Available online at www.sciencedirect.com

ScienceDirect

J. Differential Equations 269 (2020) 1037–1077

Journal of Differential Equations

www.elsevier.com/locate/jde

An averaging principle for two-time-scale stochastic functional differential equations

Fuke Wu [a,1], George Yin [b,2]

[a] School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, PR China

[b] Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

Received 1 July 2019; revised 18 December 2019; accepted 28 December 2019
Available online 8 January 2020

Abstract

Delays are ubiquitous, pervasive, and entrenched in everyday life, thus taking it into consideration is necessary. Dupire recently developed a functional Itô formula, which has changed the landscape of the study of stochastic functional differential equations and encouraged a reconsideration of many problems and applications. Based on the new development, this work examines functional diffusions with two-time scales in which the slow-varying process includes path-dependent functionals and the fast-varying process is a rapidly-changing diffusion. The gene expression of biochemical reactions occurring in living cells in the introduction of this paper is such a motivating example. This paper establishes mixed functional Itô formulas and the corresponding martingale representation. Then it develops an averaging principle using weak convergence methods. By treating the fast-varying process as a random "noise", under appropriate conditions, it is shown that the slow-varying process converges weakly to a stochastic functional differential equation whose coefficients are averages of that of the original slow-varying process with respect to the invariant measure of the fast-varying process.
© 2020 Elsevier Inc. All rights reserved.

MSC: 34K50; 60G44; 60H10; 60J60; 92D25; 93Exx

E-mail addresses: wufuke@hust.edu.cn (F. Wu), gyin@math.wayne.edu (G. Yin).

[1] The research of this author was supported in part by the National Natural Science Foundation of China (Grant Nos. 61873320 and 11761130072).

[2] The research of this author was supported in part by the National Science Foundation under grant DMS-1710827.

https://doi.org/10.1016/j.jde.2019.12.024
0022-0396/© 2020 Elsevier Inc. All rights reserved.

Keywords: Path-dependent functional; Two-time scale; Weak convergence; Martingale method; Ergodicity; Functional Itô formula; Stochastic functional differential equation

1. Introduction and motivation

In the seminal work [13], Dupire introduced a functional Itô formula using pathwise functional derivatives, which quantify the sensitivity of functional variations at the endpoint of a path. This approach was further carefully worked out and much extended by Cont and Fournié [10], and Bally, Caramellino, and Cont [2]. The newly developed functional Itô formula changed the landscape of the study of stochastic functional equations. It opens up hopes for solving many problems that we were not able to deal with due to the lack of machinery for a long time. Motivated by the new development, aiming at substantially reducing computational complexity for systems involving functional diffusions, this work develops averaging methods for two-time-scale functional diffusions. It presents an effort from two distinct angles. The first one stems from the formulation of functional diffusion systems thanks to the recent development; see Dupire [13] and also Cont and Fournié [10]. The second perspective is on two-time-scale formulation, which has a long illustrated history that may be traced back to the original work of Khasminskii [20]. The two features together produce a fresh new look of two-time-scale functional diffusions.

Delays are ubiquitous, pervasive, and entrenched in everyday life. As a result, they have received considerable attention in a wide range of applications in process control, automotive systems, biomedical sciences, epidemics, transport, communication networks, and population dynamics [15,18,26,29]. The motivation stems from non-instant transmission phenomena, for example, high velocity fields in wind tunnel experiments, or other memory processes, or biological applications. In general, delay or more general functional differential equations exhibit much more complicated dynamics than that of ordinary differential equations because of the infinite dimensionality; see [3,34,37,44]. Systems involving uncertainty and delay are often described by stochastic delay or functional differential equations, which are frequently the sources of instability [25]. In recent years, such systems have become an important focal point of research and investigation. It is well known that the solutions of stochastic functional or delay differential equations are non-Markov because of the dependence of history. Thus none of the properties of solutions based on Markov property are applicable; see [3,37,44]. Despite the effort, treating stochastic systems with delays and functional systems remains a rather difficult task. Although there were many excellent works on stochastic delay equations, until very recently, there were virtually no bona fide operators and functional Itô formulas except for some general setup in a Banach space [37].

Many applications in science and engineering contain random processes that can be modeled by fast-slow motions or involving two-time-scale random processes; see [4,45,47,48]. There are numerous applications of systems under consideration, especially because of the recent progress in networked systems. Before proceeding further, let us mention a couple of motivational examples below. We shall return to Example 1.2 in Section 6.

Example 1.1. We consider a feedback control of a functional diffusion given by

$$\begin{aligned} d\xi^\varepsilon(t) &= \frac{1}{\varepsilon}h(\xi^\varepsilon(t))dt + \frac{1}{\sqrt{\varepsilon}}\sigma(\xi^\varepsilon(t))dw(t), \\ dx^\varepsilon(t) &= f(x^\varepsilon(t), x^\varepsilon_t, \xi^\varepsilon(t), u(t), t)dt + \sigma_1(x^\varepsilon(t), x^\varepsilon_t, \xi^\varepsilon(t), t)dw_1(t), \end{aligned} \tag{1.1}$$

where $h : \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times C([0,T];\mathbb{R}^r) \times [0,T] \mapsto \mathbb{R}^{d\times d}$, $f : \mathbb{R}^r \times C([0,T];\mathbb{R}^r) \times \mathbb{R}^d \times U \times [0,T] \mapsto \mathbb{R}^r$, $u(t)$ is a feedback control living in a compact subset U of $\mathbb{R}^r$, $\sigma_1 : \mathbb{R}^r \times C([0,T];\mathbb{R}^r) \times \mathbb{R}^d \times [0,T] \mapsto \mathbb{R}^{r\times r}$, $w(\cdot)$ and $w_1(\cdot)$ are d-dimensional and r-dimensional independent standard Brownian motions, respectively, x_t^ε, denoting the dependence of history, is to be specified later, and $\varepsilon > 0$ is a small parameter. The smaller the ε is, the more variation is encountered in $\xi^\varepsilon(t)$. The control objective is to find a feedback control $u = u(x^\varepsilon(t))$ so that certain objective function is minimized. The system is difficult to deal with. One is thus interested in finding an approximate solution using averaging principles. If one can show that the slowly varying component $x^\varepsilon(\cdot)$ has a limit, one hopes to develop strategies based on the limit system. Using such strategies in the original more complex system leads to nearly optimal controls. The current paper provides a key approach to realize this plan.

Example 1.2. Gene expression is a complex process involving many biochemical reactions with proteins being the final products. It is a challenging task to develop a systematic and rigorous treatment of stochastic dynamics with time delays and to investigate combined effects of stochasticity and delays in concrete models. In a deterministic approximation, one often models it by a system with a number of differential kinetic rate equations describing transcription, translation, and degradation. It is usually assumed that all these processes are instantaneous. However, some reacting processes are rather slow, for example, the average translation speed is only about 2 codons/s; see [1]. Time delays or general memory in biological system are usually ascribed for such processes.

In many biochemical reactions occurring in living cells, the number of various molecules might be low with significant stochastic fluctuations. In addition, most reactions are not instantaneous, so there exist natural time delays in the evolution of cell states [5,36]. For example, the process of degradation of both mRNA [9] and proteins [9,33,41] often consist of several steps and can naturally be modeled using time delays. Delayed degradation of JAK2 protein in signaling pathways was considered in [9,33,35,41], and delayed protein degradation was studied in [6]. Let us consider the following delayed protein degradation model; see also [36].

