t_{p+2}) - B_l(\tau_1^{p+1} \wedge t_{p+2})\big|\mathcal{G}_{p,7}\big] = 0$ because of the independence between $B_l(\cdot)$ and $\alpha(\cdot)$. Therefore,

$$\begin{aligned}
&\mathbb{E}\Big[R_{k,p+1,7} - R_{k,p,7}\Big|\mathcal{G}_{p,7}\Big] \\
&= \mathbb{E}\Big\{\mathbb{1}(N_{p+1} = 1)\Big[\Big(\sigma_{k,l}(\tau_1^{p+1}, X(\tau_1^{p+1}), \alpha_{p+2}) - \sigma_{k,l}(\tau_1^{p+1}, X(\tau_1^{p+1}), \alpha_{p+1})\Big) \\
&\qquad - \Big(\sigma_{k,l}(t_{p+1}, X_{p+1}^h, \alpha_{p+2}) - \sigma_{k,l}(t_{p+1}, X_{p+1}^h, \alpha_{p+1})\Big)\Big]\Big[B_l(t_{p+2}) - B_l(\tau_1^{p+1} \wedge t_{p+2})\Big]\Big|\mathcal{G}_{p,7}\Big\} \\
&= \mathbb{1}(N_{p+1} = 1)\Big[\Big(\sigma_{k,l}(\tau_1^{p+1}, X(\tau_1^{p+1}), \alpha_{p+2}) - \sigma_{k,l}(\tau_1^{p+1}, X(\tau_1^{p+1}), \alpha_{p+1})\Big) \\
&\qquad - \Big(\sigma_{k,l}(t_{p+1}, X_{p+1}^h, \alpha_{p+2}) - \sigma_{k,l}(t_{p+1}, X_{p+1}^h, \alpha_{p+1})\Big)\Big]\mathbb{E}\Big[B_l(t_{p+2}) - B_l(\tau_1^{p+1} \wedge t_{p+2})\Big|\mathcal{G}_{p,7}\Big] \\
&= 0.
\end{aligned}$$

This implies that $(R_{k,p,7}, \mathcal{G}_{p,7}, p \geq 0)$ is a martingale. By Lemma 2.1 and Assumption (B), it is a square integrable martingale. Hence, by the Burkholder-Davis-Gundy inequality, and then conditioning on $\mathcal{F}_{\tau_1^i\wedge t_{i+1}}\vee\mathcal{F}_T^\alpha$, we have

$$\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}\Big|R_{k,p,7}\Big|^2\\
&\quad\le\sum_{i=0}^n\mathbb{E}\Big\{\Big|\sum_{l=1}^m \mathbb{1}(N_i=1)\Big[\Big(\sigma_{k,l}(\tau_1^i,X(\tau_1^i),\alpha_{i+1})-\sigma_{k,l}(\tau_1^i,X(\tau_1^i),\alpha_i)\Big)\\
&\qquad\quad-\Big(\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})-\sigma_{k,l}(t_i,X_i^h,\alpha_i)\Big)\Big]\Big[B_l(t_{i+1})-B_l(\tau_1^i)\Big]\Big|^2\Big\}\\
&\quad\le m\sum_{i=0}^n\sum_{l=1}^m\mathbb{E}\Big\{\Big|\mathbb{1}(N_i=1)\Big[\Big(\sigma_{k,l}(\tau_1^i,X(\tau_1^i),\alpha_{i+1})-\sigma_{k,l}(\tau_1^i,X(\tau_1^i),\alpha_i)\Big)\\
&\qquad\quad-\Big(\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})-\sigma_{k,l}(t_i,X_i^h,\alpha_i)\Big)\Big]\Big[B_l(t_{i+1})-B_l(\tau_1^i)\Big]\Big|^2\Big\}\\
&\quad= m\sum_{i=0}^n\sum_{l=1}^m\mathbb{E}\Big\{\mathbb{1}(N_i=1)\Big[\Big(\sigma_{k,l}(\tau_1^i\wedge t_{i+1},X(\tau_1^i\wedge t_{i+1}),\alpha_{i+1})-\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})\Big)\\
&\qquad\quad-\Big(\sigma_{k,l}(\tau_1^i\wedge t_{i+1},X(\tau_1^i\wedge t_{i+1}),\alpha_i)-\sigma_{k,l}(t_i,X_i^h,\alpha_i)\Big)\Big]^2\\
&\qquad\quad\times\mathbb{E}\Big\{\Big[B_l(t_{i+1})-B_l(\tau_1^i\wedge t_{i+1})\Big]^2\Big|\mathcal{F}_{\tau_1^i\wedge t_{i+1}}\vee\mathcal{F}_T^\alpha\Big\}\Big\}.
\end{aligned}$$

Using (2.39), the Cauchy-Schwarz inequality, and Lemma 2.1 and Assumption (B), we can estimate further that

$$\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}\Big|R_{k,p,7}\Big|^2\\
&\quad\le Ch\sum_{i=0}^n\sum_{l=1}^m\mathbb{E}\Big\{\mathbb{1}(N_i=1)\Big[\Big(\sigma_{k,l}(\tau_1^i\wedge t_{i+1},X(\tau_1^i\wedge t_{i+1}),\alpha_{i+1})-\sigma_{k,l}(t_i,X_i^h,\alpha_{i+1})\Big)\\
&\qquad\quad-\Big(\sigma_{k,l}(\tau_1^i\wedge t_{i+1},X(\tau_1^i\wedge t_{i+1}),\alpha_i)-\sigma_{k,l}(t_i,X_i^h,\alpha_i)\Big)\Big]^2\Big\}\\
&\quad= Ch\sum_{i=0}^n\sum_{l=1}^m\mathbb{E}\Big\{\mathbb{E}\Big\{\mathbb{1}(N_i=1)\Big[\Big|X(\tau_1^i\wedge t_{i+1})-X_i^h\Big|^2\\
&\qquad\qquad+\big(1+\sup_{0\le t\le T}|X(t)|^2\big)\big|\tau_1^i\wedge t_{i+1}-t_i\big|\Big]\Big|\mathcal{F}_T^\alpha\Big\}\Big\}\\
&\quad\le Ch\sum_{i=0}^n\mathbb{E}\Big\{\mathbb{1}(N_i=1)\mathbb{E}\Big\{\Big[\sup_{t_i\le t\le t_{i+1}}\Big|X(t)-X(t_i)\Big|^2+h\Big(1+\sup_{0\le t\le T}|X(t)|^2\Big)\Big]\Big|\mathcal{F}_T^\alpha\Big\}\Big\}\\
&\quad\le Ch^2\sum_{i=0}^n\mathbb{E}\mathbb{1}(N_i=1)\le Cnh^3\le Ch^2
\end{aligned}\tag{2.43}$$

for any $0 \le n \le T/h$, which is the desired result. $\square$

Lemma 2.10. *Assume that Assumption* (B) *holds and* $h < 1/(2q)$*. Then, for* $k = 1, 2, \ldots, d$*, there is a constant* C *independent of* h *such that* $\mathbb{E}\sup_{0\le p\le T/h}\left|R_{k,p,8}\right|^2 \le Ch^2$.

Proof. According to (2.30) and (2.33), for $p = 1, 2, \ldots,$ we have $R_{k,p,8} = \sum_{i=0}^{p}\sum_{l=1}^{m}\sum_{i_0\neq j_0}\mathbb{1}(N_i \ge 2)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s}\Big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\Big)d[M_{i_0j_0}](u)dB_l(s)$. Denote $\mathcal{G}_{p,8} = \mathcal{G}_{p,8}^h = \mathcal{F}_{t_{p+1}} \vee \mathcal{F}_T^\alpha$. Since N_p is $\mathcal{G}_{p,8}$-measurable, we can show that $(R_{k,p,8}, \mathcal{G}_{p,8}, p \ge 0)$ is a square integrable martingale.

By the Burkholder-Davis-Gundy inequality, the Cauchy-Schwarz inequality, and Lemma 2.7, for any $n \ge 0$ we obtain

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,8}\right|^2\\
&\le C\sum_{i=0}^{n}\mathbb{E}\bigg(\Big|\sum_{l=1}^{m}\sum_{i_0\neq j_0}\mathbb{1}(N_i\ge 2)\int_{t_i}^{t_{i+1}}\int_{t_i}^{s}\Big(\sigma_{k,l}(u,X(u),j_0)\\
&\qquad\qquad - \sigma_{k,l}(u,X(u),i_0)\Big)d[M_{i_0j_0}](u)dB_l(s)\Big|^2\bigg)\\
&\le C\sum_{i=0}^{n}\sum_{l=1}^{m}\sum_{N=2}^{\infty}\mathbb{E}\bigg(\Big|\mathbb{1}(N_i=N)\sum_{i_0\neq j_0}\int_{t_i}^{t_{i+1}}\int_{t_i}^{s}\Big(\sigma_{k,l}(u,X(u),j_0)\\
&\qquad\qquad - \sigma_{k,l}(u,X(u),i_0)\Big)d[M_{i_0j_0}](u)dB_l(s)\Big|^2\bigg)\\
&= C\sum_{i=0}^{n}\sum_{l=1}^{m}\sum_{N=2}^{\infty}\mathbb{E}\bigg(\Big|\mathbb{1}(N_i=N)\sum_{j=1}^{N}\Big[\sigma_{k,l}(\tau_j^i,X(\tau_j^i),\alpha(\tau_j^i)) - \sigma_{k,l}(\tau_j^i,X(\tau_j^i),\alpha(\tau_{j-1}^i))\Big]\\
&\qquad\qquad \times\Big(B_l(t_{i+1}) - B_l(\tau_j^i)\Big)\Big|^2\bigg)\\
&\le C\sum_{i=0}^{n}\sum_{l=1}^{m}\sum_{N=2}^{\infty}N\sum_{j=1}^{N}\mathbb{E}\bigg(\mathbb{1}(N_i=N)\Big|\Big[\sigma_{k,l}(\tau_j^i,X(\tau_j^i),\alpha(\tau_j^i)) - \sigma_{k,l}(\tau_j^i,X(\tau_j^i),\alpha(\tau_{j-1}^i))\Big]\\
&\qquad\qquad \times\Big(B_l(t_{i+1}) - B_l(\tau_j^i)\Big)\Big|^2\bigg).
\end{aligned}
$$

Conditioning on $\mathcal{F}_{\tau_j^i\wedge t_{i+1}} \vee \mathcal{F}_T^\alpha$, using Lemma 2.7 and the fact that $\tau_j^i = \tau_j^i \wedge t_{i+1}$ on the set

$\{N_i = N\}$ for $0 \le i \le n$, $j = 1, 2, \ldots, N$, and $N \ge 2$, we have

$$
\begin{aligned}
&\mathbb{E} \sup_{0 \le p \le n} \big| R_{k,p,8} \big|^2 \\
&\le C \sum_{i=0}^{n} \sum_{l=1}^{m} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \Big[\sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_j^i \wedge t_{i+1})) \\
&\qquad\qquad - \sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_{j-1}^i \wedge t_{i+1})) \Big]^2 \\
&\qquad\qquad\qquad\qquad \times \Big(B_l(t_{i+1}) - B_l(\tau_j^i \wedge t_{i+1}) \Big)^2 \Big| \mathcal{F}_{\tau_j^i \wedge t_{i+1}} \vee \mathcal{F}_T^\alpha \Big\} \Big\} \\
&= C \sum_{i=0}^{n} \sum_{l=1}^{m} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \Big[\sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_j^i \wedge t_{i+1})) \\
&\qquad\qquad - \sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_{j-1}^i \wedge t_{i+1})) \Big]^2 \\
&\qquad\qquad\qquad\qquad \times \mathbb{E}\Big\{ \Big(B_l(t_{i+1}) - B_l(\tau_j^i \wedge t_{i+1}) \Big)^2 \Big| \mathcal{F}_{\tau_j^i \wedge t_{i+1}} \vee \mathcal{F}_T^\alpha \Big\} \Big\} \\
&\le Ch \sum_{i=0}^{n} \sum_{l=1}^{m} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \Big[\sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_j^i \wedge t_{i+1})) \\
&\qquad\qquad - \sigma_{k,l}(\tau_j^i \wedge t_{i+1}, X(\tau_j^i \wedge t_{i+1}), \alpha(\tau_{j-1}^i \wedge t_{i+1})) \Big]^2 \Big\} \\
&\le Ch \sum_{i=0}^{n} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \Big[1 + \sup_{0 \le t \le T} |X(t)|^2 \Big] \Big\}.
\end{aligned}
$$

Note that we have used Assumption (B) in the last inequality. Next, by Lemma 2.1 and Lemma 2.5, we obtain

$$
\begin{aligned}
\mathbb{E} \sup_{0 \le p \le n} \big| R_{k,p,8} \big|^2 &\le Ch \sum_{i=0}^{n} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \Big[1 + \sup_{0 \le t \le T} |X(t)|^2 \Big] \Big\} \\
&= Ch \sum_{i=0}^{n} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E}\Big\{ \mathbb{1}(N_i = N) \mathbb{E}\Big[1 + \sup_{0 \le t \le T} |X(t)|^2 \Big| \mathcal{F}_T^\alpha \Big] \Big\} \\
&\le Ch \sum_{i=0}^{n} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} \mathbb{E} \mathbb{1}(N_i = N) \le Ch \sum_{i=0}^{n} \sum_{N=2}^{\infty} N \sum_{j=1}^{N} (qh)^N \\
&= C \sum_{i=0}^{n} \sum_{N=2}^{\infty} N^2 (qh)^{N+1} \le Cnh^3 \le Ch^2
\end{aligned}
$$

for any $0 \le n \le T/h$ and $h < 1/(2q)$. We have used the fact that N_i is $\mathcal{F}_T^\alpha$−measurable in the first equation and Lemma 2.1 in the third inequality. This completes the proof. □

Lemma 2.11. *Assume that Assumption* (B) *holds and* $h < 1/(2q)$*. Then there is a constant* C *independent of* h *such that* $\sum_{j=1}^{6} \mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,j}\right|^2 \le Ch^2$ *for any* $0 \le n \le T/h$ *and* $k = 1, 2, \ldots, d$.

Proof. The proof is carried out by establishing 4 claims.

Claim 1: Prove that $E\sup_{0\le p\le n}\left|R_{k,p,1}\right|^2 + E\sup_{0\le p\le n}\left|R_{k,p,4}\right|^2 \le Ch^2$. Note that from (2.8) we have

$$\mathcal{L}b_k(t,x,i_0) = L^0 b_k(t,x,i_0) + \frac{1}{2}\sum_{l=1}^{m} L^l L^l b_k(t,x,i_0) + Qb_k(t,x,\cdot)(i_0).$$

It follows from Assumption (B) that for $0 \le t \le T$,

$$|\mathcal{L}b_k(t,x,i_0)| \le C(1+|x|). \tag{2.44}$$

Next, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,1}\right|^2 &\le n\sum_{p=0}^{n}\mathbb{E}\Big|\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathcal{L}b_k(u,X(u),\alpha(u))duds\Big|^2\\
&\le n\sum_{p=0}^{n}h^2\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathbb{E}\big|\mathcal{L}b_k(u,X(u),\alpha(u))\big|^2 duds\\
&\le Cnh^2\sum_{p=0}^{n}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\big(1+\mathbb{E}\sup_{0\le t\le T}|X(t)|^2\big)duds \le Cn^2h^4 = Ch^2.
\end{aligned}$$

We have used (2.44) in the third inequality and Lemma 2.1 in the last inequality. By a similar way, we can use Burkholder-Davis-Gundy to prove that $\mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,4}\right|^2 \le Ch^2$. Claim 1 is therefore proved.

Claim 2: Prove that $\mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,2}\right|^2 + \mathbb{E}\sup_{0\le p\le n}\left|R_{k,p,3}\right|^2 \le Ch^2$. Since $M_{i_0 j_0}(t)$ is a martingale, we can prove that for each fixed k, $(R_{k,p,3}, \mathcal{G}_{p,3}, p \ge 0)$ is also a square integrable martingale, where $\mathcal{G}_{p,3} = \mathcal{G}^h_{p,3} = \mathcal{F}_{t_{p+1}}$. Thus, by the Burkholder-Davis-Gundy and the Cauchy-Schwarz inequalities we have

$$
\begin{aligned}
&\mathbb{E} \sup_{0\le p\le n} \left|R_{k,p,3}\right|^2 \\
&\le \sum_{p=0}^{n} \mathbb{E}\Big| \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \big(b_k(u,X(u),j_0) - b_k(u,X(u),i_0)\big) dM_{i_0j_0}(u) ds\Big|^2 \\
&\le C \sum_{p=0}^{n} \sum_{i_0\neq j_0} \mathbb{E}\Big| \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \big(b_k(u,X(u),j_0) - b_k(u,X(u),i_0)\big) dM_{i_0j_0}(u) ds\Big|^2 \\
&\le Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E}\Big| \int_{t_p}^{s} \big(b_k(u,X(u),j_0) - b_k(u,X(u),i_0)\big) dM_{i_0j_0}(u)\Big|^2 ds \\
&= Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E} \int_{t_p}^{s} \Big|\big(b_k(u,X(u),j_0) - b_k(u,X(u),i_0)\big)\Big|^2 d[M_{i_0j_0}](u) ds \\
&\le Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E} \int_{t_p}^{s} \Big[1 + \sup_{0\le t\le T} |X(t)|^2\Big] d[M_{i_0j_0}](u) ds \\
&= Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E} \int_{t_p}^{s} \Big[1 + E\big(\sup_{0\le t\le T} |X(t)|^2 \big| \mathcal{F}^\alpha_T\big)\Big] d[M_{i_0j_0}](u) ds \\
&\le Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E}\Big\{[M_{i_0j_0}](s) - [M_{i_0j_0}](t_p)\Big\} ds \\
&\le Ch \sum_{p=0}^{n} \sum_{i_0\neq j_0} \int_{t_p}^{t_{p+1}} \mathbb{E}\big[N_p\big] ds \le Cnh^3 \le Ch^2.
\end{aligned}
$$

Similarly, we obtain $E \sup_{0\le p\le n} \left|R_{k,p,2}\right|^2 \le Ch^2$.

Claim 3: Prove that $\mathbb{E} \sup_{0\le p\le n} \left|R_{k,p,5}\right|^2 \le Ch^2$. By the definition of $\langle M_{i_0j_0}\rangle$, the Cauchy-

Schwarz inequality, and Assumption (B), we have

$$
\begin{aligned}
&\mathbb{E}\Big|\int_{t_p}^{s}\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)d\langle M_{i_0j_0}\rangle(u)\Big|^2\\
&\quad=\mathbb{E}\Big|\int_{t_p}^{s}\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)q_{i_0j_0}\mathbb{1}(\alpha(u-)=i_0)du\Big|^2\\
&\quad\le(s-t_p)\mathbb{E}\int_{t_p}^{s}\Big|\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)q_{i_0j_0}\mathbb{1}(\alpha(u-)=i_0)\Big|^2du\\
&\quad\le C(s-t_p)\mathbb{E}\Big|\int_{t_p}^{s}\big(1+\sup_{0\le t\le T}|X(t)|^2\big)du\Big|^2\le C(s-t_p)^2\le Ch^2 \qquad (2.45)
\end{aligned}
$$

for $t_p\le s\le t_{p+1}$. To proceed, we observe that for fixed k, $(R_{k,p,5},\mathcal{G}_{p,5},p\ge 0)$ is a square integrable martingale, where $\mathcal{G}_{p,5}=\mathcal{G}^h_{p,5}=\mathcal{F}^B_{t_{p+1}}\vee\mathcal{F}^\alpha_T$. Thus, by the Burkholder-Davis-Gundy inequality, the Cauchy-Schwarz inequality and (2.45)

$$
\begin{aligned}
&\mathbb{E}\sup_{0\le p\le n}\big|R_{k,p,5}\big|^2\\
&\le C\sum_{p=1}^{n}\mathbb{E}\Big|\sum_{l=1}^{m}\sum_{i_0\ne j_0}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)d\langle M_{i_0j_0}\rangle(u)dB_l(s)\Big|^2\\
&\le Cmm_0^2\sum_{p=1}^{n}\sum_{l=1}^{m}\sum_{i_0\ne j_0}\mathbb{E}\Big|\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)d\langle M_{i_0j_0}\rangle(u)dB_l(s)\Big|^2\\
&=Cmm_0^2\sum_{p=1}^{n}\sum_{l=1}^{m}\sum_{i_0\ne j_0}\int_{t_p}^{t_{p+1}}\mathbb{E}\Big|\int_{t_p}^{s}\big(\sigma_{k,l}(u,X(u),i_0)-\sigma_{k,l}(u,X(u),j_0)\big)d\langle M_{i_0j_0}\rangle(u)\Big|^2ds\\
&\le C\sum_{p=1}^{n}\sum_{l=1}^{m}\sum_{i_0\ne j_0}\int_{t_p}^{t_{p+1}}h^2ds\le Cnh^3=Ch^2.
\end{aligned}
$$

Claim 4: Prove that $\mathbb{E}\sup_{0\le p\le n}\big|R_{k,p,6}\big|^2\le Ch^2$. It is clear that $(R_{k,p,6},\mathcal{G}_{p,6},p\ge 0)$ is a square integrable martingale where $\mathcal{G}_{p,6}=\mathcal{G}^h_{p,6}=\mathcal{F}^B_{t_{p+1}}\vee\mathcal{F}^\alpha_T$. By the Burkholder-Davis-Gundy and the Cauchy-Schwarz inequalities, we have

$$
\begin{aligned}
&\mathbb{E} \sup_{0\le p\le n} \big|R_{k,p,6}\big|^2 \\
&\le C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \mathbb{E}\Big| \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \Big(L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\Big) dB_{l_2}(u) dB_{l_1}(s)\Big|^2 \\
&= C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \mathbb{E}\Big| \int_{t_p}^{s} L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p) dB_{l_2}(u)\Big|^2 ds \\
&= C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2 duds \\
&= C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p=0)\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2\Big] duds \\
&\quad + C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p\ge 1)\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2\Big] duds.
\end{aligned} \tag{2.46}
$$

To proceed, note that on $\{N_p = 0\}$, $\alpha(u) = \alpha_p$ for $t_p < u < t_{p+1}$. Thus, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p=0)\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u)) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2\Big] duds \\
&= \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p=0)\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha_p) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2\Big] duds \\
&\le 2 \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p=0)\big| L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha_p) - L^{l_2}\sigma_{k,l_1}(t_p,X(u),\alpha_p)\big|^2\Big] duds \\
&\quad + 2 \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\Big[\mathbb{1}(N_p=0)\big| L^{l_2}\sigma_{k,l_1}(t_p,X(u),\alpha_p) - L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\big|^2\Big] duds \\
&\le C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} (u-t_p)\big[1 + \mathbb{E} \sup_{0\le t\le T} |X(t)|^2\big] duds \\
&\quad + C \sum_{p=1}^{n} \sum_{l_1,l_2=1}^{m} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \mathbb{E}\big|X(u) - X_p^h\big|^2 duds \\
&\le C \sum_{p=1}^{n} \int_{t_p}^{t_{p+1}} \int_{t_p}^{s} \Big\{(u-t_p)\big[1 + \mathbb{E} \sup_{0\le t\le T} |X(t)|^2\big] + (u-t_p)\Big\} duds \le Cnh^3 \le Ch^2.
\end{aligned} \tag{2.47}
$$

We have used the Cauchy-Schwarz inequality in the first inequality, Assumption (B) in the second inequality, and Lemma 2.1 in the third inequality. Next, by Assumption (B) and Lemma 2.1, we have

$$\begin{aligned}
&C\sum_{p=1}^{n}\sum_{l_1,l_2=1}^{m}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathbb{E}\Big[\mathbb{1}(N_p\geq 1)\Big|L^{l_2}\sigma_{k,l_1}(u,X(u),\alpha(u))-L^{l_2}\sigma_{k,l_1}(t_p,X_p^h,\alpha_p)\Big|^2\Big]duds\\
&\leq C\sum_{p=1}^{n}\sum_{l_1,l_2=1}^{m}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathbb{E}\Big[\mathbb{1}(N_p\geq 1)\big(1+\sup_{0\leq t\leq T}|X(t)|^2\big)\Big]duds\\
&= C\sum_{p=1}^{n}\sum_{l_1,l_2=1}^{m}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathbb{E}\Big\{\mathbb{1}(N_p\geq 1)\mathbb{E}\Big[1+\sup_{0\leq t\leq T}|X(t)|^2\Big|\mathcal{F}_T^{\alpha}\Big]\Big\}duds\\
&\leq C\sum_{p=1}^{n}\sum_{l_1,l_2=1}^{m}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}\mathbb{E}\mathbb{1}(N_p\geq 1)duds\\
&= C\sum_{p=1}^{n}\sum_{l_1,l_2=1}^{m}\int_{t_p}^{t_{p+1}}\int_{t_p}^{s}hduds\leq Cnh^3=Ch^2. \qquad (2.48)
\end{aligned}$$

Claim 4 now follows by combining (2.46), (2.47), and (2.48). □

2.4.4 Proof of Main Theorem

Proof. By Lemma 2.1 and Lemma 2.6, it suffices to prove the inequality in Theorem 2.4 for $0<h<1/(2q)$ where $q=\max\{-q_{i_0i_0}:i_0\in\mathcal{M}\}$. By (2.35) and the results of Lemmas 2.8, 2.9, 2.10 and 2.11, we have

$$\begin{aligned}
\mathbb{E}\Big[\sup_{0\leq p\leq n+1}\big|X_p^h-Y_p^h\big|^2\Big] &\leq CE|X_0-Y_0^h|^2+Ch\sum_{i=0}^{n}\mathbb{E}\Big[\sup_{0\leq p\leq i}|X_p^h-Y_p^h|^2\Big]+Ch^2\\
&\leq Ch^2+Ch\sum_{i=0}^{n}\mathbb{E}\Big[\sup_{0\leq p\leq i}\big|X_p^h-Y_p^h\big|^2\Big]
\end{aligned}$$

for $0\leq n\leq T/h$ and $0<h<1/(2q)$. Thus, by Gronwall's inequality we conclude that $\mathbb{E}\Big[\sup_{0\leq p\leq T/h}\big|X_p^h-Y_p^h\big|^2\Big]\leq Ch^2e^{C(T/h)h}\leq Ch^2$, which is the desired claim. □

2.5 Numerical Examples

In this section, we consider several examples. They are for demonstration purposes.

Example 2.12. The first example is a modification of the Ornstein-Uhlenbeck process which is used, as one of several approaches, to model the interest rates, currency exchange rates, and commodity prices under the influence of the randomness.

$$\begin{aligned} dX(t) &= \left(\alpha(t) - \frac{1}{2}\right)(1 - X(t))dt + \frac{1}{2}\left(\alpha(t) - \frac{1}{2}\right)X(t)dB(t), \\ X(0) &= 0.5,\ \alpha(0) = 1,\ 0 \le t \le 1, \end{aligned} \tag{2.49}$$

where $B(t)$ is a one-dimensional standard Brownian motion and $\alpha(\cdot)$ is a Markov chain whose the generator matrix Q of α is given by $Q = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$, and the α is independent with B. The state space of the Markov chain is $\mathcal{M} = \{0, 1\}$. The solution to (2.49) can be written as a closed-form expression involving a stochastic integral. However, for the sake of simplicity, we construct the Milstein-type scheme with $h = \delta = 2^{-17}$ to be a good approximation of the exact solution and compare this with the Milstein-type approximations using $h = 128\delta, h = 64\delta, h = 32\delta$, and $h = 16\delta$. The simulation was carried out by first generating 200 realizations of the Markov chains α and then, for each fixed realization of α, we simulated 500 sample paths of X. The resulting solid line log-log error plot is shown with a reference line of slope 1 (the dashed one) in Fig. 2.12. For the plotted trial, the empirical rate of convergence which is the slope of the least squares regression line for the solid line is 1.0496 and is closed to the expected one. The least squares residual is 0.0751.

Example 2.13. In the second exam, we consider the following highly non-linear switching diffusion

$$\begin{aligned} dX(t) &= b\left(X(t), \alpha(t)\right)dt + \sigma(X(t), \alpha(t))dB(t), \\ X(0) &= 5, \quad \alpha(0) = 1, \quad 0 \le t \le 1, \end{aligned} \tag{2.50}$$

where $B(\cdot)$ is a one-dimensional standard Brownian motion and $\alpha(\cdot)$ is a Markov chain whose

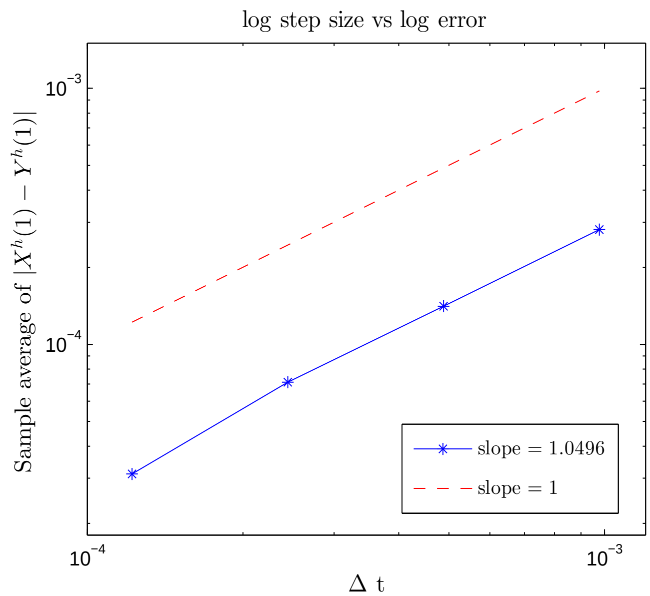

Figure 1: log-log plot for Example 1

generator Q and state space are given as in Example 2.12; α is independent with B. The drift $b(\cdot,\cdot)$ and the diffusion coefficient $\sigma(\cdot,\cdot)$ are given as follows

$$b(x,0) = 3(2+\sin x),\ b(x,1) = 2(2+\sin x\cos x),\ \sigma(x,0) = -0.2x,\ \sigma(x,1) = -0.3x.$$

It is easy to verify that the Assumption (B) holds for this equation. Due to the nonlinear nature of equation (2.50), a closed-form solution to it is hardly, if not impossible, to obtain. We therefore, as already did in the previous example, construct the Milstein-type scheme with $h = \delta = 2^{-17}$ to be a good approximation of the exact solution and compare this with the Milstein approximations using $h = 128\delta, h = 64\delta, h = 32\delta$, and $h = 16\delta$. The simulation procedure was carried out in similar manners as in Example 2.12. The resulting log-log error plot is shown in Fig. 2.13. For the plotted trial, the least squares regression gives the empirical rate of convergence to be 0.9995 and the least squares residual to be 0.0382. Hence, our results are consistent with a strong order of convergence equal to one. We observe that

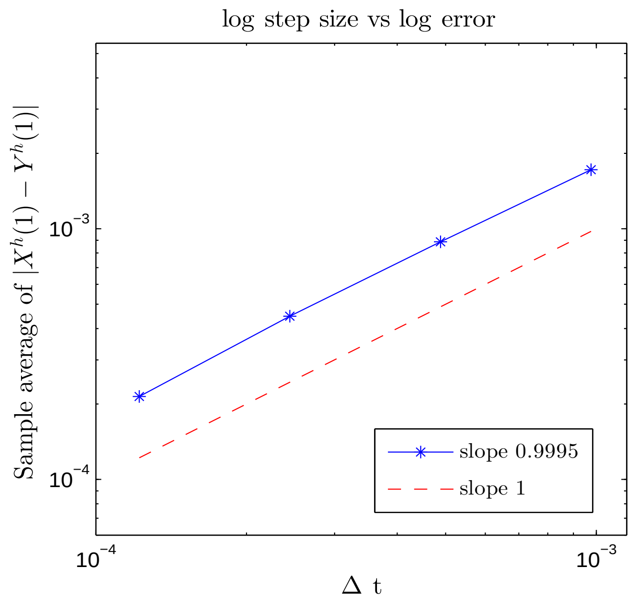

Figure 2: log-log plot for Example 2

the proposed scheme is quite sensitive to the high oscillation in the diffusion coefficient.

