AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(1): 353–387.
DOI: 10.3934/math.2025017
Received: 18 October 2024
Revised: 26 December 2024
Accepted: 28 December 2024
Published: 08 January 2025

Research article

Averaging principle for two-time-scale stochastic functional differential equations with past-dependent switching

Minyu Wu, Xizhong Yang, Feiran Yuan and Xuyi Qiu*

National Key Laboratory of Avionics Integration and Aviation System-of-Systems Synthesis, Shanghai 201100, China

* **Correspondence:** Email: qiuxy02@126.com.

Abstract: This paper aimed to establish the averaging principle for the two-timescale stochastic functional differential equations with past-dependent switching. Initially, the existence and uniqueness of solutions, as well as moment estimates, were obtained by using classical interlacing techniques. Furthermore, the interaction between the fast and slow processes was derived based on the properties of the Poisson random measure. Subsequently, employing the coupling method and the integration by parts formula for the generator, the exponential ergodicity of the frozen Markov chain and the Lipschitz continuity of its invariant measures were proved. In addition to the challenges posed by the dependence on history, the Lipschitz condition under uniform norms that the generator satisfies also introduced computational and proof difficulties. Therefore, more refined estimates were provided for the segment process. Based on these results, together with weak convergence and martingale methods, the averaging principle for the original system was established. Finally, two examples were provided to illustrate the differences between the results presented here and the classical results.

Keywords: stochastic functional differential equation; past-dependent switching; averaging principle; coupling method; weak convergence; martingale method
Mathematics Subject Classification: 34K26, 34K40, 60H10

1. Introduction

The phenomenon of two timescales is commonly found in complex systems, appearing in fields such as materials science, chemistry, fluid dynamics, control engineering, biology, ecology, financial economics, climate dynamics, and other applications. For example, see [1] and the references therein. In classical gene expression models, Messenger Ribonucleic Acid (mRNA) molecules are produced from Deoxyribo Nucleic Acid (DNA) through the transcription process, while protein molecules are generated from mRNA through the translation process. Both types of molecules are subject to degradation; however, the kinetic behavior of proteins is much slower than that of mRNA. Proteins can exist for several weeks, whereas mRNA may only last a few minutes. The processes by which protein molecules acquire their functional structures and conformations also exhibit different timescales, with the vibrational timescale of covalent bonds on the order of femtoseconds (10^{-15} seconds), while protein folding likely occurs on the order of seconds. When mathematically characterizing these phenomena that span different timescales, it is often necessary to introduce fast-varying and slow-varying processes, thereby forming a two-timescale system. Due to the coupling between the fast and slow processes, directly addressing the original system is often extremely challenging. A common approach is to average the fast variables in the slow-varying equations to obtain an averaged equation, which no longer depends on the fast-varying processes. This averaged equation can then serve as a bridge to designing feasible processes to address the original system. For instance, reference [2] introduced a reduction method grounded in chemical Langevin equations with two timescales, utilizing the stochastic averaging principle to derive a limit averaging system. This limit averaging system serves as an approximation for the slow-reacting process. This reduction method not only significantly enhances computational speed during numerical simulations, but also provides accurate error bounds.

In recent years, the averaging principles for stochastic systems with regime-switching involving two timescales have garnered considerable attention from scholars. This interest arises from the fact that in control engineering, finance, biology, and information transmission, the current state of a system is influenced not only by intrinsic uncertainty factors but also by random factors in the external environment. When both influences occur simultaneously, traditional stochastic differential equations or stochastic functional differential equations are insufficient to characterize such systems. To effectively express the impacts of both internal and external factors, researchers have introduced stochastic systems with switching. A significant feature of these systems is the coexistence of discrete events and continuous dynamics, which interact differently across various models. This characteristic yields results that differ from those of traditional stochastic differential equations. For example, reference [3] discussed the stability of stochastic systems with regime-switching, while reference [4] explored numerical methods for these systems. Furthermore, for the stochastic models with regime-switching mentioned above, when drastic changes in the external environment lead to a significant disparity in the frequency of changes both inside and outside the system, it becomes necessary to introduce a two-timescale structure to describe this phenomenon. Yin and his collaborators established a comprehensive asymptotic expansion theory related to nonhomogeneous Markov chains and their generators under various conditions in reference [5], obtaining stationary distributions and convergence rates while investigating the central limit theorem for occupation measures. Building on this theoretical foundation, the reference [6] examined the long-term behavior and stochastic persistence properties of population models driven by rapidly switching Markov chains. Notably, in the aforementioned two-timescale models, the rapidly switching Markov chains do not depend on the slow-varying process. However, in practical applications, fluctuations in the external environment affect the internal system, and, conversely, changes within the internal system can influence the external environment's development. Consequently, state-dependent regime-switching models have attracted significant scholarly interest. Generally speaking, for Markov chains that do not depend on the system state, they can be treated as exogenous noise. The challenge in dealing with regime-switching models that depend on the current state of the system lies in the coupling relationship between regime-switching and continuous states. Reference [7] used weak convergence and martingale methods to prove numerical methods for stochastic differential equations with state-dependent switching; reference [8] investigated small perturbation large deviation for diffusion systems with state-dependent rapid switching, where the diffusion coefficients may be degenerate; reference [9] established an asymptotic expansion theory for state-dependent switching diffusion systems, thereby deriving averaging principles. It is significant to point out that these studies have only considered regime-switching dependent on the current state of the system. Recently, references [10–12] proposed diffusion systems with regime-switching dependent on the historical state of the system and explored recurrence and ergodicity. To the best of our knowledge, research on stochastic differential equations with past-dependent switching is still in its early stages. Therefore, issues such as averaging principle, numerical computation, numerical simulation, and stability analysis of this model are all worthy of consideration. This paper mainly aims to introduce two time scales into the model and establish the corresponding averaging principle, thereby providing a theoretical foundation for further research on the model.

To date, there are four main methods for establishing averaging principles in two-timescale models. The first method is based on asymptotic expansion techniques of partial differential equations. It begins by demonstrating that the density function of the solution to the original system satisfies a Fokker-Planck-Kolmogorov equation (FPK equation) with singular perturbation under suitable conditions. Next, an asymptotic expansion is performed on this equation, and by taking the limit of the expansion, one obtains the limit function of the density function, which remains a valid density function. Finally, integrating the limit density function with respect to the stationary density function of the fast-varying process yields a density function that satisfies an FPK equation. The corresponding stochastic differential equation for this FPK equation is the limit equation for the slow-varying process in the original system. It should be noted that this method requires strong smoothness conditions on the coefficients of the original system, with more detailed explanations available in reference [13]. The second method is based on certain properties and estimates of the solutions to the Poisson equation defined on the entire space, including the existence and uniqueness of solutions, growth estimates, and estimates of the growth of partial derivatives. The foundational theory for this approach was established by Pardoux [14–16]. Subsequently, a substantial amount of literature has utilized methods involving the Poisson equation to obtain richer results, such as mentioned in [17, 18]. In particular, one of the advantages of this method is its ability to determine the convergence rate of the slow-varying process. The third method is the perturbation test function method, which was first proposed by Khasminskii in the 1960s [19] and later developed and refined by Kushner [20]. The fourth method is based on the technique of time discretization. If the diffusion coefficient of the slow-varying equation does not depend on the fast-varying process, one can seek the strong convergence (in the L^p sense) limit of the slow-varying equation. More detailed explanations can be found in reference [21]. If the diffusion coefficient of the slow-varying process depends on the fast-varying process, an example in [22] illustrates that the slow-varying process no longer possesses a strong convergence limit; in this case, one can only look for its weak convergence limit using time discretization techniques. Detailed discussions can be found in reference [20].

In summary, this paper aims to develop an averaging principle for stochastic functional differential equations with past-dependent switching that incorporate two time scales. To this end, this paper employs the aforementioned time discretization technique and is organized as follows. In the next section, we first provide definitions and assumptions of the model presented in this paper and prove that the switching process can be represented as a stochastic differential equation with respect to the Poisson random measure. Based on this, we can obtain the existence and uniqueness of the solution as well as the moment estimation. Section 3 mainly studies the interaction between fast-varying and slow-varying processes. Section 4 demonstrates through the coupling method that the exponential ergodicity of the frozen Markov chain does not depend on the fixed parameter, and obtains the Lipschitz continuity of invariant measures with respect to fixed parameters. The tightness of the slow-varying process $X^\varepsilon(t)$ and the moment estimation of the segment process are discussed in Section 5. Based on this preparatory work, Section 5 presents the core Theorem 5.5 of this article, which illustrates the main limit theorem in the sense of weak convergence. Finally, Section 6 provides several different types of examples to illustrate the results of this article.

2. Formulation and preliminaries

This paper considers the two-component process $(X^\varepsilon(t), \alpha^\varepsilon(t))$ where $X^\varepsilon(t)$ satisfies

$$dX^\varepsilon(t) = b(X_t^\varepsilon, \alpha^\varepsilon(t))dt + \sigma(X_t^\varepsilon, \alpha^\varepsilon(t))dW(t), \tag{2.1}$$

z' defines the transpose of z, $C([a, b]; \mathbb{R}^n)$ denotes the family of continuous function ν from $[a, b]$ to $\mathbb{R}^d$ with the norm

$$\|\nu\|_\infty = sup_{a\le t\le b}|\nu(t)|,$$

τ denotes the delay length,

$$b(\cdot, \cdot) = (b_1, b_2, \cdots, b_n)' : C([-\tau, 0]; \mathbb{R}^n) \times S \longmapsto \mathbb{R}^n,$$
$$\sigma(\cdot, \cdot) = [\sigma_{ij}]_{n\times q} : C([-\tau, 0]; \mathbb{R}^n) \times S \longmapsto \mathbb{R}^{n\times q},$$

$\alpha^\varepsilon(t)$ is a pure jump process taking value in

$$S = \{1, 2, \ldots, N\},$$

and the set of positive integers with a finite N and $W(t)$ is a standard Brownian motion defined on the complete probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbb{P})$, taking values in $\mathbb{R}^q$ and independent of $\alpha^\varepsilon(t)$. We assume that the switching intensity of $\alpha^\varepsilon(t)$ depends on the segment process X_t^ε, that is,

$$\mathbb{P}(\alpha^\varepsilon(t+\delta) = j|\alpha^\varepsilon(t) = i, X_s^\varepsilon, \alpha^\varepsilon(s), s \le t) = \frac{1}{\varepsilon}q_{ij}(X_t^\varepsilon)\delta + o(\delta), \quad \text{if} \quad i \neq j, \tag{2.2a}$$

$$\mathbb{P}(\alpha^\varepsilon(t+\delta) = i|\alpha^\varepsilon(t) = i, X_s^\varepsilon, \alpha^\varepsilon(s), s \le t) = 1 + \frac{1}{\varepsilon}q_{ii}(X_t^\varepsilon)\delta + o(\delta), \tag{2.2b}$$

for $\delta > 0$, $i, j \in S$ and $\varepsilon > 0$ is a small positive parameter. For convenience of notation, we define

$$q_i = -q_{ii}, \ i \in S.$$

The main objective of this paper is to establish the averaging principle for the aforementioned model. The highlights and major contributions of this paper are reflected in the subsequent key aspects:

(1) To the best of our knowledge, this paper establishes for the first time the averaging principle for stochastic functional differential equations with past-dependent switching involving two timescales. Since the diffusion coefficient depends on the fast-varying process, the counterexample mentioned in the introduction indicates that this model does not possess a strong convergence limit. Therefore, this paper employs weak convergence methods and martingale methods to address this difficulty.

(2) Since $X^\varepsilon(t)$ and $\alpha^\varepsilon(t)$ depend on X_t^ε, the existence and uniqueness of solutions, the interaction between the fast-varying and slow-varying processes, and the invariant measure from classical literature are no longer applicable. Therefore, this article utilizes the method in [23, 24], which represents the switching process as a stochastic differential equation with respect to the Poisson random measure. The advantage of doing so is that we can apply techniques related to stochastic differential equations to the switching process. Furthermore, based on this, we obtain the interaction between fast-varying and slow-varying processes, which will be repeatedly used in the martingale method. At the same time, this article also discusses the moment estimation of the segment process, obtaining an order that is sufficiently close to half, which will be used in the implementation of martingale methods together with inequality (3.3) to form control over the estimation term of martingales. In the following, the assumptions used in this paper are presented along with some explanations.

The following are the assumptions that will be used in this paper. Throughout this paper, K denotes a generic positive constant, whose value may change for different usage, so

$$K + K = K \quad \text{and} \quad KK = K$$

are understood in an appropriate sense. K_β represents the generic constant depending on parameters β. Let us begin by introducing some conditions on the two-timescale system $(X^\varepsilon(t), \alpha^\varepsilon(t))$, which will be used throughout this article.

(A1) Assume that the initial value

$$X_0^\varepsilon = \xi \in C([-\tau, 0]; \mathbb{R}^n)$$

is nonrandom and satisfies the Lipschitz property, and the initial value

$$\alpha^\varepsilon(0) = i_0 \in S$$

is independent of ε.

(A2) For any $\phi_1, \phi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ and any $i \in S$, there exists a positive constant L_1 such that

$$|b(\phi_1, i) - b(\phi_2, i)|^2 \vee |\sigma(\phi_1, i) - \sigma(\phi_2, i)|^2 \le L_1(\|\phi_1 - \phi_2\|_\infty^2). \tag{2.3}$$

(A3) For each $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$Q(\phi) = (q_{ij}(\phi))_{i,j\in S}$$

is a conservative transition rate matrix.

(A4) Assume that

$$\beta_{ij} := \inf_{\phi\in C([-\tau,0],\mathbb{R}^n)} q_{ij}(\phi) > 0$$

for any $i, j \in S$.

(A5) Assume that

$$M := \sup_{i\in S} \sum_{j\in S, j\neq i} \sup_{\phi\in C([-\tau,0],\mathbb{R}^n)} q_{ij}(\phi) < \infty.$$

(A6) For any $\phi_1, \phi_2 \in C([-\tau, 0]; \mathbb{R}^n)$, there exists a constant $K > 0$ such that

$$\|Q(\phi_1) - Q(\phi_2)\|_{l_1} := \sup_{i\in S} \sum_{j\neq i} |q_{ij}(\phi_1) - q_{ij}(\phi_2)| \leq K\|\phi_1 - \phi_2\|_\infty.$$

Owing to the fact that $b(\cdot, i)$ and $\sigma(\cdot, i)$ are independent of t, the condition (**A2**) yields the linear growth condition, that is, for any $i \in S$ and any $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$|b(\phi, i)| \vee |\sigma(\phi, i)| \leq K(1 + \|\phi\|_\infty). \tag{2.4}$$

Thanks to the condition (**A4**) and the finite state space, each state has a strictly positive arrival probability. According to the definition of irreducibility, we can conclude that

$$Q(\phi) = (q_{ij}(\phi))_{i,j\in S}$$

is a irreducible transition rate matrix for each $\phi \in C([-\tau, 0]; \mathbb{R}^n)$. To proceed, we construct $\alpha^\varepsilon(t)$ as the solution to a stochastic differential equation with respect to a Poisson random measure. For $\phi \in C([-\tau, 0], \mathbb{R}^n)$ and $\varepsilon > 0$, let $\{\Gamma^\varepsilon_{ij}(\phi), i, j \in S\}$ be a family of consecutive left-closed and right-open intervals on the half-line, each of length $q_{ij}(\phi)/\varepsilon$, that is

$$\Gamma^\varepsilon_{12}(\phi) = \Big[0, \frac{q_{12}(\phi)}{\varepsilon}\Big), \ldots, \Gamma^\varepsilon_{1N}(\phi) = \Big[\sum_{l=2}^{N-1} \frac{q_{1l}(\phi)}{\varepsilon}, \frac{q_1(\phi)}{\varepsilon}\Big),$$

$$\ldots$$

$$\Gamma^\varepsilon_{ij}(\phi) = \Big[\sum_{l=1}^{i-1} \frac{q_l(\phi)}{\varepsilon} + \sum_{l=1}^{j-1} \frac{q_{il}(\phi)}{\varepsilon}, \sum_{l=1}^{i-1} \frac{q_l(\phi)}{\varepsilon} + \sum_{l=1}^{j} \frac{q_{il}(\phi)}{\varepsilon}\Big), \quad i, j \in S, i \geq 1, j \neq i,$$

$$\ldots$$

$$\Gamma^\varepsilon_{N(N-1)}(\phi) = \Big[\sum_{l=1}^{N-1} \frac{q_l(\phi)}{\varepsilon} + \sum_{l=1}^{N-1} \frac{q_{il}(\phi)}{\varepsilon}, \sum_{l=1}^{N-1} \frac{q_l(\phi)}{\varepsilon} + \sum_{l=1}^{N} \frac{q_{il}(\phi)}{\varepsilon}\Big).$$

For convenience of notation, we set

$$\Gamma^\varepsilon_{ii}(\phi) = \emptyset \quad \text{and} \quad \Gamma^\varepsilon_{ij}(\phi) = \emptyset$$

if

$$q_{ij}(\phi) = 0.$$

Before stating the construction of a Poisson random measure, we give the following lemma to facilitate a subsequent proof.