In a classical model of gene expression [42], molecules of mRNA are produced from DNA in the process of transcription and then give rise to production of protein molecules in the process of translation. Both types of molecules may degrade. Because the mRNA dynamics are faster than the protein dynamics, we have a two-time-scale system; see [32,43]. Denote the intensities of the biochemical reactions by k_r/ε, k_p, γ_r/ε, and γ_p, respectively,

$$\begin{aligned} &\text{DNA} \xrightarrow{k_r/\varepsilon} \text{mRNA}, \qquad &&\text{mRNA} \xrightarrow{\gamma_r/\varepsilon} \emptyset, \\ &\text{mRNA} \xrightarrow{k_p} \text{Protein}, \qquad &&\text{Protein} \xrightarrow{\gamma_p} \emptyset, \end{aligned} \tag{1.2}$$

where the small parameter ε highlights the mRNA dynamics to be a fast-varying process. Denote concentrations of mRNA and proteins by ξ^ε and x^ε, respectively. Then the equations of chemical kinetics read

$$\begin{aligned} \frac{d\xi^\varepsilon(t)}{dt} &= \frac{1}{\varepsilon}(k_r - \gamma_r \xi^\varepsilon(t)), \\ \frac{dx^\varepsilon(t)}{dt} &= k_p \xi^\varepsilon(t) - \gamma_p x^\varepsilon(t). \end{aligned} \tag{1.3}$$

Following the work [6], the authors of [36] took into account the process of protein degradation with time delay. In [6] and [36], to simplify the mathematical models, only fixed time delay is considered, whereas distributed delays treated as memory were considered in [5]. When the complete memory is considered, integral delay from 0 to t is more suitable. Then system (1.2) may be rewritten as

$$\begin{aligned} &\text{DNA} \xrightarrow{k_r/\varepsilon} \text{mRNA}, \qquad \text{mRNA} \xrightarrow{\gamma_r/\varepsilon} \emptyset, \\ &\text{mRNA} \xrightarrow{k_p} \text{Protein}, \qquad \text{Protein} \xrightarrow{\gamma_p} \emptyset, \qquad \text{Protein} \overset{\gamma_d(s)}{\Longrightarrow} \emptyset, \end{aligned} \tag{1.4}$$

where in the last reaction, "$\Longrightarrow$" shows the reaction intensity depends on the complete memory. Then the equation about $x^\varepsilon(t)$ in (1.3) may be rewritten as

$$\frac{dx^\varepsilon(t)}{dt} = k_p\xi^\varepsilon(t) - \gamma_p x^\varepsilon(t) - \int_0^t \gamma_d(s)x^\varepsilon(s)ds, \tag{1.5}$$

where $\gamma_d(s)$ can be seen as the degradation intensity of the protein produced at time $s \in [0, t]$. This integral delay can also be seen as the approximation of multiple delays of the form

$$\sum_{i=1}^{n} \gamma_i x^\varepsilon(t - \tau_i),$$

where $\gamma_i = \gamma_d(t_i)\Delta_i$, $\tau_i = t - t_i$ is the delay, $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$, and $\Delta_i = t_{i+1} - t_i$. When more general path-dependent delay $x_t^\varepsilon = \{x^\varepsilon(\theta \wedge t) : 0 \le \theta \le T\}$ is considered, one may introduce functional $\pi(x_t^\varepsilon)$ leading to a functional diffusion system with two-time scales as follows:

$$\begin{aligned} \frac{d\xi^\varepsilon(t)}{dt} &= \frac{1}{\varepsilon}(k_r - \gamma_r\xi^\varepsilon(t)), \\ \frac{dx^\varepsilon(t)}{dt} &= k_p\xi^\varepsilon(t) - \gamma_p x^\varepsilon(t) - \pi(x_t^\varepsilon). \end{aligned} \tag{1.6}$$

Along with the development, when the intrinsic noises are considered as in [7,8,17], there exist two Brownian motions $w_1(t)$ and $w_2(t)$ such that the chemical Langevin equation corresponding to the system (1.6) can be described by

$$\begin{aligned} d\xi^\varepsilon(t) &= \frac{1}{\varepsilon}(k_r - \gamma_r\xi^\varepsilon(t))dt + \frac{1}{\sqrt{\varepsilon}}\sqrt{k_r + \gamma_r\xi^\varepsilon(t)}\,dw_1(t), \\ dx^\varepsilon(t) &= [k_p\xi^\varepsilon(t) - \gamma_p x^\varepsilon(t) - \pi(x_t^\varepsilon)]dt + \sqrt{k_p\xi^\varepsilon(t) + \gamma_p x^\varepsilon(t) + \pi(x_t^\varepsilon)}\,dw_2(t). \end{aligned} \tag{1.7}$$

The formulation is a two-time-scale diffusion system based on the chemical Langevin equation. In [36], the stochastic system with two-time scales stems from the master equation and the stochastic simulation algorithm originally proposed by Gillespie [14]. Nevertheless, the stochastic simulation algorithms are often computationally expensive and slow. One of the more efficient ways to reduce the computational load by using two-time-scale formulation in the chemical Langevin equation, which was proposed and developed in our work Wu et al. [45] for diffusions without delays. It is certainly important to establish a complexity reduction method (an averaging principle) for the delay chemical Langevin equation since no existing results are available to date.

One of the main features of the above examples is that the original systems are complex and difficult to deal with, but one may obtain much simpler limit dynamic systems. Using the simpler limit systems as a bridge, one may proceed to design feasible procedure to treat the original systems; see [20–24,27,28,38–40,46]. To deal with the two-time-scale Markovian systems, one may consider the associated transition probabilities through Kolmogorov-Fokker-Planck equations; see for example, Khasminskii and Yin [21–23] (see also related reference [24]). Such an approach is essentially analytic. Another method is to use stochastic averaging to obtain certain limit results, for example, Khasminskii [20], Kushner [27,28], and Pardoux and Veretennikov [38]. Note that in the last reference above, partial differential equations were used as a bridge for the averaging, whereas in [20,27,28,39,40,46] probabilistic methods were used as a primary tool. In spite of the notable progress and effort, an averaging principle for two-time-scale stochastic delay or functional systems has not been established to date, to the best of our knowledge.

With the recent advent of functional Itô formula, this paper aims to analyze functional diffusions with two-time scales. Our main effort is to reduce the complexity. The original systems that we are dealing with are difficult due to the fast varying processes and the noise influence. We aim to obtain limit or averaged systems that are much simpler to deal with compared with the original systems. We examine asymptotic properties of diffusion systems involving the path-dependent functionals with two-time scales described by the following stochastic functional differential equations

$$\begin{cases} d\xi^\varepsilon(t) = \dfrac{1}{\varepsilon}h(\xi^\varepsilon(t))dt + \dfrac{1}{\sqrt{\varepsilon}}\phi(\xi^\varepsilon(t))dw_1(t), \\ \\ dx^\varepsilon(t) = b(x_t^\varepsilon, \xi^\varepsilon(t))dt + \psi(x_t^\varepsilon, \xi^\varepsilon(t))dw_2(t), \end{cases} \tag{1.8}$$

with non-random initial data $\xi(0) \in \mathbb{R}^m$ and $x(0) \in \mathbb{R}^n$, where ε is a small parameter, $x_t^\varepsilon := \{x^\varepsilon(u \wedge t) : 0 \le u \le T\}$, $h = (h_1, h_2, \dots, h_m)' : \mathbb{R}^m \to \mathbb{R}^m$, z' denotes the transpose of z, $\phi = [\phi_{ij}]_{m\times l_1} : \mathbb{R}^m \to \mathbb{R}^{m\times l_1}$, $b = (b_1, b_2, \dots, b_n)' : C([0,T]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^n$, $\psi = [\psi_{ij}]_{n\times l_2} : C([0,T]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^{n\times l_2}$, and $w_1(t)$ and $w_2(t)$ are independent standard Brownian motions taking values in $\mathbb{R}^{l_1}$ and $\mathbb{R}^{l_2}$, respectively.

Because the functional diffusion system (1.8) involves the path-dependent functionals (the total past history), its solution is not a Markov process. Thus the techniques in the literature for treating Markov processes are not applicable. Moreover, the weak convergence methods of Kushner [27,28] cannot be applied directly either since we have to consider the differential of the delay term. New approaches have to be developed. The system that we are interested in has two parts. One part is a fast varying diffusion, and the other part is a slowly changing functional diffusion. We develop an averaging principle to overcome the difficulties. A key of our approach is the use of the newly developed functional Itô formula.

The rest of the paper is arranged as follows. Section 2 provides necessary notation, assumptions, and some preliminaries. Section 3 examines the invariant measure and the exponential ergodicity of the fast-varying process $\xi^\varepsilon(t)$. Section 4 recalls the functional Itô formula and develops the mixed Itô formula and the solution of the corresponding martingale problem. Using these results, Section 5 derives the weak convergence of the slow-varying process $x^\varepsilon(\cdot)$ as $\varepsilon \to 0$. Section 6 extends our results using weaker conditions; a two-time-scale stochastic functional differential equation involving the affine noise in the slow-varying subsystem is treated. Moreover, a stochastic integro-differential equation with two-time scales is examined. In addition, Example 1.2 is examined by using the two-time-scale analysis. Section 7 presents some concluding remarks. Finally, an appendix is provided at the end of the paper as technical complements.