It seems that the Assumption (B) might be weakened in some practical situations where the equation of interest has some special structures. The last example gives an instance of such a situation.

Example 2.14. In the last example, we consider a variant of the Lotka-Volterra equation that arises in population dynamics and is given as follows

$$\begin{aligned} & dX(t) = (\alpha(t) - \frac{1}{2})X(t)(1 - X(t))dt + \frac{1}{2}(\alpha(t) - \frac{1}{2})X(t)dB(t), \\ & X(0) = 0.5, \alpha(0) = 1,\ 0 \leq t \leq 1. \end{aligned} \tag{2.51}$$

In this equation the Markov chain α represents the random changes in the biology system and has the state space $\mathcal{M} = \{0, 1\}$. One may consult [52] for a discussion on this model. In this example we construct the Milstein-type scheme with $h = \delta = 2^{-21}$ to be a good approximation of the exact solution and compare this with the Milstein approximations using $h = 2^{13}\delta, h = 2^{10}\delta, h = 2^{7}\delta$, and $h = 2^{4}\delta$. The log-log error plot is given in Fig. 2.14.

The empirical rate of convergence for the above plotted trial is 0.9871 which is a little bit

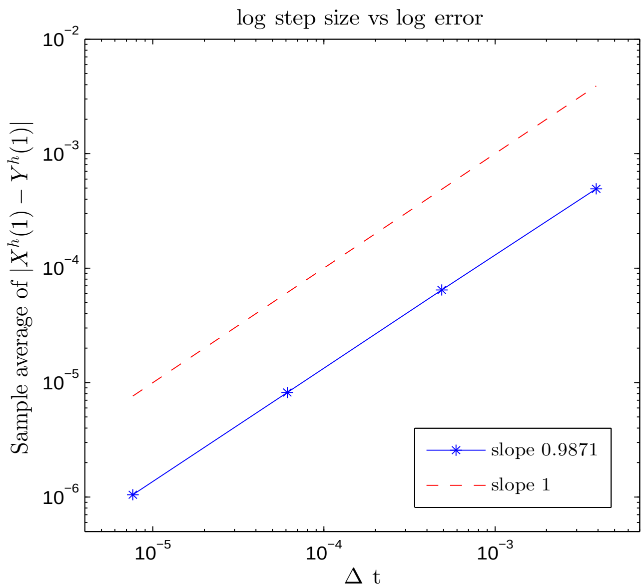

Figure 3: log-log plot for Example 3

lower than one. The least squares residual in this case is 0.0157. We can easily see that some Assumptions in (B) do not hold true for the above equation. The performance of the proposed scheme is however still very good in this case. This is probably due to the negative nonlinear term $-x^2$ in the drift that helps pull the system back to stable and somehow helps to ensure the convergence of the numerical scheme as well.

2.6 Concluding Remarks

This chapter developed Milstein-type numerical procedures for diffusions modulated by a Markovian switching process. We demonstrated order 1 rate of convergence, which shows that in case of regime-switching diffusion processes, Milstein-type schemes still outperform the well-known Euler-Maruyama method. The appearance of the discrete component represented by the Markovian switching process makes the analysis as well as the actual computation much more difficult. In constructing the algorithm and proving the convergence, we use the martingale characteristics of the Markovian switching process. We also rely on the fact that more than one jumps occur in a small interval being negligible to approximate the stochastic integrals. We hope that the numerical procedures will provide viable alternatives for many applications and open up new avenue for further investigation.

2.7 Appendix

We now give a proof for the Lemma 2.2 and Lemma 2.7.

Proof of Lemma 2.2. Let $s < \rho_1 < \ldots < \rho_\nu < t$ be the jump times of the Markov chain $\alpha(u)$ in the interval (s,t). For convenience, we denote $\rho_0 = s$ and $\rho_{\nu+1} = t$. Note that for $u \in (\rho_n, \rho_{n+1})$, $0 \le n \le \nu$, $X(u)$ behaves as a diffusion with drift term $b(\cdot,\cdot,\alpha(\rho_n))$ and diffusion term $\sigma(\cdot,\cdot,\alpha(\rho_n))$. Thus, applying the Itô formula to $f(\cdot,\cdot,\alpha(\rho_n))$ on the interval (ρ_n, ρ_{n+1}), $0 \le n \le \nu$, we have

$$f(\rho_{n+1}, X(\rho_n), \alpha(\rho_n)) - f(\rho_n, X(\rho_n), \alpha(\rho_n)) = \int_{\rho_n}^{\rho_{n+1}} \left[\frac{\partial}{\partial u} + \mathcal{L}_{\alpha(\rho_n)}\right] f(u, X(u), \alpha(\rho_n)) du$$
$$+ \sum_{l=1}^{m} \int_{\rho_k}^{\rho_{n+1}} \langle \nabla_x f(u, X(u), \alpha(\rho_n)), \sigma_l(u, X(u), \alpha(\rho_n)) \rangle \, dB_l(u).$$

Since $\alpha(u) = \alpha(\rho_n)$ for $u \in (\rho_n, \rho_{n+1})$, $0 \le n \le \nu$, adding the above equations we get that

$$\begin{aligned} f(t, X(t), \alpha(t)) - f(s, X(s), \alpha(s)) = & \int_s^t \left[\frac{\partial}{\partial u} + \mathcal{L}_{\alpha(u)}\right] f(u, X(u), \alpha(u)) du \\ & + \sum_{l=1}^{m} \int_s^t \langle \nabla_x f(u, X(u), \alpha(u)), \sigma_l(u, X(u), \alpha(u)) \rangle \, dB_l(u) \\ & + \sum_{n=1}^{\nu} \Big[f(\rho_n, X(\rho_n), \alpha(\rho_n)) - f(\rho_n, X(\rho_n), \alpha(\rho_n -)) \Big]. \end{aligned} \tag{2.52}$$

We now work on the last term of the above sum. Let $g : (s,t] \times \mathcal{M} \times \mathcal{M} \times \Omega \to \mathbb{R}$ be a random function defined by $g(u, i_0, j_0, \omega) = f(u, X(u,\omega), j_0) - f(u, X(u,\omega), i_0)$, for any

$(u, i_0, j_0, \omega) \in (s, t] \times \mathcal{M} \times \mathcal{M} \times \Omega$. Recall from (2.4)

$$M_{i_0 j_0}(t) = \sum_{0<u\le t} \mathbb{1}_{i_0}(\alpha(u-))\mathbb{1}_{j_0}(\alpha(u)) - \int_0^t q_{\alpha(u-),j_0} \mathbb{1}_{i_0}(\alpha(u-))du$$

and take the integral we have that

$$\sum_{i_0\neq j_0} \int_s^t g(u, i_0, j_0) dM_{i_0 j_0}(u) = \sum_{s<u\le t} g(u, \alpha(u-), \alpha(u)) - \sum_{j_0} \int_s^t q_{\alpha(u-),j_0} g(u, \alpha(u-), j_0) du.$$

Using the definition of g and note that the first sum on the right-hand side of the above equation is exactly the last term in the equation (2.52) we can rewrite the above equation as

$$\begin{aligned}
&\sum_{n=1}^{\nu} \big[f(\rho_n, X(\rho_n), \alpha(\rho_n)) - f(\rho_n, X(\rho_n), \alpha(\rho_n-))\big] \\
&\quad = \sum_{j_0\in\mathcal{M}} \int_s^t q_{\alpha(u-),j_0} \big(f(u, X(u), j_0) - f(u, X(u), \alpha(u-))\big) du \\
&\qquad\qquad + \sum_{i_0\neq j_0} \int_s^t \big(f(u, X(u), j_0) - f(u, X(u), i_0)\big) dM_{i_0 j_0}(u) \\
&\quad = \int_s^t Qf(u, X(u), \cdot)(\alpha(u)) du + \sum_{i_0\neq j_0} \int_s^t \big(f(u, X(u), j_0) - f(u, X(u), i_0)\big) dM_{i_0 j_0}(u).
\end{aligned}$$

Substituting this equation into equation (2.52) we have the desired result. □

Proof of Lemma 2.7. It is trivial that on the event $\{N_n = 0\}$,

$$\sum_{i_0\neq j_0} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} \big(\sigma_{k,l}(u, X(u), j_0) - \sigma_{k,l}(u, X(u), i_0)\big) d[M_{i_0 j_0}](u) dB_l(s) = 0.$$

Next, denote $\hat{\tau}_i^n = \tau_i^n$ for $1 \le i \le N_n$ and $\hat{\tau}_{N_n+1}^n = t_{n+1}$. Note that on the event $\{N_n \ge 1\}$

we have

$$
\begin{aligned}
&\sum_{i_0\neq j_0}\int_{t_n}^{s}\big(\sigma_{k,l}(u,X(u),j_0)-\sigma_{k,l}(u,X(u),i_0)\big)d[M_{i_0j_0}](u)\\
&\quad=\begin{cases}0 & \text{if } \ t_n\le s<\hat{\tau}_1^n,\\[2ex] \displaystyle\sum_{j=1}^{i}\Big[\sigma_{k,l}(\hat{\tau}_j^n,X(\hat{\tau}_j^n),\hat{\tau}_j^n)-\sigma_{k,l}(\hat{\tau}_j^n,X(\hat{\tau}_j^n),\hat{\tau}_{j-1}^n)\Big] & \text{if } \ \hat{\tau}_i^n\le s<\hat{\tau}_{i+1}^n, i\ge 1.\end{cases}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sum_{i_0\neq j_0}\int_{t_n}^{t_{n+1}}\int_{t}^{s}\big(\sigma_{k,l}(u,X(u),j_0)-\sigma_{k,l}(u,X(u),i_0)\big)d[M_{i_0j_0}](u)dB_l(s)\\
&\quad=\sum_{i=0}^{N_n}\sum_{i_0\neq j_0}\int_{\hat{\tau}_i^n}^{\hat{\tau}_{i+1}^n}\int_{t_n}^{s}\big(\sigma_{k,l}(u,X(u),j_0)-\sigma_{k,l}(u,X(u),i_0)\big)d[M_{i_0j_0}](u)dB_l(s)\\
&\quad=\sum_{i=1}^{N_n}\Big[\sum_{j=1}^{i}\big(\sigma_{k,l}(\hat{\tau}_j^n,X(\hat{\tau}_j^n),\hat{\tau}_j^n)-\sigma_{k,l}(\hat{\tau}_j^n,X(\hat{\tau}_j^n),\hat{\tau}_{j-1}^n)\big)\Big]\Big(B_l(\hat{\tau}_{i+1}^n)-B_l(\hat{\tau}_i^n)\Big)\\
&\quad=\sum_{i=1}^{N_n}\Big[\sigma_{k,l}(\hat{\tau}_i^n,X(\hat{\tau}_i^n),\alpha(\hat{\tau}_i^n))-\sigma_{k,l}(\hat{\tau}_i^n,X(\hat{\tau}_i^n),\alpha(\hat{\tau}_{i-1}^n))\Big]\Big(B_l(\hat{\tau}_{N_n+1}^n)-B_l(\hat{\tau}_i^n)\Big)\\
&\quad=\sum_{i=1}^{N_n}\Big[\sigma_{k,l}(\tau_i^n,X(\tau_i^n),\alpha(\tau_i^n))-\sigma_{k,l}(\tau_i^n,X(\tau_i^n),\alpha(\tau_{i-1}^n))\Big]\Big(B_l(t_{n+1})-B_l(\tau_i^n)\Big)
\end{aligned}
$$

since $\hat{\tau}_{N_n+1}^n=t_{n+1}$ and $\hat{\tau}_i^n=\tau_i^n$ for $1\le i\le N_n$. This implies (2.27). $\square$

CHAPTER 3 A LIMIT THEOREM FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH MARKOVIAN REGIME SWITCHING

3.1 Introduction

Owing to their wide range of applications, Markov hybrid switching diffusions or Markov switching diffusions have become more popular recently and have been drawn growing attention. In many applications, we have to deal with stochastic hybrid systems in which the dynamics are given by the following stochastic differential equations

$$\begin{aligned} &dX(t) = b(t, X(t-), \Lambda(t-))dt + \sigma(t, X(t-), \Lambda(t-))dW(t), \\ &\mathbb{P}(\Lambda(t+\delta) = j | \Lambda(t) = i, X(s), \Lambda(s), 0 \le s \le t) = q_{ij}(t)\delta + o(\delta), \end{aligned} \tag{3.1}$$

where Λ is a continuous-time Markov chain with generator $Q(t) = (q_{i,j}(t))$ representing the configuration change of the system. We assume that Λ and the Brownian motion W are independent throughout this chapter. Under suitable conditions, the coupled process $Z = (X, \Lambda)$ is a Markov process and possesses many interesting features. Due to the interactions between the discrete and continuous components, the coupled process is highly nonlinear and it is very difficult, if not impossible, to obtain an analytic solution of the system in most of the cases. It is therefore important to be able to use some kind of approximations to the process with the hope that the approximations give us valuable insights about the process itself. Many approximation schemes for (3.1) can be put into the form

$$\begin{aligned} &dX^n(t) = b_n(t, X^n(t-), \Lambda^n(t-))dA^n(t) + \sigma_n(t, X^n(t-), \Lambda^n(t-))dM^n(t), \\ &\mathbb{P}(\Lambda^n(t+\delta) = j | \Lambda^n(t) = i, X^n(s), \Lambda^n(s), 0 \le s \le t) = q_{ij}(t)\delta + o(\delta), \end{aligned} \tag{3.2}$$

where $\{M^n\}$ are square integrable martingales and $\{A^n\}$ are non-negative continuous increasing processes. The approximating process $Z^n = (X^n, \Lambda^n)$ is then used to assist us in studying the original process. The scheme (3.5) is indeed quite general and covers many cases in practice. We emphasize here that Λ^n does not depend on X^n, the first component.

The purpose of this chapter is to give some sufficient conditions under which Z^n converges in the weak sense to a stochastic process Z. Our work is inspired by [46] for the work on diffusion processes. In [46], the coefficient of the systems (3.2) and (3.1) are assumed to be uniformly bounded. We are able to relax these strict assumptions by replacing them with the Lipschitzian of the coefficients. It appears that assumptions can be relaxed further but we do not pursuit this direction to avoid complicating the proofs. Moreover, due to the presence of the discrete component, the analysis here is more delicate and difficult.

The rest of the chapter is arranged as follows. Section 3.2 begins with the formulation, preliminaries and states our main results. Section 3.3 presents the proof of the main results. Finally, the chapter is concluded with Section 3.4 on some concluding remarks.

3.2 Formulation and Preliminaries

Throughout this chapter, we use $\mathbb{1}_A(\cdot)$ to denote the characteristic function of the set A. We use K to denote generic constants whose value may change from appearance to appearance. We denote by $\mathbb{D}([0,T],\mathbb{R}^d)$ the space of all functions $\alpha : [0,T] \mapsto \mathbb{R}^d$ that are right continuous and have left limit. We equip $\mathbb{D}([0,T],\mathbb{R}^d)$ with Skorokhod J_1 topology. As usual, if $\alpha \in \mathbb{D}([0,T],\mathbb{R}^d)$ we denote by $\alpha(t)$ the value of α at time t and by $\alpha(t-)$ its left-hand limit at time t. We use $\mathscr{D}_t^0(\mathbb{R}^d)$ to denote the σ-field generated by all maps $\alpha \mapsto \alpha(s)$, for $s \le t$, and $\mathscr{D}_t(\mathbb{R}^d) = \bigcap_{s>t} \mathscr{D}_t^0(\mathbb{R}^d)$ and thus $\mathbf{D}(\mathbb{R}^d) = \big(\mathscr{D}_t(\mathbb{R}^d)\big)_{0\le t\le T}$ is a filtration.

All the processes are assumed to be realized in $\mathbb{D}([0,T],\mathbb{R}^d)$ for suitable d. A stochastic process Z with its trajectories belong to $\mathbb{D}([0,T],\mathbb{R}^d)$ can be considered as a $\mathbb{D}([0,T],\mathbb{R}^d)$-valued random variable. The law $\mathscr{L}(Z)$ is defined by $\mathbb{Q} \doteq \mathscr{L}(Z|\mathbb{P}) = (\mathbb{P}) \circ (Z)^{-1}$. We say that a sequence of stochastic processes $\{Z^n\}_{n\geq 1}$ converges in distribution or weakly to Z, denoted by $Z^n \Rightarrow Z$, if $\mathbb{Q}^n \to \mathbb{Q}$ weakly in $\mathscr{P}\left(\mathbb{D}([0,T],\mathbb{R}^d)\right)$, the space of all probability measures on $\mathbb{D}([0,T],\mathbb{R}^d)$ with the topology of weak convergence. A sequence $\{Z^n\}$ is tight if the sequence of distribution $\mathbb{Q}^n = \mathscr{L}(Z^n)$ is tight.

We assume that the iterations of approximation are carried out on a sequences of probability spaces $(\Omega^n, \mathbf{F}^n = \{\mathcal{F}^n_t\}_{0\leq t\leq T}, \mathbb{P}^n), n \geq 1$ where each of them is a complete probability space with the families of increasing sub-σ-fields $\{\mathcal{F}^n_t\}, t \geq 0, n = 1, 2, \ldots$, which satisfies the usual hypothesis, i.e., $\{\mathcal{F}^n_0\}$ contain all $\mathbb{P}^n$-negligible sets and $\mathcal{F}^n_t = \mathcal{F}^n_{t+}$. We also assume that, on each of the above-mentioned probability spaces, we have a family of $\mathcal{F}^n_t$-adapted processes

$$\begin{aligned}&\Big\{A^n, M^n, \{N^n_{i_0,j_0}\} : i_0, j_0 \in \mathcal{M}, i_0 \neq j_0\Big\}\\ &\quad = \Big\{A^n(t), M^n(t), \{N^n i_0, j_0(t)\} : i_0, j_0 \in \mathcal{M}, i_0 \neq j_0, t \geq 0\Big\}.\end{aligned}$$

Where $\{M^n\}$ are square integrable $\mathcal{F}^n_t$-martingales and $\{A^n\}$ are continuous increasing $\mathcal{F}^n_t$-mesurable processes. For simplicity we assume that $A^n(0) = M^n(0) = 0$ for all n. We shall denote by $\mu^n = \mu^n(dt, dx)$ the integral random measure of jumps of the martingale M^n

$$\mu^n\left((0,t],\Gamma\right) = \sum_{0<s\leq t} \mathbb{1}_\Gamma(\Delta M^n(s)), \Gamma \in \mathcal{B}(\mathbb{R}\backslash\{0\}), \tag{3.3}$$

where $\Delta M^n(s) = M^n(s) - M^n(s-)$. We use $\nu^n = \nu^n(dt, dx)$ to denote the compensator, or the dual predictable projection of the random measure μ^n. For each $n \in \mathbb{N}$, $i_0, j_0 \in \mathcal{M}$, $i_0 \neq j_0$, $N^n_{i_0,j_0}$ is a counting point process with intensity $q_{i_0,j_0}(t)$ where $q_{i_0,j_0} : \mathbb{R} \to \mathbb{R}_+$ is a bounded continuous function. This implies that for each fixed (i_0, j_0) the process

$$L^n_{i_0,j_0}(t) = N^n_{i_0,j_0}(t) - \int_{(0,t]} q_{i_0,j_0}(u)du \tag{3.4}$$

is a martingale. Moreover, we require that $N^n_{i_0,j_0}(0) = 0$ and the process $N^n_{i_0,j_0}$ does not have common jump with M^n, i.e.

$$\Delta M^n(t)\Delta N^n_{i_0,j_0}(t) = 0, \quad \mathbb{P}^n \text{ a.s.},$$

and for all $(i_1, j_1) \neq (i_2, j_2)$ we have

$$\Delta N^n_{i_1,j_1}(t)\Delta N^n_{i_2,j_2}(t) = 0, \quad \mathbb{P}^n \text{ a.s.}$$

The approximation sequence is a family of $\mathcal{F}^n_t$-adapted processes $\{Z^n = (X^n, \Lambda^n), n \geq 1\}$ satisfy the following stochastic differential equations:

$$\begin{aligned} dX^n(t) &= b_n(t, X^n(t-), \Lambda^n(t-))dA^n(t) + \sigma_n(t, X^n(t-), \Lambda^n(t-))dM^n(t), \\ d\Lambda^n(t) &= \sum_{i_0,j_0=1}^{m} (j_0 - i_0)\mathbb{1}_{\{i_0\}} dN^n_{i_0,j_0}(t), \end{aligned} \tag{3.5}$$

with $\sup \mathbb{E}^n((X^n)(0))^2 < \infty$, and the functions $b_n(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$, $\sigma_n(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$ satisfy some conditions that will be specified later. The limit process lives on the probability space $(\Omega, \mathbf{F} = \{\mathcal{F}_t\}, \mathbb{P})$ which is also a complete probability supporting a $\{\mathcal{F}_t\}$-Brownian motion W and jump processes $N^{i_0,j_0}, i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$. The limit process $Z = (X, \Lambda)$ satisfies the following stochastic differential equation

$$\begin{aligned} dX(t) &= b(t, X(t-), \Lambda(t-))dt + \sigma(t, X(t-), \Lambda(t-))dW(t), \\ d\Lambda(t) &= \sum_{i_0,j_0=1}^{m} (j_0 - i_0)\mathbb{1}_{\{i_0\}} dN_{i_0,j_0}(t). \end{aligned} \tag{3.6}$$

The jump processes $N_{i_0,j_0}, i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$ are counting point processes with intensities $q_{i_0,j_0}(\cdot)$. We emphasize here that q_{i_0,j_0} is the same for N_{i_0,j_0} and $N^n_{i_0,j_0}, n \in \mathbb{N}$. Similar to

(3.4), this also gives us that the process

$$L_{i_0,j_0}(t) = N_{i_0,j_0}(t) - \int_{(0,t]} q_{i_0,j_0}(u)du \tag{3.7}$$

is a martingale for each fixed (i_0, j_0). We also require that for all $(i_1, j_1) \neq (i_2, j_2)$ we have

$$\Delta N_{i_1,j_1}(t)\Delta N_{i_2,j_2}(t) = 0, \quad \mathbb{P} \text{ a.s.}$$

From (3.5) and (3.6), it is clear that the trajectories of Z^n and Z belong to $\mathbb{D}([0,T],\mathbb{R}^2)$ and as a result Z^n and Z are $\mathbb{D}([0,T],\mathbb{R}^2)$ random variables. Note that (3.5) and (3.6) are just another interpretations of (3.2) and (3.1) respectively. The current interpretations and the proof of Lemma 3.5 are inspired by the papers [3] and [19]. We prefer this representation in the current chapter since it makes the analysis more convenient.

We assume the following assumptions throughout the chapter.

(A1) The quadratic variation process $\langle M^n \rangle$ of M^n satisfies that for any $t \geq 0, \varepsilon > 0$, $\mathbb{P}^n\left(|\langle M^n \rangle(t) - t| > \varepsilon\right) \to 0$ as $n \to \infty$ and $\sup_n \mathbb{E}^n \langle M^n \rangle(t) < \infty$.

(A2) For any $t > 0, \varepsilon > 0$ and $\delta > 0$, $\mathbb{P}^n\left(\left|\int_0^t \int_{\{|x|>\varepsilon\}} x^2 \nu^n(dsdx)\right| > \delta\right) \to 0$ as $n \to \infty$.

(A3) One of the following conditions (A3-i) or (A3-ii) holds:

(A3-i) For any $t > 0$ and $\varepsilon > 0$, $\mathbb{P}^n\left(\left|\int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dsdx) - t\right| > \varepsilon\right) \to 0$ as $n \to \infty$.

(A3-ii) For any $t > 0$ and $\varepsilon > 0$, $\mathbb{P}^n\left(\left|\int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dsdx)\right| > \varepsilon\right) \to 0$ as $n \to \infty$.

(A4) For measurable functions $b_n(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$, $\sigma_n(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$, there exists a constant K such that

$$|b_n(s,x,i_0) - b_n(t,y,i_0)| + |\sigma_n(s,x,i_0) - \sigma_n(t,y,i_0)| \leq K|t-s| + |y-x|, \tag{3.8}$$

for each $i_0 \in \mathcal{M}, n \geq 1$ and there exists functions $b(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$, $\sigma(\cdot,\cdot,\cdot) : [0,\infty) \times \mathbb{R} \times \mathcal{M} \to \mathbb{R}$ satisfying that for any $i_0 \in \mathcal{M}$

$$\sigma_n(s^n, x^n, i_0) \to \sigma(s,x) \text{ if } s^n \to s \text{ and } x^n \to x,$$

$$b_n(s^n, x^n, i_0) \to b(s,x) \text{ if } s^n \to s \text{ and } x^n \to x.$$

(A5) $\{A^n\}$ is a sequence of continuous increasing $\mathcal{F}_t^n$-measurable processes such that for any $\varepsilon > 0$,

$$\mathbb{P}^n \left(\sup_{0 \leq t \leq T} |A^n(t) - t| > \varepsilon \right) \to 0, \text{ for any } T > 0.$$

(A6) For $i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$, the functions $q_{i_0,j_0}(\cdot)$ are bounded continuous.

We have following remarks regarding the assumptions we made above

Remark 3.1. Since

$$\langle M^n \rangle(t) = \langle (M^n)^c \rangle(t) + \int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dsdx), \tag{3.9}$$

where $\{(M^n)^c\}$ is the continuous part of M^n (Jacod [15, Proposition 3.77, p105]), when $\{M^n\}$ are purely discontinuous martingales, assumptions (A1) and (A3-i) are equivalent. It is also clear from the above equation that (A3-ii) holds when $\{M^n\}$ are continuous martingales.

Remark 3.2. Conditions (A1) and (A2) were used in Liptser and Shiryayev [20] to show weak convergence of $\{M^n\}$ to W, the standard Brownian motion.

Remark 3.3. Condition (A4) also implies that the function b and σ are Lipschitz continuous.

From Remark 3.3 we can show that there exists a unique global solutions for (3.5) and (3.6). Moreover, the solution of (3.6) and satisfies the following property.

Proposition 3.4. *Assume that (A1)-(A5) hold. Then there exists a constant K, depends only on T and $X(0)$, such that*

$$\mathbb{E}\left[\sup_{0\le s\le T}|X(s)|^2\right]\le K.$$

The reader may consult [52, Proposition 3.2, page 31] for a proof of the above result. This result then implies that

$$\mathbb{P}\left(|X(t)|>N\right)\to 0 \quad \text{as} \quad N\to\infty. \tag{3.10}$$

Let $\mathcal{L}$ denote the generator of system (3.6). For a function $f(\cdot,\cdot):\mathbb{R}\times\mathcal{M}\to\mathbb{R}$ such that for each $i_0\in\mathcal{M}$, $f(\cdot,i_0)\in C_c^2(\mathbb{R})$, the set of twice continuously differentiable functions with compact support

$$\mathcal{L}f(x,i_0)=\mathcal{L}_{t,i_0}f(x,i_0)+Q(t)f(x,\cdot)(i_0) \tag{3.11}$$

for all $(x,i_0)\in\mathbb{R}\times\mathcal{M}$, where

$$\begin{aligned}\mathcal{L}_{t,i_0}f(x,i_0)&=b(t,x,i_0)\frac{\partial}{\partial x}f(x,i_0)+\frac{1}{2}\frac{\partial^2}{\partial x^2}f(x,i_0)\sigma^2(t,x,i_0),\\ Q(t)f(x,\cdot)(i_0)&=\sum_{j_0\in\mathcal{M}}q_{i_0j_0}(t)\big(f(x,j_0)-f(x,i_0)\big).\end{aligned}$$

Here, $\frac{\partial}{\partial x}f$ and $\frac{\partial^2}{\partial x^2}$ denote the first and second derivative of f with respect to x, respectively. We will use the following form of Itô's lemma to find the differential of functionals of the solution of (3.5). A proof of it will be given in the Appendix. Let (X,Λ) be a solution of the stochastic differential equation

$$\begin{aligned}dX(t)&=b(t,X(t-),\Lambda(t-))dA(t)+\sigma(t,X(t-),\Lambda(t-))dM(t),\\ d\Lambda(t)&=\sum_{i_0,j_0\in\mathcal{M}}(j_0-i_0)\mathbb{1}_{\{i_0\}}dN_{i_0,j_0}(t).\end{aligned} \tag{3.12}$$

Then the following lemma holds.

Lemma 3.5. *For a function $f(\cdot,\cdot) : \mathbb{R} \times \mathcal{M} \to \mathbb{R}$ such that for each $i \in \mathcal{M}$, $f(\cdot, i) \in C_c^2([0,T] \times \mathbb{R})$ we have*

$$\begin{aligned}
f(X(t),\Lambda(t)) =& f(X(s),\Lambda(s)) \\
&+ \int_s^t \frac{\partial}{\partial x} f(X(u-),\Lambda(u-)) b(u, X(u-), \Lambda(u-)) dA(u) \\
&+ \int_s^t \frac{\partial}{\partial x} f(X(u-),\Lambda(u-)) \sigma(u, X(u-), \Lambda(u-)) dM(u) \\
&+ \frac{1}{2} \int_s^t \frac{\partial^2}{\partial x^2} f(X(u-),\Lambda(u-)) \sigma^2(u, X(u-), \Lambda(u-)) d\langle M^c \rangle(u) \\
&+ \sum_{s<u\le t} \left[f(X(u),\Lambda(u-)) - f(X(u-),\Lambda(u-)) - \frac{\partial}{\partial x} f(X(u-),\Lambda(u-)) \Delta X(u) \right] \\
&+ \sum_{j\in\mathcal{M}} \int_s^t \left[f(X(u-), j) - f(X(u-),\Lambda(u-)) \right] q_{\Lambda(u-),j}(u) du \\
&+ \sum_{j\in\mathcal{M}} \int_s^t \left[f(X(u-), j) - f(X(u-),\Lambda(u-)) \right] dL_{\Lambda(u-),j}(u).
\end{aligned} \tag{3.13}$$

Equation (3.12) represents the common form for the equations that the approximation sequences satisfy and the coupled process (X, Λ) in Lemma 3.5 should not be confused with the limit in (3.6).