Lemma 2.1. *Assume that the condition (**A5**) holds. For any $\phi_1, \phi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ and any $i, j \in S$,*

$$\mathbf{m}(\Gamma^\varepsilon_{ij}(\phi_1)\Delta\Gamma^\varepsilon_{ij}(\phi_2)) \leq \frac{2N}{\varepsilon}\|Q(\phi_1) - Q(\phi_2)\|_{l_1},$$

where $\mathbf{m}$ *denotes the Lebesgue measure on* $\mathbb{R}$.

Proof. According to the definition of $\{\Gamma^{\varepsilon}_{ij}(\phi), i, j \in S\}$,

$$\begin{aligned}
\mathbf{m}(\Gamma^{\varepsilon}_{ij}(\phi_1)\Delta\Gamma^{\varepsilon}_{ij}(\phi_2)) \leq & \frac{1}{\varepsilon}\Big|\sum_{k=1}^{i-1} q_k(\phi_1) + \sum_{k=1,k\neq i}^{j-1} q_{ik}(\phi_1) - \sum_{k=1}^{i-1} q_k(\phi_2) - \sum_{k=1,k\neq i}^{j-1} q_{ik}(\phi_2)\Big| \\
& + \frac{1}{\varepsilon}\Big|\sum_{k=1}^{i-1} q_k(\phi_1) + \sum_{k=1,k\neq i}^{j} q_{ik}(\phi_1) - \sum_{k=1}^{i-1} q_k(\phi_2) - \sum_{k=1,k\neq i}^{j} q_{ik}(\phi_2)\Big| \\
\leq & \frac{2}{\varepsilon}\sum_{k=1}^{i-1} |q_k(\phi_1) - q_k(\phi_2)| + \frac{2}{\varepsilon}\sum_{k=1,k\neq i}^{j} |q_{ik}(\phi_1) - q_{ik}(\phi_2)| \\
\leq & \frac{2}{\varepsilon}\sum_{k=1}^{i-1}\sum_{l=1,l\neq k}^{N} |q_{kl}(\phi_1) - q_{kl}(\phi_2)| + \frac{2}{\varepsilon}\sum_{k=1,k\neq i}^{j} |q_{ik}(\phi_1) - q_{ik}(\phi_2)| \\
\leq & \frac{2N}{\varepsilon}\|Q(\phi_1) - Q(\phi_2)\|_{l_1}.
\end{aligned}$$

This completes the proof. □

Next, we provide an explicit construction of the Poisson random measure as in [25, p.42] or [26, p.26]. Under condition **(A5)**, we set

$$H^{\varepsilon} = N(N-1)M/\varepsilon$$

as an upper bound on the total length of $\{\Gamma^{\varepsilon}_{ij}(\phi), i, j \in S\}$ for fixed $\varepsilon > 0$. Let $\xi^{\varepsilon}_l, l = 1, 2, \ldots$ be a sequence of random variables on $[0, H^{\varepsilon}]$ with

$$\mathbb{P}(\xi^{\varepsilon}_l \in dx) = \frac{\mathbf{m}(dx)}{H^{\varepsilon}} \tag{2.5}$$

and $\tau^{\varepsilon}_k, k = 1, 2, \ldots$ be nonnegative random variables such that

$$\mathbb{P}(\tau^{\varepsilon}_k > t) = exp(-H^{\varepsilon}t). \tag{2.6}$$

Suppose that $\{\xi^{\varepsilon}_l, \tau^{\varepsilon}_k\}_{l,k\geq 1}$ are all mutually independent. Set

$$\zeta^{\varepsilon}_1 = \tau^{\varepsilon}_1, \ldots, \quad \zeta^{\varepsilon}_k = \tau^{\varepsilon}_1 + \tau^{\varepsilon}_2 + \cdots + \tau^{\varepsilon}_k, \quad k \in \mathbb{N}.$$

Put

$$\mathbf{D}^{\varepsilon}_{p^{\varepsilon}} = \bigcup_{k\in\mathbb{N}}\{\zeta^{\varepsilon}_k\}$$

and

$$p^{\varepsilon}(\zeta^{\varepsilon}_k) = \xi^{\varepsilon}_k, \quad k \in \mathbb{N}.$$

Correspondingly, introduce a counting measure as follows:

$$N^{\varepsilon}_{p^{\varepsilon}}((0,t] \times A) = \#\{s \in \mathbf{D}^{\varepsilon}_{p^{\varepsilon}} : 0 < s \leq t, p^{\varepsilon}(s) \in A\}, \quad t > 0, A \in \mathcal{B}([0,\infty)),$$

where # means the number of $\cdot$ counting in the $\{\cdot\}$. Then, $\{p^{\varepsilon}(t)\}_{t\geq 0}$ is a Poisson point process that satisfies the jump height ξ^{ε}_k at the jump time ζ^{ε}_k, and its corresponding Poisson random measure is $N^{\varepsilon}_{p^{\varepsilon}}(dt, dz)$ with intensity measure $dt \times \mathbf{m}(dz)$, which is independent of $\{W(t)\}_{t\geq 0}$. Define

$$V^{\varepsilon} : C([-\tau, 0]; \mathbb{R}^n) \times Z_+ \times \mathbb{R}$$

by

$$V^{\varepsilon}(\phi, i, z) = \sum_{j\in S, j\neq i} (j-i)\mathbb{I}_{\{\Gamma^{\varepsilon}_{ij}(\phi)\}}(z)$$

and the systems of equations

$$\begin{cases} dX^{\varepsilon}(t) = b(X^{\varepsilon}_t, \alpha^{\varepsilon}(t))dt + \sigma(X^{\varepsilon}_t, \alpha^{\varepsilon}(t))dW(t), \\ d\alpha^{\varepsilon}(t) = \int_{\mathbb{R}} V^{\varepsilon}(X^{\varepsilon}_t, \alpha^{\varepsilon}(t-), z)N^{\varepsilon}_{p^{\varepsilon}}(dt, dz). \end{cases} \tag{2.7}$$

If the above systems of equations has a solution, the solution $(X^{\varepsilon}(t), \alpha^{\varepsilon}(t))$ to (2.7) satisfies Eqs (2.1) and (2.2). In fact, we only need to demonstrate that the solution satisfies (2.2). Due to the property of independent increment of the Poisson random measure, for any $A \in \mathcal{B}([0,\infty))$ and $t, \delta > 0$,

$$\mathbb{P}(N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times A) \geq 2) = 1 - e^{-\delta \mathbf{m}(A)}\delta \mathbf{m}(A) - e^{-\delta\mathbf{m}(A)} = o(\delta)$$

and from

$$d\alpha^{\varepsilon}(t) = \int_{\mathbb{R}} V^{\varepsilon}(X^{\varepsilon}_t, \alpha(t-), z)N^{\varepsilon}_{p^{\varepsilon}}(dt, dz),$$

we have

$$\alpha^{\varepsilon}(t+\delta) - \alpha^{\varepsilon}(t) = \sum_{t\leq s\leq t+\delta, s\in \mathbf{D}^{\varepsilon}_{p^{\varepsilon}}} \sum_{j\in S} (j - \alpha^{\varepsilon}(s-))\mathbb{I}_{\Gamma^{\varepsilon}_{\alpha^{\varepsilon}(s-)j}(X^{\varepsilon}_s)}(p^{\varepsilon}(s)). \tag{2.8}$$

In particular, if $\alpha^{\varepsilon}(t) = i$,

$$\alpha^{\varepsilon}(\zeta^{\varepsilon,t}_1) = i + \sum_{j\in S}(j-i)\mathbb{I}_{\Gamma^{\varepsilon}_{ij}(X^{\varepsilon}_{\zeta^{\varepsilon,t}_1})}(p^{\varepsilon}(\zeta^{\varepsilon,t}_1)) = i + \sum_{j\in S}(j-i)\mathbb{I}_{\Gamma^{\varepsilon}_{ij}(X^{\varepsilon}_{\zeta^{\varepsilon,t}_1})}(\xi^{\varepsilon,t}_1), \tag{2.9}$$

where $\zeta^{\varepsilon,t}_1$ and $\xi^{\varepsilon,t}_1$ denote the first jump time and jump height of $p^{\varepsilon}(t)$ after time t. For $j \neq i$ and $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$\begin{aligned} \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = j|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &= \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = j, \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) = 1|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &+ \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = j, \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) \geq 2|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &= \mathbb{P}(p^{\varepsilon}(\zeta^{\varepsilon}_1) \in \Gamma^{\varepsilon}_{ij}(\phi), \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) = 1|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) + o(\delta) &= \frac{1}{\varepsilon} e^{-\frac{q_{ij}(\phi)}{\varepsilon}\delta} q_{ij}(\phi)\delta + o(\delta) \\ &= \frac{1}{\varepsilon} q_{ij}(\phi)\delta + o(\delta). \end{aligned}$$

For $j = i$ and $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$\begin{aligned} \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = i|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &= \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = i, \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) = 0|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &+ \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = i, \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) = 1|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &+ \mathbb{P}(\alpha^{\varepsilon}(t+\delta) = i, \\ N^{\varepsilon}_{p^{\varepsilon}}((t, t+\delta] \times [0, H^{\varepsilon}]) \geq 2|\alpha^{\varepsilon}(t) = i, X^{\varepsilon}_t = \phi) &= e^{\frac{q_{ii}(\phi)}{\varepsilon}\delta} + o(\delta) \\ &= 1 + \frac{1}{\varepsilon} q_{ii}(\phi)\delta + o(\delta). \end{aligned}$$

Using (2.7), the following theorem gives the existence and uniqueness of solution and the moment estimation which is independent of the small parameter ε.

Theorem 2.2. *Suppose that **(A1)–(A6)** hold. Then, for $\varepsilon > 0$ and any initial value*

$$X_0^\varepsilon = \xi(t) \in C([-\tau, 0]; \mathbb{R}^n)$$

and

$$\alpha^\varepsilon(0) = i_0 \in S,$$

there exists a unique global strong solution $(X^\varepsilon(t), \alpha^\varepsilon(t))$ of (2.1) and (2.2). Moreover, for every $T > 0$, there exists a positive constant K_T such that

$$\sup_{0<\varepsilon<1} \mathbb{E}\Big(\sup_{-\tau\le t\le T} |X^\varepsilon(t)|^4 \Big) \le K_T. \tag{2.10}$$

We prepare a lemma to prove this theorem (see [27, Theorem 2.2]).

Lemma 2.3. *Assume that **(A2)** holds. Then, for any $i \in S$, there exists a unique gobal solution $X(t)$ to the following equation:*

$$dX(t) = b(X_t, i)dt + \sigma(X_t, i)dW(t)$$

with initial time t_0 and initial value

$$X_{t_0} = \phi \in C([-\tau, 0]; \mathbb{R}^n).$$

Moreover, the initial time t_0 can be a random variable provided that it is a stopping time.

Remark 2.4. *For fixed state $i \in S$, under the global Lipschitz condition, the Picard iterative sequence can approximate a unique solution $X(t)$ ([27, Theorem 2.2]). When the initial time is a stopping time, it does not affect the techniques used in the proof, such as martingales isometry ([28, Remark 3.10]). This proof is omitted here.*

Proof of Theorem 2.2. To construct the solution to (2.1) and (2.2) with initial value (ξ, i_0) for fixed $\varepsilon > 0$, we use the interlacing procedure similar to [11, Theorem 3.1] or [28, Theorem 3.13]. Let $\tilde{X}^{0,\varepsilon}(t), t \ge 0$ be the solution to

$$d\tilde{X}^{0,\varepsilon}(t) = b(\tilde{X}_t^{0,\varepsilon}, i_0)dt + \sigma(\tilde{X}_t^{0,\varepsilon}, i_0)dW(t)$$

with initial value

$$\tilde{X}^{0,\varepsilon}(t) = \xi(t), \quad t \in [-\tau, 0].$$

Here, for the sake of uniformity of notation, we superscript $\tilde{X}^{0,\varepsilon}(t)$ with ε, which is virtually not related to ε. Let

$$\sigma_1^\varepsilon = \inf\Big\{ t > 0 : \int_0^t \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{0,\varepsilon}, i_0, z) N_{p^\varepsilon}^\varepsilon(ds, dz) \ne 0 \Big\}$$

and

$$i_1 = i_0 + \int_0^{\sigma_1^\varepsilon} \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{0,\varepsilon}, i_0, z) N_{p^\varepsilon}^\varepsilon(ds, dz).$$

To proceed, let $\tilde{X}^{1,\varepsilon}(t), t \ge \sigma_1^\varepsilon$ be the solution to

$$d\tilde{X}^{1,\varepsilon}(t) = b(\tilde{X}_t^{1,\varepsilon}, i_1)dt + \sigma(\tilde{X}_t^{1,\varepsilon}, i_1)dW(t)$$

with initial value

$$\tilde{X}^{1,\varepsilon}(t) = \tilde{X}^{0,\varepsilon}(t), \quad t \in [\sigma_1^\varepsilon - \tau, \sigma_1^\varepsilon].$$

Define

$$\sigma_2^\varepsilon = \inf\left\{t > \sigma_1^\varepsilon : \int_0^t \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{1,\varepsilon}, i_1, z) N_{p^\varepsilon}^\varepsilon(ds, dz) \neq 0\right\}$$

and

$$i_2 = i_1 + \int_{\sigma_1^\varepsilon}^{\sigma_2^\varepsilon} \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{1,\varepsilon}, i_0, z) N_{p^\varepsilon}^\varepsilon(ds, dz).$$

For the convenience of notation, set

$$\sigma_0^\varepsilon = 0.$$

When we have already defined $\tilde{X}^{m-2,\varepsilon}(t)$ on $[\sigma_{m-2}^\varepsilon - \tau, \sigma_{m-1}^\varepsilon]$, $m \geq 2$, let $\tilde{X}^{m-1,\varepsilon}(t), t \geq \sigma_{m-1}^\varepsilon$ be the solution to

$$d\tilde{X}^{m-1,\varepsilon}(t) = b(\tilde{X}_t^{m-1,\varepsilon}, i_{m-1})dt + \sigma(\tilde{X}_t^{m-1,\varepsilon}, i_{m-1})dW(t)$$

with initial value

$$\tilde{X}^{m-1,\varepsilon}(t) = \tilde{X}^{m-2,\varepsilon}(t), \quad t \in [\sigma_{m-1}^\varepsilon - \tau, \sigma_{m-1}^\varepsilon].$$

Define

$$\sigma_m^\varepsilon = \inf\left\{t > \sigma_{m-1}^\varepsilon : \int_0^t \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{m-1,\varepsilon}, i_{m-1}, z) N_{p^\varepsilon}^\varepsilon(ds, dz) \neq 0\right\}$$

and

$$i_m = i_{m-1} + \int_{\sigma_{m-1}^\varepsilon}^{\sigma_m^\varepsilon} \int_{\mathbb{R}} V^\varepsilon(\tilde{X}_s^{m-1,\varepsilon}, i_{m-1}, z) N_{p^\varepsilon}^\varepsilon(ds, dz). \tag{2.11}$$