2. Notation, assumptions, and preliminaries

Throughout this paper, unless otherwise specified, we use the following notation. Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space with the Euclidean norm $|\cdot|$, and $\mathcal{B}(\mathbb{R}^n)$ be the Borel set of $\mathbb{R}^n$. For each $N > 0$, let $S_N = \{x : |x| \le N\}$ be a ball with radius N centered at the origin. For a vector or matrix A, denote its transpose by A'; for a matrix A, denote its trace norm by $|A| = \sqrt{\mathrm{Tr}(A'A)}$. For $a, b \in \mathbb{R}^n$, $\langle a, b\rangle = a'b$ represents the inner product of a and b. Denote by $D([0, T]; \mathbb{R}^n)$ the space of functions defined on $[0, T]$ with values in $\mathbb{R}^n$ that are right continuous with left limits. Denote by $C([0, T]; \mathbb{R}^n)$ the family of continuous functions from $[0, T]$ to $\mathbb{R}^n$, by $C^l(\mathbb{R}^n; \mathbb{R})$ the family of functions on $\mathbb{R}^n$ that have continuous partial derivatives up to the lth order, by $C_0^l(\mathbb{R}^n; \mathbb{R})$ the family of $C^l(\mathbb{R}^n; \mathbb{R})$ functions with compact support. Throughout the paper, K denotes a generic positive constant, whose value may change for different usage. Thus, $K + K = K$ and $KK = K$ are understood in an appropriate sense. Similarly, K_N denotes the generic positive constant depending on N. We use $\varepsilon > 0$ to represent a small parameter.

In this paper, if $x(t)$ is a stochastic process, denote by $\mathcal{F}_t^x = \sigma\{x(s) : s \le t\}$ the filtration generated by $\{x(s) : s \le t\}$, and $\mathbb{E}_t^x$ the corresponding conditional expectation. Based on $x(t)$ for $t \in [0, T]$, define the process $x_t = \{x(u \wedge t) : 0 \le u \le T\}$ as a function on $[0, T]$, that is, the efficient information is still in $[0, t]$. This implies that x_t is an $\mathcal{F}_t^x$-adapted stochastic process. For the stochastic processes $\xi^\varepsilon(\cdot)$ and $x^\varepsilon(\cdot)$ dependent on ε, we define $\mathcal{F}_t^\varepsilon$ as the σ-algebra generated by $\{\xi^\varepsilon(s), x^\varepsilon(s) : s \le t\}$, and $\mathbb{E}_t^\varepsilon$ the conditional expectation on $\mathcal{F}_t^\varepsilon$. In what follows, we assume the initial data of $\xi^\varepsilon(\cdot)$ and $x^\varepsilon(\cdot)$ are non-random. Then $\mathcal{F}_t^\varepsilon$ is contained in the σ-algebra generated by $\{w_1(s) : s \le t/\varepsilon; w_2(u) : u \le t\}$, which reflects the two-time-scale feature. Let $\mathcal{M}$ denote the set of real-valued progressively measurable functions that are nonzero only on a bounded t-interval and

$$\overline{\mathcal{M}}^\varepsilon = \Big\{ f \in \mathcal{M} : \sup_t \mathbb{E}|f(t)| < \infty \text{ and } f(t) \text{ is } \mathcal{F}_t^\varepsilon\text{-measurable} \Big\}. \tag{2.1}$$

Using [27,30], let us recall the definitions of the p-lim and the infinitesimal operator $\hat{\mathcal{L}}^\varepsilon$ as follows. Note that the p-lim was first introduced by Rishel and further worked out by Kurtz in the 1970s. Using conditional expectation, it in fact, introduces certain operators. The essence is that the operator provided is a kind of infinitesimal operator. When we use the p-lim in the actual calculations, some of them involve the classical Itô formula, whereas others use the functional Itô formula. For example, in (5.26) of this paper, the first term is from the classical Itô formula and the second term is from the functional Itô established in this paper. Because these Itô formulas are in the almost sure sense, they also work in the sense of “p-lim.

Definition 2.1. Let $f, f^\delta \in \overline{\mathcal{M}}^\varepsilon$ for each $\delta > 0$. We say $f =$p-lim$_\delta f^\delta$ if and only if

$$\begin{cases} \sup_{t,\delta} \mathbb{E}|f^\delta(t)| < \infty, \\ \lim_{\delta\to 0} \mathbb{E}|f^\delta(t) - f(t)| = 0 \quad \text{for each } t. \end{cases}$$

This definition implies that p-lim$_\varepsilon f^\varepsilon = 0$ if $f(\cdot) = 0$ and ε replaces δ and $f^\varepsilon \in \overline{\mathcal{M}}^\varepsilon$ for each $\varepsilon > 0$.

Definition 2.2. We say that $f(\cdot) \in \mathcal{D}(\hat{\mathcal{L}}^\varepsilon)$, the domain of $\hat{\mathcal{L}}^\varepsilon$, and $\hat{\mathcal{L}}^\varepsilon f = g$ if $f, g \in \overline{\mathcal{M}}^\varepsilon$ and

$$\text{p-}\lim_{\delta \downarrow 0} \Big(\frac{\mathbb{E}_t^\varepsilon f(t+\delta) - f(t)}{\delta} - g(t) \Big) = 0.$$

Thus $\hat{\mathcal{L}}^\varepsilon$ is a type of infinitesimal operator. The following lemma was proved in Kurtz [30].

Lemma 2.1. *If $f \in \mathcal{D}(\hat{\mathcal{L}}^\varepsilon)$, then*

$$M_f(t) = f(t) - \int_0^t \hat{\mathcal{L}}^\varepsilon f(u)du$$

is a martingale, and

$$\mathbb{E}_t^\varepsilon f(t+s) - f(t) = \mathbb{E}_t^\varepsilon \int_t^{t+s} \hat{\mathcal{L}}^\varepsilon f(u)du \quad \text{w.p.1.}$$

3. Invariant measure of the fast-varying process

To obtain the weak convergence of the slow-varying $x^\varepsilon(t)$ as $\varepsilon \to 0$, certain ergodicity is crucial. In (1.8), $\xi^\varepsilon(t)$ is rapidly varying in contrast to $x^\varepsilon(t)$. To proceed, we first consider asymptotic properties of $\xi^\varepsilon(t)$ as $\varepsilon \to 0$. Let us define the process $\tilde{\xi}^\varepsilon(t) = \xi^\varepsilon(\varepsilon t)$. The equation for ξ^ε may be rewritten as

$$d\tilde{\xi}^\varepsilon(t) = h(\tilde{\xi}^\varepsilon(t))dt + \phi(\tilde{\xi}^\varepsilon(t))d\tilde{w}(t), \tag{3.1}$$

where $\tilde{w}(t) = w_1(\varepsilon t)/\sqrt{\varepsilon}$ is a standard Brownian motion. Letting $\xi^\varepsilon(t) = \xi(t/\varepsilon)$, $\xi(t)$ satisfies the following stochastic differential equation

$$d\xi(t) = h(\xi(t))dt + \phi(\xi(t))d\tilde{w}(t). \tag{3.2}$$

For the limit problem that we are interested in, we need the following condition.

(A1) Equation (3.2) has a unique strong solution $\xi(t)$ for any $t \in [0, T]$. The solution is $\mathcal{F}^{w_1}_{\varepsilon t}$-adapted and is a strong homogeneous Markov process. In addition, $\mathbb{E}\Big(\sup_{t\in[0,T]} |\xi(t)|^2\Big) \le C_T$, where C_T is a constant depending on T. Moreover, the process given by (3.2) is exponentially ergodic with a unique invariant measure $\mu(\cdot)$.

Remark 3.1. For our purposes of averaging, (A1) is all we need. It is standard nowadays that under suitable conditions, one first obtains a local existence of solution and then extend the solution to a global solution using regularity; the proof of ergodicity follows from the non-degeneracy of the diffusion and certain stopping time argument. One may consult [24, Section 3.4, Section 4.4] for some details. It is also known that even for certain degenerate diffusions, we can still have the desired exponential ergodicity. However, it is not clear that one can give general conditions to cover every possible case. Nevertheless, for given systems, they can clearly be handled in a case-by-case fashion. Owing to this consideration, we simply state (A1) as it is. One may wonder what conditions will ensure (A1). To illustrate, we provide some sufficient conditions below together with a theorem stating the implications of the conditions. Note that the conditions in the following theorem are only sufficient but not necessary. Even if these conditions are not satisfied, (A1) may still hold. The proof of the following theorem is outlined in Appendix A; see also [44] for a stochastic functional differential equation counterpart.