3.3 Proof of Main Results

In this section, we prove our main results. We first note that with the given assumptions we can always reduce our problem to the case that $b_n(\cdot,\cdot,\cdot)$ and $\sigma_n(\cdot,\cdot,\cdot)$ are uniformly bounded for all n by using the truncation method (see, for example, [45]). In our case the method works as follows. For an arbitrary $N \ge 0$, we define a function $\mathbb{I}_N(x)$ by

$$\mathbb{I}_N(x) = \begin{cases} 0, & |x| > N+1, \\ 1, & |x| \le N, \end{cases}$$

and $\mathbb{I}_N$ smooth on $\mathbb{R}$. We define on $\mathbb{D}([0,T],\mathbb{R}^2)$ the stopping times ϱ_N given by $\varrho_N(\alpha) = \inf\{t : |\alpha(t)| \geq N\}$ with the usual convention $\inf\{\emptyset\} = \infty$. We also define the sequences of stopping times τ_N^n on $(\Omega^n, \mathbf{F}^n, \mathbb{P}^n)$, where

$$\tau_N^n(\omega) = \inf\{t : |X^n(t,\omega)| \geq N\}, \quad n \geq 1.$$

We define the truncated process $Z_N^n = Z^n(t \wedge \tau_N^n)$. Then $Z_N^n = (X_N^n, \Lambda_N^n)$ satisfies the following stochastic equation

$$\begin{aligned} dX_N^n(t) &= \mathbb{I}_N(X_N^n(t-))b_n(t, X_N^n(t-), \Lambda_N^n(t-))dA^n(t) \\ &\qquad + \mathbb{I}_N(X_N^n(t-))\sigma_n(t, X_N^n(t-), \Lambda_N^n(t-))dM^n(t), \\ d\Lambda_N^n(t) &= \sum_{i_0,j_0\in\mathcal{M}} (j_0 - i_0)\mathbb{1}_{\{i_0\}} dN_{i_0,j_0}^n(t). \end{aligned} \tag{3.14}$$

The procedure for proving our main result is: first verify the tightness of $\{Z_N^n, n \geq 1\}$ and then show that any weak limit of $Z_N = (X_N, \Lambda_N)$ of the sequence satisfies the equation

$$\begin{aligned} dX_N(t) &= b(t, X_N(t-), \Lambda_N(t-))dt + \sigma_n(t, X_N(t-), \Lambda_N(t-))dW(t), \\ d\Lambda_N(t) &= \sum_{i_0,j_0\in\mathcal{M}} (j_0 - i_0)\mathbb{1}_{\{i_0\}} dN_{i_0,j_0}(t), \end{aligned} \tag{3.15}$$

where $t \leq \tau_N$ and $W(t)$ is a standard Brownian motion and $\tau_N = \inf\{t : |X(t,\omega)| \geq N\}$. Denote the law of Z_N^n and Z_N by $\mathbb{Q}_N^n$ and $\mathbb{Q}_N$, respectively. Then by the uniqueness of the equation (3.5) and (3.6), the process $Z_N^n(\cdot)$ and $Z_N(\cdot)$ coincide in the law with the process $Z^n(\cdot)$ and $Z(\cdot)$ respectively until $X_N^n(\cdot)$ and $X_N(\cdot)$ hit the boundary of the set $\{x : |x| \leq N\}$. Thus the $\mathbb{Q}_N^n$ and $\mathbb{Q}_N$ agree with $\mathbb{Q}^n$ and $\mathbb{Q}$ respectively on $\mathcal{F}_{\varrho^N}$. Moreover, since the sequence ϱ_N is a non-decreasing, lower semicontinuous stopping times. By the virtue of Lemma 11.1.1 in [45], we get that $\mathbb{Q}^n \to \mathbb{Q}$ weakly. Following the above procedure, we note that $\tilde{b}_n(s,x) = \mathbb{I}_N(x)b_n(s,x)$ and $\tilde{\sigma}_n(s,x) = \mathbb{I}_N(x)\sigma(s,x)$ are bounded uniformly in n and, as long as $|x| \le N$, $\tilde{b}_n(s,x) = b_n(s,x)$ and $\tilde{\sigma}_n(s,x) = \sigma(s,x)$. For these reasons it suffices to prove our results in the case that b_n and σ_n are uniformly bounded.

3.3.1 Tightness of the Approximation Sequence

We first obtain the following lemma.

Lemma 3.6. *Let*

$$\begin{aligned} Y^n = &\Bigg\{ \Bigg(X^n(t), \Lambda^n(t), A^n(t), \int_0^t b_n(s, X^n(s-), \Lambda^n(s-))dA^n(s), \\ &\qquad M^n(t), \int_0^t \sigma_n(s, X^n(s-), \Lambda^n(s-))dM^n(s)) \Bigg) : t \ge 0 \Bigg\}, n = 1, 2, \ldots. \end{aligned} \tag{3.16}$$

Then for each $T > 0$, $\{Y^n\}$ is tight in $\mathbb{D}([0,T], \mathbb{R}^6)$.

Proof. It suffices to verify that the following conditions are satisfied:

(a) for any $\varepsilon > 0$, there is $a > 0$ such that

$$\sup_n \mathbb{P}^n \left(\sup_{0 \le t \le T} |Y^n(t)| > a \right) \le \varepsilon,$$

(b) for any $\varepsilon > 0, \eta > 0$, there are n_0 and $\delta > 0$ such that for any $\mathcal{F}_t^n$-adapted stopping time τ^n with $\tau^n \le T$ a.s.

$$\sup_{n \ge n_0} \mathbb{P}^n \left(\sup_{0 \le s \le \delta} |Y^n(\tau^n + s) - Y^n(\tau^n)| \ge \eta \right) \le \varepsilon.$$

But for this, it suffices to show (a) and (b) for each component of Y^n and we will verify them for the processes Λ^n and X^n only since the same argument applies to the other components of Y^n.

We now prove that the processes Λ^n satisfies (a) and (b). Since $\mathcal{M}$ is finite (a) trivially holds for Λ^n. To verify (b) we note that for $\eta > 0$

$$\begin{aligned}
\mathbb{P}^n\left(\sup_{0\le s\le\delta}|\Lambda^n(\tau^n+s)-\Lambda(\tau^n)|\ge\eta\right) \\
&= \mathbb{P}^n\Big(\exists t\in(0,\delta]:\Lambda^n(\tau^n+t)\neq\Lambda^n(\tau^n)\Big)\\
&= 1-\mathbb{P}^n\Big(\Lambda^n(\tau^n+\cdot)\text{ does not jump in the interval }(0,\delta]\Big) \qquad (3.17)\\
&= 1-\sum_{i\in\mathcal{M}}\exp\left\{\int_0^\delta q_{ii}(\tau^n+s)ds\right\}\mathbb{P}^n\left(\Lambda^n(\tau^n)=i\right)\\
&\le 1-e^{q\delta},
\end{aligned}$$

where q is the lower bound of $q_{ii}(\cdot)$. Since the last quantity can be made arbitrarily small by choosing δ small enough we conclude that $\{\Lambda^n\}$ is tight.

We next verify that $\{X^n\}$ satisfy (a) and (b). For any $\mathcal{F}_t^n$-stopping time τ, we have

$$\begin{aligned}
\mathbb{E}^n\left(\int_0^\tau\sigma_n(s,X^n(s-),\Lambda^n(s-))dM^n(s)\right)^2 &\le \mathbb{E}^n\int_0^\tau\sigma_n(s,X^n(s-),\Lambda^n(s-))d\langle M^n\rangle(s)\\
&\le K\mathbb{E}^n\langle M^n\rangle(\tau),
\end{aligned} \qquad (3.18)$$

where K is the bound for σ_n^2. Then it is easy to see that

$$\mathbb{E}^n|X^n(\tau)|^2\le 3\sup_n\mathbb{E}^n(X^n)^2(0)+3K\mathbb{E}^n(A^n)^2(\tau)+3K\mathbb{E}^n\langle M^n\rangle(\tau).$$

Thus the process $\bar{X}^n(\cdot)$ defined by

$$\bar{X}^n(t)=K+3K\mathbb{E}^n(A^n)^2(t)+3K\mathbb{E}^n\langle M^n\rangle(t)$$

is a process dominating $(X^n)^2(\cdot)$ in the sense of Lenglart (see [18, page 35, lemma 3.30]) and hence it holds that

$$\mathbb{P}^n\left(\sup_{0\le t\le T}(X^n)^2(t)>a^2\right)\le\frac{1}{a^2}\mathbb{E}^n\bar{X}^n(T)\wedge b+\mathbb{P}^n\left(\bar{X}^n(T)\ge b\right),$$

for any $b > 0$. Here $a \wedge b = \min\{a, b\}$. By assumption (A1) and (A5),

$$\bar{X}^n(T) \to K_1 + 3KT^2 + 3KT,$$

if we take $b = K_1 + 3KT^2 + 3KT$, there exists n_1 such that $\mathbb{P}^n(\bar{X}^n(T) \geq b) \leq \varepsilon$ for all $n > n_1$. Hence,

$$\sup_n \mathbb{P}^n \left(\sup_{0\leq t\leq T} |X^n(t)| > a \right) \leq \sum_{n=1}^{n_1} \mathbb{P}^n \left(\sup_{0\leq t\leq T} |X^n(t)| > a \right) + \frac{b}{a^2} + \frac{\varepsilon}{2}$$

and thus (a) holds for $\{X^n\}$ by taking a sufficiently large. We next show that (b) holds for $\{X^n\}$. We can verify that

$$|X^n(\tau^n + t) - X^n(\tau^n)|^2 \leq 2K^2 \left(A^n(\tau^n + t) - A^n(\tau^n)\right)^2 + 2K^2 \left(\langle M^n\rangle(\tau^n + t) - \langle M^n\rangle(\tau^n)\right).$$

Repeat the argument in verifying (a), we have, for any $b > 0$,

$$\begin{aligned}
&\mathbb{P}^n \left(\sup_{0\leq s\leq \delta} |X^n(\tau^n + s) - X^n(\tau^n)|^2 > \eta^2 \right) \\
&\quad \leq \frac{1}{\eta^2}\mathbb{E}^n \left[2K^2 \left(A^n(\tau^n + \delta) - A^n(\tau^n)\right)^2 + 2K^2 \left(\langle M^n\rangle(\tau^n + \delta) - \langle M^n\rangle(\tau^n)\right)\right] \wedge b \\
&\qquad + \mathbb{P}^n \left[2K^2 \left(A^n(\tau^n + \delta) - A^n(\tau^n)\right)^2 + 2K^2 \left(\langle M^n\rangle(\tau^n + \delta) - \langle M^n\rangle(\tau^n)\right) \geq b\right] \\
&\quad \leq \frac{1}{\eta^2}\mathbb{E}^n \left[2K^2 \left\{ \sup_{0\leq t\leq T} \left(A^n(t + \delta) - A^n(t)\right) \right\}^2 + 2K^2 \left\{ \sup_{0\leq t\leq T} \left(\langle M^n\rangle(t + \delta) - \langle M^n\rangle(t)\right) \right\} \right] \wedge b \\
&\qquad + \mathbb{P}^n \left[2K^2 \left\{ \sup_{0\leq t\leq T} \left(A^n(t + \delta) - A^n(t)\right) \right\}^2 + 2K^2 \left\{ \sup_{0\leq t\leq T} \left(\langle M^n\rangle(t + \delta) - \langle M^n\rangle(t)\right) \right\} \geq b \right].
\end{aligned} \tag{3.19}$$

Noting that (A1) implies that, for any $\varepsilon > 0$

$$\mathbb{P}^n \left(\sup_{0\leq t\leq T} |\langle M^n\rangle(t) - t| > \varepsilon \right) \to 0, \text{ as } n \to \infty.$$

This is proved by imitating the proof of [35, Lemma 1] with only slightly changes. It then implies

$$\mathbb{P}^n\left(\left|\sup_{0\le t\le T}\left(\langle M^n\rangle(t+\delta)-\langle M^n\rangle(t)\right)-\delta\right|>\varepsilon\right)\to 0 \text{ as } n\to\infty.$$

We also have

$$\mathbb{P}^n\left(\left|\sup_{0\le t\le T}\left(A^n(t+\delta)-A^n(t)\right)-\delta\right|>\varepsilon\right)\to 0 \text{ as } n\to\infty.$$

Combining these two facts and (3.19) we conclude that (b) holds for X^n. □

We will use frequently the following result in what follows. One may consult Liptser and Shiryaev [20] for a proof.

Lemma 3.7. *Under (A1) and (A2), $M^n \Rightarrow W$ in $\mathbb{D}([0,T],\mathbb{R}), T$ arbitrary, where W is a standard Brownian motion.*

The following lemma establishes the continuity of some limit processes.

Lemma 3.8. *Let*

$$\begin{aligned}U^n &= (\Lambda^n(t), X^n(t), R_1^n(t), R_2^n(t))\\ &= \left(\Lambda^n(t), X^n(t), \int_0^t b_n(s, X^n(s-), \Lambda^n(s-))dA^n(s), \int_0^t \sigma_n(s, X^n(s-), \Lambda^n(s-))dM^n(s))\right).\end{aligned} \tag{3.20}$$

Then Y^n converges weakly to a process $U=(\Lambda, X, R_1, R_2)$, where X, R_1 and R_2 are continuous processes and satisfy $X = X(0)+R_1+R_2$.

Proof. As a consequence of the previous lemma $\{U^n\}$ is tight in $\mathbb{D}([0,T],\mathbb{R}^4)$ and thus there exists a sub-sequence, still indexed by n for notation simplicity, such that $U^n \Rightarrow U = (\Lambda, X, R_1, R_2)$ in $\mathbb{D}([0,T],\mathbb{R}^4)$. Let $g:\mathbb{D}([0,T],\mathbb{R})\to\mathbb{R}$ defined by $g(\alpha)=\sup_{0\le t\le T}|\Delta\alpha(s)|$ where $\Delta\alpha(s)=\alpha(s)-\alpha(s-)$. By Proposition 2.4 in [18, page 339], g is continuous. Noting that

$$\Delta R_2^n(t) = \sigma(t, X^n(t-), \Lambda^n(t-))\Delta M^n(t)$$

and combining with the fact that σ is bounded, we deduce that $g(R_2^n) \leq Kg(M^n)$. Since $M^n(t) \Rightarrow W(t)$ and $R_2^n \Rightarrow R_2$, we have that $g(R_2) \leq Kg(W) = 0$ and this implies that R_2 is continuous. Using the same arguments, we also get that R_1 is continuous. Because $X^n \Rightarrow X(0) + R_1 + R_2 = X$, X is continuous. □

3.3.2 Weak Convergence of the Approximation Sequence

In this section we prove the following main convergence theorem

Theorem 3.9. *Assume conditions* (A1)-(A6) *and suppose that* $Z^n(0) = (X^n(0), \Lambda^n(0)) \Rightarrow Z(0) = (X(0), \Lambda(0))$. *Then* $Z^n = (X^n, \Lambda^n) \Rightarrow Z = (X, \Lambda)$ *in* $\mathbb{D}([0,T], \mathbb{R}^2), T$ *arbitrary.*

Before proceeding further, let us make two remarks. First, by Lemma 3.6 the sequence of process Z^n is tight thus by Prohorov's theorem we can select a convergent subsequence. For notation simplicity, we still denote the subsequence by Z^n with the limit denoted by $\tilde{Z} = (\tilde{X}, \tilde{\Lambda})$. Therefore, to prove the weak convergence of Z^n to Z, we need only prove that the law of Z and $\tilde{Z}$ coincide. Second, by Skorohod's representation [6, Theorem 1.8, page 102], we can assume that there is a common probability space $(\Omega^0, \mathbf{F}^0, \mathbb{P}^0)$ and there are sequence of $\mathbb{D}([0,T], \mathbb{R}^6)$-valued variables $\bar{Y^n}$ and $\bar{Y}$ with same distributions as those of Y^n and Y, i.e.,

$$\mathbb{P}^0(\bar{Y}^n \in \cdot) = \mathbb{P}^n(Y^n \in \cdot), \quad \mathbb{P}^0(\bar{Y} \in \cdot) = \mathbb{P}(Y \in \cdot).$$

To keep the notation simple, we still use $\{Y^n\}$ and Y to denote $\bar{Y}^n$ and $\bar{Y}$. Moreover, $Y^n(\cdot)$ converges to $Y(\cdot)$ with respect to Prokhorov's metric on $\mathbb{D}([0,T], \mathbb{R}^6)$ $\mathbb{P}^0$ almost surely.

In view of [18, Lemma 1.31, page 331], there exists a sequence of continuous functions $\lambda_n : \mathbb{R}_+ \to \mathbb{R}_+$ that are strictly increasing, with $\lambda_n(0) = 0, \lambda_n(t) \uparrow T$ as $t \to T$ and satisfy

$$\sup_{0\le s\le T} |\lambda_n(s) - s| \to 0, \quad \text{and} \quad \sup_{0\le s\le T} |Y^n(\lambda_n(s)) - Y(s)| \to 0, \quad \mathbb{P}^0 \text{ a.s.}$$

We will fix the sequence of functions λ_n and use the two previous properties without explicitly mentioning them in the subsequent proofs.

Proof. By the uniqueness of equation (3.6), we just need to prove that the law of the coupled process $\tilde{Z}$ is the same as the law of Z, the solution of (3.6). We shall prove that for any $f(\cdot,\cdot) : \mathbb{R} \times \mathcal{M} \to \mathbb{R}$ such that for each $i_0 \in \mathcal{M}, f(\cdot, i_0) \in C_c^2(\mathbb{R})$,

$$\begin{aligned}\Xi(t) = f(\tilde{X}(t), \tilde{\Lambda}(t)) - f(\tilde{X}(s), \tilde{\Lambda}(s)) - &\int_s^t \mathcal{L}_{u,\tilde{\Lambda}(u)} f(\tilde{X}(u), \tilde{\Lambda}(u))du \\ &- \int_s^t Q(u) f(\tilde{X}(u), \cdot)(\tilde{\Lambda}(u))du\end{aligned}$$

is a martingale. To do this we shall prove that, for each n,

$$\begin{aligned}\Xi^n(t) =& f(X^n(t), \Lambda^n(t)) - f(X^n(s), \Lambda^n(s)) \\ &- \int_s^t \frac{\partial}{\partial x} f(X^n(u-), \Lambda^n(u-)) b_n(u, X^n(u-), \Lambda^n(u-)) dA^n(u) \\ &- \frac{1}{2}\int_s^t \frac{\partial^2}{\partial x^2} f(X^n(u-), \Lambda^n(u-)) \sigma_n^2(u, X^n(u-), \Lambda^n(u-)) d\langle (M^n)^c \rangle(u) \\ &- \int_s^t \int_{\mathbb{R}\setminus 0} \Big[f\Big(X^n(u-) + \sigma_n(u, X^n(u-), \Lambda^n(u-))x, \Lambda^n(u-)\Big) - f(X^n(u-), \Lambda^n(u-)) \\ &\qquad - \frac{\partial}{\partial x} f(X^n(u-), \Lambda^n(u-)) \sigma_n(u, X^n(u-), \Lambda^n(u-))x \Big] \nu^n(dudx) \\ &- \sum_{j\in\mathcal{M}} \int_s^t \left[f(X^n(u), j) - f(X^n(u-), \Lambda^n(u-)) \right] q_{\Lambda^n(u-),j}(u) du\end{aligned} \tag{3.21}$$

is a martingale, $\Xi^n(t)$ converges to $\Xi(t)$ in probability at each fixed t and the sequence $\{\Xi^n\}$ is uniformly integrable and thus the limit Ξ is indeed a martingale. These steps will be carried out in the several following lemmas. □

Lemma 3.10. *Let Ξ^n be defined as in* (3.21)*. Then Ξ^n is an $\mathbf{F}^n$ martingale.*

Proof. By Lemma 3.5, we have

$$\begin{aligned}
f(X^n(t),\Lambda^n(t)) =& f(X^n(s),\Lambda^n(s)) \\
&+\int_s^t \frac{\partial}{\partial x} f(X^n(u-),\Lambda^n(u-))b_n(u,X^n(u-),\Lambda^n(u-))dA^n(u) \\
&+\int_s^t \frac{\partial}{\partial x} f(X^n(u-),\Lambda^n(u-))\sigma_n(u,X^n(u-),\Lambda^n(u-))dM^n(u) \\
&+\frac{1}{2}\int_s^t \frac{\partial^2}{\partial x^2} f(X^n(u-),\Lambda^n(u-))\sigma_n^2(u,X^n(u-),\Lambda^n(u-))d\langle (M^n)^c\rangle(u) \\
&+\sum_{s<u\le t}\Big[f(X^n(u),\Lambda^n(u-)) - f(X^n(u-),\Lambda^n(u-)) \\
&\qquad - \frac{\partial}{\partial x} f(X^n(u-),\Lambda^n(u-))\Delta X^n(u)\Big] \\
&+\sum_{j\in\mathcal{M}}\int_s^t \left[f(X^n(u-),j) - f(X^n(u-),\Lambda^n(u-))\right] q_{\Lambda^n(u-),j}(u)du \\
&+\sum_{j\in\mathcal{M}}\int_s^t \left[f(X^n(u-),j) - f(X^n(u-),\Lambda^n(u-))\right] dL^n_{\Lambda^n(u-),j}(u),
\end{aligned} \tag{3.22}$$

where $(M^n)^c$ is the continuous part of the martingale M^n. Since f is twice continuous differentiable function with a compact support and $\sup_n \mathbb{E}^0\langle M^n\rangle(t)<\infty$, the integral with respect to M^n is a martingales. Similarly, we also have the stochastic integrals with respect to $L^n_{i_0,j_0}, i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$ are martingales. Therefore, in order to prove that $\Xi^n(t)$ is a martingale, it is sufficient to prove that

$$\Gamma^n(t) = \sum_{s<u\leq t}\left[f(X^n(u),\Lambda^n(u-)) - f(X^n(u-),\Lambda^n(u-)) - \frac{\partial}{\partial x}f(X^n(u-),\Lambda^n(u-))\Delta X^n(u)\right]$$
$$- \int_s^t\int_{\mathbb{R}\setminus 0}\left[f\Big(X^n(u-)+\sigma_n(u,X^n(u-),\Lambda^n(u-))x,\Lambda^n(u-)\Big) - f(X^n(u-),\Lambda^n(u-))\right.$$
$$\left.-\frac{\partial}{\partial x}f(X^n(u-),\Lambda^n(u-))\sigma_n(u,X^n(u-),\Lambda^n(u-))x\right]\nu^n(dudx) \tag{3.23}$$

is a martingale. Let us define

$$G^n(u,x,\omega) = f\Big(X^n(u-)+\sigma_n(u,X^n(u-),\Lambda^n(u-))x,\Lambda^n(u-)\Big) - f(X^n(u-),\Lambda^n(u-))$$
$$-\frac{\partial}{\partial x}f(X^n(u-),\Lambda^n(u-))\sigma_n(u,X^n(u-),\Lambda^n(u-))x. \tag{3.24}$$

Note that

$$G^n(u,x,\omega) = \frac{1}{2}\frac{\partial^2}{\partial x^2}f\Big(X^n(u-)+\theta^n(\omega)\sigma_n(u,X^n(u-),\Lambda^n(u-))x,\Lambda^n(u-)\Big)\sigma_n^2(s,X^n(u-),\Lambda^n(u-))x,$$

where θ^n is a random variable taking value in $[0,1]$. By the assumption $\sup_n \mathbb{E}^0\langle M^n\rangle(t) < \infty$ and the fact that

$$\langle M^n\rangle(t) = \langle (M^n)^c\rangle(t) + \int_0^t\int_{\mathbb{R}\setminus 0} x^2\nu^n(dudx),$$

we have

$$\mathbb{E}^0\int_s^t\int_{\mathbb{R}\setminus 0} x^2\nu^n(dsdx) < \infty,$$

for each fixed t. By the boundedness of σ_n and $\frac{\partial^2}{\partial x^2}f$, this implies

$$\mathbb{E}^0\int_s^t\int_{\mathbb{R}\setminus 0}|G^n(u,x,\omega)|\nu^n(dudx) < \infty. \tag{3.25}$$

Since

$$\mathbb{E}^0 \int_0^t \int_{\mathbb{R}\backslash 0} |G^n(u,x,\omega)|\mu^n(dsdx) = \mathbb{E}^0 \int_0^t \int_{\mathbb{R}\backslash 0} |G^n(u,x,\omega)|\nu^n(dudx)$$

and note that $\Delta X^n(t) = \sigma_n(t, X^n(t-), \Lambda^n(t-))\Delta M^n(t)$ we have

$$\begin{aligned}
&\int_s^t \int_{\mathbb{R}\backslash 0} G^n(u,x,\omega)\mu^n(dudx) \\
&\quad = \sum_{s<u\le t} \left[f(X^n(u), \Lambda^n(u-)) - f(X^n(u-), \Lambda^n(u-)) - \frac{\partial}{\partial x} f(X^n(u-), \Lambda^n(u-))\Delta X^n(u) \right].
\end{aligned} \tag{3.26}$$

The finiteness in (3.25) also implies that

$$\Gamma^n(t) = \int_s^t \int_{\mathbb{R}\backslash 0} G^n(u,x,\omega)\mu^n(dudx) - \int_s^t \int_{\mathbb{R}\backslash 0} G^n(u,x,\omega)\nu^n(dudx)$$

is an $\mathbf{F}^n$-martingale. the desired result is obtained. $\square$

Lemma 3.11. *Assume conditions* (A1)–(A5)*. Then*

$$f(X^n(t), \Lambda^n(t)) - f(X^n(s), \Lambda^n(s)) \to f(\tilde{X}(t), \tilde{\Lambda}(t)) - f(\tilde{X}(s), \tilde{\Lambda}(s))$$

in probability for each $0 \le s \le t \le T$.

Proof. It is sufficient to prove that $f(X^n(u), \Lambda^n(u)) \to f(\tilde{X}(u), \tilde{\Lambda}(u))$ in probability for each $0 \le s \le u \le t \le T$. We have

$$\begin{aligned}
&\Big| f\big(X^n(u), \Lambda^n(u)\big) - f\big(\tilde{X}(u), \tilde{\Lambda}(u)\big)\Big| \\
&\quad \le \Big| f\big(X^n(u), \Lambda^n(u)\big) - f\big(X^n \circ \lambda_n(u), \Lambda^n(u)\big)\Big| + \Big| f\big(X^n \circ \lambda_n(u), \Lambda^n(u)\big) - f\big(\tilde{X}(u), \Lambda^n(u)\big)\Big| \\
&\qquad + \Big| f\big(\tilde{X}(u), \Lambda^n(u)\big) - f\big(\tilde{X}(u), \Lambda^n \circ \lambda_n(u)\big)\Big| + \Big| f\big(\tilde{X}(u), \Lambda^n \circ \lambda_n(u)\big) - f\big(\tilde{X}(u), \tilde{\Lambda}(u)\big)\Big|.
\end{aligned} \tag{3.27}$$

By the continuity of f and the property of λ_n, the second and fourth term on the right-hand side of the last expression converges to zero in probability as $n \to \infty$. For the third term, one has

$$
\begin{aligned}
&\mathbb{P}^0\left(|\Lambda^n(u)-\Lambda^n\circ\lambda_n(u)|>\varepsilon\right)\\
&\qquad\le \mathbb{P}^0\Big(|\Lambda^n(u)-\Lambda^n\circ\lambda_n(u)|>\varepsilon/2, |\lambda_n(u)-u|\le\delta\Big)+\mathbb{P}^0\big(|\lambda_n(u)-u|>\delta\big)\\
&\qquad\le 1-\mathbb{P}^0\big(\Lambda^n(u+\cdot)\text{ has no jump on the interval of length }\delta\big)+\mathbb{P}^0\big(|\lambda_n(u)-u|>\delta\big).
\end{aligned}
\tag{3.28}
$$

By first choosing δ to make the first term in the last expression small enough and then selecting n large enough, we can make the left-hand side as mall as we want. Using the continuity of f, we can make the third term of (3.27) to be arbitrarily small in probability. Using similar arguments, we can also prove that the second term in (3.27) converges to zero in probability. Therefore $f\big(X^n(u),\Lambda^n(u)\big)\to f\big(\tilde{X}(u),\tilde{\Lambda}(u)\big)$ in probability. □

Lemma 3.12. *Assume conditions* (A1)–(A5)*. Then*

$$
\begin{aligned}
&\int_s^t \frac{\partial}{\partial x} f(X^n(u-),\Lambda^n(u-))b_n(u,X^n(u-),\Lambda^n(u-))dA^n(u)\\
&\qquad\to\int_s^t\frac{\partial}{\partial x}f(\tilde{X}(u-),\tilde{\Lambda}(u-))b(u,\tilde{X}(u-),\tilde{\Lambda}(u-))du
\end{aligned}
\tag{3.29}
$$

in probability for each $0\le s\le t\le T$.