Clearly, continuing this procedure, we can construct a process

$$X^\varepsilon(t) = \tilde{X}^{m,\varepsilon}(t), \alpha^\varepsilon(t) = i_m,$$

when

$$\sigma_m^\varepsilon \leq t \leq \sigma_{m+1}^\varepsilon, \quad m \geq 0,$$

which satisfies that

$$X^\varepsilon(t) = \xi(t), \quad t \in [-\tau, 0]$$

and that

$$\begin{cases} X^\varepsilon(t \wedge \sigma_m^\varepsilon) = \xi(0) + \int_0^{t\wedge\sigma_m^\varepsilon} b(X_s^\varepsilon, \alpha^\varepsilon(s))ds + \int_0^{t\wedge\sigma_m^\varepsilon} \sigma(X_s^\varepsilon, \alpha^\varepsilon(s))dW(s), \\ \alpha^\varepsilon(t \wedge \sigma_m^\varepsilon) = i_0 + \int_0^{t\wedge\sigma_m^\varepsilon} \int_{\mathbb{R}} V^\varepsilon(X_s^\varepsilon, \alpha^\varepsilon(s-), z) N_{p^\varepsilon}^\varepsilon(ds, dz). \end{cases}$$

Define

$$\sigma_\infty^\varepsilon = \lim_{m\to\infty} \sigma_m^\varepsilon.$$

To imply that $(X^\varepsilon(t), \alpha^\varepsilon(t))$ is the unique global solution, it is only necessary to obtain

$$\sigma_\infty^\varepsilon = \infty.$$

For any $T > 0$, one has

$$\begin{aligned}\mathbb{P}(\sigma_m^\varepsilon \le T) &= \mathbb{P}\Big(\int_0^{\sigma_m^\varepsilon \wedge T}\int_{\mathbb{R}} \mathbf{I}_{\{z\in\big[0,\sum_{j=1}^N q_j(X_s^\varepsilon)/\varepsilon\big]\}} N_{p^\varepsilon}^\varepsilon(ds,dz) = m\Big)\\ &\le \mathbb{P}\Big(\int_0^T\int_{\mathbb{R}} \mathbf{I}_{\{z\in[0,H^\varepsilon]\}} N_p(ds,dz) \ge m\Big)\\ &= \sum_{k=m}^{\infty} e^{-H^\varepsilon T}\frac{(H^\varepsilon T)^k}{k!},\end{aligned}$$

which implies that

$$\mathbb{P}(\sigma_m^\varepsilon \le T) \to 0,$$

as $m \to \infty$. It follows that

$$\sigma_\infty^\varepsilon = \infty.$$

The uniqueness of $\tilde{X}^{m,\varepsilon}(t)$ and the uniqueness of i_m defined by (2.11) on $[\sigma_m^\varepsilon, \sigma_{m+1}^\varepsilon]$ can derive the uniqueness of $(X^\varepsilon(t), \alpha^\varepsilon(t))$. Finally, for the resulting solution, it can be given that for every $t \in [0, T]$

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s)|^4\Big) &\le K\Big[\|\xi\|_\infty^4 + \mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^s b(X_r^\varepsilon, \alpha^\varepsilon(r))dr\Big|^4\Big) + \mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^s \sigma(X_r^\varepsilon, \alpha^\varepsilon(r))dW(r)\Big|^4\Big)\Big]\\ &\le K[\|\xi\|_\infty^4 + \mathbb{E}\Big(\int_0^s |b(X_r^\varepsilon, \alpha^\varepsilon(r))|^4 dr\Big) + \mathbb{E}\Big(\int_0^s |\sigma(X_r^\varepsilon, \alpha^\varepsilon(r))|^4 dr\Big)\Big]\\ &\le K\|\xi\|_\infty^4 + K\int_0^t \Big[1 + \mathbb{E}\Big(\sup_{-\tau\le u\le r}|X^\varepsilon(u)|^4\Big)\Big]dr.\end{aligned}$$

It is easy to observe that

$$\mathbb{E}\Big(\sup_{-\tau\le s\le t}|X^\varepsilon(s)|^4\Big) \le \|\xi\|_\infty^4 + \mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s)|^4\Big).$$

This, together with Gronwall inequality, yields that

$$\mathbb{E}\Big(\sup_{-\tau\le s\le t}|X^\varepsilon(s)|^4\Big) \le K_T.$$

Letting $t = T$ concludes the proof. □

3. Interaction between fast-varying and slow-varying processes

Applying (2.7), consider the solution $(X^\varepsilon(t), \alpha^{1,\varepsilon}(t))$ and $(Y^\varepsilon(t), \alpha^{2,\varepsilon}(t))$ respectively to the following systems of equations:

$$\begin{cases} dX^\varepsilon(t) = b^1(X_t^\varepsilon, \alpha^{1,\varepsilon}(t))dt + \sigma^1(X_t^\varepsilon, \alpha^{1,\varepsilon}(t))dW(t),\\ d\alpha^{1,\varepsilon}(t) = \displaystyle\int_{\mathbb{R}} V^\varepsilon(X_t^\varepsilon, \alpha^{1,\varepsilon}(t-), z)N_{p^\varepsilon}^\varepsilon(dt,dz),\\ X_0 = \phi_1, \quad \alpha^{1,\varepsilon}(0) = i_0 \in S, \end{cases} \tag{3.1}$$

and

$$\begin{cases} dY^{\varepsilon}(t) = b^{2}(Y_t^{\varepsilon}, \alpha^{2,\varepsilon}(t))dt + \sigma^{2}(Y_t^{\varepsilon}, \alpha^{2,\varepsilon}(t))dW(t), \\ d\alpha^{2,\varepsilon}(t) = \int_{\mathbb{R}} V^{\varepsilon}(Y_t^{\varepsilon}, \alpha^{2,\varepsilon}(t-), z)N_{p^{\varepsilon}}^{\varepsilon}(dt, dz), \\ Y_0 = \phi_2, \ \ \alpha^{2,\varepsilon}(0) = i_0 \in S, \end{cases} \tag{3.2}$$

where we assume that $\phi_1, \phi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ are nonrandom and $W(t)$ is a standard Brownian motion.

Lemma 3.1. *Suppose that $b_i, \sigma_i, i = 1, 2$ satisfy the condition (**A2**), replacing b and σ, respectively. If*

$$\alpha^{1,\varepsilon}(s) = \alpha^{2,\varepsilon}(s)$$

for any $s, t \in [0, T]$, $s < t$, there exist the solutions to (3.1) and (3.2) such that

$$\frac{1}{t-s}\int_s^t \mathbb{E}\Big[\mathbb{I}_{\{\alpha^{1,\varepsilon}(r) \neq \alpha^{2,\varepsilon}(r)\}}|\mathcal{F}_s^{\varepsilon}\Big]dr \leq 2N(N-1)\frac{1}{\varepsilon}\int_s^t \mathbb{E}(\|Q(X_r^{\varepsilon}) - Q(Y_r^{\varepsilon})\|_{l_1}|\mathcal{F}_s^{\varepsilon})dr, \tag{3.3}$$

where

$$\mathcal{F}_t^{\varepsilon} = \sigma\{W(s), N_{p^{\varepsilon}}^{\varepsilon}((0, s] \times [0, H^{\varepsilon}]) : s \leq t\}.$$

Proof. Obviously,

$$1 - e^{-xH^{\varepsilon}} - e^{-xH^{\varepsilon}} \cdot (xH^{\varepsilon}) = e^{-xH^{\varepsilon}}\Big[\frac{(xH^{\varepsilon})^2}{2} + o(x^2)\Big]$$

for any $x > 0$. Choose a $\delta > 0$ so that when $x \in (0, \delta]$,

$$1 - e^{-xH^{\varepsilon}} - e^{-xH^{\varepsilon}} \cdot (xH^{\varepsilon}) \leq (xH^{\varepsilon})^2$$

holds. Divide $[s, t]$ by δ. Let $t_k = s + k\delta$, $k = 0, 1, 2, \ldots, \bar{K}$, where we denote

$$\bar{K} = [\frac{t-s}{\delta}],$$

the integer part of $\frac{t-s}{\delta}$, and

$$t_{\bar{K}+1} = t.$$

For the interval $[t_0, t_1]$,

$$\begin{aligned} \mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1)|\mathcal{F}_s^{\varepsilon}) = \ & \mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), N_{p^{\varepsilon}}^{\varepsilon}((t_0, t_1] \times [0, H^{\varepsilon}]) = 1|\mathcal{F}_{t_0}^{\varepsilon}) \\ & + \mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), N_{p^{\varepsilon}}^{\varepsilon}((t_0, t_1] \times [0, H^{\varepsilon}]) \geq 2|\mathcal{F}_{t_0}^{\varepsilon}). \end{aligned} \tag{3.4}$$

According to the definition of the Poisson random measure and its property of independent increment,

$$\begin{aligned} & \mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), N_{p^{\varepsilon}}^{\varepsilon}((t_0, t_1] \times [0, H^{\varepsilon}]) \geq 2|\mathcal{F}_{t_0}^{\varepsilon}) \\ & \leq \mathbb{P}(N_{p^{\varepsilon}}^{\varepsilon}((t_0, t_1] \times [0, H^{\varepsilon}]) \geq 2) \\ & = 1 - e^{-H^{\varepsilon}\delta} - e^{-H^{\varepsilon}\delta} \cdot (H^{\varepsilon}\delta) \\ & \leq (H^{\varepsilon}\delta)^2. \end{aligned} \tag{3.5}$$

Below, we estimate the first term on the right side of Eq (3.4). Recall that $\zeta_k^{\varepsilon,s}$ and $\xi_k^{\varepsilon,s}$ denote the kth jump time and jump height after time s, respectively. From (2.8) and (2.9),

$$\begin{aligned}
&\mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), N^{\varepsilon}_{p^\varepsilon}((t_0,t_1]\times[0,H^\varepsilon]) = 1|\mathcal{F}^\varepsilon_{t_0})\\
&\quad= \mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), \zeta_1^{\varepsilon,t_0} \in (t_0,t_1], \zeta_2^{\varepsilon,t_0} > t_1|\mathcal{F}^\varepsilon_{t_0})\\
&\quad= \mathbb{P}\Big(\alpha^{1,\varepsilon}(t_0) + \sum_{l\in S}(l-\alpha^{1,\varepsilon}(t_0))\mathbb{I}_{\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(X^\varepsilon_{\zeta_1^{\varepsilon,t_0}})}(\xi_1^{\varepsilon,t_0})\\
&\quad\neq \alpha^{2,\varepsilon}(t_0) + \sum_{l\in S}(l-\alpha^{2,\varepsilon}(t_0))\mathbb{I}_{\Gamma^\varepsilon_{\alpha^{2,\varepsilon}(t_0)l}(Y^\varepsilon_{\zeta_1^{\varepsilon,t_0}})}(\xi_1^{\varepsilon,t_0}), \zeta_1^{\varepsilon,t_0} \in (t_0,t_1], \zeta_2^{\varepsilon,t_0} > t_1\Big|\mathcal{F}^\varepsilon_{t_0}\Big)\\
&\quad= \mathbb{P}\Big(\xi_1^{\varepsilon,t_0} \in \bigcup_{l\neq\alpha^{1,\varepsilon}(s)}\{\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(X^\varepsilon_{\zeta_1^{\varepsilon,t_0}})\Delta\Gamma^\varepsilon_{\alpha^{2,\varepsilon}(t_0)l}(Y^\varepsilon_{\zeta_1^{\varepsilon,t_0}})\}, \zeta_1^{\varepsilon,t_0} \in (t_0,t_1], \zeta_2^{\varepsilon,t_0} > t_1\Big|\mathcal{F}^\varepsilon_{t_0}\Big).
\end{aligned} \tag{3.6}$$

Note that on $[t_0, \zeta_1^{\varepsilon,t_0}]$, the solutions $X^\varepsilon(u)$ and $Y^\varepsilon(u)$ of (3.1) and (3.2) are respectively determined by

$$\begin{aligned}
dX^\varepsilon(u) &= b^1(X^\varepsilon_u, \alpha^{1,\varepsilon}(t_0))du + \sigma^1(X^\varepsilon_u, \alpha^{1,\varepsilon}(t_0))dW(u),\\
dY^\varepsilon(u) &= b^2(Y^\varepsilon_u, \alpha^{2,\varepsilon}(t_0))du + \sigma^2(Y^\varepsilon_u, \alpha^{2,\varepsilon}(t_0))dW(u),
\end{aligned}$$

where the initial values are, respectively, $X^\varepsilon_{t_0}$ and $Y^\varepsilon_{t_0}$. Therefore, due to the mutual independence of $N^\varepsilon_{p^\varepsilon}(dt,dz)$ and $(W(t))_{t\geq 0}$, it follows that $\{X^\varepsilon_u\}_{u\in[t_0,\zeta_1^{\varepsilon,t_0}]}$, $\{Y^\varepsilon_u\}_{u\in[t_0,\zeta_1^{\varepsilon,t_0}]}$, and $\zeta_1^{\varepsilon,t_0}, \zeta_2^{\varepsilon,t_0}$ are mutual conditional independent with respect to $\mathcal{F}^\varepsilon_{t_0}$. This, together with (2.5), (2.6), and (3.6) yields that

$$\begin{aligned}
&\mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1), N^{\varepsilon}_{p^\varepsilon}((t_0,t_1]\times[0,H^\varepsilon]) = 1|\mathcal{F}^\varepsilon_{t_0})\\
&\quad= \int_{t_0}^{t_1}\int_\Omega \frac{1}{H^\varepsilon}\mathbf{m}\Big(\bigcup_{l\neq\alpha^{1,\varepsilon}(t_0)}\{\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(X^\varepsilon_v)\Delta\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(Y^\varepsilon_v)\}\Big)\mathbb{P}(d\omega|\mathcal{F}^\varepsilon_{t_0}) \times e^{-H^\varepsilon(r-u)}\mathbb{P}(\zeta_1^{\varepsilon,t_0} \in du)\\
&\quad\leq \int_{t_0}^{t_1}\mathbb{E}\Big(\mathbf{m}\Big(\bigcup_{l\neq\alpha^{1,\varepsilon}(t_0)}\{\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(X^\varepsilon_v)\Delta\Gamma^\varepsilon_{\alpha^{1,\varepsilon}(t_0)l}(Y^\varepsilon_v)\}\Big)\Big|\mathcal{F}^\varepsilon_{t_0}\Big)du.
\end{aligned} \tag{3.7}$$

Substituting (3.5) and (3.7) into (3.4) yields that

$$\mathbb{P}(\alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1)|\mathcal{F}^\varepsilon_{t_0}) \leq (H^\varepsilon\delta)^2 + 2N(N-1)\int_{t_0}^{t_1}\mathbb{E}(\|Q(X^\varepsilon_u) - Q(Y^\varepsilon_u)\|_{l_1}|\mathcal{F}^\varepsilon_{t_0})du. \tag{3.8}$$

To proceed, we estimate

$$\begin{aligned}
\mathbb{P}(\alpha^{1,\varepsilon}(t_2) \neq \alpha^{2,\varepsilon}(t_2)|\mathcal{F}^\varepsilon_{t_0}) &= \mathbb{P}(\alpha^{1,\varepsilon}(t_2) \neq \alpha^{2,\varepsilon}(t_2), \alpha^{1,\varepsilon}(t_1) \neq \alpha^{2,\varepsilon}(t_1)|\mathcal{F}^\varepsilon_{t_0})\\
&\quad+ \mathbb{P}(\alpha^{1,\varepsilon}(t_2) \neq \alpha^{2,\varepsilon}(t_2), \alpha^{1,\varepsilon}(t_1) = \alpha^{2,\varepsilon}(t_1)|\mathcal{F}^\varepsilon_{t_0}).
\end{aligned} \tag{3.9}$$