Theorem 3.2. *Assume that $h(\cdot)$ is locally Lipschitz continuous and there exists λ_1 such that for any $y_1, y_2 \in \mathbb{R}^m$,*

$$\langle y_1 - y_2, h(y_1) - h(y_2)\rangle \le -\lambda_1 |y_1 - y_2|^2$$

and $\phi(\cdot)$ is globally Lipschitz continuous, i.e., there exists λ_2 such that

$$|\phi(y_1) - \phi(y_2)|^2 \le \lambda_2 |y_1 - y_2|^2.$$

Then equation (3.2) admits a unique strong solution $\xi(t)$ globally for any $t \in [0, T]$, which is $\mathcal{F}^{w_1}_{\varepsilon t}$-adapted and satisfies the following properties:

(i) *the solution is a strong homogeneous Markov process;*
(ii) $\mathbb{E}\Big(\sup\limits_{t\in[0,T]} |\xi(t)|^2\Big) \le C_T$, *where C_T is a constant depending on T;*
(iii) *if $2\lambda_1 > \lambda_2$, then (3.2) has a unique invariant measure $\mu(\cdot)$, which is exponentially ergodic.*

In lieu of the sufficient conditions above, other conditions guaranteeing these results can also be provided. For example, for the first equation in (1.7) of Example 1.2, letting $\xi^\varepsilon(t) = \xi(t/\varepsilon)$ gives

$$d\xi(t) = [k_r - \gamma_r \xi(t)]dt + \sqrt{k_r + \gamma_r \xi(t)}\, d\tilde{w}(t). \tag{3.3}$$

Moreover, define $\hat{\xi}(t) = k_r + \gamma_r \xi(t)$. Then,

$$d\hat{\xi}(t) = \gamma_r (2k_r - \hat{\xi}(t))dt + \gamma_r \sqrt{\hat{\xi}(t)}\, d\tilde{w}(t),$$

which does not satisfy the local Lipschitz condition. However it can be shown that it has a unique global solution, known as the mean-reverting square root process, and this solution is pth moment bounded for any $p \geq 0$ and exponentially ergodic with the stationary distribution being the noncentral chi-square distribution; see [11,34].

4. Functional Itô formula and martingale problem

In this section, we consider the following stochastic functional differential equation

$$dX(t) = B(t, X_t)dt + \Psi(t, X_t)dW(t) \tag{4.1}$$

with the deterministic initial value $X(0) \in \mathbb{R}^n$, where $X_t := \{X(u \wedge t) : 0 \leq u \leq T\}$, $B = (B_1, B_2, \dots, B_n)' : \mathbb{R}_+ \times D([0,T]; \mathbb{R}^n) \to \mathbb{R}^n$, $\Psi = [\Psi_{ij}]_{n\times l_2} : \mathbb{R}_+ \times D([0,T]; \mathbb{R}^n) \to \mathbb{R}^{n\times l_2}$, $W(t)$ is an l_2-dimensional standard Brownian motion. We assume that there exists an $\mathcal{F}_t^W$-adapted solution $x(t)$ on $[0,T]$ and $B_1(t) = B(t, X_t) \in \mathbf{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ and $\Psi_1(t) = \Psi(t, X_t) \in \mathbf{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times l_2})$, where $\mathcal{F}_t^W$ is the filtration generated by the Brownian motion $W(t)$, $\mathbf{L}^1(\mathbb{R}_+; \mathbb{R}^n)$ denotes the family of all $\mathbb{R}^n$-valued measurable $\mathcal{F}_t^W$-adapted processes $F_1(t)$ such that $\int_0^T |F_1(t)|dt < \infty$ almost surely for every $T > 0$, and $\mathbf{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times l_2})$ denotes the family of all $\mathbb{R}^{n\times l_2}$-valued measurable $\mathcal{F}_t^W$-adapted processes $F_2(t)$ such that $\int_0^T |F_2(t)|^2 dt < \infty$ almost surely for every $T > 0$.

In an insightful work, Dupire [13] proposed a method for defining a non-anticipative calculus, which extends the Itô formula to path-dependent functionals of stochastic processes. To proceed, we first give the definitions of the *horizontal* and the *vertical* derivatives, respectively; see also [2,10].

Definition 4.1. For $x_t = \{x(u \wedge t) : 0 \leq u \leq T\}$, a non-anticipative functional $F : [0,T] \times D([0,T]; \mathbb{R}^n) \to \mathbb{R}$ is said to be horizontally differentiable at $(t, x) \in [0,T] \times D([0,T]; \mathbb{R}^n)$ if the limit

$$\mathcal{D}F(t, x) = \lim_{\delta \to 0^+} \frac{F(t+\delta, x_t) - F(t, x_t)}{\delta}$$

exists. The $\mathcal{D}F(t, x)$ is called horizontal derivative of F at (t, x).

Definition 4.2. For $x_t = \{x(u \wedge t) : 0 \leq u \leq T\}$, a non-anticipative functional $F : [0,T] \times D([0,T]; \mathbb{R}^n) \to \mathbb{R}$ is said to be vertically differentiable at $(t, x) \in [0,T] \times D([0,T]; \mathbb{R}^n)$ if the functional map

$$e \to F(t, x_t + e\mathbf{1}_{[t,T]})$$

is differentiable at 0. Its gradient at 0 is called the vertical derivative of F at (t, x):

$$\nabla_x F(t, x) = (\partial_1 F(t, x), \partial_2 F(t, x), \dots, \partial_n F(t, x)),$$

where

$$\partial_i F(t, x) = \lim_{h \to 0} \frac{F(t, x_t + he_i\mathbf{1}_{[t,T]}) - F(t, x_t)}{h}.$$

If F is vertically differentiable at all $(t, x) \in [0, T] \times D([0, T]; \mathbb{R}^n)$, then $\nabla_x F$ is a non-anticipative functional called the vertical derivative of F. Repeating this procedure leads to the definition of the second derivative, or $\nabla_x^2 F(t, x) = [\partial_{ij}^2 F(t, x)]_{n\times n}$, the derivative of the gradient at 0 (if it exists) of the map

$$e \to \nabla_x F(t, x_t + e\mathbf{1}_{[t,T]}).$$

Remark 4.1. In fact, to consider the derivative of functionals, we have to consider $x_{t+\delta} - x_t$. Noting that $x_t = \{x(t \wedge u) : 0 \le u \le T\}$, roughly speaking,

$$x_{t+\delta} - x_t = 0 \cdot \mathbf{1}_{[0,t)} + [x(u) - x(t)]\mathbf{1}_{[t,t+\delta)}(u) + [x(t+\delta) - x(t)]\mathbf{1}_{[t+\delta,T]}.$$

As $\delta \to 0$, we need to consider the perturbation $x_t + e\mathbf{1}_{[t,T]}$. In view of these definitions, although x_t may be a continuous function if $x(t)$ is a continuous process, $x_t + e\mathbf{1}_{[t,T]}$ is right continuous and has left limit, that it is in $D([0, T]; \mathbb{R}^n)$. Thus we need to have F be defined on $[0, T] \times D([0, T]; \mathbb{R}^n)$.

Now let us define the continuity for non-anticipative functionals; see [2].

Definition 4.3 *(Joint continuity in (t, x)).* A continuous non-anticipative functional is a continuous map $F : [0, T] \times D([0, T]; \mathbb{R}^n) \to \mathbb{R}$ if, for any $(t, x) \in [0, T] \times D([0, T]; \mathbb{R}^n)$, any $\Delta > 0$, there exists an $\eta > 0$, for any $(\widetilde{t}, \widetilde{x}) \in [0, T] \times D([0, T]; \mathbb{R}^n)$ satisfying

$$d_\infty((t, x), (\widetilde{t}, \widetilde{x})) = \sup_{u\in[0,T]} \{|x(t \wedge u) - \widetilde{x}(t \wedge u)| + |t - \widetilde{t}|\} < \eta$$

such that

$$|F(t, x) - F(\widetilde{t}, \widetilde{x})| < \Delta.$$

The set of jointly continuous non-anticipative functionals is denoted by $\mathbb{C}^{0,0}([0, T] \times D(0, T); \mathbb{R}^n)$.

Next, we introduce the notion of "local boundedness" for functionals. We call a functional F "boundedness preserving" if it is bounded on each bounded set of paths [2,10].

Definition 4.4. A non-anticipative functional $F : [0, T] \times D([0, T]; \mathbb{R}^n) \to \mathbb{R}$ is said to be boundedness preserving if for any compact $\mathbb{K} \in \mathbb{R}^n$ and $t_0 < T$, there exists a $K_{\mathbb{K},t_0} > 0$, for all $t \le t_0$ and $x \in D([0, T]; \mathbb{K})$ such that $|F(t, x)| \le K_{\mathbb{K},t_0}$.