Proof. By the change variable formula of the Lebesgue-Stieljes integral, we have that

$$
\begin{aligned}
&\int_{\lambda_n(s)}^{\lambda_n(t)} f(X^n(u-),\Lambda^n(u-))b_n(u,X^n(u-),\Lambda^n(u-))dA^n(u)\\
&=\int_s^t f(X^n\circ\lambda_n(u-),\Lambda^n\circ\lambda_n(u-))b_n(\lambda_n(u),X^n\circ\lambda_n(u-),\Lambda^n\circ\lambda_n(u-))dA^n\circ\lambda_n(u-).
\end{aligned}
\tag{3.30}
$$

Let $\gamma_n(t) = \inf\{s : A^n(\lambda_n(s)) > t\}$, by [5, page 91], we have that

$$\int_s^t f(X^n \circ \lambda_n(u-), \Lambda^n \circ \lambda_n(u-)) b_n(\lambda_n(u), X^n \circ \lambda_n(u-), \Lambda^n \circ \lambda_n(u-)) dA^n \circ \lambda_n(u-)$$
$$= \int_{A^n \circ \lambda_n(s)}^{A^n \circ \lambda_n(t)} f(X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) b_n(\gamma_n(u), X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) du. \tag{3.31}$$

We note that with probability one, $\gamma_n(u) \to u$ for each u. Then by convergence of $X^n \circ \lambda_n$ and $\Lambda^n \circ \lambda_n$ we have that $X^n \circ \lambda_n \circ \gamma_n(u-) \to \tilde{X}(u-)$ and $\Lambda^n \circ \lambda_n \circ \gamma_n(u-) \to \tilde{\Lambda}(u-)$. Combine with assumptions (A4-5) we get

$$f(X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) b_n(\gamma_n(u), X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-))$$
$$\to f(\tilde{X}(u-), \tilde{\Lambda}(u-)) b(s, \tilde{X}(u-), \tilde{\Lambda}(u-)).$$

Then by bounded convergence theorem it follows that with probability one

$$\int_s^t f(X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) b_n(\gamma_n(u), X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) du$$
$$\to \int_s^t f(\tilde{X}(u-), \tilde{\Lambda}(u-)) b(s, \tilde{X}(u-), \tilde{\Lambda}(u-)) du.$$

We also have that with probability one

$$\int_t^{A^n(\lambda_n(t))} f(X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) b_n(\gamma_n(u), X^n \circ \lambda_n \circ \gamma_n(u-), \Lambda^n \circ \lambda_n \circ \gamma_n(u-)) du \to 0.$$

Thus the claim is proved. $\square$

Lemma 3.13. *Assume* (A1) – A(5). *Then*

$$\frac{1}{2} \int_s^t \frac{\partial^2}{\partial x^2} f(X^n(u-), \Lambda^n(u-)) \sigma_n^2(u, X^n(u-), \Lambda^n(u-)) d\langle (M^n)^c \rangle(u) \to 0,$$

in probability for $0 \le s \le t \le T$.

Proof. We prove the claim under the Assumption (A3-i). The proof under Assumption (A3-

ii) can be carried out in the similar manner. We first note that, by the uniform boundedness of $\sigma_n(\cdot)$, we have

$$\left| \int_s^t \frac{\partial^2}{\partial x^2} f(X^n(u-), \Lambda^n(u-)) \sigma_n^2(u, X^n(u-), \Lambda^n(u-)) d\langle (M^n)^c \rangle(u) \right| \leq K \langle (M^n)^c \rangle(t).$$

Since

$$\langle M^n \rangle(t) = \langle (M^n)^c \rangle(t) + \int_0^t \int_{\mathbb{R}\setminus 0} x^2 \nu^n(dsdx),$$

assumptions (A1) and (A3-i) imply that $\langle (M^n)^c \rangle(t) \to 0$ in probability for each t and the above estimate gives us the desired result. □

Lemma 3.14. *Assume* (A1) – (A6). *Then*

$$\begin{aligned} I^n(t) = \int_s^t \int_{\mathbb{R}\setminus 0} \Bigg[& f\Big(X^n(u-) + \sigma_n(u, X^n(u-), \Lambda^n(u-))x, \Lambda^n(u-)\Big) - f(X^n(u-), \Lambda^n(u-)) \\ & - \frac{\partial}{\partial x} f(X^n(u-), \Lambda^n(u-)) \sigma_n(u, X^n(u-), \Lambda^n(u-)) x \Bigg] \nu^n(dudx) \\ \to \frac{1}{2} \int_s^t & \frac{\partial^2}{\partial x^2} f(\tilde{X}(s), \tilde{\Lambda}(s)) \sigma^2(s, \tilde{X}(s), \tilde{\Lambda}(s)) ds \end{aligned}$$

in probability for $0 \leq s \leq t \leq T$.

Proof. We shall prove the result under the Assumption (A3-i). The proof under the assumption (A3-ii) can be done similarly. Using Taylor's expansion, we have

$$\left| I^n(t) - \frac{1}{2} \int_s^t \frac{\partial^2}{\partial x^2} f(\tilde{X}(s), \tilde{\Lambda}(s)) \sigma^2(s, \tilde{X}(s), \tilde{\Lambda}(s)) \right| \leq \sum_{i=1}^6 J_i^n(t),$$

where

$$
\begin{aligned}
J_1^n(t) =& \frac{1}{2}\int_s^t \int_{\mathbb{R}\backslash 0} \Big| \frac{\partial^2}{\partial x^2} f\big(X^n(u-) + \theta^n(\omega)\sigma_n(u, X^n(u-), \Lambda^n(u-))x, \Lambda^n(u-)\big) \\
& \qquad - \frac{\partial^2}{\partial x^2} f\big(X^n(u-) + \theta^n(\omega)\sigma_n(u, X^n(u-), \Lambda^n(u-))x, \tilde{\Lambda}(u-)\big)\Big| \\
& \qquad \times \sigma_n^2(s, X^n(u), \Lambda^n(u-))x^2 \nu^n(dudx), \\
J_2^n(t) =& \frac{1}{2}\int_s^t \int_{\mathbb{R}\backslash 0} \Big| \frac{\partial^2}{\partial x^2} f\big(X^n(u-) + \theta^n(\omega)\sigma_n(u, X^n(u-), \Lambda^n(u-))x, \tilde{\Lambda}(u-)\big) \\
& \qquad - \frac{\partial^2}{\partial x^2} f(X^n(u-), \tilde{\Lambda}(u-))\Big| \times \sigma_n^2(s, X^n(s), \Lambda^n(s-))x^2 \nu^n(dudx), \\
J_3^n(t) =& \frac{1}{2}\int_s^t \int_{\mathbb{R}\backslash 0} \Big| \frac{\partial^2}{\partial x^2} f(X^n(u-), \tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| \\
& \qquad \times \sigma_n^2(u, X^n(u-), \Lambda^n(u-))x^2 \nu^n(dudx), \\
J_4^n(t) =& \frac{1}{2}\int_s^t \int_{\mathbb{R}\backslash 0} \Big| \Big(\sigma_n^2(s, X^n(u-), \Lambda^n(u-)) - \sigma^2(s, \tilde{X}(u-), \Lambda^n(u-))\Big)\Big| \\
& \qquad \times \Big|\frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| x^2 \nu^n(dudx), \\
J_5^n(t) =& \frac{1}{2}\int_s^t \int_{\mathbb{R}\backslash 0} \Big| \Big(\sigma_n^2(s, \tilde{X}(u-), \Lambda^n(u--)) - \sigma^2(s, \tilde{X}(u-), \tilde{\Lambda}(u-))\Big)\Big| \\
& \qquad \times \Big|\frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| x^2 \nu^n(dudx), \\
J_6^n(t) =& \frac{1}{2}\Big| \int_s^t \int_{\mathbb{R}\backslash 0} \sigma^2(s, \tilde{X}(u-), \tilde{\Lambda}(u-)) \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-)) x^2 \nu^n(dudx) \\
& \qquad - \int_s^t \sigma^2(s, \tilde{X}(u-), \tilde{\Lambda}(u-)) \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-)) du\Big|.
\end{aligned}
\tag{3.32}
$$

We will prove that each of those above terms converges to zero in probability for each fixed t and the desired claim then follows.

The proof for J_1^n can be carried out in a similar manner to that of Lemma 3.11. We now give a proof for the convergence to zero of J_2^n. Since $\frac{\partial^2}{\partial x^2} f(\cdot)$ is uniformly continuous and $\sigma_n(\cdot)$ is uniformly bounded, for an arbitrary $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that, for all x, if $|x| \le \delta(\varepsilon)$ then

$$\Big|\frac{\partial^2}{\partial x^2} f\big(X^n(u-) + \theta^n(\omega)\sigma_n(u, X^n(u-), \Lambda^n(u-))x, \tilde{\Lambda}(u-)\big) - \frac{\partial^2}{\partial x^2} f(X^n(u-), \tilde{\Lambda}(u-))\Big| \le \varepsilon.$$

We therefore can estimate J_2^n by

$$\begin{aligned} J_2^n &\le \frac{1}{2} K \left(\int_0^t \int_{0<|x|\le\delta(\varepsilon)} \varepsilon x^2 \nu^n(dsdx) + 2K \int_0^t \int_{|x|>\delta(\varepsilon)} x^2 \nu^n(dsdx) \right) \\ &\le \frac{\varepsilon}{2} K \int_0^t \int_{0<|x|\le\delta(\varepsilon)} x^2 \nu^n(dsdx) + K^2 \int_0^t \int_{|x|>\delta(\varepsilon)} x^2 \nu^n(dsdx) \to \frac{\varepsilon}{2} K \end{aligned} \tag{3.33}$$

in probability by assumption (A2). Since ε is arbitrary we conclude that J_2^n converges to 0 in probability for each t. For $J_3^n(t)$, we have

$$\begin{aligned} J_3^n(t) &\le K \sup_{s\le u\le t} \Big|\frac{\partial^2}{\partial x^2} f(X^n(u-), \tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| \int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dudx) \\ &\le K \sup_{s\le u\le \lambda_n^{-1}(t)} \Big|\frac{\partial^2}{\partial x^2} f(X^n \circ \lambda_n(u-), \tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| \int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dudx) \\ &\quad + K \sup_{s\le u\le \lambda_n^{-1}(t)} \Big|\frac{\partial^2}{\partial x^2} f(X \circ \lambda_n(u-), \tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2} f(\tilde{X}(u-), \tilde{\Lambda}(u-))\Big| \int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dudx) \\ &\to 0 \end{aligned} \tag{3.34}$$

in probability. Using quite similar arguments we can also prove that $J_4^n(t)$ and $J_5^n(t)$ are all convergent to zero with probability one. Indeed, we have

$$J_4^n(t) \le K \sup_{s\le u\le t} \big|\sigma_n^2(u, X^n(u-), \Lambda^n(u-)) - \sigma^2(u, \tilde{X}(u-), \Lambda^n(u-))\big| \int_0^t \int_{\mathbb{R}\setminus\{0\}} x^2 \nu^n(dudx),$$

and

$$
\begin{aligned}
&\sup_{s\le u\le t}\left|\sigma_n^2(u,X^n(u-),\Lambda^n(u-))-\sigma^2(u,\tilde X(u-),\Lambda^n(u-))\right|\\
&\quad\le \sup_{s\le u\le t}\left|\sigma_n^2(u,X^n(u-),\Lambda^n(u-))-\sigma_n^2(u,\tilde X(u-),\Lambda^n(u-))\right|\\
&\qquad+\sup_{s\le u\le t}\left|\sigma_n^2(u,\tilde X(u-),\Lambda^n(u-))-\sigma^2(u,\tilde X(u-),\Lambda^n(u-))\right|\\
&\quad\le \sup_{s\le u\le \lambda_n^{-1}(t)}\left|\sigma_n^2(u,X^n\circ\lambda^n(u-),\Lambda^n(u-))-\sigma_n^2(u,\tilde X(u),\Lambda^n(u-))\right|\\
&\qquad+\sup_{s\le u\le \lambda_n^{-1}(t)}\left|\sigma_n^2(u,\tilde X(u-),\Lambda^n(u-))-\sigma_n^2(u,\tilde X\circ\lambda_n(u-),\Lambda^n(u-))\right|\\
&\qquad+\sup_{s\le u\le t}\left|\sigma_n^2(u,\tilde X(u-),\Lambda^n(u-))-\sigma^2(u,\tilde X(u),\Lambda^n(u-))\right|\to 0
\end{aligned}
\tag{3.35}
$$

in probability.

We now work with the last term $J_6^n(t)$. Let

$$h(u,\omega)=\sigma^2(u,\tilde X(u-,\omega),\tilde\Lambda(u-,\omega))\frac{\partial^2}{\partial x^2}f(\tilde X(u-,\omega),\tilde\Lambda(u-,\omega)),$$

and for each $N>0$ we define h_N by

$$h_N(u,\omega)=h(u_m,\omega),u_m\le u<u_{m+1}\text{ where }\quad u_m=s+\frac{t-s}{N}m,m=0,1,2,\dots,N.$$

We have

$$
\begin{aligned}
J_6^n(t)&\le\int_s^t\int_{\mathbb{R}\setminus\{0\}}|h(u,\omega)-h_N(u,\omega)|x^2\nu^n(dudx)\\
&\quad+\left|\int_s^t\int_{\mathbb{R}\setminus\{0\}}h_N(u,\omega)x^2\nu^n(dudx)-\int_s^t h_N(u,\omega)du\right|+\int_s^t|h_N(u,\omega)-g(u,\omega)|du\\
&\le\sup_{s\le u\le t}|h(u,\omega)-h_N(u,\omega))|\int_s^t\int_{\mathbb{R}\setminus\{0\}}x^2\nu^n(dudx)\\
&\quad+\left|\int_s^t\int_{\mathbb{R}\setminus\{0\}}h_N(u,\omega)x^2\nu^n(dudx)-\int_s^t h_N(u,\omega)du\right|+\sup_{s\le u\le t}|h(u,\omega)-h_N(u,\omega))|t\\
&=U_{6,1}^n+U_{6,2}^n+U_{6,3}^n.
\end{aligned}
\tag{3.36}
$$

We first prove that for each n, $U_{6,1}^n$ and $U_{6,3}^n$ converge to zero in probability as $N \to \infty$. By assumption (A3), to do so it is sufficient to prove that $\sup_{s\le u\le t}|h(u,\omega) - h_N(u,\omega))|$ converges to zero in probability when N goes to infinity. Note that

$$\sup_{s\le u\le t}|h(u,\omega) - h_N(u,\omega))| = \max_{0\le m\le N-1}\sup_{u_m\le u\le u_{m+1}}|h(u,\omega) - h_N(u_m,\omega)|, \tag{3.37}$$

and

$$\begin{aligned}
&\sup_{u_m\le u\le u_{m+1}}|h(u) - h_N(u_m)| \\
&\quad\le \sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u,\tilde{X}(u-),\tilde{\Lambda}(u-)) - \sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u-))\right|\frac{\partial^2}{\partial x^2}f(\tilde{X}(u-),\tilde{\Lambda}(u-)) \\
&\qquad+ \sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u-)) - \sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right|\frac{\partial^2}{\partial x^2}f(\tilde{X}(u-),\tilde{\Lambda}(u-)) \\
&\qquad+ \sup_{u_m\le u\le u_{m+1}}\left|\frac{\partial^2}{\partial x^2}f(\tilde{X}(u-),\tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2}f(\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right|\sigma^2(u_m,\tilde{X}(u_m),\tilde{\Lambda}(u_m)) \\
&\quad\le K\sup_{u_m\le u\le u_{m+1}}|\tilde{X}(u-) - \tilde{X}(u_m-)| \\
&\qquad+ K\sup_{u_m\le u\le u_{m+1}}\left|\frac{\partial^2}{\partial x^2}f(\tilde{X}(u-),\tilde{\Lambda}(u-)) - \frac{\partial^2}{\partial x^2}f(\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right| \\
&\qquad+ K\sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u-)) - \sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right|,
\end{aligned} \tag{3.38}$$

where we have used the Lipschiz continuity of σ and the boundedness of f and σ to derive the last inequality. Therefore,

$$\begin{aligned}
&\sup_{s\le u\le t}|h(u)-h_N(u))| \\
&\qquad \le K\max_{0\le m\le N-1}\sup_{u_m\le u\le u_{m+1}}|\tilde{X}(u-)-\tilde{X}(u_m-)| \\
&\qquad\quad + K\max_{0\le m\le N-1}\sup_{u_m\le u\le u_{m+1}}\left|\frac{\partial^2}{\partial x^2}f(\tilde{X}(u-),\tilde{\Lambda}(u-))-\frac{\partial^2}{\partial x^2}f(\tilde{X}(s_m-),\tilde{\Lambda}(s_m-))\right| \\
&\qquad\quad + K\max_{0\le m\le N-1}\sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u-))-\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right| \\
&\qquad = V_1+V_2+V_3.
\end{aligned} \tag{3.39}$$

By Lemma 3.8, $\tilde{X}$ is a continuous processes and thus uniformly continuous. Therefore V_1 converges to zero a.e and hence in probability when N large enough. For V_3, we have

$$\sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u-))-\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right|\le K\mathbb{1}_\Delta,$$

where

$$\Delta=\{\exists u\in[u_m,u_{m+1})\text{ such that }\tilde{\Lambda}(u)\ne\tilde{\Lambda}(u_m-)\}.$$

Hence by the Tchebyshev inequality,

$$\begin{aligned}
&\mathbb{P}^0\left\{\sup_{u_m\le u\le u_{m+1}}\left|\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u))-\sigma^2(u_m,\tilde{X}(u_m-),\tilde{\Lambda}(u_m-))\right|\ge\varepsilon\right\} \\
&\qquad \le\frac{K}{\varepsilon}\left\{1-\mathbb{P}^0\left(\tilde{\Lambda}(u_m+\cdot)\text{ does not jump in the interval }\left[0,\frac{t-s}{N}\right)\right)\right\} \\
&\qquad =\frac{K}{\varepsilon}\left\{1-\sum_{i\in\mathcal{M}}\exp\left\{\int_0^{(t-s)/N}q_{ii}(u_m+u)du\right\}\mathbb{P}^0\left(\tilde{\Lambda}(u_m)=i\right)\right\} \\
&\qquad \le\frac{K}{\varepsilon}\left(1-e^{q(t-s)/N}\right).
\end{aligned} \tag{3.40}$$

This implies

$$\mathbb{P}^0\{V_3>\varepsilon\}\le\frac{NK}{\varepsilon}\left(1-e^{q(t-s)/N}\right),$$

which can be arbitrarily small when N is sufficiently large enough, thus V_3 converges to zero in probability. The proof that V_2 converges in probability to zero can be carried out by combining the arguments in the proofs that V_1 and V_3 converge in probability to zero. Therefore $\sup_{s\le u\le t}|h(u,\omega)-h_N(u,\omega))|$ converges to zero in probability.

We now prove that $U^n_{6,2}$ converges to zero in probability as $n\to\infty$. Indeed, for an arbitrary $\varepsilon>0$,

$$
\begin{aligned}
&\mathbb{P}^0\left\{\left|\int_s^t\int_{\mathbb{R}\setminus\{0\}} h_N(u,\omega)x^2\nu^n(dudx)-\int_s^t h_N(u,\omega)du\right|\ge e\right\}\\
&\quad=\mathbb{P}^0\left\{\left|\sum_{i=0}^{N-1} h_N(u_i,\omega)\left[\int_s^t\int_{\mathbb{R}\setminus\{0\}} x^2\nu^n(dudx)-(u_{i+1}-u_i)\right]\right|\ge\varepsilon\right\}\\
&\quad\le\mathbb{P}^0\left\{K\sum_{i=0}^{N-1}\left|\int_s^t\int_{\mathbb{R}\setminus\{0\}} x^2\nu^n(dudx)-(u_{i+1}-u_i)\right|\ge\varepsilon\right\}\\
&\quad\le\sum_{i=0}^{N-1}\mathbb{P}^0\left\{\left|\int_s^t\int_{\mathbb{R}\setminus\{0\}} x^2\nu^n(dudx)-(u_{i+1}-u_i)\right|\ge\frac{\varepsilon}{KN}\right\},
\end{aligned}
\tag{3.41}
$$

and the desired convergence follows. Since for each n, $U^n_{6,1}$ and $U^n_{6,3}$ converge to zero in probability when $N\to\infty$, for each $\varepsilon>0$, by first choose n and then N, we can make $\mathbb{P}\{J^n_6\ge\varepsilon\}$ arbitrarily small. Hence J^n_6 converges to zero in probability as desired. This finished the proof of the lemma. $\square$

Lemma 3.15. *Assume conditions* (A1)–(A5). *Then*

$$
\begin{aligned}
&\sum_{j_0\in\mathcal{M}}\int_s^t\left[f(X^n(u-),j_0)-f(X^n(u-),\Lambda^n(u-))\right]q_{\Lambda^n(u-),j_0}(u)du\\
&\qquad\to\int_s^t\sum_{i_0,j_0\in\mathcal{M}} q_{i_0,j_0}\left(\tilde{X}(u-)\right)\mathbb{1}_{i_0}\left(\tilde{\Lambda}(u-)\right)\left(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\right)du
\end{aligned}
$$

in probability for each $0\le s\le t\le T$.

Proof. We first note that

$$\begin{aligned}\sum_{j_0\in\mathcal{M}}\int_s^t &\left[f(X^n(u-),j_0)-f(X^n(u-),\Lambda^n(u-))\right]q_{\Lambda^n(u-),j_0}(u)du\\ &=\int_s^t\sum_{i_0,j_0\in\mathcal{M}}q_{i_0,j_0}\left(X^n(u-)\right)\mathbb{1}_{i_0}\left(\Lambda^n(u-)\right)\Big(f(X^n(u-),j_0)-f\left(X^n(u-),i_0\right)\Big)du.\end{aligned}$$

In addition, using similar arguments as we did in Lemma 3.11, we can prove that

$$f(X^n(u-),j_0)-f(X^n(u-),\Lambda^n(u-))\to f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)$$

in probability for $i_0,j_0\in\mathcal{M},i_0\neq j_0$. Moreover, for each $i_0\in\mathcal{M}$

$$\mathbb{P}^0\Big(|\mathbb{1}_{i_0}(\Lambda^n(s))-\mathbb{1}_{i_0}(\tilde{\Lambda}(s))|>\varepsilon\Big)\le\mathbb{P}^0\Big(|\Lambda^n(s)-\tilde{\Lambda}(s)|>\varepsilon\Big),$$

thus $\mathbb{1}_{i_0}(\Lambda^n(s))\to\mathbb{1}_{i_0}(\tilde{\Lambda}(s))$ in probability. These imply, for each $i_0,j_0\in\mathcal{M}$,

$$\begin{aligned}&q_{i_0,j_0}\left(u\right)\mathbb{1}_{i_0}\left(\Lambda^n(u-)\right)\Big(f(X^n(u-),j_0)-f\left(X^n(u-),i_0\right)\Big)\\ &\qquad\to q_{i_0,j_0}\left(u\right)\mathbb{1}_{i_0}\Big(\tilde{\Lambda}(u-)\Big)\Big(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\Big)\end{aligned}\tag{3.42}$$

in probability. By the uniformly integrable of the family

$$\Big\{q_{i_0,j_0}\left(u\right)\mathbb{1}_{i_0}\left(\Lambda^n(u-)\right)\Big(f(X^n(u-),j_0)-f\left(X^n(u-),i_0\right)\Big)\Big\},$$

we then have

$$\begin{aligned}\mathbb{E}^0\Big|&q_{i_0,j_0}\left(u\right)\mathbb{1}_{i_0}\left(\Lambda^n(u-)\right)\Big(f(X^n(u-),j_0)-f\left(X^n(u-),i_0\right)\Big)\\ &-q_{i_0,j_0}\left(u\right)\mathbb{1}_{i_0}\Big(\tilde{\Lambda}(u-)\Big)\Big(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\Big)\Big|\to 0.\end{aligned}\tag{3.43}$$

By virtue of Chebyshev's inequality and the bounded convergence theorem we obtain

$$
\begin{aligned}
&\mathbb{P}^0\Big(\Big|\int_s^t q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}(\Lambda^n(u-))\Big(f(X^n(u-),j_0)-f(X^n(u-),i_0)\Big)du \\
&\qquad\qquad -\int_s^t q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}\Big(\tilde{\Lambda}(u-)\Big)\Big(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\Big)du\Big|>\varepsilon\Big) \\
&\quad\le \frac{1}{\varepsilon}\mathbb{E}^0\Big|\int_s^t q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}(\Lambda^n(u-))\Big(f(X^n(u-),j_0)-f(X^n(u-),i_0)\Big)du \\
&\qquad\qquad -\int_s^t q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}\Big(\tilde{\Lambda}(u-)\Big)\Big(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\Big)du\Big| \\
&\quad\le \frac{1}{\varepsilon}\int_s^t \mathbb{E}^0\Big|q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}(\Lambda^n(u-))\Big(f(X^n(u-),j_0)-f(X^n(u-),i_0)\Big) \\
&\qquad\qquad - q_{i_0,j_0}(u)\,\mathbb{1}_{i_0}\Big(\tilde{\Lambda}(u-)\Big)\Big(f(\tilde{X}(u-),j_0)-f(\tilde{X}(u-),i_0)\Big)\Big|\to 0.
\end{aligned}
\tag{3.44}
$$

From this we can easily get the claim and finish the proof. □

Lemma 3.16. *Under assumptions* (A1)–(A6), $(\Xi(t),\mathbf{F},\mathbb{P}^0)$ *is a martingale.*

Proof. Thanks to the result of Lemma 3.10–Lemma 3.15, to show that $(\Xi(t),\mathbf{F},\mathbb{P}^0)$ is a martingale, we need only prove that the sequence $\{\Xi^n\}$ is uniformly integrable. From the definition of Ξ^n, the boundedness of b_n and σ_n and the fact that f is in $C_c^2(\mathbb{R})$, we can easily deduce that

$$|\Xi^n(t)|\le K+K\langle (M^n)^c\rangle(t)+K\int_0^t\int_{\mathbb{R}\setminus\{0\}} x^2\nu^n(dsdx)$$

for a positive constant K independent of n. This implies

$$|\Xi^n(t)|\le K+K\langle (M^n)^c\rangle(t). \tag{3.45}$$

By virtue of (3.45) and note that $\sup_n \mathbb{E}^0\langle M^n\rangle(t)<\infty$, we have the desired result and thus complete the proof. □

3.4 Final Remarks

In this chapter we work with one dimensional set up for simplicity in presentation. The result in multidimensional set up can be obtained by repeating the arguments presented here.

3.5 Appendix

We give here a proof of Lemma 3.5.

Proof. Let us define

$$\begin{aligned} H_{i_0}(t) &= \mathbb{1}_{\{\Lambda(t)=i_0\}}, \\ H_{i_0,j_0}(t) &= \sum_{s<u\leq t} \mathbb{1}_{\{\Lambda(u-)=i_0\}}\mathbb{1}_{\{\Lambda(u)=j_0\}} = \sum_{s<u\leq t} H_{i_0}(u-)H_{j_0}(u), \end{aligned} \tag{3.46}$$

for $i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$. The random variable $H_{i_0}(t)$ indicates whether or not Λ resides at i_0 at the time t, and $H_{i_0,j_0}(t)$ counts the number of jumps of Λ from i_0 to j_0 in the time interval $(s,t]$. Using the definition of H_{i_0} and H_{i_0,j_0} and the integration by part formula, we have

$$d(f(X(t),\Lambda(t))) = \sum_{i_0\in\mathcal{M}} d(f(X(t),\Lambda(t))H_{i_0}(t)) = I_1 + I_2 + I_3,$$

where

$$I_1 = \sum_{i_0\in\mathcal{M}} H_{i_0}(t-)df(X(t),i_0), \quad I_2 = \sum_{i_0\in\mathcal{M}} f(X(t-),i_0)dH_{i_0}(t), \quad I_3 = \sum_{i_0\in\mathcal{M}} \Delta f(X(t),i_0)\Delta H_{i_0}(t),$$

and $\Delta f(X(t),i_0) = f(X(t),i_0) - f(X(t-),i_0)$. We first work with I_2. Noting that

$$H_{i_0}(t) = \sum_{j_0, j_0\neq i_0} H_{j_0,i_0}(t) - \sum_{j_0,j_0\neq i_0} H_{i_0,j_0}(t),$$

hence we have

$$\begin{aligned}
I_2 &= \sum_{i\in\mathcal{M}} f(X(t-),i)d\left(\sum_{j_0,j_0\neq i_0}(H_{j_0,i_0}(t)-H_{i_0,j_0}(t))\right)\\
&= \sum_{i_0,j_0\in\mathcal{M};j_0\neq i_0} f(X(t-),i_0)dH_{j_0,i_0}(t) - \sum_{i_0,j_0\in\mathcal{M};j_0\neq i_0} f(X(t-),i_0)dH_{i_0,j_0}(t)\\
&= \sum_{i_0,j_0;j_0\neq i_0} f(X(t-),j_0)dH_{i_0,j_0}(t) - \sum_{i_0,j_0,j_0\neq i_0} f(X(t-),i_0)dH_{i_0,j_0}(t)\\
&= \sum_{i_0,j_0;j_0\neq i_0} \left(f(X(t-),j_0)-f(X(t-),i_0)\right)dH_{i_0,j_0}(t).
\end{aligned}\tag{3.47}$$

Next, from the fact that H_{i_0,j_0} and M do not have common jump, for $i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$, A is continuous and

$$\Delta X(t) = b(X(t-),\Lambda(t-))\Delta A + \sigma(X(t-),\Lambda(t-))\Delta M = \sigma(X(t-),\Lambda(t-))\Delta M.$$

we deduce $\Delta X(t)\Delta H_{i_0,j_0} = 0$ for $i_0, j_0 \in \mathcal{M}, i_0 \neq j_0$. Therefore, using similar calculations as in the case of I_1, we have

$$\begin{aligned}
I_3 &= \sum_{i_0\in\mathcal{M}} \Delta f(X(t),i_0)\Delta H_{i_0}(t)\\
&= \sum_{i_0,j_0\in\mathcal{M};i_0\neq j_0} \left(\Delta f(X(t),j_0)-\Delta f(X(t),i)\right)\Delta H_{i_0}(t) = 0.
\end{aligned}\tag{3.48}$$

For the term I_1, using Itô's lemma [42, page 79], we have

$$\begin{aligned}
df(X(t),i_0) =&\frac{\partial}{\partial x}f(X(t-),i_0)dX(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}f(X(t-),i_0)d\langle X^c\rangle(t)\\
&+\left[f(X(t),i_0)-f(X(t-),i_0)-\frac{\partial}{\partial x}f(X(t-),i_0)\Delta X(t)\right].
\end{aligned}\tag{3.49}$$

Thus,

$$
\begin{aligned}
H_{i_0}(t-)df(X(t),i_0) = & H_{i_0}(t-)\frac{\partial}{\partial x}f(X(t-),i_0)dA(t) + H_{i_0}(t-)\frac{\partial}{\partial x}f(X(t-),i_0)dM(t) \\
& + \frac{1}{2}H_{i_0}(t-)\frac{\partial^2}{\partial x^2}f(X(t-),i_0)d\langle M^c\rangle(t) \\
& + H_{i_0}(t-)\left[f(X(t),i_0) - f(X(t-),i_0) - \frac{\partial}{\partial x}f(X(t-),i_0)\Delta X(t)\right].
\end{aligned}
\tag{3.50}
$$

Combining (3.47), (3.48), and (3.50), we arrive at

$$
\begin{aligned}
df(X(t),\Lambda(t)) = & \sum_{i_0\in\mathcal{M}} H_{i_0}(t-)\frac{\partial}{\partial x}f(X(t-),i_0)b(t,X(t-),i_0)dA(t) \\
& + \sum_{i_0\in\mathcal{M}} H_{i_0}(t-)\frac{\partial}{\partial x}f(X(t-),i_0)\sigma(t,X(t-),i_0)dM(t) \\
& + \frac{1}{2}\sum_{i_0\in\mathcal{M}} H_{i_0}(t-)\frac{\partial^2}{\partial x^2}f(X(t-),i_0)\sigma^2(t,X(t-),i_0)d\langle M^c\rangle(t) \\
& + \sum_{i_0\in\mathcal{M}} H_{i_0}(t-)\left[f(X(t),i_0) - f(X(t-),i_0) - \frac{\partial}{\partial x}f(X(t-),i_0)\Delta X(t)\right] \\
& + \sum_{i_0,j_0;j_0\neq i_0} \Big(f(X(t-),j_0) - f(X(t-),i_0).\Big)dH_{i_0,j_0}(t).
\end{aligned}
\tag{3.51}
$$

Using (3.4) and the definition of H_{i_0}, H_{i_0,j_0}, we can rewrite

$$
\begin{aligned}
& \Big(f(X(t-),j_0) - f(X(t-),i_0)\Big)dH_{i_0,j_0}(t) \\
& \quad = \Big(f(X(t-),j_0) - f(X(t-),i_0)\Big)H_{i_0}(t-)dL^{i_0,j_0}(t) \\
& \quad\quad + \Big(f(X(t-),j_0) - f(X(t-),i_0)\Big)H_{i_0}(t-)q_{i_0,j_0}(X(t-))dt.
\end{aligned}
$$

Substituting this into equation (3.51) and rewriting the resulted equation in the integral form we get the desired claim. $\square$

CHAPTER 4 NEAR OPTIMALITY AND NEAR EQUILIBRIUM FOR CONTROLLED SYSTEMS WITH WIDEBAND NOISE FOR HYBRID SYSTEMS

4.1 Introduction

This chapter focuses on controlled hybrid systems being good approximations to controlled switching diffusion processes. Even though Brownian motion based models are good approximation to the real models, and are easily dealt with in terms of analysis. In real applications, the noise is often non-Markovian and the so-called "white noise" is only an idealization and simplification. So in lieu of the true "white noise", one may have an approximation of the Brownian motion. In lieu of a Brownian motion, we use a wide-band noise formulation, which facilitates the treatment of non-Markovian models. The wide-band noise is one whose spectrum has band width wide enough. We work with a basic stationary mixing type process. We conveniently introduce a small parameter ε so that as $\varepsilon \to 0$, the band width goes to that of the white noise. On top of this wide-band noise process, we allow the system to be subject to random discrete event influence. The discrete event process is a continuous-time Markov chain with a finite state space. Although the state space is finite, we assume that the state space is rather large and the Markov chain is irreducible. Using a two-time-scale formulation and assuming the Markov chain also subjects to fast variations, we obtain a limit controlled process.