On one hand, (3.8) gives the estimation of the first term on the right side of the above equation. On the other hand, clearly,

$$\begin{aligned}
&\mathbb{P}(\alpha^{1,\varepsilon}(t_2) \neq \alpha^{2,\varepsilon}(t_2), \alpha^{1,\varepsilon}(t_1) = \alpha^{2,\varepsilon}(t_1)|\mathcal{F}^\varepsilon_{t_0})\\
&\leq \mathbb{P}(\alpha^{1,\varepsilon}(t_2) \neq \alpha^{2,\varepsilon}(t_2), \alpha^{1,\varepsilon}(t_1) = \alpha^{2,\varepsilon}(t_1), N^{\varepsilon}_{p^\varepsilon}((t_1,t_2]\times[0,H^\varepsilon]) = 1|\mathcal{F}^\varepsilon_{t_0}) + (H^\varepsilon\delta)^2.
\end{aligned} \tag{3.10}$$

Similar to (3.6) and (3.7),

$$\begin{aligned}&\mathbb{P}(\alpha^{1,\varepsilon}(t_2)\neq\alpha^{2,\varepsilon}(t_2),\alpha^{1,\varepsilon}(t_1)=\alpha^{2,\varepsilon}(t_1),N^{\varepsilon}_{p^{\varepsilon}}((t_1,t_2]\times[0,H^{\varepsilon}])=1|\mathcal{F}^{\varepsilon}_{t_0})\\&\quad=\mathbb{E}(\mathbb{I}_{\alpha^{1,\varepsilon}(t_1)=\alpha^{2,\varepsilon}(t_1)}\mathbb{P}(\alpha^{1,\varepsilon}(t_2)\neq\alpha^{2,\varepsilon}(t_2),N^{\varepsilon}_{p^{\varepsilon}}((t_1,t_2]\times[0,H^{\varepsilon}])=1|\mathcal{F}^{\varepsilon}_{t_1})|\mathcal{F}^{\varepsilon}_{t_0}).\end{aligned}\tag{3.11}$$

Restricted to the set $\{\alpha^{1,\varepsilon}(t_1)=\alpha^{2,\varepsilon}(t_1)\}$, it can be deduced that

$$\begin{aligned}&\mathbb{P}(\alpha^{1,\varepsilon}(t_2)\neq\alpha^{2,\varepsilon}(t_2),N^{\varepsilon}_{p^{\varepsilon}}((t_1,t_2]\times[0,H^{\varepsilon}])=1|\mathcal{F}^{\varepsilon}_{t_1})\\&\quad=\mathbb{P}(\alpha^{1,\varepsilon}(t_2)\neq\alpha^{2,\varepsilon}(t_2),\zeta^{\varepsilon,t_1}_1\in(t_1,t_2],\zeta^{\varepsilon,t_1}_2>2\delta|\mathcal{F}^{\varepsilon}_{t_1})\\&\quad\leq\mathbb{P}\Big(\xi^{\varepsilon,t_1}_1\in\bigcup_{l\neq\alpha^{1,\varepsilon}(t_1)}\{\Gamma^{\varepsilon}_{\alpha^{1,\varepsilon}(t_1)l}(X^{\varepsilon}_{\zeta^{\varepsilon,t_1}_1})\Delta\Gamma^{\varepsilon}_{\alpha^{2,\varepsilon}(t_1)l}(Y^{\varepsilon}_{\zeta^{\varepsilon,t_1}_1})\},\zeta^{\varepsilon,t_1}_1\in(t_1,t_2],\zeta^{\varepsilon,t_1}_2>t_2\Big|\mathcal{F}^{\varepsilon}_{t_1}\Big)\\&\quad=\int_{t_1}^{t_2}\int_{\Omega}\frac{1}{H^{\varepsilon}}\mathbf{m}\Big(\bigcup_{l\neq\alpha^{1,\varepsilon}(t_1)}\{\Gamma^{\varepsilon}_{\alpha^{1,\varepsilon}(t_1)l}(X^{\varepsilon}_u))\Delta\Gamma^{\varepsilon}_{\alpha^{2,\varepsilon}(t_1)l}(Y^{\varepsilon}_u)\}\Big)\mathbb{P}(d\omega|\mathcal{F}^{\varepsilon}_{t_1})e^{-H^{\varepsilon}(t_2-u)}\mathbb{P}(\zeta^{\varepsilon,t_1}_1\in du)\\&\quad\leq 2N(N-1)\frac{1}{\varepsilon}\int_{t_1}^{t_2}\mathbb{E}(\|Q(X^{\varepsilon}_u)-Q(Y^{\varepsilon}_u)\|_{l_1}|\mathcal{F}^{\varepsilon}_{t_1})du.\end{aligned}\tag{3.12}$$

Substituting (3.8), (3.10), and (3.12) into (3.9) yields that

$$\mathbb{P}(\alpha^{1,\varepsilon}(t_2)\neq\alpha^{2,\varepsilon}(t_2)|\mathcal{F}^{\varepsilon}_{t_0})\leq 2(H^{\varepsilon}\delta)^2+2N(N-1)\frac{1}{\varepsilon}\int_{t_0}^{t_2}\mathbb{E}(\|Q(X^{\varepsilon}_u)-Q(Y^{\varepsilon}_u)\|_{l_1}|\mathcal{F}^{\varepsilon}_{t_0})du.$$

Deducing inductively, we obtain

$$\mathbb{P}(\alpha^{1,\varepsilon}(t_k)\neq\alpha^{2,\varepsilon}(t_k)|\mathcal{F}^{\varepsilon}_{t_0})\leq k(H^{\varepsilon}\delta)^2+K_{N,\varepsilon}\int_{t_0}^{t_k}\mathbb{E}(\|Q(X^{\varepsilon}_u)-Q(Y^{\varepsilon}_u)\|_{l_1}|\mathcal{F}^{\varepsilon}_{t_0})du,\tag{3.13}$$

where

$$K_{N,\varepsilon}=2N(N-1)\frac{1}{\varepsilon}.$$

Finally, applying (3.13) gives that

$$\begin{aligned}\int_s^t\mathbb{P}(\alpha^{1,\varepsilon}(r)\neq\alpha^{2,\varepsilon}(r)|\mathcal{F}^{\varepsilon}_s)&=\sum_{k=0}^{\bar{K}}\int_{t_k}^{t_{k+1}}\mathbb{P}(\alpha^{1,\varepsilon}(r)\neq\alpha^{2,\varepsilon}(r)|\mathcal{F}^{\varepsilon}_s)\\&=\sum_{k=0}^{\bar{K}}\int_{t_k}^{t_{k+1}}\mathbb{P}(\alpha^{1,\varepsilon}(r)\neq\alpha^{2,\varepsilon}(r),\alpha^{1,\varepsilon}(t_k)=\alpha^{2,\varepsilon}(t_k)|\mathcal{F}^{\varepsilon}_s)\\&\quad+\sum_{k=0}^{\bar{K}}\int_{t_k}^{t_{k+1}}\mathbb{P}(\alpha^{1,\varepsilon}(r)\neq\alpha^{2,\varepsilon}(r),\alpha^{1,\varepsilon}(t_k)\neq\alpha^{2,\varepsilon}(t_k)|\mathcal{F}^{\varepsilon}_s)\\&\leq\sum_{k=0}^{\bar{K}}\int_{t_k}^{t_{k+1}}\mathbb{P}(N^{\varepsilon}_{p^{\varepsilon}}(t_k,t_{k+1})\times[0,H^{\varepsilon}]\geq 1)dr\\&\quad+\sum_{k=0}^{\bar{K}}\int_{t_k}^{t_{k+1}}\mathbb{P}(\alpha^{1,\varepsilon}(t_k)\neq\alpha^{2,\varepsilon}(t_k)|\mathcal{F}^{\varepsilon}_s)dr\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{k=0}^{\bar{K}} \int_{t_k}^{t_{k+1}} \Big[(1-e^{-H^\varepsilon\delta}) + k(H^\varepsilon\delta)^2 \\
&\quad + K_{N,\varepsilon}\int_{t_0}^{t_k} \mathbb{E}(\|Q(X_u^\varepsilon)-Q(Y_u^\varepsilon)\|_{l_1}|\mathcal{F}_s^\varepsilon)du\Big]dr \\
&= (\bar{K}+1)\delta(1-e^{-H^\varepsilon\delta}) + (H^\varepsilon)^2\delta^3\frac{\bar{K}(\bar{K}+1)}{2} \\
&\quad + \tilde{K}\int_s^t \mathbb{E}(\|Q(X_u^\varepsilon)-Q(Y_u^\varepsilon)\|_{l_1}|\mathcal{F}_s^\varepsilon)du,
\end{aligned}
$$

where

$$\tilde{K} = (\bar{K}+1)\delta K_{\varepsilon,N}.$$

Letting $\delta \to 0$ together with the fact that $(\bar{K}+1)\delta \to (t-s)$, as $\delta \to 0$, gives that (3.3) holds.

The proof is completed. □

4. Invariant probability measures for frozen Markov chains

In this section, for any probability measure

$$\mu = (\mu_i)_{i\in S} \quad \text{and} \quad \nu = (\nu_i)_{i\in S},$$

the total variation distance between μ and ν is defined by

$$\|\mu-\nu\|_{var} = \sup_{|f|\leq 1}|\mu(f)-\nu(f)|,$$

where

$$\mu(f) = \sum_{i\in S}\mu_i f(i) \quad \text{and} \quad \nu(f) = \sum_{i\in S}\nu_i f(i).$$

To proceed, we will use the following coupling methods; for details, see [29, Chapter 5]. For any $\phi \in C([-\tau,0];\mathbb{R}^n)$, let $(\alpha_1^\phi(t), \alpha_2^\phi(t))$ be a coupling Markov process on phase space $S\times S$ with marginal distributed as $\alpha^\phi(t)$ and $\alpha_l^\phi(0) = i_l$ $(l = 1,2)$, where the process $\alpha_\cdot^\phi(\cdot)$ with the parameter variable ϕ, when ϕ is fixed, is referred to as a frozen Markov process, which can be handled using the conclusions of Markov processes. Denote the classical coupling operator

$$
\begin{aligned}
\Omega^\phi f(k_1,k_2) = \mathbb{I}_{\Delta^c}(k_1,k_2)\Big(\sum_{l_1\in S} q_{k_1l_1}(\phi)(f(l_1,k_2)-f(k_1,k_2)) + \sum_{l_2\in S} q_{k_2l_2}(\phi)(f(k_1,l_2)-f(k_1,k_2))\Big) \\
+ \mathbb{I}_\Delta(k_1,k_2)\sum_{l_1\in S} q_{k_1l_1}(\phi)\Big(f(l_1,l_1)-f(k_1,k_2)\Big), \qquad (4.1)
\end{aligned}
$$

where f is a bounded function on $S\times S$ and

$$\Delta := \{(k_1,k_2)\in S\times S : k_1 = k_2\}.$$

This classical coupling means that the marginals evolve independently until they meet. After they meet, they will move together at rate $Q(\phi)$. Define the coupling time

$$T = \inf\{t \geq 0 : \alpha_1^\phi(t) = \alpha_2^\phi(t)\}.$$

Then, by the well-known coupling inequality (see [29, p.195]), we have

$$\|P^{\phi}(t,i_1,\cdot)-P^{\phi}(t,i_2,\cdot)\|_{var}\le 2\mathbb{E}_{\phi}^{i_1,i_2}(\mathbb{I}_{\alpha_1^{\phi}(t)\neq\alpha_2^{\phi}(t)})=2\mathbb{P}_{\phi}^{i_1,i_2}(T>t). \tag{4.2}$$

where $\mathbb{P}_{\phi}^{i_1,i_2}$ and $\mathbb{E}_{\phi}^{i_1,i_2}$ denote the probability and expectation for the coupling process $(\alpha_1^{\phi}(t),\alpha_2^{\phi}(t))$ starting from (i_1,i_2). Furthermore, let π^{ϕ} be invariant probability measure associated to Markov chain $\alpha^{\phi}(t)$ with fixed $\phi\in C([-\tau,0];\mathbb{R}^n)$. When π^{ϕ} is assumed to exist, then from the fact

$$\pi^{\phi}=\pi^{\phi}P^{\phi}$$

and (4.2), it can be obtained that:

$$\begin{aligned}\|P^{\phi}(t,i_1,\cdot)-\pi^{\phi}\|_{var}&=\Big\|P^{\phi}(t,i_1,\cdot)-\sum_{i_2\in S}P^{\phi}(t,i_2,\cdot)\pi^{\phi}(i_2)\Big\|_{var}\\&\le\sum_{i_2\in S}\pi(i_2)\|P^{\phi}(t,i_1,\cdot)-P^{\phi}(t,i_2,\cdot)\|_{var}\\&\le 2\sum_{i_2\in S}\pi(i_2)\mathbb{P}_{\phi}^{i_1,i_2}(T>t).\end{aligned} \tag{4.3}$$

From the inequality above, it can be seen that a coupling time gives us some information about the convergence rate. To obtain the strong ergodicity uniformly in $\phi\in C([-\tau,0];\mathbb{R}^n)$, we just estimate $\mathbb{P}_{\phi}^{i_1,i_2}(T>t)$.

Theorem 4.1. *Under assumptions* **(A3)–(A5)***, P_t^{ϕ} is strongly ergodic and uniformly in $\phi\in C([-\tau,0];\mathbb{R}^n)$, that is, there exist constants $L_3,\lambda>0$ such that*

$$\sup_{i\in S}\|P_t^{\phi}(i,\cdot)-\pi^{\phi}\|_{var}\le L_3e^{-\lambda t}. \tag{4.4}$$

Proof. For fixed $\phi\in C([-\tau,0];\mathbb{R}^n)$, from classic conclusions (see [5, Lemma A.2]), it can be concluded that

$$\sup_{i\in S}\|P_t^{\phi}(i,\cdot)-\pi^{\phi}\|_{var}\le Kexp(-\hat{\kappa}(\phi)s).$$

where the convergence rate $\tilde{\kappa}(\phi)$ depends on ϕ. To proceed, we show that $\tilde{\kappa}(\phi)$ has a uniform lower bound. Set

$$\begin{aligned}\hat{\kappa}_0(\phi)&=\min\Big\{\min\Big\{(1-e^{-q_k(\phi)})\frac{q_{kl}(\phi)}{q_k(\phi)}:l\in S\backslash\{k\}\Big\},e^{-q_k(\phi)}:k\in S\Big\},\\\kappa_1&=\inf_{\phi\in C([-\tau,0];\mathbb{R}^n)}\hat{\kappa}_0(\phi),\\\kappa_2&=\sup_{\phi\in C([-\tau,0];\mathbb{R}^n)}\hat{\kappa}_0(\phi).\end{aligned}$$

Under assumptions **(A4)** and **(A5)**, we can obtain

$$\kappa_1\ge\min\Big\{\min\Big\{(1-e^{-\beta_{kk}})\frac{\beta_{kl}}{M}:l\in S\backslash\{k\}\Big\},e^{-\beta_{kk}}:k\in S\Big\}>0 \tag{4.5}$$

and

$$
\begin{aligned}
\kappa_2 &\leq \sup_{\phi\in C([-\tau,0];\mathbb{R}^n)} \sum_{l\in S\setminus\{k\}} (1-e^{-q_k(\phi)})\frac{q_{kl}(\phi)}{q_k(\phi)} \\
&\leq \sup_{\phi\in C([-\tau,0];\mathbb{R}^n)} (1-e^{-q_k(\phi)}) \\
&\leq 1-e^{-M}.
\end{aligned} \tag{4.6}
$$

By the classical coupling $(\alpha_1^\phi(t),\alpha_2^\phi(t))$ constructed by (4.1), together with the definition of $\hat{\kappa}_0(\phi)$ one has for $i_1\neq i_2$,

$$
\begin{aligned}
\mathbb{P}_\phi^{i_1,i_2}(T\leq 1) \geq& \sum_{l\neq k_1,k_2} (1-e^{-q_{k_1}(\phi)})\frac{q_{k_1 l}(\phi)}{q_{k_1}(\phi)}(1-e^{-q_{k_2}(\phi)})\frac{q_{k_2 l}(\phi)}{q_{k_2}(\phi)} \\
&+(1-e^{-q_{k_1}(\phi)})\frac{q_{k_1k_2}(\phi)}{q_{k_1}(\phi)}e^{-q_{k_2}(\phi)}+(1-e^{-q_{k_2}(\phi)})\frac{q_{k_2k_1}(\phi)}{q_{k_1}(\phi)}e^{-q_{k_1}(\phi)} \\
\geq& \hat{\kappa}_0(\phi)\sum_{l\neq k_1,k_2}(1-e^{-q_{k_1}(\phi)})\frac{q_{k_1 l}(\phi)}{q_{k_1}(\phi)}+\hat{\kappa}_0(\phi)(1-e^{-q_{k_1}(\phi)})\frac{q_{k_1k_2}(\phi)}{q_{k_1}(\phi)}+\hat{\kappa}_0(\phi)e^{-q_{k_1}(\phi)} \\
\geq& \hat{\kappa}_0(\phi).
\end{aligned}
$$