Let $X(t)$ be the solution to (4.1). Then $X(t)$ is a semimartingale whose quadratic variation process can be represented by

$$[X](t) = \int_0^t |\Psi(u, X_u)|^2 du, \tag{4.2}$$

which is a finite quadratic variation process since $\Psi \in \mathbf{L}^2(\mathbb{R}_+; \mathbb{R}^{n\times l_2})$. To consider the derivative of functionals, let us give the following definition.

Definition 4.5 *($\mathbb{C}^{1,2}$ functionals).* Define $\mathbb{C}^{1,2}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$ as the family of continuous non-anticipative functionals $V \in \mathbb{C}^{0,0}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$ such that

(i) V admits a horizontal derivative $\mathcal{D}V(t,x)$ for all $t,x \in [0,T]\times D([0,T];\mathbb{R}^n)$, and $\mathcal{D}V(t,\cdot)$ is continuous for any $t\in[0,T)$;
(ii) $\nabla_x V$ and $\nabla_x^2 V$ are jointly continuous;
(iii) $\mathcal{D}V$, $\nabla_x V$ and $\nabla_x^2 V$ are boundedness preserving.

For any $V \in \mathbb{C}^{1,2}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$, Dupire and Cont et al. established the following functional Itô formula; see [2,10,13].

Theorem 4.2 *(Functional Itô formula). Let $X(t)$ be the solution of (4.1). For any $V \in \mathbb{C}^{1,2}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$,*

$$\begin{aligned} V(t,X_t) &= V(0,X_0) + \int_0^t \mathcal{D}V(u,X_u)du + \int_0^t \nabla_x V(u,X_u)dX(u) \\ &\quad + \frac{1}{2}\int_0^t |\nabla_x^2 V(u,X_u)| d[X](u) \\ &= V(0,X_0) + \int_0^t \mathcal{L}V(u,X_u)du + \int_0^t \nabla_x V(u,X_u)\Psi(u,X_u)dW(u) \ \ a.s., \end{aligned} \tag{4.3}$$

where

$$\mathcal{L} = \mathcal{D} + \sum_{i=1}^n B_i(t,x)\partial_i + \frac{1}{2}\sum_{i,j=1}^n \Psi_i(t,x)\Psi_j(t,x)\partial_{ij}^2 \tag{4.4}$$

is an infinitesimal generator for any $(t,x)\in[0,T]\times D([0,T];\mathbb{R}^n)$. In particular, $Y(t) = V(t,X_t)$ is a continuous semimartingale and

$$V(t,X_t) - V(0,X_0) - \int_0^t \mathcal{L}V(u,X_u)du \tag{4.5}$$

is a local martingale with respect to the σ-algebra $\mathcal{F}_t^W$.

If V is assumed to be bounded (for example, $V \in \mathbb{C}^{1,2}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$ with compact support), (4.5) is in fact a martingale and for any $\delta > 0$,

$$\mathbb{E}_t^W V(t+\delta, X_{t+\delta}) - V(t, X_t) = \int_t^{t+\delta} \mathcal{L}V(u, X_u)du \quad \text{a.s.}, \tag{4.6}$$

where $\mathbb{E}_t^W$ is the conditional expectation with respect to $\mathcal{F}_t^W$.

For $v \in C^{1,2}([0,T] \times \mathbb{R}^n; \mathbb{R})$ with compact support, the generator $\mathcal{L}$ becomes L according to the standard Itô formula. That is,

$$L(t,x)\cdot = \frac{\partial \cdot}{\partial t} + \sum_{i=1}^n B_i(t,x)\frac{\partial \cdot}{\partial y_i} + \frac{1}{2}\sum_{i,j=1}^n \Psi_i(t,x)\Psi_j(t,x)\frac{\partial^2 \cdot}{\partial y_i \partial y_j} \tag{4.7}$$

for $x \in C([0,T]; \mathbb{R}^n)$ and $y = (y_1, y_2, \ldots, y_n)' \in \mathbb{R}^n$. Theorem 4.2 implies that

$$v(t, X(t)) - v(0, X(0)) - \int_0^t L(u, X_u)v(u, X(u))du$$

is a martingale with respect to the filtration $\mathcal{F}_t^W$, and for any $\delta > 0$,

$$\mathbb{E}_t^W v(t+\delta, X(t+\delta)) - v(t, X(t)) = \int_t^{t+\delta} L(u, X_u)v(u, X(u))du \quad \text{a.s.} \tag{4.8}$$

Remark 4.3. Let us emphasize that here v is a function defined on $[0,T] \times \mathbb{R}^n$, but the operator L is a functional operator on $[0,T] \times D([0,T]; \mathbb{R}^n)$, and the argument of X_u in $L(u, X_u)$ is from that of the coefficients of B and Ψ. In fact, the two generators $\mathcal{L}$ and L are the same when the horizontal derivative in functionals is the partial derivative w.r.t. time t and the vertical derivative is the partial derivative w.r.t. the state variable of the functions.

In this paper, we need to consider the derivative of $V(t,y,x)$ for $(t,y,x) \in [0,T] \times \mathbb{R}^n \times D([0,T]; \mathbb{R}^n)$, so we need to extend the existing functional Itô formula to include mixed derivatives. First, we give the definition of $\mathbb{C}^{0,0,0}([0,T] \times \mathbb{R}^n \times D([0,T]; \mathbb{R}^n); \mathbb{R})$ similar to Definition 4.3 by choosing

$$d_\infty((t,y,x), (\widetilde{t}, \widetilde{y}, \widetilde{x})) = \sup_{u \in [0,T]} |x(t \wedge u) - \widetilde{x}(t \wedge u)| + |y - \widetilde{y}| + |t - \widetilde{t}| < \eta.$$

Then we can define $\mathbb{C}^{1,2,2}$ functionals followed by mixed functional Itô formula.

Definition 4.6 *($\mathbb{C}^{1,2,2}$ functionals).* Define $\mathbb{C}^{1,2,2}([0,T] \times \mathbb{R}^n \times D([0,T]; \mathbb{R}^n); \mathbb{R})$ as the family of the jointly continuous non-anticipative functional $V \in \mathbb{C}^{0,0,0}([0,T] \times \mathbb{R}^n \times D([0,T]; \mathbb{R}^n); \mathbb{R})$ such that

(i) V admits a horizontal derivative $\mathcal{D}V(t,y,x)$ for all $t,y,x \in [0,T] \times \mathbb{R}^n \times D([0,T]; \mathbb{R}^n)$, and $\mathcal{D}V(t,\cdot,\cdot)$ is continuous for any $t \in [0,T)$;
(ii) $\nabla_x V$, $\nabla_x^2 V$, V_y, V_{yy}, and $\nabla_x(V_y)$ are jointly continuous;
(iii) $\mathcal{D}V$, $\nabla_x V$, $\nabla_x^2 V$, V_y, V_{yy}, and $\nabla_x(V_y)$ are boundedness preserving.

Theorem 4.4 *(Mixed functional Itô formula). Let $X(t)$ be the solution of (4.1). For any $V \in \mathbb{C}^{1,2,2}([0,T] \times \mathbb{R}^n \times D([0,T];\mathbb{R}^n);\mathbb{R})$,*

$$\begin{aligned}
&V(t, X(t), X_t)\\
&= V(0, X(0), X_0) + \int_0^t \mathcal{D}V(u, X(u), X_u)du\\
&\quad + \int_0^t [V_y(u, X(u), X_u) + \nabla_x V(u, X(u), X_u)]dX(u)\\
&\quad + \frac{1}{2}\int_0^t |[V_{yy}(u, X(u), X_u) + \nabla_x^2 V(u, X(u), X_u) + 2\nabla_x V_y(u, X(u), X_u)]|d[X](u)\\
&= V(0, X(0), X_0) + \int_0^t \mathbb{L}(u, X_u)V(u, X(u), X_u)du\\
&\quad + \int_0^t [V_y(u, X(u), X_u) + \nabla_x V(u, X(u), X_u)]\Psi(u, X_u)dW(u) \ \ a.s.,
\end{aligned} \tag{4.9}$$

where $\mathcal{D}$, ∇_x, and ∇_x^2 were defined previously,

$$V_y(t, y, x) = \Big(\frac{\partial V(t, y, x)}{\partial y_1}, \frac{\partial V(t, y, x)}{\partial y_2}, \dots, \frac{\partial V(t, y, x)}{\partial y_2}\Big),$$

$$V_{yy} = \Big[\frac{\partial^2 V(t, y, x)}{\partial y_i \partial y_j}\Big]_{n\times n} \ \text{and} \ \nabla_x V_y(t, y, x) = \Big[\partial_i\Big(\frac{\partial V(t, y, x)}{\partial y_j}\Big)\Big]_{n\times n},$$

and

$$\mathbb{L}(t, x)\cdot = \mathcal{D}\cdot + \sum_{i=1}^n B_i(t, x)\Big[\frac{\partial}{\partial y_i} + \partial_i\Big]\cdot + \frac{1}{2}\sum_{i,j=1}^n \Psi_i(t, x)\Psi_j(u, x)\Big[\frac{\partial^2}{\partial y_i \partial y_j} + 2\partial_i\Big(\frac{\partial}{\partial y_j}\Big) + \partial_{ij}^2\Big]\cdot$$

is an infinitesimal generator for any $(t, x) \in [0, T] \times D([0, T]; \mathbb{R}^n)$. In particular, $Y(t) = V(t, X(t), X_t)$ is a continuous semimartingale and

$$V(t, X(t), X_t) - V(0, X(0), X_0) - \int_0^t \mathbb{L}(u, X_u)V(u, X(u), X_u)du$$

is a local martingale with respect to the filtration $\mathcal{F}_t^W$.