Working with the original process is rather involved.

(1) The original process is non-Markov. Thus, the usual stochastic control techniques cannot be used. For example, in optimal stochastic control problems, we normally

obtain the associate Hamilton-Jaccobi-Bellman (HJB) equations using a dynamic programming approach. Because it is non-Markovian, it is not clear how we can get the corresponding HJB equations.

(2) Because the random environment, the Markov chain, has a large state space, the computational complexity is a main issue.

(3) The wide-band noise process makes the noise part varying fast. It makes handle the original problem even more difficult.

To face the challenges, we show that as the small parameter goes to 0, we obtain a limit system that is a controlled diffusion. Assuming that the limit problem has controls of desired type, we then plug such controls into the original system and show the resulting system is nearly optimal.

In this chapter, we first provide the formulation of the problem followed by some preliminary results regarding relaxed controls, weak convergence, and perturbed test functions method. We next establish weak convergence of the wide-band width noise control process to the suitable controlled diffusion. This result is interesting in its own right. This work is different from [28] in three aspects. First, our controlled process is not homogeneous in time. This leads to the use of different perturbed test functions. Second, our controlled process is also governed by a random discrete component. The third one is that, beside the optimal control problem, we also consider the equilibrium problem. Similar to [28], we first obtain a limit controlled system. Based on the optimal or near-optimal controls of the limit system, we then construct controls for the original problem. We further show that the controls so constructed lead to near optimality. Section 4.2 begins with the formulation of our problem.

Section 4.3 gives some preliminary results. Section 4.4 presents the weak convergence and near-optimal controls. Section 4.6 concentrates on linear quadratic problems.

4.2 Problem Setup

This section formulate an optimal control problem and a stochastic game involving a perturbed Markov chain and a wideband noise. Let $\beta(t), t \geq 0$, be an ergodic Markov chain on the state space $\mathcal{M} = \{1, 2, \ldots, m_0\}$ with the invariant measure $\nu = (\nu^1, \nu^2, \ldots, \nu^{m_0})$ and the generator $Q = (q_{i_0 j_0})_{1 \leq i_0, j_0 \leq m_0}$ satisfying $q_{i_0 j_0} \geq 0$ if $i_0 \neq j_0$ and $\sum_{j_0=1}^{m_0} q_{i_0 j_0} = 0$ for each $1 \leq i_0 \leq m_0$. For each $\varepsilon > 0$ denote $\beta^\varepsilon(t), t \geq 0$, the perturbed Markov chain with the same the state space $\mathcal{M}$ and generator $\frac{Q}{\varepsilon}$. Let r and N be positive integers, T a positive real number. In this chapter, the vectors are always column vectors and for a given vector or matrix M, M' denotes its transpose.

Let $\xi(t), t \geq 0$, be a stationary zero mean process taking values in $\mathbb{R}^r$ and $\mathcal{U}$ is a compact set in $\mathbb{R}^r$. In general, the dimension of ξ and the Euclidean space that contains $\mathcal{U}$ can be arbitrary. However, for conveniently keeping track the dimensions, we assume here that they both live in $\mathbb{R}^r$. The process $\xi(t)$ is bounded, right continuous and strongly mixing with the mixing rate function $\phi(t)$ defined by

$$\begin{aligned}\phi(t) = \sup \Big\{ \big|\mathbb{P}(B|A) - \mathbb{P}(B)\big| : A \in \sigma(\xi(u) : u \leq s),& \\ B \in \sigma(\xi(u) : u \geq s + t), s \geq 0 \Big\},& \quad t \geq 0.\end{aligned}$$

For each $\varepsilon > 0$ denote $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$, $t \geq 0$ the wideband noise process (the band width tends to infinity as $\varepsilon \to 0$). Throughout this chapter, we assume that the perturbed Markov chain $\beta^\varepsilon(t)$ and the wideband noise process $\xi^\varepsilon(t)$ are independent. Denote $\mathcal{P}(\mathbb{R})$ the space of all probability measures on $\mathbb{R}$. We are now in a position to setup the two problems.

Control problem ($\mathbf{CP}_\varepsilon$). Let $b(\cdot,\cdot,\cdot,\cdot):[0,T]\times\mathbb{R}^r\times\mathcal{M}\times\mathcal{U}\to\mathbb{R}^r$ and $g(\cdot,\cdot,\cdot):[0,T]\times\mathbb{R}^r\times\mathbb{R}^r\to\mathbb{R}^r$ be two functions of the forms

$$b(t,x,\beta,u)=\big(b^1(t,x,\beta,u),b^2(t,x,\beta,u),\dots,b^r(t,x,\beta,u)\big)',$$
$$g(t,x,\xi)=\big(g^1(t,x,\xi),g^2(t,x,\xi),\dots,g^r(t,x,\xi)\big)',$$

where $t\in[0,T]$, $x\in\mathbb{R}^r$, $\beta\in\mathcal{M}$, $\xi\in\mathbb{R}^r$, and $u\in\mathcal{U}$.

We consider the system of the following type

$$\begin{aligned} dx^\varepsilon(t)&=\Big[b\big(t,x^\varepsilon(t),\beta^\varepsilon(t),u^\varepsilon(t)\big)+\frac{1}{\varepsilon}g\big(t,x^\varepsilon(t),\xi^\varepsilon(t)\big)\Big]dt,\quad 0\le t\le T,\\ x^\varepsilon(0)&=x_0,\end{aligned}\tag{4.1}$$

where $i_0,j_0\in\mathcal{M}$, the initial condition x_0, the Markov chain $\beta^\varepsilon(t)$ and $\mathbb{R}^r$-valued stationary process $\xi^\varepsilon(t)$ defined above are independent, and the control $u^\varepsilon(t)\in\mathcal{U}$ for all $t\in[0,T]$. The control $u^\varepsilon(\cdot)$ is said to be admissible if $u^\varepsilon(t)\in\mathcal{U}$ for $0\le t\le T$ and $u^\varepsilon(\cdot)$ is progressively measurable with respect to the $\sigma-$algebras $\sigma\{x_0,\beta^\varepsilon(s),\xi^\varepsilon(s),s\le t\}$. It can be shown that under some mild conditions that are given in the next section and are assumed throughout this chapter, this ODE system has a unique solution.

The finite time horizon objective function is given by

$$J^\varepsilon(u^\varepsilon)=\mathbb{E}\left[\int_0^T h(x^\varepsilon(s),u^\varepsilon(s))ds+k(x^\varepsilon(T),u^\varepsilon(T))\right],\tag{4.2}$$

where u^ε is an admissible control, and $h(\cdot,\cdot)$ and $k(\cdot,\cdot):\mathbb{R}^r\times\mathcal{U}\to\mathbb{R}$ are bounded and continuous functions. The stochastic optimal control problem to be studied is to choose u_t to minimize the objective function over the finite time horizon $[0,T]$, subject to (4.1). Next, we consider a stochastic mean-field game with N-players whose dynamics involve a perturbed Markov chain and a wideband noise.

Game problem (GP$_\varepsilon$). The problem setup is motivated by recent advances in mean-field game problems (see [4, 29, 30, 39, 41] among others). Let $\hat{b}(\cdot,\cdot,\cdot,\cdot,\cdot) : [0,T]\times\mathbb{R}\times\mathcal{P}(\mathbb{R})\times\mathcal{M}\times\mathbb{R}\to\mathbb{R}$, $\hat{g}(\cdot,\cdot,\cdot,\cdot) : [0,T]\times\mathbb{R}\times\mathcal{P}(\mathbb{R})\times\mathcal{M}\to\mathbb{R}$ and $\mathcal{U}=\mathcal{U}^1\times\mathcal{U}^2\times\ldots\times\mathcal{U}^N$ be a compact set in $\mathbb{R}^N$. Assume that the dynamics of player i, $i=1,2,\ldots,N$, is given by

$$dx^{i,\varepsilon}(t)=\Big[\hat{b}\big(t,x^{i,\varepsilon}(t),\mu^{N,\varepsilon}(t),\beta^\varepsilon(t),u^{i,\varepsilon}(t)\big)+\frac{1}{\varepsilon}\hat{g}\big(t,x^{i,\varepsilon}(t),\mu^{N,\varepsilon}(t),\xi^{i,\varepsilon}(t)\big)\Big]dt, \tag{4.3}$$

for $0\le t\le T$ with the initial condition $x^{i,\varepsilon}(0)=x_0^i$, where

$$\mu^{N,\varepsilon}(t)=\frac{1}{N}\sum_{j=1}^{N}\delta_{x^{j,\varepsilon}(t)}\in\mathcal{P}(\mathbb{R}),\quad 0\le t\le T$$

is the mean-field coupled term that shows the weak interactions between the players, δ_x denotes the Dirac measure centered at x for each $x\in\mathbb{R}$, $\beta^\varepsilon(t)$ is a perturbed Markov chain, $\xi^\varepsilon(t)=\big(\xi^{1,\varepsilon}(t),\xi^{2,\varepsilon}(t),\ldots,\xi^{N,\varepsilon}(t)\big)'$ is an $\mathbb{R}^N$-valued stationary process, and the strategy $u^{i,\varepsilon}(t)\in\mathcal{U}^i$. We assume that the initial conditions x_0^i, $i=1,2,,\ldots,N$ are independent, identically distributed with bounded second moments and that $\xi^\varepsilon(\cdot)$ and $\beta^\varepsilon(\cdot)$ are independent. In addition, we assume that the components of $\xi^\varepsilon(\cdot)$ (i.e., $\xi^{1,\varepsilon}(\cdot),\xi^{2,\varepsilon}(\cdot),\ldots,\xi^{N,\varepsilon}(\cdot)$) are also independent.

A strategy (control) $u^{i,\varepsilon}(\cdot)$ of player i is said to be *admissible* if $u^{i,\varepsilon}(t)\in\mathcal{U}^i$ for $0\le t\le T$ and $u^{i,\varepsilon}(\cdot)$ is progressively measurable with respect to the σ-algebras $\sigma\{\beta^\varepsilon(s),\xi^\varepsilon(s),s\le t;\ x_0^j,1\le j\le N\}$. The set of strategies $u^\varepsilon(\cdot)=\big(u^{1,\varepsilon}(\cdot),u^{2,\varepsilon}(\cdot),\ldots,u^{N,\varepsilon}(\cdot)\big)$ is said to be admissible if $u^{i,\varepsilon}(\cdot)$ is admissible for each $i=1,2,\ldots,N$. For each set of strategies $u^\varepsilon=\big(u^{1,\varepsilon},u^{2,\varepsilon},\ldots,u^{N,\varepsilon}\big)'$ and each strategy $v^{i,\varepsilon}$ we denote $u^{-i,\varepsilon}=\big(u^{1,\varepsilon},\ldots,u^{i-1,\varepsilon},u^{i+1,\varepsilon},\ldots,u^{N,\varepsilon}\big)'$ and $(v^{i,\varepsilon},u^{-i,\varepsilon})=\big(u^{1,\varepsilon},\ldots,u^{i-1,\varepsilon},v^{i,\varepsilon},u^{i+1,\varepsilon},\ldots,u^{N,\varepsilon}\big)'$.

Define the cost functional of the ith player by

$$J^{i,\varepsilon}(u^\varepsilon) = \mathbb{E}\Big[\int_0^T \hat{h}\big(x^{i,\varepsilon}(s), \mu^{N,\varepsilon}(s), u^{i,\varepsilon}(s)\big)ds + \hat{k}\left(x^{i,\varepsilon}(T), \mu^{N,\varepsilon}(T), u^{i,\varepsilon}(T)\right)\Big] \tag{4.4}$$

where $u^\varepsilon(t) = \big(u^{1,\varepsilon}(t), u^{2,\varepsilon}(t), \ldots, u^{N,\varepsilon}(t)\big)'$ is a set of admissible strategies, and $\hat{h}(\cdot,\cdot,\cdot)$ and $\hat{k}(\cdot,\cdot,\cdot) : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ are continuous functions.

A set of admissible strategies $u^\varepsilon(t) = \big(u^{1,\varepsilon}(t), u^{2,\varepsilon}(t), \ldots, u^{N,\varepsilon}(t)\big)'$ is called a δ-Nash equilibrium if for any admissible strategy $v^{i,\varepsilon}$,

$$J^{i,\varepsilon}(u^\varepsilon) \leq J^{i,\varepsilon}(v^{i,\varepsilon}, u^{-i,\varepsilon}) + \delta, \quad i = 1, 2, \ldots, N.$$

The game problem to be studied in this case is to find a decentralized set of admissible strategies which is a δ-Nash equilibrium.

Because those two problems are non-Markovian, the usual stochastic control techniques do not work. Similar to [28] (see also [1, 50, 56]), our approach to solve these problems is based on the observation that if $u^\varepsilon(\cdot)$ is a sequence of "nice" admissible controls in the optimal control problem or a set of "nice" strategies in the game problem then as ε tends to 0 (and N is fixed in the game problem), the corresponding dynamics $x^\varepsilon(t)$ (which is $(x^{1,\varepsilon}(t), x^{2,\varepsilon}(t), \ldots, x^{r,\varepsilon}(t))'$ in the optimal control problem and $(x^{1,\varepsilon}(t), x^{2,\varepsilon}(t), \ldots, x^{N,\varepsilon}(t))'$ in the game problem) converges weakly to that of controlled diffusion process. The convergence of corresponding costs also hold. The convergence of $x^\varepsilon(t)$ in both problems can be proved using the relaxed controls and perturbed test function method. The following remark shows the similarity between the dynamics and cost functions in two problems in many cases which leads to only one proof of the weak convergence of $x^\varepsilon(t)$ in both problems and we emphasize that we only discuss the problems when those similarities are applied.

Remark 4.1. We note the following points.

(i) Suppose for the moment that $r = N$, $u^\varepsilon(t) = \big(u^{1,\varepsilon}(t), u^{2,\varepsilon}(t), \ldots, u^{N,\varepsilon}(t)\big)'$ is an admissible set of strategies, $x^\varepsilon(t) = \big(x^{1,\varepsilon}(t), x^{2,\varepsilon}(t), \ldots, x^{N,\varepsilon}(t)\big)'$ is the corresponding dynamics, $\xi^\varepsilon(t) = \big(\xi^{1,\varepsilon}(t), \xi^{2,\varepsilon}(t), \ldots, \xi^{N,\varepsilon}(t)\big)'$ are the wideband noises and

$$b^i\big(t, x^\varepsilon(t), \beta^\varepsilon(t), u^\varepsilon(t)\big) = \hat{b}\big(t, x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), \beta^\varepsilon(t), u^{i,\varepsilon}(t)\big), \tag{4.5}$$

$$g^i\big(t, x^\varepsilon(t), \xi^\varepsilon(t)\big) = \hat{g}\big(t, x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), \xi^{i,\varepsilon}(t)\big), \tag{4.6}$$

$$h\big(x^\varepsilon(t), u^\varepsilon(t)\big) = \hat{h}\big(x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), u^{i,\varepsilon}(t)\big),$$

$$k\big(x^\varepsilon(t), u^\varepsilon(t)\big) = \hat{k}\big(x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), u^{i,\varepsilon}(t)\big),$$

for $i = 1, 2, \ldots, N$ then (4.3) and (4.4) are respectively represented in exactly the same way as (4.1) and (4.2).

(ii) A typical situation of (4.3) and (4.4) is the case that the functions $\hat{b}(\cdot,\cdot,\cdot,\cdot,\cdot)$, $\hat{g}(\cdot,\cdot,\cdot,\cdot)$, $\hat{h}(\cdot,\cdot,\cdot)$ and $\hat{k}(\cdot,\cdot,\cdot)$ have the following forms (see [13, 14, 23, 41]).

$$\begin{aligned}
\hat{b}(t, x, \mu, \beta, u) &= \int_{\mathbb{R}} \tilde{b}(t, x, y, \beta, u)\mu(dy), \\
\hat{g}(t, x, \mu, \xi) &= \int_{\mathbb{R}} \tilde{g}(t, x, y, \xi)\mu(dy), \\
\hat{h}(x, \mu, u) &= \int_{\mathbb{R}} \tilde{h}(x, y, u)\mu(dy), \\
\hat{k}(x, \mu, u) &= \int_{\mathbb{R}} \tilde{k}(x, y, u)\mu(dy)
\end{aligned}$$

for $(t, x, \mu, \beta, u, \xi) \in [0,T] \times \mathbb{R} \times \mathcal{P}(\mathbb{R}) \times \mathcal{M} \times \mathbb{R} \times \mathbb{R}$, where $\tilde{b}(\cdot,\cdot,\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathcal{M} \times \mathbb{R} \to \mathbb{R}$, $\tilde{g}(\cdot,\cdot,\cdot,\cdot) : [0,T] \times \mathbb{R} \times \mathcal{P}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$, $\tilde{h}(\cdot,\cdot,\cdot) : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\tilde{k}(\cdot,\cdot,\cdot) : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. It is easy to see that in this case the dynamics and cost functional of player i becomes

$$dx^{i,\varepsilon}(t) = \Big[\frac{1}{N}\sum_{j=1}^{N} \tilde{b}(t, x^{i,\varepsilon}(t), x^{j,\varepsilon}(t), \beta^{\varepsilon}(t), u^{i,\varepsilon}(t)) + \\ + \frac{1}{N\varepsilon}\sum_{j=1}^{N} \tilde{g}\left(t, x^{i,\varepsilon}(t), x^{j,\varepsilon}(t), \xi^{i,\varepsilon}(t)\right)\Big]dt,$$

$$J^{i,\varepsilon}(u^{\varepsilon}) = \mathbb{E}\Big[\int_0^T \frac{1}{N}\sum_{j=1}^{N} \tilde{h}\big(x^{i,\varepsilon}(s), x^{j,\varepsilon}(s), u^{i,\varepsilon}(s)\big)ds + \\ + \frac{1}{N}\sum_{j=1}^{N} \tilde{k}\left(x^{i,\varepsilon}(T), x^{j,\varepsilon}(T), u^{i,\varepsilon}(T)\right)\Big].$$

To give a heuristic explanation for our approach, let us suppose that as ε tends to 0, the dynamics $x^{\varepsilon}(t)$ of the optimal control problem $(\mathbf{CP}_{\varepsilon})$ given by (4.1) are "close" to a controlled diffusion process (modeled by (4.7) below) in the sense that if $u^{\varepsilon}(\cdot)$ is a sequence of "nice" controls for (4.1), then there is a control $u(\cdot)$, and a corresponding controlled diffusion $x(u,\cdot)$ such that (as ε tends to 0) $x^{\varepsilon}(u^{\varepsilon},\cdot)$ converges weakly to $x(u,\cdot)$ that satisfies the following equation

$$dx(t) = \bar{b}(t, x(t), u(t))dt + \bar{\sigma}(t, x(t))dw(t), \tag{4.7}$$

where $\bar{b}$ and $\bar{\sigma}$ will be specified later (in (4.15) and (4.14)). Let $\bar{u}^{\delta}(\cdot)$, $\delta > 0$, be a "smooth" δ-optimal feedback control for the limit diffusion (4.7). Now apply $\bar{u}^{\delta}(\cdot)$ to (4.1) in the problem $(\mathbf{CP}_{\varepsilon})$. We will show that under quite broad conditions

$$\inf_{u\in\mathcal{R}^{\varepsilon}} J^{\varepsilon}(u) \geq J^{\varepsilon}(\bar{u}^{\delta}) - \delta$$

for $\varepsilon > 0$ small enough, where we used $\mathcal{R}^{\varepsilon}$ to denote the admissible relaxed controls for (4.1).

Similarly, for the mean field game problem, using some known results in mean field game theory, we will be able to construct $\hat{u}^{\delta}(\cdot)$, $\delta > 0$, a set of δ-Nash equilibrium feedback strategies for the limiting mean field game problem. Applying $\hat{u}^\delta(\cdot)$ to (4.1) in the problem $(\mathbf{GP}_\varepsilon)$, we can show that

$$J^{i,\varepsilon}(v^i, \hat{u}^{-i,\delta}) \geq J^{i,\varepsilon}(\hat{u}^\delta) - \delta$$

for $i = 1, 2, \ldots, N$, sufficiently small $\varepsilon > 0$, sufficiently large N, and any admissible feedback strategy v^i. Since $\bar{u}^\delta, \hat{u}^\delta$ are only functions of x and t, it would be considerably simple to find a nearly optimal control for (4.1) or nearly equilibrium for (4.3).

4.3 Preliminaries

In this section we present some preliminary results regarding the relaxed control and weak convergence needed for our problem.

4.3.1 Relaxed Controls

Let $\mathcal{U}$ denote the set of controls and $\mathcal{B}([0,\infty) \times \mathcal{U})$ the σ-algebra of Borel subsets of $[0,\infty) \times \mathcal{U}$. We assume that $\mathcal{U}$ is a compact set in some Euclidean space and let $\mathcal{F}_t$ be any given filtration, for example, $\mathcal{F}_t = \sigma\{w(s) : 0 \leq s \leq t\}$ where $w(\cdot)$ is a Brownian motion. Let

$$\begin{aligned} \mathcal{R}([0,\infty) \times \mathcal{U}) = \big\{ & m(\cdot) : m(\cdot) \text{ is a measure on } \mathcal{B}([0,\infty) \times \mathcal{U}) \\ & \text{and } m([0,t] \times \mathcal{U}) = t \text{ for all } t \big\}. \end{aligned}$$

A random $\mathcal{R}([0,\infty) \times \mathcal{U})$-valued measure $m(\cdot)$ is called an *admissible relaxed control* if for each $B \in \mathcal{B}(\mathcal{U})$, the function defined by $m(t, B) = m([0,t] \times B)$ is $\mathcal{F}_t-$adapted. An equivalent formulation reads that $m(\cdot)$ is a relaxed control if

$$\int_0^t h(s,\alpha) m(ds \times d\alpha)$$

is progressively measurable with respect to $\mathcal{F}_t$ for each bounded and continuous function $h(\cdot)$.

If $m(\cdot)$ is an admissible relaxed control and $B \in \mathcal{U}$ then the mapping $t \mapsto m([0,t] \times B)$ is absolutely continuous and hence differentiable almost everywhere. Since $\mathcal{B}(\mathcal{U})$ is countably generated, the time derivative of m exists almost everywhere and is a measurable mapping $m_t(\cdot) = \frac{d}{dt} m(t, \cdot)$, the "derivative" process, such that $m_t(d\alpha)dt = m(dt \times d\alpha)$ and $m_{\cdot}(B)$ is $\mathcal{F}_t-$adapted for each $B \in \mathcal{B}(\mathcal{U})$ and for nice (for example smooth) function $h(\cdot)$,

$$\int_{[0,\infty)\times\mathcal{U}} h(s,\alpha) m(ds \times d\alpha) = \int_0^\infty ds \int_{\mathcal{U}} h(s,\alpha) m_s(d\alpha).$$

As it is easier to work with $[0,\infty)$, we have defined above the relaxed controls on the interval $[0,\infty)$. If the control problem is of interest on the finite interval $[0,T]$ only, then we define $m(\cdot)$ in any admissible way on $[T,\infty)$. We topologize $\mathcal{R}([0,\infty)\times\mathcal{U})$ as follow (see [27]). For a bounded continuous function f on $[0,\infty] \times \mathcal{U}$ and a measure m in $\mathcal{R}([0,\infty)\times\mathcal{U})$ denote

$$\langle f, m \rangle = \int f(s,\alpha) m(ds \times d\alpha).$$

For each positive integer n let $\{f_{n_i}(\cdot) : i = 1, 2, \ldots\}$ be a countable dense set (under the sup-norm) of continuous functions on $[0,n] \times \mathcal{U}$ and denote

$$d_n(m_1, m_2) = \sum_{i=1}^\infty \frac{1}{2^i} \frac{|\langle f_{n_i}, m_1 - m_2 \rangle|}{1 + |\langle f_{n_i}, m_1 - m_2 \rangle|}.$$

We can now define a metric on $\mathcal{R}([0,\infty)\times\mathcal{U})$. Define

$$d(m_1, m_2) = \sum_{n=1}^\infty \frac{1}{2^n} d_n(m_1, m_2), \quad m_1, m_2 \in \mathcal{R}([0,\infty)\times\mathcal{U}).$$

Then the weak convergence in $\mathcal{R}([0,\infty)\times\mathcal{U})$ is equivalent to the convergence in this metric (i.e., $m_k(\cdot) \Rightarrow m(\cdot)$ if and only if $d(m_k, m) \to 0$ for any sequence $\{m_k, k \geq 1\}$ and m in

$\mathcal{R}([0,\infty)\times\mathcal{U})$).

4.3.2 Formulation Using Relaxed Controls

For $\varepsilon > 0, t \geq 0$, denote

$$\mathcal{F}_t^\varepsilon = \sigma\{x^\varepsilon(0), x^\varepsilon(s), \beta^\varepsilon(s), \xi^\varepsilon(s) : s \leq t\}. \tag{4.8}$$

Following the notion of Section 4.3.1, an admissible relaxed control for ($\mathbf{OP}_\varepsilon$) or ($\mathbf{GP}_\varepsilon$) is any $\mathcal{R}([0,\infty)\times\mathcal{U})$-valued function $m(\cdot)$ such that for any collection $\{h_\gamma(\cdot)\}$ of bounded and continuous functions $h_\gamma(\cdot)$ and for each $t > 0$, $\{\int_0^t h_\gamma(s,\alpha)m(ds\times d\alpha)\}$ is progressively measurable with respect to $\mathcal{F}_t^\varepsilon$.

We denote the set of admissible controls for the underlying problem with $\mathcal{F}_t = \mathcal{F}_t^\varepsilon$ by $\mathcal{R}^\varepsilon$,

$$\mathcal{R}^\varepsilon = \Big\{m^\varepsilon(\cdot) \in \mathcal{R}([0,\infty)\times\mathcal{U}) : \ m^\varepsilon(\cdot) \text{ is } \mathcal{F}_t^\varepsilon\text{−adapted}\Big\}.$$

Owning to the relaxed control formulation, instead of ($\mathbf{CP}_\varepsilon$) and ($\mathbf{GP}_\varepsilon$), we consider the following two relaxed problems.

Control problem with relaxed control representation ($\mathbf{CPR}_\varepsilon$). Minimizing

$$J^\varepsilon(m^\varepsilon) = \mathbb{E}\left[\int_0^T \int_{\mathcal{U}} h\big(x^\varepsilon(s),\alpha\big)m_s^\varepsilon(d\alpha)ds + \int_{\mathcal{U}} k\big(x^\varepsilon(T),\alpha\big)m_T^\varepsilon(d\alpha)\right], \tag{4.9}$$

where $x^\varepsilon(\cdot)$ subjects to

$$\begin{aligned} dx^\varepsilon(t) &= \left[\int_{\mathcal{U}} b\big(t, x^\varepsilon(t), \beta^\varepsilon(t), \alpha\big)m_t^\varepsilon(d\alpha) + \frac{1}{\varepsilon}g\big(t, x^\varepsilon(t), \xi^\varepsilon(t)\big)\right]dt, \\ x^\varepsilon(0) &= x_0, \quad \beta^\varepsilon(0) = \beta_0, \end{aligned} \tag{4.10}$$

and $m^\varepsilon(\cdot) \in \mathcal{R}^\varepsilon$.