Note that

$$
\mathbb{P}_\phi^{i_1,i_2}(T\leq 1)=1\geq\hat{\kappa}_0(\phi)
$$

as $i_1=i_2$, which implies

$$
\mathbb{P}_\phi^{i_1,i_2}(T\leq 1)\geq\hat{\kappa}_0(\phi)
$$

for any $i_1,i_2\in S$. This, together with Markov property, can derive that

$$
\mathbb{P}_\phi^{i_1,i_2}(T>t)\leq(1-\hat{\kappa}_0(\phi))^{[t]}
$$

for any $i_1,i_2\in S$ by the induction method, where $[t]$ denotes the integral part of t. In fact, suppose that

$$
\mathbb{P}_\phi^{i_1,i_2}(T>[t]-1)\leq(1-\hat{\kappa}_0(\phi))^{[t]-1}
$$

for $t\geq 1$. Then

$$
\begin{aligned}
\mathbb{P}_\phi^{i_1,i_2}(T>t) &\leq \mathbb{E}_\phi^{i_1,i_2}(T>[t]) \\
&= \mathbb{E}_\phi^{i_1,i_2}(\mathbb{I}_{(T>1)}\mathbb{E}_\phi^{\alpha_1^\phi(1),\alpha_2^\phi(1)}\mathbb{I}_{(T>[t]-1)}) \\
&\leq (1-\hat{\kappa}_0(\phi))^{[t]-1}\mathbb{P}_\phi^{i_1,i_2}(T>1) \\
&\leq (1-\hat{\kappa}_0(\phi))^{[t]}.
\end{aligned} \tag{4.7}
$$

Substituting (4.7) into (4.3) yields that

$$
\begin{aligned}
\|P^\phi(t,i_1,\cdot)-\pi^\phi\|_{var} &\leq 2(1-\hat{\kappa}_0(\phi))^{[t]} \\
&= 2e^{-\{t\}ln(1-\hat{\kappa}_0(\phi))}e^{tln(1-\hat{\kappa}_0(\phi))} \\
&\leq 2e^{-\{t\}ln(1-\kappa_2)}e^{tln(1-\kappa_1)} \\
&\leq 2e^{M}e^{tln(1-\kappa_1)},
\end{aligned}
$$

where $\{\cdot\}$ denotes the decimal part.

This concludes the proof. □

Theorem 4.2. *Under assumptions (A3)–(A6), the functional $\phi \mapsto \pi^{\phi}$ from $C([-\tau,0];\mathbb{R}^n)$ to $\mathscr{P}(S)$, endowed with the total variation norm, is Lipschitz continuous, i.e., there exists a constant L_4 such that*

$$\|\pi^{\phi_1} - \pi^{\phi_2}\|_{var} \leq L_4 \|\phi_1 - \phi_2\|_{\infty} \tag{4.8}$$

for any ϕ_1 and $\phi_2 \in C([-\tau,0],\mathbb{R}^n)$.

Proof. For ϕ_1 and $\phi_2 \in C([-\tau,0],\mathbb{R}^n)$, by the integration by parts formula for continuous Markov chains (see [29, Thoerem 13.40]),

$$P_t^{\phi_1} f(i) - P_t^{\phi_2} f(i) = \int_0^t P_{t-s}^{\phi_1}(Q(\phi_1) - Q(\phi_2)) P_s^{\phi_2} f(i) ds, t > 0, f \in \mathcal{B}(S). \tag{4.9}$$

For any $|f| \leq 1$ and any $0 \leq s < t$,

$$\begin{aligned}\sup_{i\in S} |P_{t-s}^{\phi_1}(Q(\phi_1) - Q(\phi_2)) P_s^{\phi_2} f(i)| &\leq \sup_{i\in S} |(Q(\phi_1) - Q(\phi_2)) P_s^{\phi_2} f(i)| \\ &= \sup_{i\in S} |(Q(\phi_1) - Q(\phi_2))(P_s^{\phi_2} - \pi^{\phi_2}) f(i)|. \end{aligned} \tag{4.10}$$

It is easy to observe that for any $|f| \leq 1$,

$$\begin{aligned}\sup_{i\in S} |(Q(\phi_1) - Q(\phi_2)) f(i)| &\leq \sup_{i\in S} \Big[|q_{ii}(\phi_1) - q_{ii}(\phi_2)| + \sum_{j\in S, j\neq i} |q_{ij}(\phi_1) - q_{ij}(\phi_2)| \Big] \\ &\leq 2\|Q(\phi_1) - Q(\phi_2)\|_{l_1}.\end{aligned}$$

This, together with (4.10), the condition (**A6**), and the fact that for any $|f| \leq 1$,

$$\sup_{i\in S} |(P_s^{\phi_2} - \pi^{\phi_2}) f(i)| \leq \sup_{i\in S} \|P_s^{\phi_2}(i,\cdot) - \pi^{\phi_2}\|_{var}$$

implies that

$$\sup_{i\in S} |P_{t-s}^{\phi_1}(Q(\phi_1) - Q(\phi_2)) P_s^{\phi_2} f(i)| \leq K \|\phi_1 - \phi_2\|_{\infty} \sup_{i\in S} \|P_s^{\phi_2}(i,\cdot) - \pi^{\phi_2}\|_{var}.$$

Combining the estimation above with (4.4), we get from (4.9) that for $i \in S$

$$|P_t^{\phi_1} f(i) - P_t^{\phi_2} f(i)| \leq K L_3 \|\phi_1 - \phi_2\|_{\infty} \int_0^t e^{-\lambda s} ds \leq K \|\phi_1 - \phi_2\|_{\infty} (1 - e^{-\lambda t}). \tag{4.11}$$

Consequently, (4.4) and (4.11) show that

$$\begin{aligned}|\pi^{\phi_1} f - \pi^{\phi_2} f| &= \Big| \sum_{i=1}^N \sum_{j=1}^N \pi^{\phi_2}(i) \pi^{\phi_1}(j) P_t^{\phi_1} f(j) - \sum_{i=1}^N \sum_{j=1}^N \pi^{\phi_1}(j) \pi^{\phi_2}(i) P_t^{\phi_2} f(i) \Big| \\ &\leq \sum_{i,j\in S} \pi^{\phi_1}(j) \pi^{\phi_2}(i) \Big| P_t^{\phi_1} f(j) - P_t^{\phi_2} f(i) \Big| \\ &\leq \sum_{i,j\in S} \pi^{\phi_1}(j) \pi^{\phi_2}(i) \Big| P_t^{\phi_1} f(j) - P_t^{\phi_2} f(j) \Big| + \sum_{i,j\in S} \pi^{\phi_1}(j) \pi^{\phi_2}(i) \Big| P_t^{\phi_2} f(j) - \pi^{\phi_2} f \Big|\end{aligned}$$

$$+\sum_{i,j\in S}\pi^{\phi_1}(j)\pi^{\phi_2}(i)\left|P_t^{\phi_2}f(i)-\pi^{\phi_2}f\right|$$
$$\le K\|\phi_1-\phi_2\|_\infty(1-e^{-\lambda t})+L_3e^{-\lambda t}.$$

Finally, letting $t\to\infty$ and taking supremum over $|f|\le 1$, it follows that

$$\|\pi^{\phi_1}-\pi^{\phi_2}\|_{var}\le K\|\phi_1-\phi_2\|_\infty.$$

This completes the proof. □

By using the invariant measure π^ϕ, let us define

$$\bar b(\phi)=\sum_{i\in S}b(\phi,i)\pi^\phi(i)\quad\text{and}\quad\bar\Sigma(\phi)=\sum_{i\in S}\sigma(\phi,i)\sigma'(\phi,i)\pi^\phi(i).$$

Obviously, $\bar b(\cdot)$ satisfies the linear growth condition. According to the definition of $\bar\Sigma(\cdot)$, we can yield that for ϕ, ϕ_1, and $\phi_2\in C([-\tau,0];\mathbb{R}^n)$

$$|\bar b(\phi_1)-\bar b(\phi_2)|\le K(1+\|\phi_1\|_\infty^2+\|\phi_2\|_\infty^2)\|\phi_1-\phi_2\|_\infty, \tag{4.12}$$
$$|\bar\Sigma(\phi)|\le\sum_{i\in S}|\sigma(\phi,i)|^2\pi^\phi(i)\le K(1+\|\phi\|_\infty^2) \tag{4.13}$$

and

$$|\bar\Sigma(\phi_1)-\bar\Sigma(\phi_2)|\le K(1+\|\phi_1\|_\infty^2+\|\phi_2\|_\infty^2)\|\phi_1-\phi_2\|_\infty. \tag{4.14}$$

Let us introduce the following equation:

(A7) The following equation,

$$dX(t)=\bar b(X_t)dt+\bar\sigma(X_t)d\tilde w(t) \tag{4.15}$$

has a solution that is unique in the weak sense (i.e., uniqueness in the sense of the distribution) on $[0,T]$ for the same initial date

$$X(t)=\xi\in C([-\tau,0],\mathbb{R}^n)$$

as Eq (2.1), where $\tilde w$ is a standard Brownian motion and

$$\bar\sigma(\cdot)\bar\sigma'(\cdot)=\bar\Sigma(\cdot).$$

5. Weak convergence and asymptotic approximation

This section will show that X^ε weakly converges to X determined by (4.15). To prove this claim, tightness of X^ε is needed.

Theorem 5.1. *Under assumptions (A1)–(A5),* $\{X^\varepsilon(\cdot)\}_{\varepsilon\in(0,1)}$ *is tight on* $C([0,T];\mathbb{R}^n)$.

To obtain tightness, we need the following sufficient condition for tightness (see [30, p.64]).

Lemma 5.2. *If* $\{X^\varepsilon\}_{\varepsilon\in(0,1)} \in C([0,T];\mathbb{R}^n)$ *satisfy that, for some positive constants* α, β, ν,

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}|X^\varepsilon(0)|^\nu < \infty, \tag{5.1a}$$

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}|X^\varepsilon(t) - X^\varepsilon(s)|^\alpha \le K|t-s|^{1+\beta};\ 0 \le s < t \le T, \tag{5.1b}$$

then the probability measures induced by X^ε *form a tight sequence.*

Proof of Theorem 5.1. Because

$$X^\varepsilon(0) = \xi(0)$$

is independent of ε, it is sufficient to check (5.1b). Hence, for any $p \ge 2$, $\delta > 0$, and $0 \le s < t \le T$, it follows from (2.1) that

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\le t\le s+\delta} |X^\varepsilon(t) - X^\varepsilon(s)|^p\Big] &\le K\Big\{\mathbb{E}\Big[\sup_{s\le t\le s+\delta}\Big|\int_s^t b(X_r^\varepsilon, \alpha^\varepsilon(r))dr\Big|^p\Big] + \mathbb{E}\Big[\sup_{s\le t\le s+\delta}\Big|\int_s^t \sigma(X_r^\varepsilon, \alpha^\varepsilon(r))dW(r)\Big|^p\Big]\Big\}\\ &\le K\delta^{p-1}\int_s^t \mathbb{E}|b(X_r^\varepsilon, \alpha^\varepsilon(r))|^p dr + K\delta^{\frac{p-2}{2}}\int_s^t \mathbb{E}|\sigma(X_r^\varepsilon, \alpha^\varepsilon(r))|^p dr,\end{aligned}$$

which, combined with assumption (**A2**) and (2.10), implies that

$$\mathbb{E}\Big[\sup_{s\le t\le s+\delta} |X^\varepsilon(t) - X^\varepsilon(s)|^p\Big] \le K\delta^{\frac{p}{2}}. \tag{5.2}$$

Then letting

$$p = 4 > 2 \quad \text{and} \quad \delta = t - s$$

gives that (5.1b) holds. □

According to tightness of $\{X^\varepsilon(\cdot)\}$, by the Prohorov theorem (see [31, p.59]), there exists $X(\cdot)$ defined on $C([-\tau,T];\mathbb{R}^n)$ and the subsequence of ε (without loss of generality, we still denote the superscript of convergent subsequence as ε) such that

$$X^\varepsilon(t) \Rightarrow X(t),$$

as $\varepsilon \to 0$. To proceed, we shall need the following version of the Arzelá-Ascoli theorem (see [30, p.63]) and the moment estimation of the segment process:

Lemma 5.3. $\{X^\varepsilon\}_{\varepsilon\in(0,1)} \in C([-\tau,T];\mathbb{R}^n)$ *is tight if and only if*

$$\lim_{\mu\uparrow\infty}\sup_{\varepsilon\in(0,1)} \mathbb{P}(|X^\varepsilon(0)| > \mu) = 0, \tag{5.3a}$$

$$\lim_{\delta\downarrow 0}\sup_{\varepsilon\in(0,1)} \mathbb{P}\Big(\Lambda_{\delta,\varepsilon} > \gamma_1\Big) = 0 \ \text{ for any }\ \gamma_1 > 0, \tag{5.3b}$$

where we define

$$\Lambda_{\delta,\varepsilon} = \sup_{\substack{s,t\in[-\tau,T]\\ |s-t|<\delta}} |X^\varepsilon(t) - X^\varepsilon(s)|.$$

Lemma 5.4. *For $p > 2$, $0 < \delta < 1$, and $\gamma_2 < \frac{p}{2} - 1$, we have*

$$\mathbb{E}\Big[\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta) - X_t^\varepsilon(\theta)|\Big] \le K\delta^{\frac{1}{2}-\frac{1+\gamma_2}{p}}.$$

Proof. Let

$$N_1 = [\frac{T+\tau}{\delta}], \quad t_m = -\tau + m\delta,$$

$m = 0, 1, \ldots, N_1$, and $t_{N_1+1} = T$. Denote

$$\Xi = \max_{1\le i\le N_1+1}\sup_{t_{i-1}\le s\le t_i}|X^\varepsilon(s) - X^\varepsilon(t_{i-1})|.$$

Due to $|s-t| < \delta$, s and t either fall into the same interval

$$I_i := [t_{i-1}, t_i]$$

or into different adjacent intervals I_i and

$$I_{i+1} := [t_i, t_{i+1}].$$

If s and t fall into the same interval I_i, then

$$|X^\varepsilon(s) - X^\varepsilon(t)| \le |X^\varepsilon(s) - X^\varepsilon(t_{i-1})| + |X^\varepsilon(t) - X^\varepsilon(t_{i-1})| \le 2\Xi.$$

If s and t fall into different adjacent intervals I_i and I_{i+1}, then

$$|X^\varepsilon(s) - X^\varepsilon(t)| \le |X^\varepsilon(s) - X^\varepsilon(t_{i-1})| + |X^\varepsilon(t_i) - X^\varepsilon(t_{i-1})| + |X^\varepsilon(t) - X^\varepsilon(t_i)| \le 3\Xi.$$

From this, it can be concluded that

$$\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}|X^\varepsilon(t) - X^\varepsilon(s)| \le 3\max_{1\le i\le N_1+1}\sup_{t_{i-1}\le s\le t_i}|X^\varepsilon(s) - X^\varepsilon(t_{i-1})|, \tag{5.4}$$

which implies that for $\beta > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta) - X_s^\varepsilon(\theta)| > \beta\Big) &\le \mathbb{P}\Big(\sup_{\substack{s,t\in[-\tau,T]\\|s-t|<\delta}}|X^\varepsilon(t) - X^\varepsilon(s)| > \beta\Big)\\
&\le \mathbb{P}\Big(\max_{1\le i\le N_1+1}\sup_{t_{i-1}\le s\le t_i}|X^\varepsilon(s) - X^\varepsilon(t_{i-1})| > \frac{\beta}{3}\Big)\\
&\le \sum_{i=1}^{N_1+1}\mathbb{P}\Big(\sup_{t_{i-1}\le s\le t_i}|X^\varepsilon(s) - X^\varepsilon(t_{i-1})| > \frac{\beta}{3}\Big)\\
&\le (N_1+1)\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t) - X^\varepsilon(s)| > \frac{\beta}{3}\Big)\\
&\le \frac{K}{\delta}\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t) - X^\varepsilon(s)| > \frac{\beta}{3}\Big).
\end{aligned} \tag{5.5}$$