Proof. Let us denote a sequence $\pi = \{\pi_N\}_{N\geq 1}$ of partitions of $[0, T]$ as

$$\pi_N = \{0 = t_0^N < t_1^N < \cdots < t_{k(N)}^N = T\},$$

and $|\pi_N| = \sup\{|t_{i+1}^N - t_i^N|, i = 1, 2, \ldots, k(N)\}$ be the mesh size of the partition π_N. For example, one can choose $t_i^N = iT/2^N$, $i = 0, 1, \ldots, k(N) = 2^N$, and $|\pi_N| = 2^{-N}$. Note that $X(t, \omega)$ represents a sample trajectory of the solution. In the proof, for notional simplicity, we omit ω and write $X(t, \omega)$ as $X(t)$ henceforth. Since any solution of (4.1) is continuous on $[0, T]$, it is uniformly continuous, which implies that for any trajectory $X(t)$,

$$\eta_N = \sup\{|X(u) - X(t_{i+1}^N)| + |t_{i+1}^N - t_i^N|, u \in [t_i^N, t_{i+1}^N)\} \to 0 \ \text{ as } \ N \to \infty.$$

Since $\mathcal{D}V$, $\nabla_x V$, $\nabla_x^2 V$, V_y, V_{yy}, and $\nabla_x(V_y)$ are boundedness preserving, for sufficiently large N and two sample paths $X(t), X^*(t^*) \in D([0, T]; \mathbb{R}^n)$ and $t, t^* < T$, if

$$d_\infty((t, X(t)), (t^*, X^*(t^*))) := \sup_{u\in[0,T]} \{|X(t \wedge u) - X^*(t^* \wedge u)| + |t - t^*|\} \leq \eta_N,$$

then $\mathcal{D}V$, $\nabla_x V$, $\nabla_x^2 V$, V_y, V_{yy}, and $\nabla_x(V_y)$ are bounded. For any $i < k(n) - 1$, consider the decomposition of increments into "horizontal", and "vertical" terms as

$$\begin{aligned}
&V(t_{i+1}^N, X(t_{i+1}^N), X_{t_{i+1}^N}) - V(t_i^N, X(t_i^N), X_{t_i^N}) \\
&\quad = V(t_{i+1}^N, X(t_{i+1}^N), X_{t_{i+1}^N}) - V(t_i^N, X(t_{i+1}^N), X_{t_{i+1}^N}) + V(t_i^N, X(t_{i+1}^N), X_{t_{i+1}^N}) \\
&\qquad - V(t_i^N, X(t_i^N), X_{t_i^N}) \\
&\quad =: I_1 + I_2.
\end{aligned} \tag{4.10}$$

Denote $\Delta_i^N = t_{i+1}^N - t_i^N$ and $\Lambda(u) = V(t_i^N + u, X(t_{i+1}^N), X_{t_{i+1}^N})$. Since $V \in \mathbb{C}^{1,2,2}([0, T] \times \mathbb{R}^n \times D([0, T]; \mathbb{R}^n); \mathbb{R})$, Λ is differentiable and

$$I_1 = \int_0^{\Delta_i^N} \mathcal{D}V(t_i^N + u, X(t_{i+1}^N), X_{t_{i+1}^N}) du.$$

Noting that when x_t becomes $x_t + e\mathbf{1}_{[t,T]}$, $x(t)$ becomes $x(t) + e$. We denote

$$\Gamma(u) = V(t_i^N, X(t_i^N) + u, X_{t_i^N} + u\mathbf{1}_{[t_i^N,T]}).$$

Since $V \in \mathbb{C}^{1,2,2}([0, T] \times \mathbb{R}^n \times D([0, T]; \mathbb{R}^n); \mathbb{R})$, $\Gamma \in C^2(\mathbb{R}^n; \mathbb{R})$,

$$\frac{d\Gamma(u)}{du} = V_y(t_i^N, X(t_i^N) + u, X_{t_i^N} + u\mathbf{1}_{[t_i^N,T]}) + \nabla_x V(t_i^N, X(t_i^N) + u, X_{t_i^N} + u\mathbf{1}_{[t_i^N,T]}),$$

and

$$\frac{d^2\Gamma(u)}{du^2} = V_{yy}(t_i^N, X(t_i^N)+u, X_{t_i^N}+u\mathbf{1}_{[t_i^N,T]}) + \nabla_x^2 V(t_i^N, X(t_i^N)+u, X_{t_i^N}+u\mathbf{1}_{[t_i^N,T]})$$
$$+2\nabla_x(V_y(t_i^N, X(t_i^N)+u, X_{t_i^N}+u\mathbf{1}_{[t_i^N,T]})).$$

Denote $\delta X_i^N = X(t_{i+1}^N) - X(t_i^N)$. Taking a Taylor expansion leads to

$$\begin{aligned} I_2 &= \frac{d\Gamma(u)}{du}\Big|_{u=0}\delta X_i^N + \frac{1}{2}\mathrm{Tr}\Big[\frac{d^2\Gamma(u)}{du^2}\Big|_{u=0}(\delta X_i^N)'\delta X_i^N\Big] + r_i^N \\ &= [V_y(t_i^N, X(t_i^N), X_{t_i^N}) + \nabla_x V(t_i^N, X(t_i^N), X_{t_i^N})]\delta X_i^N \\ &\quad + \frac{1}{2}|[V_{yy}(t_i^N, X(t_i^N), X_{t_i^N}) + \nabla_x^2 V(t_i^N, X(t_i^N), X_{t_i^N}) \\ &\quad +2\nabla_x(V_y(t_i^N, X(t_i^N), X_{t_i^N}))](\delta X_i^N)'\delta X_i^N| + r_i^N, \end{aligned}$$

where r_i^N is the remainder of the order $o(|\delta X_i^N|^2)$. Applying the same method as [2,10] gives the first equation in (4.9). Substituting the finite quadratic variation process (4.2) into the first equation gives the second equation in (4.9). In addition, (4.9) reveals that $Y(t)$ is a continuous semimartingale and

$$V(t, X(t), X_t) - V(0, X(0), X_0) - \int_0^t \mathbb{L}(u, X_u)V(u, X(u), X_u)du$$

is a local martingale with respect to the filtration $\mathcal{F}_t^W$ as desired. □

Similar to (4.6) and (4.8), this theorem yields that if V is further bounded, for any $\delta > 0$,

$$\mathbb{E}_t^W V(t+\delta, X(t+\delta), X_{t+\delta}) - V(t, X(t), X_t) = \mathbb{E}_t^W \int_t^{t+\delta} \mathbb{L}(u, X_u)V(u, X(u), X_u)du \ \text{ a.s.} \tag{4.11}$$

Remark 4.5. In this theorem, if $V \in \mathbb{C}^{1,2}([0,T]\times D([0,T];\mathbb{R}^n);\mathbb{R})$, (4.9) becomes (4.3) (resp., $v \in C^{1,2}([0,T]\times\mathbb{R}^n;\mathbb{R})$, and the formula corresponding to (4.8) holds.

5. Weak convergence and averaged system

In this section, we show that the sequence $x^\varepsilon(\cdot)$ converges weakly to a stochastic process that is the solution of an appropriate stochastic functional differential equation. In order to obtain the desired weak convergence, we need to prove tightness first.