Game problem with relaxed control representation ($\mathbf{GPR}_\varepsilon$). Find a set of admissible relaxed strategies $m^\varepsilon = (m^{1,\varepsilon}, \dots, m^{N,\varepsilon})$ such that the following inequality holds

$$J^{i,\varepsilon}(v^i, m^{-i,\varepsilon}) \geq J^{i,\varepsilon}(m^\varepsilon),$$

for any admissible relaxed strategy v^i and $i = 1, 2, \dots, N$, where

$$\begin{aligned} J^{i,\varepsilon}(m^\varepsilon) =& \mathbb{E}\Bigg[\int_0^T \int_{\mathcal{U}} h\big(x^{i,\varepsilon}(s), \mu^{N,\varepsilon}(s), \alpha\big) m_s^{i,\varepsilon}(d\alpha) ds \\ & \qquad + \int_{\mathcal{U}} k\big(x^{i,\varepsilon}(T), \mu^{N,\varepsilon}(T), \alpha\big) m_T^{i,\varepsilon}(d\alpha) \Bigg], \end{aligned} \tag{4.11}$$

where $x^\varepsilon(\cdot)$ subjects to

$$\begin{aligned} dx^{i,\varepsilon}(t) =& \Bigg[\int_{\mathcal{U}} b\big(t, x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), \beta^\varepsilon(t), \alpha\big) m_t^{i,\varepsilon}(d\alpha) \\ & \qquad + \frac{1}{\varepsilon} g\big(t, x^{i,\varepsilon}(t), \mu^{N,\varepsilon}(t), \xi^{i,\varepsilon}(t)\big) \Bigg] dt, \end{aligned} \tag{4.12}$$

$$x^{i,\varepsilon}(0) = x_0^i, \quad \beta^\varepsilon(0) = \beta_0,$$

and $m^\varepsilon(\cdot) \in \mathcal{R}^\varepsilon$.

We use the following assumptions throughout this chapter

Assumption (A).

(A1) $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$, $t \geq 0$, where $\xi(t)$ is a stationary zero mean process that is strong mixing, right continuous and bounded, with the mixing rate function $\phi(t)$ satisfying $\int_0^\infty \phi^{1/2}(s) ds < \infty$.

(A2) The following conditions hold.

(a) $\mathbb{E}\big[g(t, x, \xi(s))\big] = 0$ for each fixed $x \in \mathbb{R}^r$ and $t, s \geq 0$.

(b) $b(\cdot, \cdot, i_0, u)$ is continuous for each fixed $i_0 \in \mathcal{M}$ and u; g is continuous, $g_x(\cdot, \cdot, y)$ is continuous for each y.

(c) $b(\cdot,\cdot,i_0,\cdot)$ and g satisfy the linear growth condition and a Lipschitz condition in x uniformly with respect to t,β and u for each fixed $i_0\in\mathcal{M}$.

(d) $h(\cdot,\cdot)$ and $k(\cdot,\cdot)$ are bounded and continuous.

(A3) As $0<T_1,T_2\to\infty$,

$$\begin{aligned}\int_{-T_1}^{T_2}\mathbb{E}g(t,x,\xi(s))g'(t,x,\xi(0))ds &\to \frac{1}{2}a(t,x),\\ \int_0^{T_2}\mathbb{E}\sum_j g^i_{x_j}(t,x,\xi(s))g^j(t,x,\xi(0))ds &\to c(t,x),\quad 1\le i\le r.\end{aligned}\tag{4.13}$$

Denote

$$\bar a(t,x)=\frac{1}{2}\left(a(t,x)+a'(t,x)\right)\ge 0,$$

then there exists a function $\bar\sigma(t,x)$ such that

$$\bar a(t,x)=\bar\sigma(t,x)\bar\sigma'(t,x).\tag{4.14}$$

For $(t,x,\alpha)\in[0,T]\times\mathbb{R}^r\times\mathcal{U}$ let

$$\bar b(t,x,\alpha)=\sum_{i_0\in\mathcal{M}}b(t,x,i_0,\alpha)\nu^i+c(t,x)\tag{4.15}$$

where $\nu=(\nu^1,\nu^2,\ldots,\nu^{m_0})$ is the invariant measure of the Markov chain $\beta(t)$ and $c(t,x)$ is defined in (4.13). Because of the scaling of the wide-band process $\xi^\varepsilon(\cdot)$ and the assumption on the mixing rate $\phi(\cdot)$ in (A1) we can obtain a Brownian motion in the averaging for the $\xi^\varepsilon(\cdot)$ process. This leads to following two limit problems.

Limit optimal control problem (LCP). Assume Assumption (**A**). Let $\bar a(\cdot,\cdot)$ and $\bar b(\cdot,\cdot,\cdot)$ be respectively defined as in (4.14) and (4.15). Minimizing

$$J(m)=\mathbb{E}\left[\int_0^T\int_{\mathcal{U}}h\big(x(s),\alpha\big)m_s(d\alpha)ds+\int_{\mathcal{U}}k\big(x(T),\alpha\big)m_T(d\alpha)\right]\tag{4.16}$$

where $x(t)$ satisfies

$$
\begin{aligned}
dx(t) &= dt \int_{\mathcal{U}} \bar{b}\big(t, x(t), \alpha\big) m_t(d\alpha) + \bar{\sigma}\big(t, x(t)\big) dw(t), \\
x(0) &= x_0
\end{aligned}
\tag{4.17}
$$

and $m(\cdot) \in \mathcal{R}^0$ where

$$
\mathcal{R}^0 = \Big\{ m(\cdot) \in \mathcal{R}([0,\infty) \times \mathcal{U}) : \ m(\cdot) \text{ is } \mathcal{F}_t^0\text{−adapted} \Big\} \tag{4.18}
$$

and $\mathcal{F}_t^0 = \sigma(x_0, w(s) : 0 \leq s \leq t)$ for each $t \geq 0$.

Next, we define the limit game problem. Let $\xi(\cdot) = (\xi^1(\cdot), \xi^2(\cdot) \ldots, \xi^N(\cdot))'$, $\hat{b}(\cdot,\cdot,\cdot,\cdot,\cdot)$, $\hat{g}(\cdot,\cdot,\cdot,\cdot)$, $\hat{h}(\cdot,\cdot,\cdot)$ and $\hat{k}(\cdot,\cdot,\cdot)$ be defined as in problem ($\mathbf{GP}_\varepsilon$). Let $r = N$, $b = (b^1, b^2, \ldots, b^N)$ and $g = (g^1, g^2, \ldots, g^N)$ where b^i and g^i are respectively defined in (4.5) and (4.6) for $i = 1, 2, \ldots, N$. We assume the Assumption (**A**) and let $\bar{a}(\cdot,\cdot)$ and $\bar{b}(\cdot,\cdot,\cdot)$ be respectively defined as in (4.14) and (4.15). Notice that as we assume that the components of $\xi(\cdot)$ are independent it follows from (4.6) and (4.13) that $\bar{a}$ and thus $\bar{\sigma}$ are diagonal matrices. We write $\bar{\sigma} = \text{diag}(\bar{\sigma}^1, \bar{\sigma}^2, \ldots, \bar{\sigma}^N)$.

Limit game problem (LGP). Find an admissible set of relaxed strategies $m = (m^1, \ldots, m^N)$ such that the following inequality holds

$$
J^i(v^i, m^{-i}) \geq J^i(m),
$$

for any admissible relaxed strategy v^i and $i = 1, 2, \ldots, N$, where

$$
\begin{aligned}
J^i(m) = \mathbb{E}\Bigg[\int_0^T \int_{\mathcal{U}} \hat{h}\big(x^i(s), \mu^{N,\varepsilon}(s), \alpha\big) m_s^i(d\alpha) ds \\
+ \int_{\mathcal{U}} \hat{k}\big(x^i(T), \mu^{N,\varepsilon}(T), \alpha\big) m_T^i(d\alpha) \Bigg],
\end{aligned}
\tag{4.19}
$$

and $x(t)$ satisfies

$$\begin{aligned} dx^i(t) &= dt \int_{\mathcal{U}} \bar{b}^i\big(t, x^i(t), \mu^{N,\varepsilon}(t), \alpha\big) m_t^i(d\alpha) + \bar{\sigma}^i\big(t, x(t), \mu^{N,\varepsilon}(t)\big) dw^i(t), \\ x^i(0) &= x_0^i, \end{aligned} \tag{4.20}$$

where $w = (w^1, w^2, \dots, w^N)'$ is an N-dimensional standard Brownian motion, $m(\cdot) \in \mathcal{R}^0$, and $\mathcal{R}^0$ is defined as in (4.18) with $w = (w^1, w^2, \dots, w^N)'$. We sometimes write the solution to (4.1) as $x(u, \cdot)$ or $x(m, \cdot)$. We have the following results regarding the limiting control problem.

Theorem 4.2. Let $m(\cdot)$ be an admissible relaxed control (with respect to a Brownian motion $w(\cdot)$). The following assertions hold.

(i) There exists an adapted solution to

$$dx(t) = dt \int_{\mathcal{U}} \bar{b}\big(t, x(t), \alpha\big) m_t(d\alpha) + \bar{\sigma}\big(t, x(t)\big) dw(t), \quad x(0) = x_0, \tag{4.21}$$

and

$$\mathbb{E}\left[\sup_{t \le T} |x(t)|^2\right] \le K(1 + |x|^2),$$

where K depends only on T and on the Lipschitz coefficient of the drift and diffusion coefficients.

(ii) Define $\{x_n^\Delta\}$ by $x_0^\Delta = x_1^\Delta = x_0$ and for $n \ge 1$,

$$x_{n+1}^\Delta = x_n^\Delta + \int_{n\Delta - \Delta}^{n\Delta} ds \int \bar{b}(n\Delta, x_n^\Delta, \alpha) m_s(d\alpha) + \bar{\sigma}(n\Delta, x_n^\Delta)\big[w(n\Delta + \Delta) - w(n\Delta)\big]. \tag{4.22}$$

Define $x^\Delta(\cdot)$ to be the piecewise constant interpolation of $\{x_n^\Delta\}$. Then there is a $K_\Delta \to 0$ as $\Delta \to 0$ such that

$$\mathbb{E}\left[\sup_{t \le T} |x^\Delta(t) - x(t)|^2\right] \le K_\Delta(1 + |x|^2).$$

(K_Δ does not depend on $m(\cdot)$).

(iii) Let $m^n(\cdot) \Rightarrow \bar{m}(\cdot)$, where the $m^n(\cdot)$ are admissible with respect to some Brownian process, and let $x^n(\cdot)$ satisfy the equation (4.17) with $m(\cdot) = m^n(\cdot)$. Then $(x^n(\cdot), m^n(\cdot))$ converges weakly to $(x(\cdot), \bar{m}(\cdot))$ where $x(\cdot), \bar{m}(\cdot)$ satisfy the equation (4.17) for some Brownian motion process $w(\cdot)$ and $m(\cdot)$ is admissible with respect to $w(\cdot)$.

Proof. Though there is t-dependence in the drift and diffusion coefficients, the proofs of (i), (ii) follow the classical approximation method with slight modifications, details can be found in [8]. The proof of (iii) can be found in [28]. $\square$

Proposition 4.3. In the class of admissible relaxed controls for the problem (**LCP**) given in (4.16) and (4.17), there is an optimal control.

Proof. The proof of this Proposition can be found in [28]. We however include it here for the completeness of the presentation. The proof essentially follows from Theorem (4.2). We chose a weak convergent sequence $m^\delta(\cdot), \delta \to 0$, such that $J(m^\delta) \to \inf_{m \in \mathcal{R}^0} J(m) = \bar{J}$. We denote the limit of $\{x(m^\delta, \cdot), m^\delta(\cdot)\}$ by $(x(\bar{m}, \cdot), \bar{m}(\cdot))$. Then by Theorem 4.2, $\bar{m}(\cdot)$ is admissible for some Brownian motion $w(\cdot)$ and $(x(\bar{m}, \cdot), \bar{m}(\cdot), w(\cdot))$ solves (4.21). By weak convergence and the boundedness of h and k, we thus have

$$\begin{aligned}
&\mathbb{E} \int_0^T \int_{\mathcal{U}} h(x^\delta(s), \alpha) m_s^\delta(d\alpha) ds + \int_{\mathcal{U}} k(x^\delta(T), \alpha) m_T^\delta(d\alpha) \\
&\quad \to \mathbb{E} \int_0^T \int_{\mathcal{U}} h(x(s), \alpha) \bar{m}_s(d\alpha) ds + \int_{\mathcal{U}} k(x(T), \alpha) m_T(d\alpha) = \bar{J} = J(\bar{m}).
\end{aligned}$$

$\square$

Since we want to show (in the following sections) that any smooth and nearly optimal feedback control for (4.7) is nearly optimal control of (4.1) for small $\varepsilon > 0$, therefore it is important to know that there is a smooth nearly optimal control for (4.7).

Proposition 4.4. [The chattering lemma] For each $\delta > 0$, there is a piece-wise constant admissible control $u^\delta(\cdot)$ for the problem (**LCP**) such that

$$J(u^\delta) \le \inf_{m \in \mathcal{R}^0} J(m) + \delta.$$

Proof. The proof follows classical lines of arguments in [7], [8]. $\square$

Proposition 4.5. For each $\delta > 0$, there is a piece-wise constant (in t) and locally Lipschitz continuous in x (uniformly in t) control $\bar{u}^\delta(\cdot)$ for the problem (**LCP**) such that $\bar{u}^\delta(t) = \bar{u}^\delta(x(i\Delta), i\Delta)$ for $t \in (i\Delta, (i+1)\Delta]$ and

$$J(\bar{u}^\Delta) \le \inf_{m \in \mathcal{R}^0} J(m) + \delta.$$

Proof. The proof of this one is well-known in relaxed control but is long and technical so we do not include the proof here. One may consult Theorem 5.2 page 59 in [26] for a detailed proof. $\square$

Remark 4.6. Although we may not have a similar result to Proposition 4.5 for the game problem (**LGP**), it would be worthy noting that general results on the existence of equilibrium for mean field games with diffusion can be found in [29, 30]. For these problems, in order to reduce the complexity when there is a large population of players, the main concern is to determine a set of $\delta-$Nash equilibrium strategies such that each player only need to know its state information. General results on the existence of this kind of strategies are given in [4, 13, 14, 23] using a powerful method called Nash certainty equivalence principle.

4.3.3 Perturbed Test Function Method

This section is devoted to the perturbed test function method to be used to prove the weak convergence result in the next section. Let $D^r[0,\infty)$ denote the space of $\mathbb{R}^r$ valued functions which are right continuous and have left limits endowed with the Skorohod topology.

Following the approach of Kushner [25], we define the notation of "p-lim" and an operator $\hat{\mathcal{L}}^\varepsilon$ as follows. Let $\mathcal{F}_t^\varepsilon$ denote the minimal σ-algebra generate by $\{x^\varepsilon(s), \beta^\varepsilon(s), \xi^\varepsilon(s) : s \leq t\}$ and let $\mathbb{E}_t^\varepsilon$ denote the conditional expectation with respect to $\mathcal{F}_t^\varepsilon$. Let $\tilde{M}$ denote the set of real valued function of (t, ω) that are nonzero only on a bounded $t-$interval. Let

$$\bar{M}^\varepsilon = \left\{ f \in \tilde{M} : \sup_t \mathbb{E}|f(t)| < \infty \text{ and } f(t) \text{ is } \mathcal{F}_t^\varepsilon \text{ measurable} \right\}.$$

Let $f(\cdot), f^\delta(\cdot) \in \bar{M}^\varepsilon$, for each $\delta > 0$. Then $f = p - \lim_\delta f^\delta$ if and only if the following to conditions hold

$$\sup_{t,\delta} \mathbb{E}|f^\delta(t)| < \infty, \quad \lim_{\delta \to 0} \mathbb{E}|f(t) - f^\delta(t)| = 0, \quad \forall t > 0.$$

We say that $f(\cdot)$ is in $D(\hat{\mathcal{L}}^\varepsilon)$, the domain of the operator $\hat{\mathcal{L}}^\varepsilon$ and $\hat{\mathcal{L}}^\varepsilon f = g$ if for each $T < \infty$

$$p - \lim_{\delta \to 0} \left(\frac{\mathbb{E}_t^\varepsilon f(t+\delta) - f(t)}{\delta} - g(t) \right) = 0.$$

If $f(\cdot) \in D(\hat{\mathcal{L}}^\varepsilon)$ then

$$f(t) - \int_0^t \hat{\mathcal{L}}^\varepsilon f(s) ds \quad \text{is a martingale}$$

and

$$\mathbb{E}_t^\varepsilon f(t+s) - f(t) = \int_t^{t+s} \mathbb{E}_t^\varepsilon \hat{\mathcal{L}}^\varepsilon f(u) du.$$

The following theorem, which is a modified version of Theorem 4 page 44 of [25], gives a criterion for tightness of singular perturbed systems via perturbed test function methods.

Theorem 4.7. Let $x^\varepsilon(\cdot)$ have paths in $D^r[0, \infty)$ and let

$$\lim_{K \to \infty} \limsup_{\varepsilon \to 0} \mathbb{P}\{\sup_{t \leq T} |x^\varepsilon(t)| \geq K\} = 0, \quad \text{for each } T < \infty.$$

For each $f(\cdot, \cdot) \in C_0^{2,3}([0,T] \times \mathbb{R}^r)$ and $T < \infty$ let there be a sequence $f^\varepsilon(\cdot) \in \mathcal{D}(\hat{\mathcal{L}}^\varepsilon)$ such that either (i) or (ii) below hold. Then $\{x^\varepsilon(\cdot)\}$ is tight in $\mathbb{D}^r[0,\infty)$.

(i) For each $T<\infty, \{\hat{\mathcal{L}}^\varepsilon f^\varepsilon(t) : \varepsilon>0, 0\le t\le T\}$ is uniformly integrable and for each $\delta>0$,

$$\lim_{\varepsilon\to 0}\mathbb{P}\{\sup_{t\le T}|f^\varepsilon(t)-f(t,x^\varepsilon(t))|\ge\delta\}=0. \tag{4.23}$$

(ii) Equation (4.23) holds and for each $T<\infty$ there is a random variable $B_T^\varepsilon(f)$ such that

$$\sup_{t\le T}|\hat{\mathcal{L}}^\varepsilon f^\varepsilon(t)|\le B_T^\varepsilon(f),\qquad \lim_{K\to\infty}\limsup_{\varepsilon\to 0}\mathbb{P}\{B_T^\varepsilon(f)\ge K\}=0.$$

The following lemma provides a sufficient condition for weak convergence of a sequence of processes using the perturbed test function techniques.

Lemma 4.8. Suppose that $\{z^\varepsilon(\cdot)\}$ is defined on $[0,T]$. Let $\{z^\varepsilon(\cdot)\}$ be tight on $\mathbb{D}^r[0,\infty)$. Suppose that for each $f(\cdot,\cdot)\in C_0^{2,3}([0,T]\times\mathbb{R}^r)$ there exists $f^\varepsilon(\cdot)\in D(\mathcal{L}^\varepsilon)$ such that

$$\mathrm{p}-\lim_{\varepsilon\to 0}(f^\varepsilon(\cdot)-f(\cdot,z^\varepsilon(\cdot)))=0,$$

and

$$\mathrm{p}-\lim_{\varepsilon\to 0}(\mathcal{L}^\varepsilon f^\varepsilon(\cdot)-\mathcal{L}f(\cdot,z^\varepsilon(\cdot)))=0.$$

Then $z^\varepsilon(\cdot)\Rightarrow z(\cdot)$.

4.4 Weak Convergence and Approximation of Optimal and Equilibrium Control for $x^\varepsilon(\cdot)$

In this section, we consider control problem $(\mathbf{CP}_\varepsilon)$ and prove the main result of this chapter. Theorem 4.9 states that the weak limit of any weak convergent sequence of admissible relaxed control for the problem $(\mathbf{CP}_\varepsilon)$ is an admissible relaxed control for the problem

(**LCP**) and that the corresponding costs converge. We show that any smooth "nearly optimal" feedback control for (**LCP**) also is "nearly optimal" for ($\mathbf{CP}_\varepsilon$) for ε small enough, and that any "nearly equilibrium" feedback strategies for (**LGP**) is also "nearly equilibrium" in some sense for ($\mathbf{GP}_\varepsilon$) for ε small enough, N large enough.

4.4.1 Weak Convergence of $x^\varepsilon(\cdot)$

Let $\delta_\varepsilon \to 0$ and let $\hat{m}^\varepsilon(\cdot)$ be a δ_ε-optimal admissible relaxed control for the problem ($\mathbf{CPR}_\varepsilon$) with the state process defined by

$$dx^\varepsilon(t) = \left[\int_{\mathcal{U}} b(t, x^\varepsilon(t), \beta^\varepsilon, \alpha^\varepsilon(t)) m_t(d\alpha) + \frac{1}{\varepsilon} g\left(t, x^\varepsilon(t), \xi^\varepsilon(t)\right)\right] dt,$$

and the cost function given in (4.9). *As we mentioned before, we define all $m(\cdot)$ on $[0,\infty)$ for convenience.* Define $\mathcal{L}_t^\alpha$, the generator of the control diffusion (4.17), by

$$\mathcal{L}_t^\alpha f(t,x) = f_t(t,x) + f_x(t,x)'\bar{b}(t,x,\alpha) + \frac{1}{2}\sum_{i,j} f_{x_i x_j}(x)\bar{a}_{ij}(t,x), \tag{4.24}$$

where

$$\bar{b}(t,x,\alpha) = \sum_{i_0 \in \mathcal{M}} b(t,x,i_0,\alpha)\nu^{i_0} + c(t,x)$$

as defined in (4.15).

Theorem 4.9. Assume (**A**). Then $\{x^\varepsilon(\hat{m}^\varepsilon,\cdot), \hat{m}^\varepsilon(\cdot)\}$ is tight in $\mathbb{D}^r[0,T] \times \mathcal{R}([0,\infty)\times\mathcal{U})$. Let $(x^\varepsilon(\hat{m}^\varepsilon,\cdot), \hat{m}^\varepsilon) \Rightarrow (x(\hat{m},\cdot), \hat{m}(\cdot))$. Then there is a $w(\cdot)$ such that $\hat{m}(\cdot)$ is admissible with respect to $w(\cdot)$ and

$$dx(t) = dt\int_{\mathcal{U}} \bar{b}(t,x(t),\alpha)\hat{m}_t(d\alpha) + \bar{\sigma}(t,x(t))dw(t).$$

Also

$$
\begin{aligned}
J^{\varepsilon}(m^{\varepsilon}) &= \mathbb{E}\left[\int_0^T \int_{\mathcal{U}} h(x^{\varepsilon}(s),\alpha)m^{\varepsilon}_s(d\alpha)ds + \int_{\mathcal{U}} k(x^{\varepsilon}(T),\alpha(T))m^{\varepsilon}_T(d\alpha)\right] \\
&\to \mathbb{E}\left[\int_0^T \int_{\mathcal{U}} h(x(s),\alpha)\hat{m}_s(d\alpha)ds + \int_{\mathcal{U}} k(x(T),\alpha(T))\hat{m}_T(d\alpha)\right] \\
&= J(\hat{m}).
\end{aligned}
$$

Proof. To prove the theorem we use the perturbed test function methods. The proof is divided into several steps as described follows. We can assume that $x(\cdot)$ is bounded. Otherwise, by using truncation methods, we can work out the details.

Step 1. Tightness of $\{x^{\varepsilon}(\cdot)\}$. To establish the tightness of $\{x^{\varepsilon}(\cdot)\}$ we verify that the conditions of Theorem (4.7) are satisfied. First we need to verify

$$
\lim_{K\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}\{\sup_{t\le T} |x^{\varepsilon}(t)| \ge K\} = 0, \quad \text{for each } T < \infty.
$$

However, direct verification of this claim is tedious; we thus instead use the truncation method to prove the convergence. The method is described as follows: For each $K > 0$, let

$$
B_K = \{x \in \mathbb{R}^r : |x| \le K\} \quad \text{be the ball of radius } K.
$$

Let $x^{\varepsilon,K}(t) = x^{\varepsilon}(t)$ up until first exit time of x^{ε} from B_K, and then it is clear that

$$
\lim_{K\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}\{\sup_{t\le T} |x^{\varepsilon,K}(t)| \ge K\} = 0, \quad \text{for each } T < \infty,
$$

$x^{\varepsilon,K}$ is said to be the $K-$truncation of x^{ε}. Let

$$
q^K(x) = \begin{cases} 1 & \text{for } x \in B_K, \\ 0 & \text{for } x \in \mathbb{R}^r - B_{K+1}, \\ \text{smooth} & \text{otherwise.} \end{cases}
$$

Define

$$b_K(t,x,i,u) = b(t,x,i,u)q^K(x) \;\text{ and } g_K(t,x,\xi) = g(t,x,\xi)q^K(x).$$

Let $x^{\varepsilon,K}$ is the solution of (4.1) corresponding to the coefficients that are truncated as above, then it is clear that $x^{\varepsilon,K}(t) = x^\varepsilon(t)$ whenever $|x^\varepsilon(t)| < K$. To avoid the heavy notation, we however will write $x^{\varepsilon,K}(t)$ as x.

Since we are working with the truncated system, our work boils down to verify (i). We also write $\beta = \beta^\varepsilon(t)$ for notation simplicity. Let $f(\cdot,\cdot) \in C_0^{2,3}([0,T]\times\mathbb{R}^r)$. then

$$\hat{\mathcal{L}}^\varepsilon f(t,x) = f_t(t,x) + f_x(t,x)' \left[\int_{\mathcal{U}} b(t,x,\beta,\alpha)\hat{m}_t^\varepsilon(d\alpha) + \frac{1}{\varepsilon} g\left(t,x,\xi^\varepsilon(t)\right)\right].$$

For arbitrary $T < \infty$ and for $t \le T$ define $f_1^\varepsilon(t) = f_1^\varepsilon(t, x^\varepsilon(t))$, where

$$\begin{aligned} f_1^\varepsilon(t,x) &= \frac{1}{\varepsilon}\int_t^T f_x(t,x)'\mathbb{E}_t^\varepsilon g(t,x,\xi^\varepsilon(s))ds \\ &= \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon f_x(t,x)' g(t,x,\xi(s))ds, \end{aligned}$$

in which we have used, and will use frequently from now on, the change variable $s \to \varepsilon^2 s$ in the second equality. Thus, using the boundedness of f_x, g and the mixing property of ξ, we have

$$\begin{aligned} \sup_{t\le T} |f_1^\varepsilon(t,x)| &\le \varepsilon K \int_{t/\varepsilon^2}^{T/\varepsilon^2} |\mathbb{E}_t^\varepsilon g(t,x,\xi(s))| ds \\ &\le \varepsilon K \int_{t/\varepsilon^2}^{T/\varepsilon^2} \left|\mathbb{E}_t^\varepsilon g(t,x,\xi(s)) - \mathbb{E} g(t,x,\xi(s))\right| ds \\ &\le \varepsilon K \int_{t/\varepsilon^2}^{T/\varepsilon^2} \phi\left(s-t\right) ds \\ &= O(\varepsilon), \end{aligned}$$

uniformly in t and thus

$$\sup_{t\le T} |f_1^\varepsilon(t,x)| \to 0 \;\text{ in probability as }\; \varepsilon \to 0.$$

Furthermore, $f_1^\varepsilon(t,x)$ is differentiable in x and

$$f_1^\varepsilon(\cdot) = f_1^\varepsilon(\cdot, x^\varepsilon(\cdot)) \in \mathcal{D}(\hat{\mathcal{L}}^\varepsilon) \quad \text{and}$$

$$\begin{aligned}
\hat{\mathcal{L}}^\varepsilon f_1^\varepsilon(t,x) = &-\frac{1}{\varepsilon} f_x(t,x)' g(t,x,\xi^\varepsilon(t)) \\
&+ \frac{1}{\varepsilon}\int_t^T \mathbb{E}_t^\varepsilon \left[f_x(t,x)' g(t,x,\xi^\varepsilon(s)) \right]_t ds \\
&+ \left(f_1^\varepsilon(t,x)\right)_x' \left[\int_{\mathcal{U}} b(t,x,\beta,\alpha) m_t(d\alpha) + \frac{1}{\varepsilon} g\left(t,x,\xi^\varepsilon(t)\right) \right],
\end{aligned}$$

where

$$\left(f_1^\varepsilon(t,x)\right)_x = \frac{1}{\varepsilon}\int_t^T \mathbb{E}_t^\varepsilon \left[f_x(t,x)' g(t,x,\xi^\varepsilon(s)) \right]_x ds.$$

We next consider the function f perturbed by f_1^ε. To be more specific, let

$$h^\varepsilon(t) = f(t, x^\varepsilon(t)) + f_1^\varepsilon(t).$$

Note that the order of magnitude of f_1^ε is small and it results in the needed cancelations.

We have

$$
\begin{aligned}
\hat{\mathcal{L}}^\varepsilon h^\varepsilon(t) &= \hat{\mathcal{L}}^\varepsilon f(t, x^\varepsilon(t)) + \hat{\mathcal{L}}^\varepsilon f_1^\varepsilon(t) \\
&= f_t(t,x) + f_x(t,x)' \int_{\mathcal{U}} b(t,x,\beta,\alpha)\hat{m}_t^\varepsilon(d\alpha) \\
&\quad + \frac{1}{\varepsilon}\int_t^T \mathbb{E}_t^\varepsilon \left[f_x(t,x)' g(t,x,\xi^\varepsilon(s))\right]_t ds \\
&\quad + \frac{1}{\varepsilon}\left[\int_t^T \mathbb{E}_t^\varepsilon \left(f_x(t,x)' \mathbb{E}_t^\varepsilon g(t,x,\xi^\varepsilon(s))\right)_x ds\right]' \int_{\mathcal{U}} b(t,x,\beta,\alpha)\hat{m}_t^\varepsilon(d\alpha) \\
&\quad + \frac{1}{\varepsilon^2}\left[\int_t^T \mathbb{E}_t^\varepsilon \left(f_x(t,x)' g(t,x,\xi^\varepsilon(s))\right)_x ds\right]' g(t,x,\xi^\varepsilon(t)) \\
&= f_t(t,x) + f_x(t,x) \int_{\mathcal{U}} b(t,x,\beta,\alpha)\hat{m}_t^\varepsilon(d\alpha) \\
&\quad + \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_x(t,x)' g(t,x,\xi(s))\right]_t ds \\
&\quad + \varepsilon \left[\int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left(f_x(t,x)' g(t,x,\xi(s))\right)_x ds\right]' \int_{\mathcal{U}} b(t,x,\beta,\alpha)\hat{m}_t^\varepsilon(d\alpha) \\
&\quad + \left[\int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left(f_x(t,x)' g(t,x,\xi(s))\right)_x ds\right]' g(t,x,\xi^\varepsilon(t)).
\end{aligned}
$$

Under our assumptions, in the last equation, the first, the second, and the last term are $O(1)$; the remaining terms are $O(\varepsilon)$. Therefore the sequence $x^\varepsilon(\cdot))$ is tight.