By using the in [32, Corollary 2] and (5.5), one can derive that

$$\begin{aligned}\mathbb{E}\Big[\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta)-X_s^\varepsilon(\theta)|^p\Big] &= p\int_0^\infty \beta^{p-1}\mathbb{P}\Big(\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta)-X_s^\varepsilon(\theta)|^p>\beta\Big)d\beta\\
&\le Kp\int_0^\infty \beta^{p-1}\frac{1}{\delta}\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|>\frac{\beta}{3}\Big)d\beta\\
&= Kp\int_0^\delta \beta^{p-1}\frac{1}{\delta}\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|>\frac{\beta}{3}\Big)d\beta\\
&\quad + Kp\int_\delta^1 \beta^{p-1}\frac{1}{\delta}\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|>\frac{\beta}{3}\Big)d\beta\\
&\quad + Kp\int_1^\infty \beta^{p-1}\frac{1}{\delta}\sup_{-\tau\le s\le T}\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|>\frac{\beta}{3}\Big)d\beta\\
&=: \Upsilon_1+\Upsilon_2+\Upsilon_3.\end{aligned}$$

It is easy to observe that

$$\Upsilon_1 \le K\delta^{-1}\int_0^\delta p\beta^{p-1}d\beta = K\delta^{p-1}.$$

Applying the Chebyshev inequality, assumption (**A1**) and (5.2) give that for $q\ge 2$

$$\mathbb{P}\Big(\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|>\frac{\beta}{3}\Big)\le\frac{K\mathbb{E}\Big[\sup_{s\le t\le s+\delta}|X^\varepsilon(t)-X^\varepsilon(s)|^q\Big]}{\beta^q}\le K\frac{\delta^{\frac{q}{2}}}{\beta^q}. \tag{5.6}$$

In the above inequality, letting $q=p$ yields that

$$\Upsilon_2\le Kp\int_\delta^1\beta^{p-1}\frac{\delta^{\frac{p}{2}-1}}{\beta^p}d\beta = K\delta^{\frac{p}{2}-1}\ln\frac{1}{\delta}.$$

Similarly, letting $q=2p$ gives

$$\Upsilon_3\le Kp\int_1^\infty\beta^{p-1}\frac{\delta^{p-1}}{\beta^{2p}}d\beta = K\delta^{p-1}.$$

Note that

$$\gamma_2<\frac{p}{2}-1$$

and

$$0<\delta<1,\quad \lim_{\delta\to 0}\delta^{\gamma_2}\ln\frac{1}{\delta}=0.$$

From this, it can be concluded that

$$\Upsilon_2\le K\delta^{\frac{p}{2}-1-\gamma_2},$$

which implies that

$$\mathbb{E}\Big[\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta)-X_s^\varepsilon(\theta)|^p\Big]\le K\delta^{\frac{p}{2}-1-\gamma_2}.$$

This, together with the Lyapunov inequlity, yields that

$$\mathbb{E}\Big[\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]}|X_t^\varepsilon(\theta)-X_s^\varepsilon(\theta)|\Big]\le K\delta^{\frac{1}{2}-\frac{1+\gamma_2}{p}}.$$

This completes the proof. □

Now, let us state the main result of this article.

Theorem 5.5. *Under assumptions **(A1)–(A7)**, the limit of any weakly convergent subsequence of the process $\{X^\varepsilon(\cdot)\}_{\varepsilon\in(0,1)}$ satisfies the Eq (4.15) with the same value*

$$X_0^\varepsilon=\xi\in C([-\tau,0];\mathbb{R}^n).$$

Proof. For any $f\in C_0^\infty(\mathbb{R}^n,\mathbb{R})$, applying the Itô formula to $f(X^\varepsilon(t))$ for Eq (2.1)yields that

$$\begin{aligned}M_f^\varepsilon(t):&=f(X^\varepsilon(t))-f(X^\varepsilon(0))-\int_0^t\mathbb{L}^\varepsilon(X_r^\varepsilon,\alpha^\varepsilon(r))f(X^\varepsilon(r))dr\\&=\int_0^t f_x(X^\varepsilon(r))\sigma(X_r^\varepsilon,\alpha^\varepsilon(r))dW(r)\end{aligned}\tag{5.7}$$

is a martingale, where

$$\mathbb{L}^\varepsilon(X_r^\varepsilon,\alpha^\varepsilon(r))f(X^\varepsilon(r))=f_x(X^\varepsilon(r))b(X_r^\varepsilon,\alpha^\varepsilon(r))+\frac{1}{2}\sum_{i,j=1}^n\Sigma_{ij}(X_r^\varepsilon,\alpha^\varepsilon(r))f_{x_ix_j}(X^\varepsilon(r))$$

and

$$\Sigma_{ij}(X_r^\varepsilon,\alpha^\varepsilon(r))=\sigma_i(X_r^\varepsilon,\alpha^\varepsilon(r))\sigma_j(X_r^\varepsilon,\alpha^\varepsilon(r)).$$

This is equivalent to

$$\mathbb{E}\Big[(h(X^\varepsilon(s_i)),i\le k)\Big(f(X^\varepsilon(t))-f(X^\varepsilon(s))-\int_s^t\mathbb{L}^\varepsilon(X_r^\varepsilon,\alpha^\varepsilon(r))f(X^\varepsilon(r))dr\Big)\Big]=0,\tag{5.8}$$

for arbitrary k, s, and t with $s_1<s_2<\ldots<s_k<s<t$, and any bounded and continuous function $h(\cdot)$. To characterize the limit process $\{X(t)\}_{t\ge0}$, it suffices to show that letting $\varepsilon\to0$ on both sides of (5.8),

$$\begin{aligned}0&=\lim_{\varepsilon\to0}\mathbb{E}\Big[(h(X^\varepsilon(s_i)),i\le k)\Big(f(X^\varepsilon(t))-f(X^\varepsilon(s))-\int_s^t\mathbb{L}^\varepsilon(X_r^\varepsilon,\alpha^\varepsilon(r))f(X^\varepsilon(r))dr\Big)\Big]\\&=\mathbb{E}\Big[(h(X(s_i)),i\le k)\Big(f(X(t))-f(X(s))-\int_s^t\mathcal{L}(X_r)f(X(r))dr\Big)\Big],\end{aligned}\tag{5.9}$$

where

$$\mathcal{L}(X_r)f(X(r))=f_x(X(r))\bar{b}(X_r)+\frac{1}{2}\sum_{i,j=1}^n\bar{\sigma}_i(X_r)\bar{\sigma}_j(X_r)f_{x_ix_j}(X(r)).$$

By the Skorohod representation theorem ([32, p.354]) and Theorem 5.1, we may assume for $s \in [0,T]$, $X^\varepsilon(s) \to X(s)$ in the sense of w.p.1 as $\varepsilon \to 0$. This, together with the Lebesgue dominated convergence theorem, yields that

$$\mathbb{E}[h(X^\varepsilon(s_i), i \le k)(f(X^\varepsilon(t)) - f(X^\varepsilon(s)))] \to \mathbb{E}[h(X(s_i), i \le k)(f(X(t)) - f(X(s)))] \tag{5.10}$$

for all $0 \le s < t$. Furthermore, by Vitali convergence theorem (refer to [33] and the related literature for further details) and Theorem 2.10, we can obtain

$$\mathbb{E}\Big[\sup_{0\le s\le T} |X^\varepsilon(s)|^4\Big] \longrightarrow \mathbb{E}\Big[\sup_{0\le s\le T} |X(s)|^4\Big] \le K. \tag{5.11}$$

Next, we only need to show

$$\lim_{\varepsilon\to 0} \mathbb{E}\Big[h(X^\varepsilon(s_i), i \le k) \int_s^t \mathbb{L}^\varepsilon(X^\varepsilon_r, \alpha^\varepsilon(r)) f(X^\varepsilon(r))dr\Big] = \mathbb{E}\Big[h(X(s_i), i \le k) \int_s^t \mathcal{L}(X_r) f(X(r))dr\Big].$$

According to the definition of $\mathbb{L}^\varepsilon$ and $\mathcal{L}$, we shall only consider

$$I_1 := \mathbb{E}\Big[\int_s^t |f_x(X^\varepsilon(r))b(X^\varepsilon_r, \alpha^\varepsilon(r)) - f_x(X(r))\bar{b}(X_r)|dr\Big]$$

and

$$I_2 := \sum_{k,l=1}^n \mathbb{E}\Big[\int_s^t |f_{x_k x_l}(X^\varepsilon(r))\sigma_k(X^\varepsilon_r, \alpha^\varepsilon(r))\sigma_l(X^\varepsilon_r, \alpha^\varepsilon(r)) - f_{x_k x_l}(X(r))\bar{\sigma}_k(X_r)\bar{\sigma}_l(X_r)|dr\Big],$$

where we define

$$I_2 =: \sum_{k,l=1}^n I_2^{k,l}.$$

For $\delta > 0$, set

$$N = [\frac{t-s}{\delta}], \quad s_m = s + m\delta$$

for $m = 0, 1, \ldots, N$ and

$$s_{N+1} = t.$$

Hence,

$$\begin{aligned}
I_1 \le\ & \mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(r))b(X^\varepsilon_r, \alpha^\varepsilon(r)) - f_x(X^\varepsilon(s_m))b(X^\varepsilon_{s_m}, \alpha^\varepsilon(r))|dr\Big] \\
& + \mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))b(X^\varepsilon_{s_m}, \alpha^\varepsilon(r)) - f_x(X^\varepsilon(s_m))\bar{b}(X^\varepsilon_{s_m})|dr\Big] \\
& + \mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))\bar{b}(X^\varepsilon_{s_m}) - f_x(X(r))\bar{b}(X_r)|dr\Big] \\
=:\ & I_{11} + I_{12} + I_{13}.
\end{aligned}$$

Recall that $f \in C_0^\infty(\mathbb{R}^n, \mathbb{R})$. This and assumption (**A2**) give

$$\begin{aligned} I_{11} &\le \mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(r))b(X_r^\varepsilon, \alpha^\varepsilon(r)) - f_x(X^\varepsilon(s_m))b(X_r^\varepsilon, \alpha^\varepsilon(r))|dr\Big] \\ &\quad + \mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))b(X_r^\varepsilon, \alpha^\varepsilon(r)) - f_x(X^\varepsilon(s_m))b(X_{s_m}^\varepsilon, \alpha^\varepsilon(r))|dr\Big] \\ &\le K\mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} (|X^\varepsilon(r) - X^\varepsilon(s_m)|(1 + \|X_r^\varepsilon\|_\infty) + \|X_r^\varepsilon - X_{s_m}^\varepsilon\|_\infty)\Big]. \end{aligned}$$

Consequently, by virtual of (2.10), (5.2), and the Hölder inequality, we arrive at

$$I_{11} \le K\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} ((\mathbb{E}|X^\varepsilon(r) - X^\varepsilon(s_m)|^2)^{\frac{1}{2}} + \mathbb{E}\|X_r^\varepsilon - X_{s_m}^\varepsilon\|_\infty)dr.$$

According to the tightness of $\{X^\varepsilon(\cdot)\}$ and (5.3b), we obtain that $\Lambda_{\delta,\varepsilon} \xrightarrow{\mathbb{P}} 0$ uniformly with respect to ε, as $\delta \to 0$. This, together with the Lebesgue dominated convergence theorem, yields that

$$I_{11} \le K(N+1)\delta(\delta^{\frac{1}{2}} + \mathbb{E}\Lambda_{\delta,\varepsilon}) \to 0, \text{ as } \delta \to 0. \tag{5.12}$$

Similar to I_{11}, together with the definition of $\bar{b}(\cdot)$ and (4.12), we have

$$\begin{aligned} I_{13} &\le \mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))\bar{b}(X_{s_m}^\varepsilon) - f_x(X^\varepsilon(s_m))\bar{b}(X_r)|dr\Big] \\ &\quad + \mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))\bar{b}(X_r) - f_x(X(r))\bar{b}(X_r)|dr\Big] \\ &\le K\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}} [(\mathbb{E}\|X_{s_m}^\varepsilon - X_r\|_\infty^2)^{\frac{1}{2}} + (\mathbb{E}|X^\varepsilon(s_m) - X(r)|^2)^{\frac{1}{2}}]dr. \end{aligned}$$

Note that

$$|X^\varepsilon(s_m) - X(r)| \le \|X_{s_m}^\varepsilon - X_r\|_\infty$$

and

$$\begin{aligned} \|X_{s_m}^\varepsilon - X_r\|_\infty &\le \sup_{\substack{s,t\in[-\tau,T] \\ |s-t|<\delta}} |X^\varepsilon(s) - X(t)| \\ &\le \sup_{\substack{s,t\in[-\tau,T] \\ |s-t|<\delta}} |X^\varepsilon(s) - X^\varepsilon(t)| + \sup_{\substack{s,t\in[-\tau,T] \\ |s-t|<\delta}} |X^\varepsilon(t) - X(t)| \\ &\le \Lambda_{\delta,\varepsilon} + \sup_{t\in[-\tau,T]} |X^\varepsilon(t) - X(t)|. \end{aligned} \tag{5.13}$$

Define

$$\Gamma_{\delta,\varepsilon} = \Lambda_{\delta,\varepsilon} + \sup_{t\in[-\tau,T]} |X^\varepsilon(t) - X(t)|.$$

The Lebesgue dominated convergence theorem and (5.11) give

$$I_{13} \leq K\delta(N+1)(\mathbb{E}\Gamma^2_{\delta,\varepsilon})^{\frac{1}{2}} \to 0, \text{ as } \delta,\varepsilon \to 0. \tag{5.14}$$

Now, we construct an auxiliary Markov chain $\{\tilde{\alpha}^m(r)\}_{r\geq s_m}$ in S satisfying the transition rate $(\frac{1}{\varepsilon}q_{ij}(X^\varepsilon_{s_m}))_{i,j\in S}$ and

$$\tilde{\alpha}^m(s_m) = \alpha^\varepsilon(s_m).$$

By virtue of Lemma 3.3, together with assumption (**A6**), we have

$$\int_{s_m}^{s_{m+1}} \mathbb{E}\Big[\mathbb{I}_{\{\alpha^\varepsilon(r)\neq\tilde{\alpha}^m(r)\}}|\mathcal{F}^\varepsilon_{s_m}\Big]dr \leq 2N(N-1)\frac{\delta}{\varepsilon}\int_{s_m}^{s_{m+1}} \mathbb{E}(\|X^\varepsilon_r - X^\varepsilon_{s_m}\|_\infty|\mathcal{F}^\varepsilon_{s_m})dr. \tag{5.15}$$

Then,

$$\begin{aligned} I_{12} &\leq K\mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))b(X^\varepsilon_{s_m},\alpha^\varepsilon(r)) - f_x(X^\varepsilon(s_m))b(X^\varepsilon_{s_m},\tilde{\alpha}^m(r))|dr\Big] \\ &\quad + K\mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |f_x(X^\varepsilon(s_m))b(X^\varepsilon_{s_m},\tilde{\alpha}^m(r)) - f_x(X^\varepsilon(s_m))\bar{b}(X^\varepsilon_{s_m})|dr\Big] \\ &=: I_{12,1} + I_{12,2}. \end{aligned}$$