To begin, we need to verify

$$\lim_{N_0\to\infty}\limsup_{\varepsilon\to 0}\mathbb{P}\Big(\sup_{t\le T}|x^\varepsilon(t)| \ge N_0\Big) = 0 \quad \text{for each} \quad T<\infty, \tag{5.1}$$

where $\mathbb{P}(A)$ denotes the probability of A. The verification of (5.1) is usually quite involved, and requires complicated calculations. To circumvent the difficulties, we use the truncation technique as follows. For each $N > 0$ sufficiently large such that $|x(0)| \le N$, consider

$$dx^{\varepsilon,N}(t) = b^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t))dt + \psi^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t))dw_2(t), \tag{5.2}$$

where $x_t^{\varepsilon,N} = \{x^{\varepsilon,N}(t \wedge u) : 0 \le u \le T\}$, $\xi^\varepsilon(t) = \xi(t/\varepsilon)$ is the solution of the first equation in (1.8), $b^N(x,\xi) = b(x,\xi)q^N(x)$, $\psi^N(x,\xi) = \psi(x,\xi)q^N(x)$, and

$$q^N(x) = \begin{cases} 1, & \text{when } x \in C([0,T]; S_N), \\ 0, & \text{when } x \in C([0,T]; \mathbb{R}^n - S_{N+1}), \\ \text{smooth}, & \text{otherwise.} \end{cases}$$

From the definition, it can be seen that $x^{\varepsilon,N}(t) = x^\varepsilon(t)$ up until the first exit from $S_N = \{x : |x| \le N\}$. Then $x^{\varepsilon,N}(t)$ is said to be the N-truncation of $x^\varepsilon(t)$. According to the definition of x_t, it is easily seen that $x_t^{\varepsilon,N} \in D([0,T]; S_N)$ if $x^{\varepsilon,N}(t) \in S_N$ since $\|x_t^{\varepsilon,N}\|_\infty := \sup_{u\in[0,T]}\{x^{\varepsilon,N}(t \wedge u) : 0 \le u \le T\} \le N$. Let $\mathcal{F}_t^{\varepsilon,N} = \sigma\{\xi^\varepsilon(s), x^{\varepsilon,N}(s) : s \le t\}$. We can also give the corresponding definitions for $\overline{\mathcal{M}}^{\varepsilon,N}$ and $\hat{\mathcal{L}}^{\varepsilon,N}$. It is clear that $\mathcal{F}_t^{\xi^\varepsilon} \subset \mathcal{F}_t^{\varepsilon,N}$. To proceed, the following assumptions are needed.

(A2) The functions $b(x,\xi)$, $\psi(x,\xi)$, $\nabla_x b_i(x,\xi)$, $\nabla_x \psi_{ij}(x,\xi)$, $\nabla_x^2 b_i(x,\xi)$ and $\nabla_x^2 \psi_{ij}(x,\xi)$, $i, j = 1, 2, \ldots, n$ are boundedness preserving with respect to $x \in D([0,T]; \mathbb{R}^n)$, and continuous and bounded with respect to $\xi \in \mathbb{R}^m$ for $x \in D([0,T]; G)$, where $G \subset \mathbb{R}^n$ is a compact set.

(A3) For $G \subset \mathbb{R}^n$ being a compact set, $x \in D([0,T]; G)$, $b(x,\cdot)$ and $\psi(x,\cdot)\psi'(x,\cdot) = A(x,\cdot) = [a_{ij}(x,\cdot)]_{n\times n}$ are integrable functionals with respect to the measure μ, and

$$\begin{cases} \displaystyle\int_{\mathbb{R}^n} b(x,\xi)\mu(d\xi) = \bar{b}(x), \\ \displaystyle\int_{\mathbb{R}^n} a_{ij}(x,\xi)\mu(d\xi) = \bar{a}_{ij}(x), \end{cases}$$

that is, $\bar{b}(x) = \mathbb{E}_\mu b(x,\xi)$ and $\bar{a}_{ij}(x) = \mathbb{E}_\mu a_{ij}(x,\xi)$, where $\mathbb{E}_\mu$ is the expectation with respect to the invariant measure μ. Moreover, $A(\cdot,\cdot)$ is nonnegative definite.

(A4) The following equation

$$dx(t) = \bar{b}(x_t)dt + \bar{\psi}(x_t)dB(t) \tag{5.3}$$

has a solution that is unique in the weak sense (i.e., uniqueness in the sense of distribution) on $[0,T]$ for each continuous deterministic initial value $x(0)$, where $B(t)$ is a standard Brownian motion, $\bar{b} = (\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_n)'$, $\bar{\psi}(\cdot)\bar{\psi}'(\cdot) = \bar{A}(\cdot)$.

Define the transition function

$$P_t(x,A) = p(x,0;A,t) = \mathbb{P}(\xi(t) \in A|\xi(0) = x) \tag{5.4}$$

by using the transition probabilities. Let us introduce the notation

$$\mathcal{P}_{t,s}F(x)=\int_{\mathbb{R}^m} p(x,s;dy,t)F(y), \quad t\geq s, x\in\mathbb{R}^m, \quad F\in B_b(\mathbb{R}^m;\mathbb{R}). \tag{5.5}$$

Then the linear operators $\mathcal{P}_t=\mathcal{P}_{t,0}$ is called the *Markovian transition semigroup* or *Markovian semigroup* associated with the transition function $P_t(x,A)$. Note that $\mathcal{P}_tF(x)=\mathbb{E}F(\xi(t;x))$ for the Markov process $\xi(t)$ with the deterministic initial value $\xi(0)=x$. If $\{\xi(t)\}_{t\geq0}$ is a time homogeneous process, then $\mathcal{P}_{t-s}=\mathcal{P}_{t,s}$

Denote by μ the invariant measure of the Markov process ξ (see Theorem 3.2 for the existence of μ). Then we can write

$$\mu b(x,\xi)=\mathbb{E}_\mu b(x,\xi)=\bar{b}(x) \quad \text{and} \quad \mu a_{ij}(x,\xi)=\mathbb{E}_\mu a_{ij}(x,\xi)=\bar{a}_{ij}(x).$$

Likewise, for an appropriate function F, we use μF to denote the integral of F with respect to the invariance measure μ. According to the Markov property of $\xi=(\xi(t))_{t\geq0}$, (A3) implies that for any bounded and continuous functional $b(x,\cdot)$ and $a_{ij}(x,\cdot)$, and deterministic initial value $\xi(0)$,

$$\begin{cases}\lim\limits_{t\to\infty}\mathcal{P}_t^\xi b(x,\xi(0))=\lim\limits_{t\to\infty}\mathbb{E}_t^\xi b(x,\xi)=\lim\limits_{t\to\infty}\mathbb{E}b(x,\xi(t))=\bar{b}(x),\\ \lim\limits_{t\to\infty}\mathcal{P}_t^\xi a_{ij}(x,\xi(0))=\lim\limits_{t\to\infty}\mathbb{E}_t^\xi a_{ij}(x,\xi)=\lim\limits_{t\to\infty}\mathbb{E}a_{ij}(x,\xi(t))=\bar{a}_{ij}(x),\end{cases}$$

where $\mathcal{P}_t^\xi$ has the same definition as $\mathcal{P}_t$ and the superscript ξ denotes the transition probabilities are for $\xi(t)$, and $\mathbb{E}_t^\xi$ denotes the conditional expectation with respect to $\mathcal{F}_t^\xi$ generated by the process $\xi(t)$. Note that the nonnegative definiteness of $A(\cdot)$ implies that $\bar{A}(\cdot)=[\bar{a}_{ij}(\cdot)]_{n\times n}$ is nonnegative definite as well. Next, we state the main weak convergence theorem of this section. Its proof will be divided into several parts.