Step 2. The martingale problem satisfied by the limit. To characterize the limit we compute one more perturbed test function. Let

$$
\hat{F}_\varepsilon(t,x) = \int_{t/\varepsilon^2}^{T/\varepsilon^2} ds \mathbb{E}_t^\varepsilon \left(f_x(t,x)' g(t,x,\xi(s))\right)_x' g(t,x,\xi^\varepsilon(t)),
$$

and define

$$
\begin{aligned}
&f_2^\varepsilon(t,x)\\
&= \int_t^T (\mathbb{E}_t^\varepsilon \hat{F}_\varepsilon(x,\tau) - \mathbb{E}\hat{F}_\varepsilon(x,\tau))d\tau\\
&= \int_t^T d\tau \Bigg\{ \int_{\tau/\varepsilon^2}^{T/\varepsilon^2} ds \Big[\mathbb{E}_t^\varepsilon \left(f_x(\tau,x)' g(\tau,x,\xi(s)) \right)'_x g(\tau,x,\xi^\varepsilon(\tau))\\
&\qquad\qquad - \mathbb{E}\left(f_x(\tau,x)' g(\tau,x,\xi(s)) \right)'_x g(\tau,x,\xi^\varepsilon(\tau)) \Big] \Bigg\}\\
&= \varepsilon^2 \int_{t/\varepsilon^2}^{T/\varepsilon^2} d\tau \Bigg\{ \int_{\tau/\varepsilon^2}^{T/\varepsilon^2} ds \Big[\mathbb{E}_t^\varepsilon \left(f_x(\tau,x)' g(\tau,x,\xi(s)) \right)'_x g(\tau,x,\xi(\tau))\\
&\qquad\qquad - \mathbb{E}\left(f_x(\tau,x)' g(\tau,x,\xi(s)) \right)'_x g(\tau,x,\xi(\tau)) \Big] \Bigg\}\\
&= O(\varepsilon^2).
\end{aligned}
$$

By our assumptions, $f_2^\varepsilon(\cdot) \in \mathcal{D}(\hat{\mathcal{L}}^\varepsilon)$. Its $\mathrm{p}-\lim_\varepsilon$ is zero and

$$
\begin{aligned}
&\hat{\mathcal{L}}^\varepsilon f_2^\varepsilon(t)\\
&\quad = -\hat{F}_\varepsilon(t,x) + \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}\left(f_x(t,x)' g(t,x,\xi(s)) \right)'_x g(t,x,\xi(\frac{t}{\varepsilon^2}))ds + O(\varepsilon^2).
\end{aligned}
$$

To simplify the notation let still denote the weak convergence subsequence by $\{x^\varepsilon(\cdot), \hat{m}^\varepsilon(\cdot)\}$ and its limit by $x(\cdot), \hat{m}(\cdot)$. By virtues of weak convergence and the absolutely continuity of $\hat{m}(\cdot)$ there is an $(\omega,t)-$measurable $\hat{m}_t$ such that $\hat{m}_t(U)=1$ and

$$
\int_0^t \int_{\mathcal{U}} f(s,\alpha)\hat{m}_s(d\alpha)ds = \int_0^t \int_{\mathcal{U}} f(s,\alpha)\hat{m}(ds\times d\alpha)
$$

for each continuous $f(\cdot)$. Let $f(\cdot) \in C_0^{2,3}([0,T]\times\mathbb{R}^r)$ and define $M_f(\cdot)$ by

$$
M_f(t) = f(t,x(t)) - f(0,x(0)) - \int_0^t \int_{\mathcal{U}} \mathcal{L}_s^\alpha f(s,x(s))m_s(d\alpha)ds.
$$

We are going to show that $M_f(\cdot)$ is a martingale with respect to $\mathcal{G}_t = \sigma\{x(s), \hat{m}(A\times[0,s]) :$ A is a Borel set $, s\le t\}$. Let $h(\cdot)$ be any real valued, bounded and continuous function of its argument. Let $\varphi(\cdot), \varphi_j(\cdot)$ below be real valued and continuous function with compact support. Define the function

$$(\varphi, \hat{m})_t = \int_0^t \int_{\mathcal{U}} \varphi(s, \alpha) \hat{m}(d\alpha \times ds).$$

Let $t_i < t < t + s$, let q_1 and q_2 be arbitrary integers, using the results of the calculations with perturbed tests functions we have

$$\begin{aligned}
&\mathbb{E} h(x^\varepsilon(t_i), (\varphi_j, \hat{m}^\varepsilon)_{t_i}, 1 \le q_1, j \le q_2) \\
&\quad \Big\{ f(t+s, x^\varepsilon(t+s)) - f(t, x^\varepsilon(t)) + f_1^\varepsilon(t+s) - f_1^\varepsilon(t) + f_2^\varepsilon(t+s) - f_2^\varepsilon(t) \\
&\qquad - \int_t^{t+s} f_t(\tau, x^\varepsilon(\tau)) d\tau \\
&\qquad + \int_t^{t+s} d\tau \int_{\mathcal{U}} f_x(\tau, x^\varepsilon(\tau))' b(\tau, x^\varepsilon(\tau), \beta^\varepsilon(\tau), \alpha) \hat{m}_\tau^\varepsilon(d\alpha) \\
&\qquad - \int_t^{t+s} d\tau \int_{\tau/\varepsilon^2}^{T/\varepsilon^2} ds \mathbb{E} \left(f_x(\tau, x^\varepsilon(\tau))' g(\tau, x^\varepsilon(\tau), \xi(s)) \right)_x' g(\tau, x^\varepsilon(\tau), \xi^\varepsilon(\tau)) \\
&\qquad\quad + \text{term which goes to 0 in mean as } \varepsilon \to 0 \Big\} = 0.
\end{aligned} \tag{4.25}$$

We now take the limit as $\varepsilon \to 0$ in (4.25) and use Skorohod representation, which allows us to define a new probability space so that the weak convergence becomes the convergence almost surely in the topology of the space $\mathbb{D}^r[0, \infty) \times \mathcal{R}([0, \infty) \times \mathcal{U})$.

The terms related to perturbed test functions f_1^ε and f_2^ε go to 0 as $\varepsilon \to 0$. By the weak convergence of x^ε and Skorohod representation we have

$$\int_t^{t+s} f_t(\tau, x^\varepsilon(\tau)) d\tau \to \int_t^{t+s} f_t(\tau, x(\tau)) d\tau, \qquad (\varphi_j, \hat{m}^\varepsilon)_{t_i} \to (\varphi_j, \hat{m})_{t_i}.$$

For the double integral in the second line of (4.25), using the estimate regarding the convergence of the transition probability to the invariant measure of the underlying Markov chain (see [51]) we have

$$\begin{aligned}
&\mathbb{E}\left[\int_t^{t+s} d\tau \int_{\mathcal{U}} f_x(\tau, x^\varepsilon(\tau))' \bar{b}(\tau, x^\varepsilon(\tau), \alpha)\hat{m}_t^\varepsilon(d\alpha)\right.\\
&\qquad \left. - \int_t^{t+s} \int_{\mathcal{U}} f_x(\tau, x^\varepsilon(\tau))' b(\tau, x^\varepsilon(\tau), \beta^\varepsilon(\tau), \alpha)\hat{m}_t^\varepsilon(d\alpha)\right]\\
&= \mathbb{E}\int_t^{t+s} d\tau \int_{\mathcal{U}} \sum_{i\in\mathcal{M}} f_x(\tau, x^\varepsilon(\tau))' b(\tau, x^\varepsilon(\tau), i, \alpha)\left[\nu_i - \mathbb{1}_{\{\beta^\varepsilon(\tau)=i\}}\right]\hat{m}_t^\varepsilon(d\alpha)\\
&= O(\varepsilon).
\end{aligned}$$

By Theorem 5.11 of [25], we have

$$\begin{aligned}
&\int_t^{t+s} d\tau \int_{\tau/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}\left(f_x(\tau, x^\varepsilon(\tau))' g(\tau, x^\varepsilon(\tau), \xi(s))\right)_x' g(\tau, x^\varepsilon(\tau), \xi^\varepsilon(\tau))ds\\
&\to \int_t^{t+s} d\tau \int_0^{\infty} \mathbb{E}\left(f_x(\tau, x^\varepsilon(\tau))' g(\tau, x^\varepsilon(\tau), \xi(s))\right)_x' g(\tau, x(\tau), \xi(0))ds.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}h(x^\varepsilon(t_i), (\varphi_j, \hat{m}^\varepsilon)_{t_i}, 1 \le q_1, j \le q_2)\\
&\quad\times \left[f(t+s, x(t+s)) - f(t, x(t)) - \int_t^{t+s} \int_{\mathcal{U}} \mathcal{L}_\tau^\alpha f(\tau, x(\tau)) m_\tau(d\alpha)d\tau\right] = 0,
\end{aligned}$$

and thus $M_f(\cdot)$ is a martingale. Note that, in the above arguments, we chose $t, t_i, t+s$ in the set of points of continuity of the limit process x. This is because we only know that $x(\cdot)$ has paths in $\mathbb{D}^r[0,T]$ but have not yet proved that the path are in $C^r[0,T]$. Since the set of points of discontinuity of x is at most countable, due to Theorem 7.8 and Theorem 8.10 in [6] it suffices to choose $t, t_i, t+s$ being the points of continuity of x.

Step 3. Representation of the limit. Since $M_f(\cdot)$ are martingales with respect to $\mathcal{G}_t$, there is a standard Brownian motion $w(\cdot)$ such that $w(t)$ is adapted to $\mathcal{G}_t$, $x(t)$ is nonanticipative with respect to $w(\cdot)$ and

$$dx(t) = \int_{\mathcal{U}} \bar{b}(t, x(t), \alpha)\hat{m}_t(d\alpha)dt + \bar{\sigma}(t, x(t))dw(t).$$

If the probability space is not rich enough (e.g., the case a is degenerate), we can augment it by adding an independent Brownian motion, see Theorem 4.5.2 in [45] for a detail discussion. Moreover, because $w(t)$ is $\mathcal{G}_t-$adapted, $\hat{m}(A \times [0,t])$ and $\hat{m}_t(A)$ are nonanticipative with respect to w. Thus $\hat{m}(\cdot)$ is an admissible relaxed control for the problem. The convergence of sequence of cost functions follows from the weak convergence of $(x^\varepsilon(\cdot), m^\varepsilon(\cdot)) \Rightarrow (x(\cdot), \hat{m}(\cdot))$, and the continuity of the process $x(\cdot)$ $\square$

Remark 4.10. Repeat the arguments of the above theorem, we have the following remark. Let $u(\cdot, x(\cdot))$ be a time-dependent feedback control which is continuous in x, uniformly in t on each bounded (x,t) set, and for which the martingale problem associated with (4.7) has a unique solution. Then $x^\varepsilon(u(\cdot, x^\varepsilon(\cdot)), \cdot) \Rightarrow x(u(\cdot, x(\cdot)), \cdot)$ and $J^\varepsilon(u) \to J(u)$. This is a useful observation and the key to obtain the nearly optimal controls or equilibriums.

4.5 Approximation of Optimal Controls and Equilibrium Controls for $x^\varepsilon(\cdot)$

In this subsection we show the following theorems which explain how we get the "nearly" desired controls for the original system from those one of the limit system.

Theorem 4.11. Assume (**A**) and let $\delta > 0$ be given. For any Lipschitz continuous (uniformly in t) δ-optimal feedback control $\bar{u}^\delta(\cdot)$ for the problem (**LCP**) (which always exists thanks to Proposition 4.3.2), we have

$$\limsup_{\varepsilon \to 0} \left[J^\varepsilon(\bar{u}^\delta) - \inf_{m \in \mathcal{R}^\varepsilon} J^\varepsilon(m) \right] \leq \delta.$$

Proof. Apply the weak convergence argument of Theorem 4.9 for the fixed control $\bar{u}^\delta(\cdot, x)$ we have $x^\varepsilon(\bar{u}^\delta(\cdot, x^\varepsilon(\cdot)), \cdot) \to x(\bar{u}^\delta(\cdot, x(\cdot)), \cdot)$ and $J^\varepsilon(\bar{u}^\delta) \to J(\bar{u}^\delta)$. Let $\delta_\varepsilon \to 0$ and let $\hat{m}^\varepsilon$ be a

δ_ε-optimal admissible relaxed control for the process (4.1) then by Theorem 4.9 we have

$$J^\varepsilon(\hat{m}^\varepsilon) \to J(\hat{m}) \geq \inf_{m \in \mathcal{R}^0} J(m) \geq J(\bar{u}^\delta) - \delta.$$

Combining the two above observations, the definition of $\hat{m}_\varepsilon$ and the following inequality

$$\begin{aligned}
\limsup_{\varepsilon \to 0} \Big[J^\varepsilon(\bar{u}^\delta) - \inf_{m \in \mathcal{R}^\varepsilon} J^\varepsilon(m) \Big] \leq & \limsup_{\varepsilon \to 0} \Big[J^\varepsilon(\bar{u}^\delta) - J(\bar{u}^\delta) \Big] \\
& + \limsup_{\varepsilon \to 0} \Big[J(\bar{u}^\delta) - J^\varepsilon(\hat{m}^\varepsilon) \Big] \\
& + \limsup_{\varepsilon \to 0} \Big[J^\varepsilon(\hat{m}^\varepsilon) - \inf_{m \in \mathcal{R}^\varepsilon} J^\varepsilon(m) \Big]
\end{aligned}$$

then we have the desired result. □

Theorem 4.12. Assume (**A**). Let $\delta > 0$ and $\hat{u}^\delta(\cdot)$ a δ-Nash Equilibrium feedback strategies for the problem (**LGP**) with corresponding state processes $\hat{x}(\cdot)$. Denote $\hat{x}^\varepsilon(\cdot)$ the corresponding state processes obtained from the problem ($\mathbf{LGP}_\varepsilon$) using the strategies $\hat{u}^\delta(\cdot)$). Assume that $\hat{u}^\delta(\cdot)$ is Lipschitz continuous. Then we have $\hat{x}^\varepsilon(\cdot)$ converges weakly to $\hat{x}(\cdot)$ and

$$\lim_{\varepsilon \to 0} \big[J^{i,\varepsilon}(\hat{u}^\delta) - J^i(\hat{u}^\delta) \big] = 0, \quad i = 1, 2, \ldots, N. \tag{4.26}$$

Moreover, for any feedback strategy $v^i(\cdot)$, $i = 1, 2, \ldots, N$, we have

$$\limsup_{\varepsilon \to 0} \Big[J^{i,\varepsilon}(\hat{u}^{i,\delta}, \hat{u}^{-i,\delta}) - J^{i,\varepsilon}(v^i, \hat{u}^{-i,\delta}) \Big] \leq \delta. \tag{4.27}$$

Proof. The weak convergence of $\hat{x}^\varepsilon(\cdot)$ and (4.26) follow from Theorem 4.9. In order to prove

(4.27) we have

$$\begin{aligned}
&\limsup_{\varepsilon\to 0}\left[J^{i,\varepsilon}(\hat{u}^{i,\delta},\hat{u}^{-i,\delta}) - J^{i,\varepsilon}(v^i,\hat{u}^{-i,\delta})\right] \\
&\quad \le \limsup_{\varepsilon\to 0}\left[J^{i,\varepsilon}(\hat{u}^{i,\delta},\hat{u}^{-i,\delta}) - J^{i}(\hat{u}^{i,\delta},\hat{u}^{-i,\delta})\right] \\
&\qquad + \left[J^{1}(\hat{u}^{i,\delta},\hat{u}^{-i,\delta}) - J^{1}(v^i,\hat{u}^{-i,\delta})\right] \\
&\qquad + \limsup_{\varepsilon\to 0}\left[J^{i}(v^i,\hat{u}^{-i,\delta}) - J^{i,\varepsilon}(v^i,\hat{u}^{-i,\delta})\right].
\end{aligned}$$

It follows from Remark 4.10 that the lim sup in the first and the third terms are zero. By the assumption on $\hat{u}^{\delta}(\cdot)$ the second term is upper bounded by δ. The inequality (4.27) therefore holds true. □

4.6 Linear Quadratic Control and Game with Wide Bandwidth Noise

In this section, we consider a linear quadratic optimal control problem and a linear quadratic game problem with N players involving a Markov switching and wideband noises. The state of each player is again a non-Markovian process so the usual stochastic control techniques do not work. It should be pointed out that "linear" is meant to be linear in the continuous state variable. In our formulation, we also have a continuous-time Markov chain in the original problem. Thus the problem is not really linear. Nevertheless, for simplifying the discussion, we call the problems linear rather than using the phrase "linear in continuous state variable" in each appearance. We use the idea proposed in [49] to obtain a "nearly desired" solution. Let us briefly describe this idea in the linear quadratic optimal control problem. As we have seen in the previous section, when $\varepsilon \to 0$, the optimal control problem with wideband noise will "converge" to the corresponding problem of diffusion. By the standard result in linear quadratic optimal control of diffusion processes, we know that the optimal control is of the linear feedback form $u(t) = \Sigma(t)x(t)$. Since the control problem with switching wideband noise is complicated, one possible thing we can do is to see how good the system performs under the control $u^\varepsilon(t) = \Sigma(t)x^\varepsilon(t)$. It turns out that with this control both the controlled process and the corresponding cost function converge to those of diffusion. Similarly, in the game problem, by considering the limit problem (as $\varepsilon \to 0$) and using the standard results of mean-field game theory, we propose a set of feedback strategies that are obtained from the form of the δ-Nash equilibriums of the mean-field problem, such that each player only needs to use the information of its own state. When $\varepsilon \to 0$, both the dynamic of players and their costs approach to those of the mean field type control limiting problem.

4.6.1 Linear Quadratic Optimal Control

In this subsection we study a linear quadratic optimal control problem which is a special case of the problem ($\mathbf{CP}_\varepsilon$) but without the assumption on the compactness of $\mathcal{U}$. Because of the advantage of the linear structure, we can still obtain results similar in spirit to those given in Section 4. Let $\beta(t)$ is a Markov chain on the state space $\mathcal{M} = \{1, \ldots, m_0\}$ with the generator Q and the invariant measure $\nu = (\nu^1, \nu^2, \ldots, \nu^{m_0})$. Let $A(\cdot,\cdot), B(\cdot,\cdot) : [0,T] \times \mathcal{M} \to \mathbb{R}^{r\times r}$ and $D, Q, R : [0,T] \to \mathbb{R}^r$ be bounded deterministic matrix-valued functions such that Q, R, G are positive definite. Assume that for fixed $i_0 \in \mathcal{M}$, $A(\cdot, i_0)$ and $B(\cdot, i_0)$ are continuous and that $D(t) = \text{diag}\big(D^1(t), \ldots, \mathbb{D}^r(t)\big)$. The noise $\xi(t) = (\xi^1(t), \xi^2(t), \ldots, \xi^r(t))' \in \mathbb{R}^r$ and $\{\xi^i(t), 1 \le i \le r\}$ are assumed to be independent identically distributed mixing second order wide-sense stationary with mixing rate $\phi(t)$ satisfying the Assumption (A1).

Linear quadratic control problem ($\mathbf{LQCP}_\varepsilon$). Minimize

$$J^\varepsilon(u^\varepsilon) = \mathbb{E}\left[\int_0^T [x^\varepsilon(t)' Q(t) x^\varepsilon(t) + u^\varepsilon(t)' R(t) u^\varepsilon(t)]dt + x^\varepsilon(T)' G x^\varepsilon(T)\right].$$

where $x^\varepsilon(\cdot)$ satisfies

$$dx^\varepsilon(t) = \left[A(t,\beta^\varepsilon(t))x^\varepsilon(t) + B(t,\beta^\varepsilon(t))u^\varepsilon(t) + D(t)\frac{\xi^\varepsilon(t)}{\varepsilon}\right]dt,$$

$$x^\varepsilon(0) = x_0,$$

where $x_0 \in \mathbb{R}^r$. As usual $u^\varepsilon(\cdot)$ is the control process and belongs to $\mathcal{U}^\varepsilon$, the set of all processes $u : [0,T]\times\Omega \longrightarrow \mathbb{R}^r$ such that $\mathbb{E}\int_0^T |u^\varepsilon(t)|^2 dt < \infty$ and $\mathcal{F}_t^\varepsilon-$adapted where $\mathcal{F}_t^\varepsilon$ is defined in (4.8).

Guided by what we have done in Section 4.4, we associate with ($\mathbf{LQCP}_\varepsilon$) a *formal* limit problem. Denote

$$\bar{A}(t) = \sum_{i_0=1}^{m_0} A(t,i_0)\nu^{i_0}, \quad \bar{B}(t) = \sum_{i_0=1}^{m_0} B(t,i_0)\nu^{i_0}, \tag{4.28}$$

and $a(\cdot) = \left(a^{i,j}(\cdot)\right)_{r\times r} : [0,T] \to \mathbb{R}^{r\times r}$ where

$$\begin{aligned} \frac{1}{2}a^{i,j}(t) &= \mathbb{E}\left[\int_{-\infty}^{\infty} g^i(\xi(s))g^j(\xi(0))ds\right] \\ &= \mathbb{E}\left[\int_{-\infty}^{\infty} D^i(s)D^j(0)\xi^i(s)\xi^j(0)ds\right], \quad 1 \le i,j \le r. \end{aligned} \tag{4.29}$$

Here, we use $g^i(t,x,\xi) = D^i(t)\xi^i(t)$ as in the Assumption (A3). Similar to (A3), assume that $\bar{a}(\cdot)$ defined by $\bar{a}(t) = \frac{1}{2}\left[a(t) + a'(t)\right]$ is positive definite an denote $\bar{\sigma}(\cdot)$ the Lipschitz square root of the symmetric matrix $\bar{a}(\cdot)$. Now we can state the limit linear quadratic control problem.

Limit linear quadratic control problem ($\mathbf{LQCP}_0$). Minimize

$$J(x,u) = \mathbb{E}\left[\int_0^T [x(t)' Q(t) x(t) + u(t)' R(t) u(t)]dt + x(T)' G x(T)\right],$$

where $x(\cdot)$ satisfies

$$dx(t) = \left[\bar{A}(t)x(t) + \bar{B}(t)u(t)\right] dt + \bar{\sigma}(t)dw(t),$$

$$x(0) = x_0,$$

where $x_0 \in \mathbb{R}^n$, $w(\cdot)$ is a standard Brownian motion in $\mathbb{R}^r$, and the control process $u(\cdot)$ belongs to $\mathcal{U}^0$, the set of all processes $u : [0,T] \times \Omega \longrightarrow \mathbb{R}^r$ such that $\mathbb{E}\int_0^T |u^\varepsilon(t)|^2 dt < \infty$ and $\mathcal{F}_t^0-$adapted where $\mathcal{F}_t^0 = \sigma(w(s) : 0 \le s \le t)$.

Note that the problem ($\mathbf{LQCP}_0$) is a standard linear quadratic optimal control with diffusion process. It is well known that the solution to ($\mathbf{LQCP}_0$) is a feedback control of the form $\bar{u}(t) = \Sigma(t)\bar{x}(t)$, where $\Sigma(\cdot)$ is the bounded continuous matrix-valued function that can be obtained by solving a Riccati equation associated with ($\mathbf{LQCP}_0$). Now, we define $\bar{u}^\varepsilon(t) = \Sigma(t)\bar{x}^\varepsilon(t)$ for $t \in [0,T]$. Plug this particular control $\bar{u}^\varepsilon(t)$ into ($\mathbf{LQCP}_\varepsilon$), then the corresponding state equation and cost function can be rewritten as:

$$d\bar{x}^\varepsilon(t) = \left[\Big(A(t,\beta^\varepsilon(t)) + B(t,\beta^\varepsilon(t))\Sigma(t)\Big)\bar{x}^\varepsilon(t) + D(t)\frac{\xi^\varepsilon(t)}{\varepsilon}\right] dt \qquad (4.30)$$

and

$$J^\varepsilon(\bar{u}^\varepsilon) = \mathbb{E}\left[\int_0^T \bar{x}^\varepsilon(t)' \left(Q(t) + \Sigma(t)'R(t)\Sigma(t)\right) \bar{x}^\varepsilon(t)dt + \bar{x}^\varepsilon(T)'G\bar{x}^\varepsilon(T)\right].$$

We now aim to prove that when $\varepsilon \to 0$, $\bar{x}^\varepsilon(\cdot)$ converges weakly to $\bar{x}(\cdot)$ and $J^\varepsilon(\bar{u}^\varepsilon)$ converges to $J(\bar{u})$ where $\bar{u}(t) = \Sigma(t)\bar{x}(t)$ and $\bar{x}(\cdot)$ satisfies

$$\begin{aligned} d\bar{x}(t) &= \left[\bar{A}(t) + \bar{B}(t)\Sigma(t)\right] \bar{x}(t)dt + \bar{\sigma}(t)dw(t), \\ \bar{x}(0) &= x_0 \end{aligned} \qquad (4.31)$$

and

$$J(\bar{u}) = \mathbb{E}\left[\int_0^T \bar{x}(t)' \left(Q(t) + \Sigma(t)'R(t)\Sigma(t)\right) \bar{x}(t)dt + \bar{x}(T)'G\bar{x}(T)\right]. \qquad (4.32)$$

We emphasize that we do not prove that $(\mathbf{LQCP}_\varepsilon)$ "converges" to $(\mathbf{LQCP}_0)$ in the sense of section 4.4; we only prove $\bar{x}^\varepsilon(\cdot) \Rightarrow \bar{x}(\cdot)$ and $J^\varepsilon(\bar{u}^\varepsilon) \to J(\bar{u})$. That is why we called $(\mathbf{LQCP}_0)$ the *formal* limit of $(\mathbf{LQCP}_\varepsilon)$. We need the following auxiliary lemma to obtain the tightness of x^ε.

Lemma 4.13. The following assertions hold.

(i) For the system given by (4.30),

$$\sup_{0\le t\le T} \mathbb{E}|\bar{x}^\varepsilon(t)|^2 \le K < \infty.$$

(ii) For any given $\varepsilon > 0, t, s \ge 0$, we have

$$\mathbb{E}|\bar{x}^\varepsilon(t+s) - \bar{x}^\varepsilon(t)|^2 = O(s). \tag{4.33}$$

Proof. To prove (i), we first note that in this proof, K is a general constant that its value may change from line to line and depends only on T and the continuity and hence the boundness of $A(t,i_0)$, $B(t,i_0)$, $\Sigma(t)$, and $D(t)$. Because of the boundedness of $A(\cdot,\cdot), B(\cdot,\cdot)$ and $D(\cdot)$, we have

$$|\bar{x}^\varepsilon(t)|^2 \le K\left[\left|\int_0^t \bar{x}^\varepsilon(s)ds\right|^2 + \left|\int_0^t \frac{\xi^\varepsilon(s)}{\varepsilon}ds\right|^2\right],$$

so by the Hölder inequality,

$$\mathbb{E}|\bar{x}^\varepsilon(t)|^2 \le KT\int_0^t |\bar{x}^\varepsilon(s)|^2 ds + K\varepsilon^2 \int_0^{t/\varepsilon^2}\int_0^{t/\varepsilon^2} \mathbb{E}|\xi(s)'\xi(\rho)|ds d\rho. \tag{4.34}$$

Using the usual mixing inequality, an application of the Gronwall's inequality leads to $\mathbb{E}|\bar{x}^\varepsilon(t)|^2 < \infty$. Taking sup over $0 \le t \le T$ yields the desired result.

For (ii), the proof can be carried by doing similar calculations as in the previous part.

□

Remark 4.14. Note that in addition to Lemma 4.13, we can also obtain the conditional second moment estimate. That is, $\sup_{0\le t\le T}\mathbb{E}_s|\bar{x}^\varepsilon(t)|^2<\infty$ for $s<t$, where $\mathbb{E}_s$ denotes the conditional expectation w.r.t. the σ-algebra generated by $\{x_0,\xi^\varepsilon(\rho),\beta^\varepsilon(\rho):\rho\le s\}$.

With the above notation and assumptions, we have the following result.

Proposition 4.15. The sequence $\{\bar{x}^\varepsilon(\cdot)\}$ defined by (4.30) is tight in $\mathbb{D}^r[0,\infty)$. As $\varepsilon\to 0$, $\bar{x}^\varepsilon(\cdot)$ converges weakly to $\bar{x}(\cdot)$ where $\bar{x}(\cdot)$ is a the solution of the stochastic differential equation

$$d\bar{x}(t)=\left[\bar{A}(t)+\bar{B}(t)\Sigma(t)\right]\bar{x}(t)dt+\bar{\sigma}(t)dw(t).$$

In addition,

$$J^\varepsilon(\bar{u}^\varepsilon)\to J(\bar{u}).$$

Proof. The proof of this proposition is quite similar to that of Theorem 4.9 though we do not use relaxed control here so we only sketch it. Due to the linear structure of the problem at hand, we do not need to utilize the truncation method.

Step 1. Tightness of $\{x^\varepsilon(\cdot)\}$. The tightness of $\bar{x}^\varepsilon$ follows from Lemma 4.13 and Theorem 3, page 47 in [25].

Step 2. The martingale problem satisfied by the limit. For each $f\in C_0^{2,3}([0,T]\times\mathbb{R}^r)$, we set $f^\varepsilon(t)=f(t,\bar{x}^\varepsilon(t))+f_1^\varepsilon(t)+f_2^\varepsilon(t)$, where

$$\begin{aligned}
f_1^\varepsilon(t)=f_1^\varepsilon(t,\bar{x}^\varepsilon(t))&=\frac{1}{\varepsilon}\int_t^T f_x(t,x)'D(t)\mathbb{E}_t^\varepsilon\xi^\varepsilon(s)ds\\
f_2^\varepsilon(t)=f_2^\varepsilon(t,\bar{x}^\varepsilon(t))&=\int_t^T(\mathbb{E}_t^\varepsilon\hat{F}_\varepsilon(\tau,x)-\mathbb{E}\hat{F}_\varepsilon(\tau,x))d\tau
\end{aligned}$$

and

$$\hat{F}_\varepsilon(t,x) = \int_{t/\varepsilon^2}^{T/\varepsilon^2} \left[\mathbb{E}_t^\varepsilon \left(f_x(t,x)' D(t)\xi(s)\right)_x ds\right]' D(t)\xi^\varepsilon(t).$$

Similar to what has been done in Theorem 4.9 (except that we do not have the integral over the space of relax control) we can prove that for each $f \in C_0^{2,3}([0,T] \times \mathbb{R}^r)$, f^ε satisfies the requirements of Lemma 4.8 and thus establish the convergence of $\bar{x}^\varepsilon$.