Applying (5.15), the Hölder inequality, assumption (**A2**), and Theorem 2.2, we obtain that

$$\begin{aligned} I_{12,1} &\leq K\mathbb{E}\Big[\sum_{m=0}^N \int_{s_m}^{s_{m+1}} |b(X^\varepsilon_{s_m},\alpha^\varepsilon(r)) - b(X^\varepsilon_{s_m},\tilde{\alpha}^m(r))|\mathbb{I}_{\{\alpha^\varepsilon(r)\neq\tilde{\alpha}^m(r)\}}dr\Big] \\ &\leq K\sum_{m=0}^N \Big(\int_{s_m}^{s_{m+1}} \mathbb{E}|b(X^\varepsilon_{s_m},\alpha^\varepsilon(r)) - b(X^\varepsilon_{s_m},\tilde{\alpha}^m(r))|^2\Big)^{\frac{1}{2}}\Big(\mathbb{E}\int_{s_m}^{s_{m+1}} \mathbb{E}[\mathbb{I}_{\{\alpha^\varepsilon(r)\neq\tilde{\alpha}^m(r)\}}|\mathcal{F}^\varepsilon_{s_m}]dr\Big)^{\frac{1}{2}} \\ &\leq K\sum_{m=0}^N \delta^{\frac{1}{2}}\Big(\frac{\delta}{\varepsilon}\int_{s_m}^{s_{m+1}} \mathbb{E}\|X^\varepsilon_r - X^\varepsilon_{s_m}\|_\infty dr\Big)^{\frac{1}{2}}. \end{aligned} \tag{5.16}$$

Next, in Lemma 5.4, letting

$$p = 4$$

and

$$\gamma_2 = \frac{1}{3}$$

gives

$$\mathbb{E}\|X^\varepsilon_r - X^\varepsilon_{s_m}\|_\infty \leq \mathbb{E}\Big[\sup_{\substack{s,t\in[0,T]\\|s-t|<\delta}}\sup_{\theta\in[-\tau,0]} |X^\varepsilon_t(\theta) - X^\varepsilon_s(\theta)|\Big] \leq K\delta^{\frac{1}{6}}. \tag{5.17}$$

Substituting (5.17) into (5.16) yields that

$$I_{12,1} \leq K(N+1)\delta\frac{\delta^{\frac{7}{12}}}{\varepsilon^{\frac{1}{2}}}. \tag{5.18}$$

According to (4.4), one can derive that

$$
\begin{aligned}
I_{12,2} &\le K\mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}}\mathbb{E}\Big(|b(X^{\varepsilon}_{s_m},\tilde{\alpha}^m(r))-\bar{b}(X^{\varepsilon}_{s_m})|\Big|\mathcal{F}^{\varepsilon}_{s_m}\Big)dr\Big]\\
&\le K\mathbb{E}\Big[\sum_{m=0}^{N}\int_{s_m}^{s_{m+1}}(1+\|X^{\varepsilon}_{s_m}\|_\infty)\mathbb{E}\Big(\Big\|P^{X^{\varepsilon}_{s_m}}_{\frac{r-s_m}{\varepsilon}}(\alpha^{\varepsilon}(s_m),\cdot)-\pi^{X^{\varepsilon}_{s_m}}\Big\|_{var}\Big|\mathcal{F}^{\varepsilon}_{s_m}\Big)dr\Big]\\
&= K\mathbb{E}\Big[\sum_{m=0}^{N}\varepsilon\int_{0}^{\frac{\delta}{\varepsilon}}(1+\|X^{\varepsilon}_{s_m}\|_\infty)\mathbb{E}(\|P^{X^{\varepsilon}_{s_m}}_{r}(\alpha^{\varepsilon}(s_m),\cdot)-\pi^{X^{\varepsilon}_{s_m}}\|_{var}|\mathcal{F}^{\varepsilon}_{s_m})dr\Big]\\
&\le K\varepsilon(N+1)\frac{L_3}{\lambda}(1-e^{-\frac{\delta}{\varepsilon}\lambda}). \qquad (5.19)
\end{aligned}
$$

For the same interval division $[s_m, s_{m+1}], m = 1, 2, \dots, N$,

$$
\begin{aligned}
I^{k,l}_2 \le &\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\sigma_k(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))\\
&-\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(s_m))|dr\Big)\\
&+\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(s_m))\\
&-\bar{\sigma}_k(X^{\varepsilon}_{s_m})\bar{\sigma}_l(X^{\varepsilon}_{s_m})f_{x_kx_l}(X^{\varepsilon}(s_m))|dr\Big)\\
&+\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\bar{\sigma}_k(X^{\varepsilon}_{s_m})\bar{\sigma}_l(X^{\varepsilon}_{s_m})f_{x_kx_l}(X^{\varepsilon}(s_m))-\bar{\sigma}_k(X_r)\bar{\sigma}_l(X_r)f_{x_kx_l}(X(r))|dr\Big)\\
=:&\ I^{k,l}_{2,1}+I^{k,l}_{2,2}+I^{k,l}_{2,3}. \qquad (5.20)
\end{aligned}
$$

By the Hölder inequality, assumption (**A2**), Theorem 2.2, and the fact $f \in C^{\infty}_0(\mathbb{R}^n,\mathbb{R})$, combined with the result that $\Lambda_{\delta,\varepsilon} \xrightarrow{\mathbb{P}} 0$ uniformly with respect to ε, as $\delta \to 0$, we have

$$
\begin{aligned}
I^{k,l}_{2,1} \le &\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\sigma_k(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))\\
&-\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))|dr\Big)\\
&+\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_r,\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))\\
&-\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))|dr\Big)\\
&+\sum_{m=0}^{N}\mathbb{E}\Big(\int_{s_m}^{s_{m+1}}|\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(r))\\
&-\sigma_k(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))\sigma_l(X^{\varepsilon}_{s_m},\alpha^{\varepsilon}(r))f_{x_kx_l}(X^{\varepsilon}(s_m))|dr\Big)\\
\le &\ K(N+1)\delta(\mathbb{E}\Lambda^2_{\delta,\varepsilon})^{\frac{1}{2}} \to 0, \qquad (5.21)
\end{aligned}
$$

as $\delta \to 0$, where the convergence is derived from the Lebesgue dominated convergence theorem.

According to (4.13) and (4.14), it follows that

$$\begin{aligned}
I_{2,3}^{k,l} &\leq \sum_{m=0}^{N} \mathbb{E}\Big(\int_{s_m}^{s_{m+1}} |\bar{\Sigma}_{k,l}(X_{s_m}^{\varepsilon}) f_{x_k x_l}(X^{\varepsilon}(s_m)) - \bar{\Sigma}_{k,l}(X_{s_m}^{\varepsilon}) f_{x_k x_l}(X(r))| dr\Big) \\
&\quad + \sum_{m=0}^{N} \mathbb{E}\Big(\int_{s_m}^{s_{m+1}} |\bar{\Sigma}_{k,l}(X_{s_m}^{\varepsilon}) f_{x_k x_l}(X(r)) - \bar{\Sigma}_{k,l}(X_r) f_{x_k x_l}(X(r))| dr\Big) \\
&\leq K \sum_{m=0}^{N} \int_{s_m}^{s_{m+1}} (1 + \mathbb{E}\|X_{s_m}^{\varepsilon}\|_{\infty}^{4} + \mathbb{E}\|X_r\|_{\infty}^{4})^{\frac{1}{2}} (\mathbb{E}\|X_{s_m}^{\varepsilon} - X_r\|_{\infty}^{2})^{\frac{1}{2}} dr,
\end{aligned}$$

which, combined with (5.13), gives

$$I_{2,3}^{k,l} \leq K\delta(N+1)(\mathbb{E}\Gamma_{\delta,\varepsilon}^{2})^{\frac{1}{2}} \to 0, \text{ as } \delta, \varepsilon \to 0. \tag{5.22}$$

As for $I_{2,2}^{k,l}$, recalling the definition of $\tilde{\alpha}^m(t)$, we have

$$\begin{aligned}
I_{2,2}^{k,l} &\leq \sum_{m=0}^{N} \mathbb{E}\Big(\int_{s_m}^{s_{m+1}} |\sigma_k(X_{s_m}^{\varepsilon}, \alpha^{\varepsilon}(r))\sigma_l(X_{s_m}^{\varepsilon}, \alpha^{\varepsilon}(r)) - \sigma_k(X_{s_m}^{\varepsilon}, \tilde{\alpha}^m(r))\sigma_l(X_{s_m}^{\varepsilon}, \tilde{\alpha}^m(r))| dr\Big) \\
&\quad + \sum_{m=0}^{N} \mathbb{E}\Big(\int_{s_m}^{s_{m+1}} |\sigma_k(X_{s_m}^{\varepsilon}, \tilde{\alpha}^m(r))\sigma_l(X_{s_m}^{\varepsilon}, \tilde{\alpha}^m(r)) - \bar{\sigma}(X_{s_m}^{\varepsilon})\bar{\sigma}(X_{s_m}^{\varepsilon})| dr\Big) \\
&=: I_{2,21}^{k,l} + I_{2,22}^{k,l}.
\end{aligned}$$

Similar to (5.16), using (5.15) yields that

$$\begin{aligned}
I_{2,21}^{k,l} &\leq \sum_{m=0}^{N} K\Big(\int_{s_m}^{s_{m+1}} \mathbb{E}(1 + \|X_{s_m}^{\varepsilon}\|_{\infty}^{2})^2\Big)^{\frac{1}{2}} \Big(\mathbb{E}\int_{s_m}^{s_{m+1}} \mathbb{E}[\mathbb{I}_{\{\alpha^{\varepsilon}(r) \neq \tilde{\alpha}^m(r)\}} | \mathcal{F}_{s_m}^{\varepsilon}] dr\Big)^{\frac{1}{2}} \\
&\leq K(N+1)\delta \frac{\delta^{\frac{7}{12}}}{\varepsilon^{\frac{1}{2}}}.
\end{aligned} \tag{5.23}$$

Furthermore, as a result of (4.4), we obtain that

$$\begin{aligned}
I_{2,22}^{k,l} &\leq \sum_{m=0}^{N} \mathbb{E}\Big(\int_{s_m}^{s_{m+1}} (1 + \|X_{s_m}^{\varepsilon}\|_{\infty})^2 \mathbb{E}\Big(\Big\|P_{\frac{r-s_m}{\varepsilon}}^{X_{s_m}^{\varepsilon}}(\alpha^{\varepsilon}(s_m), \cdot) - \pi^{X_{s_m}^{\varepsilon}}\Big\|_{var} \Big| \mathcal{F}_{s_m}^{\varepsilon}\Big) dr\Big) \\
&= K\mathbb{E}\Big[\sum_{m=0}^{N} \varepsilon \int_0^{\frac{\delta}{\varepsilon}} (1 + \|X_{s_m}^{\varepsilon}\|_{\infty})^2 \mathbb{E}(\|P_r^{X_{s_m}^{\varepsilon}}(\alpha^{\varepsilon}(s_m), \cdot) - \pi^{X_{s_m}^{\varepsilon}}\|_{var} | \mathcal{F}_{s_m}^{\varepsilon}) dr\Big] \\
&\leq K\varepsilon(N+1)\frac{L_3}{\lambda}(1 - e^{-\frac{\delta}{\varepsilon}\lambda}).
\end{aligned} \tag{5.24}$$

Finally, according to (5.12), (5.14), (5.18), (5.19), and (5.21)–(5.24), we can get

$$I_1 + I_2 \leq K(N+1)\delta[(\mathbb{E}\Lambda_{\delta,\varepsilon}^{2})^{\frac{1}{2}} + (\mathbb{E}\Gamma_{\delta,\varepsilon}^{2})^{\frac{1}{2}} + \delta^{\frac{7}{12}}\varepsilon^{-\frac{1}{2}}] + K\varepsilon(N+1)\frac{L_3}{\lambda}(1 - e^{-\frac{\delta}{\varepsilon}\lambda}).$$

To obtain the desired conclusion, let $\delta = \varepsilon^{\frac{11}{12}}$ in the inequality above. □

6. Examples

The two examples provided in this section cannot be validated using results from the classical literature due to the presence of past-dependent switching. We will verify them one by one to ensure they meet the assumptions proposed in this paper, allowing us to derive the averaged equations from Theorem 5.5. To proceed, consider a special two-state switching process with generator Q, that is,

$$S = \{1, 2\}.$$

Set

$$Q = \begin{pmatrix} -b_1 & b_1 \\ a_1 & -a_1 \end{pmatrix}.$$

It is easy to obtain that the stationary distribution is

$$\nu := (\frac{a_1}{a_1 + b_1}, \frac{b_1}{a_1 + b_1}).$$

Example 6.1. *Consider the following one-dimensional two-timescale stochastic integral differential equation:*

$$dX^\varepsilon(t) = A(\alpha^\varepsilon(t)) f\Big(\int_{-\tau}^{0} X^\varepsilon(t+\theta)d\theta\Big)dt + B(\alpha^\varepsilon(t)) g\Big(\int_{-\tau}^{0} X^\varepsilon(t+\theta)d\theta\Big)dB_1(t), \tag{6.1}$$

where $B_1(t)$ is a standard Brownian motion,

$$X_0^\varepsilon = \xi \in C([-\tau, 0]; \mathbb{R}^n)$$

is nonrandom and satisfies the Lipschitz property,

$$\alpha^\varepsilon(t) = \alpha(t/\varepsilon),$$
$$\alpha^\varepsilon(0) = 1,$$

and $\alpha(t)$ is a pure jump process taking value in $\{1, 2\}$ and equipping generator

$$\tilde{Q}(\phi) = \begin{pmatrix} -\Big(a + b\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big)\Big) & a + b\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big) \\ c + d\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big) & -\Big(c + d\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big)\Big) \end{pmatrix},$$

for $\phi \in C([-\tau, 0]; \mathbb{R}^n)$ and $a, b, c, d > 0$.

Clearly, for

$$\phi, \psi \in C([-\tau, 0]; \mathbb{R}^n),$$

$$\begin{aligned} \|\tilde{Q}(\phi) - \tilde{Q}(\psi)\|_{l_1} &\le \max\{b, d\}\Big|\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big) - \cos^2\Big(\int_{-\tau}^{0}\psi(\theta)d\theta\Big)\Big| \\ &\le 2\tau\max\{b, d\}\|\phi - \psi\|_\infty, \end{aligned}$$

which implies that $\tilde{Q}(\phi)$ is Lipschitz continuous. According to the beginning of this section, the stationary distribution corresponding to $\tilde{Q}(\phi)$ is

$$\begin{aligned}\tilde{\nu}(\phi) &= (\tilde{\nu}_1(\phi), \tilde{\nu}_2(\phi)) \\ &:= \Bigg(\frac{c + d\cos^2\left(\int_{-\tau}^{0}\phi(\theta)d\theta\right)}{a + c + (b+d)\cos^2\left(\int_{-\tau}^{0}\phi(\theta)d\theta\right)}, \frac{a + b\cos^2\left(\int_{-\tau}^{0}\phi(\theta)d\theta\right)}{a + c + (b+d)\cos^2\left(\int_{-\tau}^{0}\phi(\theta)d\theta\right)} \Bigg).\end{aligned}$$

Furthermore, let

$$f : \mathbb{R} \longmapsto \mathbb{R} \quad \text{and} \quad g : \mathbb{R} \longmapsto \mathbb{R}$$

be Borel measurable and $A(1)$, $A(2)$, $B(1)$, and $B(2)$ be both constants and

$$B(1) = -B(2).$$

Assume that there exists a positive K such that for any $x, y \in \mathbb{R}$,

$$|f(x) - f(y)| \vee |g(x) - g(y)| \leq K|x - y|$$

and

$$|f(x)| \leq K.$$

By virtue of (4.15), we can get

$$dX(t) = F(X_t)dt + G(X_t)dB_2(t), \tag{6.2}$$

where $B_2(t)$ is a standard Brownian motion. Meanwhile, for $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$F(\phi) = (A(1)\tilde{\nu}_1(\phi) + A(2)\tilde{\nu}_2(\phi))f\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big)$$

and

$$G(\phi) = |B(1)|\Big|g\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big)\Big|.$$

It is easy to obtain that $\psi, \phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$|G(\phi) - G(\psi)| \leq |B(1)|\Big|\int_{-\tau}^{0}(\phi(\theta) - \psi(\theta))d\theta\Big| \leq K\|\phi - \psi\|_\infty. \tag{6.3}$$

By calculation, it can be concluded that for $\psi, \phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$\begin{aligned}|\tilde{\nu}_1(\phi) - \tilde{\nu}_1(\psi)| &\leq \frac{c(b+d) + d(a+c)}{(a+c)^2}\Big|\cos^2\Big(\int_{-\tau}^{0}\phi(\theta)d\theta\Big) - \cos^2\Big(\int_{-\tau}^{0}\psi(\theta)d\theta\Big)\Big| \\ &\leq \tau\frac{c(b+d) + d(a+c)}{(a+c)^2}\|\phi - \psi\|_\infty\end{aligned}$$

and

$$|\tilde{\nu}_1(\phi)| \leq \frac{c+d}{a+c}.$$

Similarly,

$$|\tilde{v}_2(\phi) - \tilde{v}_2(\psi)| \le \tau \frac{a(b+d) + b(a+c)}{(a+c)^2} \|\phi - \psi\|_\infty$$

and

$$|\tilde{v}_2(\phi)| \le \frac{a+b}{a+c}.$$

Due to the boundedness and the Lipschitz property of $f(\cdot)$, for $\psi, \phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$\begin{aligned} |F(\phi) - F(\psi)| \le & \Big| f\Big(\int_{-\tau}^{0} \phi(\theta) d\theta \Big) \Big| [|A(1)||\tilde{v}_1(\phi) - \tilde{v}_2(\psi)| + |A(2)||\tilde{v}_2(\phi) - \tilde{v}_2(\psi)|] \\ & + |A(1)\tilde{v}_1(\psi) + A(2)\tilde{v}_2(\psi)| \Big| f\Big(\int_{-\tau}^{0} \phi(\theta) d\theta \Big) - f\Big(\int_{-\tau}^{0} \psi(\theta) d\theta \Big) \Big| \\ \le & K \|\phi - \psi\|_\infty. \end{aligned} \tag{6.4}$$

This, togeher with (6.3) and (6.4) yields the existence and uniqueness of the solution to (4.15). Finally, according to Theorem 5.5, the limit of any weakly convergent subsequence of the solution to (6.1) satisfies (6.2).

Example 6.2. *Consider a one-dimensional two-timescale stochastic delay differential equation:*

$$\begin{aligned} dX^\varepsilon(t) =& (A(\alpha^\varepsilon(t)) + B(\alpha^\varepsilon(t)) h(X^\varepsilon(t-\tau)))dt \\ & + (C(\alpha^\varepsilon(t)) + D(\alpha^\varepsilon(t)) r(X^\varepsilon(t-\tau)))dB_3(t), \end{aligned} \tag{6.5}$$

where $B_3(t)$ is a standard Brownian motion,

$$X_0^\varepsilon = \xi \in C([-\tau, 0]; \mathbb{R}^n)$$

is nonrandom and satisfies the Lipschitz property,

$$\alpha^\varepsilon(t) = \alpha(t/\varepsilon), \quad \alpha^\varepsilon(0) = 1$$

and $\alpha(t)$ is a pure jump process taking value in $\{1, 2\}$ and equipping generator

$$\hat{Q}(\phi) = \begin{pmatrix} -(a + b \cdot e^{-\sup_{-\tau\le\theta\le 0} |\phi(\theta)|}) & a + b \cdot e^{-\sup_{-\tau\le\theta\le 0} |\phi(\theta)|} \\ c + d \cdot e^{-\sup_{-\tau\le\theta\le 0} |\phi(\theta)|} & -(c + d \cdot e^{-\sup_{-\tau\le\theta\le 0} |\phi(\theta)|}) \end{pmatrix},$$

for $\phi \in C([-\tau, 0]; \mathbb{R}^n)$ and $a, b, c, d > 0$.

Clearly, $\hat{Q}(\cdot)$ satisfies assumputions (**A3**) and (**A4**). Next, let us verify that (6.5) satisfies the assumption (**A5**). For $\phi, \psi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$\begin{aligned} \|\hat{Q}(\phi) - \hat{Q}(\psi)\|_{l_1} \le & \max\{b, d\} \cdot \Big| e^{-\sup_{-\tau\le\theta\le 0} |\phi(\theta)|} - e^{-\sup_{-\tau\le\theta\le 0} |\psi(\theta)|} \Big| \\ \le & \max\{b, d\} \cdot \Big| \sup_{-\tau\le\theta\le 0} |\phi(\theta)| - \sup_{-\tau\le\theta\le 0} |\phi(\theta)| \Big| \end{aligned}$$

$$\leq \max\{b, d\} \cdot \sup_{-\tau\leq\theta\leq 0} |\phi(\theta) - \psi(\theta)|.$$

According to the beginning of this section, the stationary distribution corresponding to $\hat{Q}(\phi)$ is

$$\begin{aligned}\hat{\nu}(\phi) &= (\hat{\nu}_1(\phi), \hat{\nu}_2(\phi)) \\ &:= \left(\frac{c + d \cdot e^{-\sup_{-\tau\leq\theta\leq 0}|\phi(\theta)|}}{a + c + (b + d) \cdot e^{-\sup_{-\tau\leq\theta\leq 0}|\phi(\theta)|}}, \frac{a + b \cdot e^{-\sup_{-\tau\leq\theta\leq 0}|\phi(\theta)|}}{a + c + (b + d) \cdot e^{-\sup_{-\tau\leq\theta\leq 0}|\phi(\theta)|}}\right).\end{aligned}$$

By calculation, it follows that for $\psi, \phi \in C([-\tau, 0]; \mathbb{R})$,

$$\begin{aligned}|\hat{\nu}_1(\phi) - \hat{\nu}_1(\psi)| &\leq \frac{c(b + d) + d(a + c)}{(a + c)^2}\left|e^{-\sup_{-\tau\leq\theta\leq 0}|\phi(\theta)|} - e^{-\sup_{-\tau\leq\theta\leq 0}|\psi(\theta)|}\right| \\ &\leq \frac{c(b + d) + d(a + c)}{(a + c)^2}\|\phi - \psi\|_\infty\end{aligned}$$

and

$$|\hat{\nu}_1(\phi)| \leq \frac{c + d}{a + c}.$$

Similarly,

$$|\hat{\nu}_2(\phi) - \hat{\nu}_2(\psi)| \leq \frac{a(b + d) + b(a + c)}{(a + c)^2}\|\phi - \psi\|_\infty$$

and

$$|\hat{\nu}_2(\phi)| \leq \frac{a + b}{a + c}.$$

Furthermore, let

$$h : C([-\tau, 0]; \mathbb{R}) \longmapsto \mathbb{R} \quad \text{and} \quad r : C([-\tau, 0]; \mathbb{R}) \longmapsto \mathbb{R}$$

be both Borel measurable and $A(i), B(i), C(i), D(i), i = 1, 2$ be both constants and

$$C(1) = -C(2), \quad D(1) = -D(2).$$

Assume that there exists a positive constant K such that for any ϕ and $\psi \in C([-\tau, 0]; \mathbb{R})$,

$$|h(\phi(-\tau)) - h(\psi(-\tau))| \vee |r(\phi(-\tau)) - r(\psi(-\tau))| \leq K\|\phi - \psi\|_\infty$$

and

$$|h(\phi(-\tau))| \leq K.$$

According to (4.15), one can derive that

$$dX(t) = H(X_t)dt + R(X_t)dB_4(t), \tag{6.6}$$

where $B_4(t)$ is a standard Brownian motion and for any $\phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$H(\phi(-\tau)) = [A(1) + B(1)h(\phi(-\tau))]\hat{\nu}_1(\phi) + [A(2) + B(2)h(\phi(-\tau))]\hat{\nu}_2(\phi)$$

and

$$R(\phi(-\tau)) = |C(1) + D(1)r(\phi(-\tau))|.$$

Similar to (6.1), according to the definition of $\hat{\nu}(\cdot)$, $r(\cdot)$, and $h(\cdot)$, for $\psi, \phi \in C([-\tau, 0]; \mathbb{R}^n)$,

$$|H(\phi) - H(\psi)| \vee |R(\phi) - R(\psi)| \leq K\|\phi - \psi\|_\infty.$$

which yields the existence and uniqueness of the solution to (6.6). Finally, according to Theorem 5.5, the limit of any weakly convergent subsequence of the solution to (6.5) satisfies (6.6).

7. Conclusions

The article overcomes the difficulties arising from past-dependent switching and the presence of delay terms in continuous dynamic equations. Under the Lipschitz condition defined by the uniform norm, it establishes for the first time the averaging principle for stochastic functional differential equations with past-dependent switching. Inspired by reference [34], there will be a focus on how to establish the averaging principle of this system under non-Lipschitz conditions in future research. In addition, the numerical simulation and stability analysis issues related to this model will also be explored in future research.

Author contributions

Minyu Wu: conceptualization, methodology, investigation, writing–original draft, writing–review and editing; Xizhong Yang: conceptualization, methodology and validation; Feiran Yuan: conceptualization, validation, writing–review and editing; Xuyi Qiu: supervision, writing–review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Here, we would like to express our sincere gratitude to Professor Shao Jinghai and Professor Wu Fuke for their guidance and care during the initial stage of writing this paper.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Y. Li, F. Wu, G. Yin, Asymptotic behavior of gene expression with complete memory and two-time scales based on the chemical Langevin equations, *Discrete Contin. Dyn. Syst. Ser. B*, **24** (2019), 4417–4443. http://doi.org/10.3934/dcdsb.2019125
2. F. Wu, T. Tian, J. B. Rawlings, Approximate method for stochastic chemical kinetics with two-time scales by chemical Langevin equations, *J. Chem. Phys.*, **144** (2016), 174112. http://doi.org/10.1063/1.4948407
3. L. Hu, X. Mao, Y. Shen, Stability and boundedness of nonlinear hybrid stochastic differential delay equations, *Syst. Control Lett.*, **62** (2013), 178–187. http://doi.org/10.1016/j.sysconle.2012.11.009
4. H. Qian, Z. Wen, G. Yin, Numerical solutions for optimal control of stochastic Kolmogorov systems with regime-switching and random jumps, *Stat. Infer. Stoch. Process.*, **25** (2022), 105–125. http://doi.org/10.1007/s11203-021-09267-z

5. G. Yin, Q. Zhang, *Continuous-time Markov chains and applications: a two-time-scale approach*, Springer, 2013. http://doi.org/10.1007/978-1-4614-4346-9
6. R. Wang, X. Li, G. Yin, Asymptotic properties of multi-species Lotka-Volterra models with regime switching involving weak and strong interactions, *J. Nonlinear Sci.*, **30** (2020), 565–601. http://doi.org/10.1007/s00332-019-09583-y
7. G. Yin, X. Mao, C. Yuan, 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. http://doi.org/10.1137/080727191
8. A. Budhiraja, P. Dupuis, A. Ganguly, Large deviations for small noise diffusions in a fast Markovian environment, *Electron. J. Probab.*, **23** (2018), 112. http://doi.org/10.1214/18-EJP228
9. D. T. Nguyen, G. Yin, Asymptotic expansions of solutions of systems of Kolmogorov backward equations for two-time-scale switching diffusions, *Q. Appl. Math.*, **71** (2013), 601–628. http://doi.org/10.1090/S0033-569X-2013-01277-X
10. D. H. Nguyen, G. Yin, Recurrence for switching diffusion with past dependent switching and countable state space, *Math. Control Relat. Fields*, **8** (2018), 879–897. http://doi.org/10.3934/mcrf.2018039
11. D. H. Nguyen, G. Yin, Modeling and analysis of switching diffusion systems: past-dependent switching with a countable state space, *SIAM J. Control Optim.*, **54** (2016), 2450–2477. http://doi.org/10.1137/16M1059357
12. D. H. Nguyen, G. Yin, Recurrence and ergodicity of switching diffusions with past-dependent switching having a countable state space, *Potential Anal.*, **48** (2018), 405–435. http://doi.org/10.1007/s11118-017-9641-y
13. R. Z. Khasminskii, G. Yin, On transition densities of singularly perturbed diffusions with fast and slow components, *SIAM J. Appl. Math.*, **56** (1996), 39–51. http://doi.org/10.1137/S0036139995282906
14. E. Pardoux, A. Y. Veretennikov, On the Poisson equation and diffusion approximation 1, *Ann. Probab.*, **29** (2001), 1061–1085. http://doi.org/10.1214/aop/1015345596
15. E. Pardoux, A. Y. Veretennikov, On the Poisson equation and diffusion approximation 2, *Ann. Probab.*, **31** (2003), 1166–1192. http://doi.org/10.1214/aop/1055425774
16. E. Pardoux, A. Y. Veretennikov, On the Poisson equation and diffusion approximation 3, *Ann. Probab.*, **33** (2005), 1111–1133. http://doi.org/10.1214/009117905000000062
17. M. Röckner, L. Xie, Averaging principle and normal deviations for multiscale stochastic systems, *Commun. Math. Phys.*, **383** (2021), 1889–1937. http://doi.org/10.1007/s00220-021-04069-z
18. M. Röckner, L. Xie, L. Yang, Averaging principle and normal deviations for multi-scale stochastic hyperbolic-parabolic equations, *Stoch. Partial Differ. Equations*, **11** (2023), 869–907. http://doi.org/10.1007/s40072-022-00248-8
19. R. Z. Khasminskii, On the principle of averaging the Itô's stochastic differential equations, *Kybernetika*, **4** (1968), 260–279.
20. H. J. Kushner, *Approximation and weak convergence methods for random processes, with applications to stochastic systems theory*, MIT Press, 1985. http://doi.org/10.2307/2288588

21. M. Han, B. Pei, An averaging principle for stochastic evolution equations with jumps and random time delays, *AIMS Math.*, **6** (2021), 39–51. http://doi.org/10.3934/math.2021003

22. D. Liu, Strong convergence of principle of averaging for multiscale stochastic dynamical systems, *Commun. Math. Sci.*, **8** (2010), 999–1020. http://doi.org/10.4310/CMS.2010.v8.n4.a11

23. J. Shao, Strong solutions and strong Feller properties for regime-switching diffusion processes in an infinite state space, *SIAM J. Control Optim.*, **53** (2025), 2462–2479. http://doi.org/10.1137/15M1013584

24. J. Shao, K. Zhao, Continuous dependence for stochastic functional differential equations with state-dependent regime-switching on initial values, *Acta Math. Sin. Engl. Ser.*, **37** (2021), 389–407. http://doi.org/10.1007/s10114-020-9205-8

25. N. Ikeda, S. Watanabe, *Stochastic differential equations and diffusion processes*, 2 Eds., North-Holland Publishing Company, 1989.

26. R. Situ, *Theory of stochastic differential equations with jumps and applications*, Springer, 2005. http://doi.org/10.1007/b106901

27. X. Mao, *Stochastic differential equations and applications*, 2 Eds., Elsevier, 2008.

28. X. Mao, C. Yuan, *Stochastic differential equations with Markovian switching*, Imperial College Press, 2006. http://doi.org/10.1142/p473

29. M. Chen, *From Markov chains to non-equilibrium particle systems*, 2 Eds., World Scientific, 1992. http://doi.org/10.1142/5513

30. I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calculus*, Springer-Verlag, 1988. http://doi.org/10.1007/978-1-4684-0302-2

31. P. Billingsley, *Convergence of probability measures*, 2 Eds., John Wiley & Sons, Inc., 1999. http://doi.org/10.1002/9780470316962

32. A. N. Shiryaev, *Probability*, 2 Eds., Springer-Verlag, 1996. http://doi.org/10.1007/978-1-4757-2539-1

33. F. Wu, G. Yin, An averaging principle for two-time-scale stochastic functional differential equations, *J. Differ. Equations*, **269** (2020), 1037–1077. http://doi.org/10.1016/j.jde.2019.12.024

34. R. Kasinathan, R. Kasinathan, D. Chalishajar, D. Baleanu, V. Sandrasekaran, The averaging principle of Hilfer fractional stochastic pantograph equations with non-Lipschitz conditions, *Stat. Probab. Lett.*, **215** (2024), 110221. http://doi.org/10.1016/j.spl.2024.110221