Step 3. Representation of the limit. This step can be done by a similar way to that of Theorem 4.9. For the last claim, though the functions h, k that appear in the cost function are not bounded, due to the linear structure of the cost function and the uniformly boundedness of second moment of $\bar{x}^\varepsilon$, the convergence of $J^\varepsilon(\bar{u}^\varepsilon)$ to $J(\bar{u})$ can be proved by using the localization and the Cantor diagonal argument. □

Remark 4.16. Let $u^\varepsilon(t) = \Sigma(t, \beta^\varepsilon(t))x^\varepsilon(t), t \in [0,T]$ be a family of controls, where $\Sigma(t,i)$ is a continuous matrix-valued function for each $i \in \mathcal{M}$. It is easy to see from the previous proposition that when plugging these controls in to ($\mathbf{LQCP}_\varepsilon$), we also obtain the same conclusion with Σ in (4.31) is replace by $\bar{\Sigma}$ where

$$\bar{\Sigma}(t) = \sum_{i_0=1}^{m_0} \Sigma(t, i_0)\nu^{i_0}.$$

4.6.2 Linear Quadratic Game Problem

We consider in this subsection a linear quadratic game problem with N players. Let $A(\cdot,\cdot), B(\cdot,\cdot) : [0,T] \times \mathcal{M} \to \mathbb{R}$ and $D, Q, \bar{Q}, R : [0,T] \to \mathbb{R}$ be bounded deterministic real-valued functions such that $Q, R, G > 0$. Assume that for fixed $i_0 \in \mathcal{M}$, $A(\cdot, i_0)$ and $B(\cdot, i_0)$ are continuous and that the noises $\xi^1(t), \xi^2(t), \ldots, \xi^N(t)$ are independent identically distributed mixing second order wide-sense stationary with mixing rate $\phi(t)$ satisfying the

Assumption (A1). Denote $\xi(t) = \big(\xi^1(t), \xi^2(t), \ldots, \xi^N(t)\big)'$ and $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$ for $\varepsilon > 0, 0 \leq t \leq T$.

A linear quadratic game problem (LQGP$_\varepsilon$). We consider a class of $N-$person stochastic differential games where the individual dynamic of player i, $x^{i,\varepsilon}$ satisfies the following stochastic differential equation with wideband noise

$$\begin{aligned}\frac{dx^{i,\varepsilon}(t)}{dt} &= \big[A(t,\beta^\varepsilon(t))x^{i,\varepsilon}(t) + B(t,\beta^\varepsilon(t))u^{i,\varepsilon}(t) \\ &\qquad + F(t,\beta^\varepsilon(t))x^{(N),\varepsilon}(t) + \frac{D(t)}{\varepsilon}\xi^{i,\varepsilon}(t)\bigg],\end{aligned} \tag{4.35}$$

with initial conditions $x^{i,\varepsilon}(0) = x_0^i \in \mathbb{R}$ for $1 \leq i \leq N$, where the control process $u^{i,\varepsilon}(\cdot)$ belongs to $\mathcal{U}^{i,\varepsilon}$, the set of all processes $u^{i,\varepsilon} : [0,T] \times \Omega \longrightarrow \mathbb{R}$ such that $\mathbb{E}\int_0^T |u^{i,\varepsilon}(t)|^2 dt < \infty$ and $\mathcal{F}_t^\varepsilon-$adapted where $\mathcal{F}_t^\varepsilon$ is defined as in (4.8), and the term

$$x^{(N),\varepsilon}(t) = \frac{1}{N}\sum_{1\leq j\leq N} x^{j,\varepsilon}(t)$$

is the mean-field coupling term. The cost function for player i is given by

$$\begin{aligned}J^{i,\varepsilon}(u^\varepsilon) = \mathbb{E}\Big[&\int_0^T \big(x^{i,\varepsilon}(t)Q(t)x^{i,\varepsilon}(t) + u^{i,\varepsilon}(t)R(t)u^{i,\varepsilon}(t)\big)dt \\ &+ x^{i,\varepsilon}(T)Q(T)x^{i,\varepsilon}(T)\Big] \\ &+ \mathbb{E}\left[\int_0^T \big(x^{i,\varepsilon}(t) - S(t)x^{(N),\varepsilon}(t)\big)\bar{Q}(t)\big(x^{i,\varepsilon}(t) - S(t)x^{(N),\varepsilon}(t)\Big)dt\right] \\ &+ \mathbb{E}\left[\big(x^{i,\varepsilon}(T) - S(T)x^{(N),\varepsilon}(T)\big)\bar{Q}(T)\big(x^{i,\varepsilon}(T) - S(T)x^{(N),\varepsilon}(T)\big)\right].\end{aligned} \tag{4.36}$$

We note that, the system of stochastic differential equations describing the dynamic of players can also be rewritten as

$$dx^\varepsilon(t) = \big[\mathbb{A}(t,\beta^\varepsilon(t))x^\varepsilon(t)dt + \mathbb{B}(t,\beta^\varepsilon(t))u^\varepsilon(t) + \frac{1}{\varepsilon}\mathbb{D}(t)\xi^\varepsilon(t)\big]dt,$$

where for $t \geq 0$ and $i_0 \in \mathcal{M}$,

$$\mathbb{A}(t, i_0) = \begin{bmatrix} A(t,i_0) + \frac{F(t,i_0)}{N} & \frac{F(t,i_0)}{N} & \cdots & \frac{F(t,i_0)}{N} \\ \frac{F(t,i_0)}{N} & A(t,i_0) + \frac{F(t,i_0)}{N} & \cdots & \frac{F(t,i_0)}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{F(t,i_0)}{N} & \frac{F(t,i_0)}{N} & \cdots & A(t,i_0) + \frac{F(t,i_0)}{N} \end{bmatrix},$$

$$\mathbb{B}(t, i_0) = \begin{bmatrix} B(t,i_0) & 0 & \cdots & 0 & 0 \\ 0 & B(t,i_0) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & B(t,i_0) \end{bmatrix},$$

$$\mathbb{D}(t) = \begin{bmatrix} D(t) & 0 & \cdots & 0 & 0 \\ 0 & D(t) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & D(t) \end{bmatrix}.$$

So the states of N players describing in (4.35) can be written in a same way as the dynamic of the linear quadratic control problem ($\mathbf{LQCP}_\varepsilon$). Using a similar idea that was used the previous subsection, we associate ($\mathbf{LQGP}_\varepsilon$) with the following *formal* limit problem.

Limit linear quadratic game problem ($\mathbf{LQGP}_0$). Similar to previous subsection, let $\bar{A}(t), \bar{B}(t)$ be determined as in (4.28) and $\bar{F}(t) = \sum_{i_0=1}^{m_0} F(t, i_0)\nu^{i_0}$. Denote $\bar{a}(t) = \mathbb{E}\int_{-\infty}^{\infty} D(s)D(0)\xi^i(s)\xi^i(0)ds$. Assume that $\bar{a}(t) \geq 0$ for $t \in [0, T]$ and let $\bar{\sigma}(t)$ be any Lipschitz square root of $\bar{a}(t)$.

Let the dynamic of player i be described by

$$\frac{dx^i(t)}{dt} = \left[\bar{A}(t)x^i(t) + \bar{B}(t)u^i(t) + \bar{F}(t)x^{(N)}(t)\right] + \bar{\sigma}dw(t),$$

where $x^{(N)}(t) = \frac{1}{N}\sum_{j=1}^{N} x^j(t)$ and $w(\cdot)$ is a standard Brownian motion in $\mathbb{R}^1$. The cost function for player i is given by

$$\begin{aligned} J^i(u) = \mathbb{E}\left[\int_0^T \big(x^i(t)Q(t)x^i(t) + u^i(t)R(t)u^i(t)\big)\,dt + x^i(T)Q(T)x^i(T)\right] \\ + \mathbb{E}\left[\int_0^T \big(x^i(t) - S(t)x^{(N)}(t)\big)\bar{Q}(t)\big(x^i(t) - S(t)x^{(N)}(t)\big)dt\right] \\ + \mathbb{E}\left[\big(x^i(T) - S(T)x^{(N)}(T)\big)\bar{Q}(T)\big(x^i(T) - S(T)x^{(N)}(T)\big)\right]. \end{aligned} \tag{4.37}$$

It should be noted that ($\mathbf{LQGP}_0$) is a well-known LQG mean-field game problem. Under some mild conditions (see [13] or [4]), there exists a δ-Nash equilibrium for this problem where $\delta = O(1/\sqrt{N})$, in which the strategies have the following feedback forms $\hat{u}^i = \Psi(t)\hat{x}^i(t)+\Lambda(t)$ where $\hat{x}^i(t)$ is the corresponding state, $\Psi(t)$ and $\Lambda(t)$ are continuous real-valued functions which can be determined from the coefficients $\bar{A}, \bar{B}, \bar{\sigma}, Q, \bar{Q}, R, S$ (see for instant, equation (9.33) in [13] or Remark 3.3 in [4]).

Consider the problem ($\mathbf{LQGP}_\varepsilon$), let $\hat{u}^\varepsilon = (\hat{u}^{1,\varepsilon}, \hat{u}^{2,\varepsilon}, \ldots, \hat{u}^{N,\varepsilon})$ be the set of strategies of the feedback form

$$\hat{u}^{i,\varepsilon}(t) = \Psi(t)\hat{x}^{i,\varepsilon}(t) + \Lambda(t), \quad i = 1, 2, \ldots, N,$$

where $\hat{x}^{i,\varepsilon}(t)$ is the corresponding state of player i. Plugging this particular control into problem ($\mathbf{LQGP}_\varepsilon$), the state equation and cost function become

$$d\hat{x}^\varepsilon(t) = \left[\big(\mathbb{A}(t, \beta^\varepsilon(t)) + \mathbb{B}(t, \beta^\varepsilon(t))\Psi(t)\big)\hat{x}^\varepsilon(t) + \Lambda(t)I_{N\times N} + \frac{1}{\varepsilon}\mathbb{D}(t)\xi^\varepsilon(t)\right]dt, \tag{4.38}$$

where $I_{N\times N}$ denotes the $N\times N$ identity matrix, and

$$\begin{aligned}J^{i,\varepsilon}(\hat{u}^{\varepsilon}) = \mathbb{E}&\left[\int_0^T \hat{x}^{i,\varepsilon}(t)\left(Q(t)+\Psi(t)R(t)\Psi(t)\right)\hat{x}^{i,\varepsilon}(t)+\Lambda(t)R(t)\Lambda(t)dt\right]\\ &+\mathbb{E}\left[\int_0^T \big(\hat{x}^{i,\varepsilon}(t)-S(t)\hat{x}^{(N),\varepsilon}(t)\big)\bar{Q}(t)\big(\hat{x}^{i,\varepsilon}(t)-S(t)\hat{x}^{(N),\varepsilon}(t)\big)dt\right]\\ &+\mathbb{E}\left[\big(\hat{x}^{i,\varepsilon}(T)-S(T)\hat{x}^{(N),\varepsilon}(T)\big)\bar{Q}(T)\big(\hat{x}^{i,\varepsilon}(T)-S(T)\hat{x}^{-i,\varepsilon}(T)\big)\right]\\ &+\mathbb{E}\hat{x}^{i,\varepsilon}(T)Q(T)\hat{x}^{i,\varepsilon}(T).\end{aligned}\tag{4.39}$$

Similar to subsection 4.6.1, we have that the $\hat{x}^{\varepsilon}(\cdot)$ converges weakly to $\hat{x}(\cdot)$ and $J^{i,\varepsilon}(\hat{u}^{\varepsilon})\to J^{i}(\hat{u})$, for $i=1,2,\dots,N$, where $\hat{x}^{\varepsilon}(\cdot)$ is the solution of (4.38), $J^{i,\varepsilon}(\hat{u}^{\varepsilon})$ is given by (4.39) and $\hat{x}(\cdot)$ is given as below

$$d\hat{x}(t)=\Big[\big(\bar{\mathbb{A}}(t)+\bar{\mathbb{B}}(t)\Psi(t)\big)\hat{x}(t)+\Lambda(t)I_{N\times N}\Big]dt+\bar{\sigma}(t)I_{N\times N}dw(t),\tag{4.40}$$

where the feedback strategy $\hat{u}(t)=\Psi(t)\hat{x}(t)+\Lambda(t)$ is used, and $\bar{\mathbb{A}}$ and $\bar{\mathbb{B}}$ are defined by

$$\bar{\mathbb{A}}(t)=\begin{bmatrix}\bar{A}(t)+\frac{1}{N}\bar{F}(t) & \frac{1}{N}\bar{F}(t) & \cdots & \frac{1}{N}\bar{F}(t)\\ \frac{1}{N}\bar{F}(t) & \bar{A}(t)+\frac{1}{N}\bar{F}(t) & \cdots & \frac{1}{N}\bar{F}(t)\\ \vdots & \vdots & \ddots & \vdots\\ \frac{1}{N}\bar{F}(t) & \frac{1}{N}\bar{F}(t) & \cdots & \bar{A}(t)+\frac{1}{N}\bar{F}(t)\end{bmatrix},$$

$$\bar{\mathbb{B}}(t)=\begin{bmatrix}\bar{B}(t) & 0 & \cdots & 0 & 0\\ 0 & \bar{B}(t) & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & \bar{B}(t)\end{bmatrix}.$$

The cost function $J^i(\hat{u})$ is given by

$$\begin{aligned} J^i(\hat{u}) &= \mathbb{E}\left[\int_0^T \hat{x}^i(t)\big(Q(t)+\Psi(t)R(t)\Psi(t)\big)\hat{x}^i(t)+\Lambda(t)R(t)\Lambda(t)dt\right] \\ &\quad+\mathbb{E}\left[\int_0^T \big(\hat{x}^i(t)-S(t)\hat{x}^{(N)}(t)\big)\bar{Q}(t)\big(\hat{x}^i(t)-S(t)\hat{x}^{(N)}(t)\big)dt\right] \\ &\quad+\mathbb{E}\left[\big(\hat{x}^i(T)-S(T)\hat{x}^{(N)}(T)\big)\bar{Q}(T)\big(\hat{x}^i(T)-S(T)\hat{x}^{(N)}(T)\big)\right] \\ &\quad+\mathbb{E}\hat{x}^i(T)Q(T)\hat{x}^i(T). \end{aligned} \tag{4.41}$$

We summarize what have been discussing so far into following result that can be proved similar to that of Proposition 4.15.

Proposition 4.17. Assume that the problem ($\mathbf{LQGP}_0$) has a $\delta-$Nash equilibrium $\hat{u}$ in the feedback form $\hat{u}^i=\Psi(t)\hat{x}^i(t)+\Lambda(t)$, $i=1,2,\dots,N$ and $\Psi(\cdot),\Lambda(\cdot)$ are continuous functions. Let $\hat{x}^\varepsilon(\cdot)$ and $\hat{x}(\cdot)$ respectively be the solutions of (4.38) and (4.40). Let $J^{i,\varepsilon}(\hat{u}^\varepsilon)$ and $J^i(\hat{u})$ be defined as in (4.39) and (4.41) respectively for $i=1,2,\dots,N$. Then, as $\varepsilon\to 0$, we have

$$\hat{x}^\varepsilon(\cdot)\Rightarrow\hat{x}(\cdot) \text{ and } J^{i,\varepsilon}(\hat{u}^\varepsilon)\to J^i(\hat{u}).$$

Remark 4.18. Notice that we can also consider the problem where the state space of each player and the control process take value in multidimensional Euclidean spaces, however, the results are essentially the same.

4.7 Further Remarks

This chapter developed near-optimal controls of hybrid systems under wideband noise perturbations. The original problems are difficult to solve because they are non-Markovian. There are no techniques readily available to treat such systems. To overcome the difficulties, we dealt with the problem from a different angle. It is shown that the underlying problems are reduced to certain limit Markovian systems under suitable scaling. Using optimal or near-optimal controls of the limit systems, we build controls for the original problems and show that such a scheme leads to desired near optimality. Several directions may be worthwhile for future investigation. Currently, for the linear (linear in the continuous state variable) problem, the control weights are assumed to be positive definite. It would be interesting to extend the current setting by considering indefinite control weights. The key idea is to use backward stochastic differential equations. Another problem is to treat Markov chains that involve multiple ergodic classes. These questions deserve further thoughts and careful consideration.

CHAPTER 5 CONCLUDING REMARKS AND FUTURE DIRECTIONS

In this dissertation, we concentrate on numerical methods, limit results, and controlled stochastic differential systems with random switching. In the first part, we study the numerical approximation to those systems. We designed a new numerical scheme in the spirit of the classical Milstein scheme to SDEs without switching and use a new approach to thoroughly investigate the convergence of the proposed scheme. Inspired by the numerical approximation procedures, we next investigate a somewhat generalized and abstracted limit theorem for stochastic systems with state-independent regime switching with quite general driving processes. The last part of the dissertation is devoted to the controlled hybrid systems that are good approximations to controlled switching diffusion processes. Motivated by applications, we study the switching systems perturbed by wide bandwidth noise and design the nearly optimal and nearly equilibrium controls for those systems. Although the dissertation is mainly concerned with quite general stochastic systems and does not focus on any specific models, the results as well as methods and techniques developed can be use in certain specific systems involving regime switching diffusions.

There are several directions that are worthwhile for further study and investigation. By mimicking the classical numerical schemes for SDEs without switching counterpart, the approach in Chapter 2 can be used to design and study various numerical schemes to Markovian switching systems. It is also interesting to extend this approach further so that it can be used to treat the state-dependent switching systems as well. One may also want to relax the assumptions on the coefficients of the systems to better suite many applications in practical situations as well. Up to now, the main focus of research on numerical methods for switching diffusions was concerned with asymptotic error estimates, whereas central limit theorems for the schemes have not received as much attention. Thus investigation in this direction is an interesting choice. In general, central limit theorems illustrate how the choice of parameters affect the efficiency of the scheme and they are a central tool for tuning the parameters.

In the last part of the dissertation we have considered a problem related to mean-field models. These models are concerned with many particle systems having weak interactions. To reduce the computational complexity of interactions due to a large number of particles (or many body problems), all interactions with each particle are replaced by a single average interaction. There has been a renew interest in Mean-field models in the past decades. Initiated independently by Huang, Malhame, and Caines [13], and Lasry and Lions [31] mean field differential games have drawn much attentions and became a very active area. Along with the renewed interest in the classical models, the studies for some other type of mean field models were also carried out. Recently, there has been some effort devoted to study the mean-field models involving regime switching. However, the investigation is far more difficult than that of the counterpart in mean-field models with SDEs. Further study in this direction deserves more attention and careful consideration.

REFERENCES

[1] G. Badowski, G. Yin and Q. Zhang, Near-optimal controls of discrete-time dynamic systems driven by singularly-perturbed Markov chains. *J. Optim. Theory Appl.* **116** (2003), no. 1, 131–166.

[2] J. Bao and J. Shao, Permanence and Extinction of Regime-Switching Predator-Prey Models, *SIAM J. Math. Anal.*, **48** (2016), 725–739.

[3] D. Becherer, M. Schweizer, Classical solutions to reaction-diffusion systems for hedging problems with interacting Itô and point processes. *Ann. Appl. Probab.* **15** (2) (2005), 1111–1144.

[4] A. Bensoussan, K. C. J. Sung, S. C. P. Yam, S.P. Yung, Linear-Quadratic Mean Field Games. *Preprint*, available online at `http://arxiv.org/pdf/1404.5741v1.pdf`.

[5] C. Dellacherie, *Capcités et processus stochastiques*. Springer-Verlag, Berlin, 1972.

[6] S. N. Ethier and T. G. Kurtz, *Markov Processes Characterization and Convergence* . Jonh Wiley & Sons, 1986.

[7] W. Fleming, Generalized solutions in optimal stochastic control. *Pro. URI Conf. on Control* (1982), 147–165.

[8] W. Fleming and M. Nisio, On stochastic relaxed controls for partially observable diffusions. *Nagoya Math. J.* **93** (1984), 71–108.

[9] T. A. Hoang and G. Yin, Properties for a Class of Multi-type Mean-field Models, *Communications in Information and Systems*, **15** (2015), no. 4, 489-519.

[10] T. A. Hoang and G. Yin, A limit theorem for stochastic differential equations with state-dependent regime switching, preprint.

[11] T. A. Hoang, S. L. Nguyen and G. Yin, Near Optimality and Near Equilibrium for Controlled Systems with Wideband Noise for Hybrid Systems, *Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis*, **23** (2016), 163-194.

[12] T.A. Hoang, G. Yin, and F. Xi, Numerical solutions of regime-switching jump diffusions. *Appl. Math. Comput.* **244** (2014), 822–835.

[13] M. Huang, R. P. Malhame and P. E. Caines, Nash equilibria for large-population linear stochastic systems of weakly coupled agents. *Analysis, control and optimization of complex dynamic systems*, 215–252, GERAD 25th Anniv. Ser., **4**, Springer, New York, 2005.

[14] M. Huang, P. E. Caines and R. P. Malhame, Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ε-Nash equilibria. *IEEE Trans. Autom. Control*, **52** (2007), 1560–1571.

[15] J. Jacod, *Calcul Stochastique et Problḿs de Martingales.* Lecture Notes in Math. 714, Springer-Verlag Berlin, 1979.

[16] J. Jacod, The Euler scheme for Lvy driven stochastic differential equations: limit theorems. *Ann. Probab.* **32** (2004), no. 3A, 18301872.

[17] J. Jacod and P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** (1998), no. 1, 267307

[18] J. Jacod and A.N. Shiryayev, *Limit Theorems for Stochastic Processes* (2nd Ed.). Springer-Verlag Berlin Heidelberg, 2003.

[19] J. Jakubowski, M. Nieweglowski, Jump-diffusion processes in random environments. *J. Differential Equations.* **257** (2014), 2671–2703.

[20] R.Sh. Liptser and A.N. Shiryayev, A functional central limit theorem for semimartingales. *Theory of Probability and its Appl.* **1** (1956), 216–290.

[21] R.Sh. Liptser and A.N. Shiryayev, *Theory of Martingales*. Kluwer Academic Publishers Group, Dordrecht, 1989.

[22] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin, 1992.

[23] V. N. Kolokoltsov, J. Li and W. Yang, Mean field games and nonlinear Markov processes. *Preprint*, available online at `http://arxiv.org/pdf/1112.3744.pdf`.

[24] T. G. Kurtz and P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** (1991), no. 3, 10351070.

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

[26] H. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control*. Birkhauser, Boston, 1990.

[27] H. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time.* Springer, New York, 1992.

[28] H. Kushner and W. Runggaldier, Nearly optimal state feedback controls for stochastic systems with wideband noise disturbances. *SIAM J. Control Optim.* **25** (1987), no. 2, 298–315.

[29] D. Lacker, Mean field games via controlled martingale problems: Existence of Markovian equilibria. *Stochastic Process. Appl.* **125** (2015), 2856–2894.

[30] D. Lacker, A general characterization of the mean field limit for stochastic differential games, *Probab. Theory Relat. Fields. To appear,* `DOI: 10.1007/s00440-015-0641-9`.

[31] J. M. Lasry and P. L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007), no. 1, 229–260.

[32] R.Sh. Liptser and A.N. Shiryayev, A functional central limit theorem for semimartingales. *Theory of Probability and its Appl.* **1** (1980), 216–290.

[33] R.Sh. Liptser and A.N. Shiryayev, *Theory of Martingales*. Kluwer Academic Publishers Group, Dordrecht, 1989.

[34] X. Mao and C. Yuan. *Stochastic differential equations with Markovian switching*. Imperial College Press, London, 2006.

[35] D. L. Mcleish, An extended martingale principle. *Ann. Probability.* **2** (1974), 620–628.

[36] G.N. Milstein, Approximate integration of stochastic differential equations. *Theory Probab. Appl.* **19** (1974), 557–562.

[37] S. L. Nguyen, T. A. Hoang, D. T. Nguyen and G. Yin, Milstein-Type Procedures for Numerical Solutions of Stochastic Differential Equations with Markovian Switching, submitted.

[38] D. T. Nguyen, T. A. Hoang, S. L. Nguyen and G. Yin, On Tamed Euler scheme for stochastic differential equations with Markov switching, preprint.

[39] S. L. Nguyen and M. Huang, Linear-quadratic-Gaussian mixed games with continuum-parametrized minor players, *SIAM J. Control Optim.* **50** (2012), 2907–2937.

[40] S.L. Nguyen and G. Yin, Pathwise convergence rates for numerical solutions of Markovian switching stochastic differential equations. *Nonlinear Anal. Real World Appl.* **13** (2012), 1170–1185.

[41] M. Nourian and P. E. Caines, ϵ-Nash mean field game theory for nonlinear stochastic dynamical systems with major and minor agents. *SIAM J. Control Optim.* **51** (2013), no. 4, 3302–3331.

[42] P. E. Protter *Stochastic Integration and Differential Equations* (2nd Ed.). Springer-Verlag Berlin Heidelberg, 2005. Jonh Wiley & Sons, 1986.

[43] L.C.G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales. Vol. 2. Itô Calculus*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

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

[45] D. W. Stroock and S.R.S Varadhan, *Multidimensional Diffusion Processes* (Reprint of the 1997 Edition). Springer-Verlag Berlin Heidelberg, 2006.

[46] K. Yamada, A Stability Theorem for Stochastic Differential Equations and Application to Stochastic Control Problems. *Stochastics* **13** (1984), 257–279.

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

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

[49] G. Yin and J. Yong, A weak convergence approach to hybrid LQG problems with infinite control weights, *J. Appl. Math. Stoch. Anal.* **15**, 2002, no. 1, 1–21.

[50] G. Yin and Q. Zhang, Nearly optimality of stochastic control in systems with unknown parameter processes. *Appl. Math. Optim.* **29** (1994), no. 2, 263–284.

[51] G. Yin and Q. Zhang, *Continuous-time Markov chains and applications. A two-time-scale approach.* Second edition. Springer, New York, 2013.

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

[53] C. Yuan and X. Mao, Convergence of the Euler-Maruyama method for stochastic differential equations with Markovian switching. *Math. Comput. Simulation* **64** (2004), 223–235.

[54] Q. Zhang, Nonlinear filtering and control of a switching diffusion with small observation noise, *SIAM J. Control Optim.*, **36** (1998), 1738–1768.

[55] Q. Zhang, Stock trading: An optimal selling rule, *SIAM J. Control Optim.*, **40** (2001), 64–87.

[56] Q. Zhang and G. Yin, On nearly optimal controls of hybrid LQG problems. *IEEE Trans. Automat. Control.* **44** (1999), no. 12, 2271–2282.

ABSTRACT

HYBRID STOCHASTIC SYSTEMS: NUMERICAL METHODS, LIMIT RESULTS, AND CONTROLS

by

TUAN ANH HOANG

May 2017

Advisor: Dr. G. George Yin

Major: Mathematics

Degree: Doctor of Philosophy

This dissertation is concerned with the so-called stochastic hybrid systems, which are featured by the coexistence of continuous dynamics and discrete events and their interactions. Such systems have drawn much needed attentions in recent years. One of the main reasons is that such systems can be used to better reflect the reality for a wide range of applications in networked systems, communication systems, economic systems, cyber-physical systems, and biological and ecological systems, among others. Our main interest is centered around one class of such hybrid systems known as switching diffusions. In such a system, in addition to the driving force of a Brownian motion as in a stochastic system represented by a stochastic differential equation (SDE), there is an additional continuous-time switching process that models the environmental changes due to random events.

In the first part, we develops numerical schemes for stochastic differential equations with Markovian switching (Markovian switching SDEs). By utilizing a special form of Itô's

formula for switching SDEs and special structural of the jumps of the switching component we derived a new scheme to simulate switching SDEs in the spirit of Milstein's scheme for purely SDEs. We develop a new approach to establish the convergence of the proposed algorithm that incorporates martingale methods, quadratic variations, and Markovian stopping times. Detailed and delicate analysis is carried out. Under suitable conditions that are natural extensions of the classical ones, the convergence of the algorithms is established. The rate of convergence is also ascertained.

The second part is concerned with a limit theorem for general stochastic differential equations with Markovian regime switching. Given a sequence of stochastic regime switching systems where the discrete switching processes are independent of on the state of the systems. The continuous-state component of these systems are governed by stochastic differential equations with driving processes that are continuous increasing processes and square integrable martingales. We establish the convergence of the sequence of systems to the one described by a Markovian regime-switching diffusion process.

The third part is concerned with controlled hybrid systems that are good approximations to controlled switching diffusion processes. In lieu of a Brownian motion noise, we use a wide-band noise formulation, which facilitates the treatment of non-Markovian models. The wide-band noise is one whose spectrum has band width wide enough. We work with a basic stationary mixing type process. On top of this wide-band noise process, we allow the system to be subject to random discrete event influence. The discrete event process is a continuous-time Markov chain with a finite state space. Although the state space is finite, we assume that the state space is rather large and the Markov chain is irreducible. Using a two-time-scale formulation and assuming the Markov chain also subjects to fast variations, using weak convergence and singular perturbation test function method we first proved that the when controlled by nearly optimal and equilibrium controls, the state and the corresponding costs of the original systems would "converge" to those of controlled diffusions systems. Using the limit controlled dynamic system as a guidance, we construct controls for the original problem and show that the controls so constructed are near optimal and nearly equilibrium.

AUTOBIOGRAPHICAL STATEMENT

Tuan Anh Hoang

Education

- *Ph.D. in Applied Mathematics*, Wayne State University, 2016 (expected)
- *M.A. in Mathematical Statistics*, Wayne State University (2016)
- *M.S. in Applied Mathematics*, University of Orléans, Pôle Universitaires Français in Ho Chi Minh City (2010)
- *B.Sc. in Mathematics and Computer Science*, University of Science, Vietnam National University - Ho Chi Minh City (2009)

Publications and Preprints

T. A. Hoang and G. Yin, A limit theorem for stochastic differential equations with state-dependent regime switching, preprint.

D. T. Nguyen, **T. A. Hoang**, S. L. Nguyen and G. Yin, On Tamed Euler scheme for stochastic differential equations with Markov switching, preprint.

S. L. Nguyen, **T. A. Hoang**, D. T. Nguyen and G. Yin, Milstein-Type Procedures for Numerical Solutions of Stochastic Differential Equations with Markovian Switching, *SIAM Journal on Numerical Analysis (in press)* .

T. A. Hoang and G. Yin, Properties for a Class of Multi-type Mean-field Models, *Communications in Information and Systems*, **15** (2015), no. 4, 489-519.

T. A. Hoang, S. L. Nguyen and G. Yin, Near Optimality and Near Equilibrium for Controlled Systems with Wideband Noise for Hybrid Systems, *Dynamics of Continuous, Discrete and Impulsive Systems Series A: Mathematical Analysis*, **23** (2016), 163-194.

T. A. Hoang, G. Yin and F. Xi, Numerical Solutions of Regime-Switching Jump Diffusions, *Applied Mathematics and Computation*, **244** (2014), 822-835.