Theorem 5.1. *If* (A1)-(A4) *hold, then* $\{x^\varepsilon(\cdot)\}$ *is tight in* $D([0,T];\mathbb{R}^n)$*, and the limit of any weakly convergent subsequence satisfies equation* (5.3) *with the same initial value as* $x^\varepsilon(0)=x(0)$ *that is non-random and independent of* ε*, that is,* $x^\varepsilon(\cdot)$ *converges weakly to* $x(\cdot)$ *determined by (5.3).*

To prove the theorem, we use the martingale problem formulation. We say that $x(t)$ of (5.3), is a solution of the martingale problem with operator L, in that for any function $f\in C^{1,2}([0,T]\times\mathbb{R}^n;\mathbb{R})$

$$M_f(t)=f(t,x(t))-f(0,x(0))-\int_0^t L(u,x_u)f(u,x(u))du \tag{5.6}$$

is a local martingale, where for $x\in D([0,T];\mathbb{R}^n)$,

$$L(t,x)\cdot = \frac{\partial \cdot}{\partial t} + \sum_{i=1}^{n} \bar{b}_i(x)\frac{\partial \cdot}{\partial y_i} + \frac{1}{2}\sum_{i,j=1}^{n} \bar{a}_{ij}(x)\frac{\partial^2 \cdot}{\partial y_i \partial y_j}. \tag{5.7}$$

As was mentioned in the beginning of this section, it is not easy to verify (5.1). We thus begin the proof of Theorem 5.1 by working with the N-truncated process. Corresponding to this truncation, we have the operators L^N, $\mathcal{L}^N$ and $\mathbb{L}^N$, which are operators L, $\mathcal{L}$ and $\mathbb{L}$ with x, y, $\bar{b}$, $\bar{\psi}$, and $\bar{A}$ replaced by x^N, y^N, $\bar{b}^N$, $\bar{\psi}^N$, and $\bar{A}^N$, respectively. Not only can assumption (A2) guarantee the existence and uniqueness of the strong solution of the stochastic functional differential truncated equation (5.2), but also the tightness. We proceed with the following theorem.

Theorem 5.2. *Under assumption* (A2)*, there exists a unique strong solution $x^{\varepsilon,N}(t)$ for the stochastic functional differential truncated equation* (5.2) *for any initial value $x^{\varepsilon,N}(0) = x(0) \in S_N$ that is non-random and independent of ε. Moreover, this solution is continuous and $\mathcal{F}_t^{\varepsilon,N}$-adapted, and it is also tight in $D([0,T];\mathbb{R}^n)$.*

To prove this theorem, we need the following lemma (see [27, Theorem 3, p.47] or [31] for a proof).

Lemma 5.3. *Let $\{X^\varepsilon(\cdot)\}$ be a sequence of $\mathcal{F}_t^\varepsilon$-measurable processes with paths in $D([0,T];\mathbb{R}^n)$ satisfying*

$$\lim_{N_0\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}\{\sup_{t\le T} |X^\varepsilon(t)| \ge N_0\} = 0 \tag{5.8}$$

for each $T < \infty$ and $\varepsilon > 0$. For any $\delta > 0$, let there be a random variable $\gamma_\varepsilon(\delta)$ such that

$$\begin{cases} \mathbb{E}_t^\varepsilon \gamma_\varepsilon(\delta) \ge \mathbb{E}_t^\varepsilon \min\{1, |X^\varepsilon(t+s) - X^\varepsilon(t)|^2\}, \ \ \textit{all}\ \ 0 \le s \le \delta, \ \ t \le T, \\ \lim\limits_{\delta\to 0} \limsup\limits_{\varepsilon\to 0} \mathbb{E}\gamma_\varepsilon(\delta) = 0, \end{cases} \tag{5.9}$$

then $X^\varepsilon(\cdot)$ is tight in $D([0,T];\mathbb{R}^n)$.

Proof of Theorem 5.2. Assumption (A2) shows that for any $x \in D([0,T];S_N)$, $b^N(x,\xi)$, $\psi^N(x,\xi)$, $\nabla_x b_i^N(x,\xi)$, $\nabla_x \psi_{ij}^N(x,\xi)$, $\nabla_x^2 b_i^N(x,\xi)$, and $\nabla_x^2 \psi_{ij}^N(x,\xi), i,j = 1,2,\dots,n$ are bounded with respect to x, and continuous and bounded for any $\xi \in \mathbb{R}^m$. Thus $b^N(\cdot,\xi)$ and $\psi^N(\cdot,\xi)$ satisfy the linear growth condition and the Lipschitz condition for any $\xi \in \mathbb{R}^m$. Modifying the argument of [34, Chapter 5], it can be shown that the truncated stochastic functional differential equation (5.2) has a unique strong solution and this solution is continuous and $\mathcal{F}_t^{\varepsilon,N}$-adapted.

From (5.2) and the elementary inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$,

$$|x^{\varepsilon,N}(t)|^2 \le 3|x(0)|^2 + 3\Big|\int_0^t b^N(x_u^{\varepsilon,N}, \xi^\varepsilon(u))du\Big|^2 + 3\Big|\int_0^t \psi^N(x_u^{\varepsilon,N}, \xi^\varepsilon(u))dw_2(u)\Big|^2.$$

Applying the Hölder inequality and the Burkholder-Divis-Gundy inequality [34, Theorem 7.3, p.40] gives

$$\mathbb{E}\Big[\sup_{t\le T}|x^{\varepsilon,N}(t)|^2\Big]\le 3\mathbb{E}|x(0)|^2+3T\int_0^T\mathbb{E}|b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))|^2du$$

$$+12\int_0^T\mathbb{E}|\psi^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))|^2du. \tag{5.10}$$

Assumption (A2) shows that $b^N(x,\xi)$ and $\psi^N(x,\xi)$ are bounded for $x\in D([0,T];S_N)$ and any $\xi\in\mathbb{R}^n$. This implies that there exists K dependent T such that

$$\mathbb{E}\Big[\sup_{t\le T}|x^{\varepsilon,N}(t)|^2\Big]\le K,$$

which shows that (5.8) holds by using the Chebyshev inequality.

Considering (5.2), for any $\delta>0$, $T>0$, and $0\le v\le\delta\wedge 1$

$$x^{\varepsilon,N}(t+v)-x^{\varepsilon,N}(t)=\int_t^{t+v}b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))du+\int_t^{t+v}\psi^N(x_u^\varepsilon,\xi^\varepsilon(u))dw_2(u).$$

Recalling the familiar elementary inequality $(a+b)^2\le 2(|a|^2+|b|^2)$, this implies that

$$\mathbb{E}_t^\varepsilon|x^{\varepsilon,N}(t+v)-x^{\varepsilon,N}(t)|^2\le K\mathbb{E}_t^\varepsilon\Big|\int_t^{t+v}b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))du\Big|^2$$

$$+K\mathbb{E}_t^\varepsilon\Big|\int_t^{t+v}\psi^N(x_u^\varepsilon,\xi^\varepsilon(u))dw_2(u)\Big|^2. \tag{5.11}$$

Note that $x^{\varepsilon,N}(t)\in S_{N+1}$ for any $t\ge 0$. According to (A2), $b^N(x,\xi)$ is bounded for any $x\in D([0,T];S_N)$ and $\xi\in\mathbb{R}^m$. This implies that there exists a random variable $\widetilde{K}_{N,1}^\varepsilon(\delta)$ such that

$$\mathbb{E}_t^\varepsilon\Big|\int_t^{t+v}b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))du\Big|^2\le\mathbb{E}_t^\varepsilon\widetilde{K}_{N,1}^\varepsilon(\delta) \tag{5.12}$$

satisfying $\limsup_{\varepsilon\to 0}\mathbb{E}\widetilde{K}_{N,1}^\varepsilon(\delta)=O(\delta^2)$. According to the martingale isometry [34, Page 28, Theorem 5.21], there exists a random variable $\widetilde{K}_{N,2}^\varepsilon(\delta)$ such that

$$\mathbb{E}_t^\varepsilon\Big|\int_t^{t+v}\psi^N(x_u^\varepsilon,\xi^\varepsilon(u))dw_2(u)\Big|^2=\mathbb{E}_t^\varepsilon\int_t^{t+v}|\psi^N(x_u^\varepsilon,\xi^\varepsilon(u))|^2du\le\mathbb{E}_t^\varepsilon\widetilde{K}_{N,2}^\varepsilon(\delta) \tag{5.13}$$

satisfying $\limsup_{\varepsilon\to 0}\mathbb{E}\widetilde{K}_{N,2}^\varepsilon(\delta)=O(\delta)$. Substituting (5.12) and (5.13) into (5.11) gives that there is a $\widetilde{K}_N^\varepsilon(\delta)$ such that $\mathbb{E}_t^\varepsilon|x^{\varepsilon,N}(t+v)-x^{\varepsilon,N}(t)|^2\le\mathbb{E}_t^\varepsilon\widetilde{K}_N^\varepsilon(\delta)$ satisfying

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\mathbb{E}\widetilde{K}^{\varepsilon}_{N}(\delta)=0.$$

Applying Lemma 5.3 yields the desired result. This completes the proof. □

Since $x^{\varepsilon,N}(\cdot)$ is tight, by Prohorov's theorem, it is sequentially compact. Thus we can extract a weakly convergent subsequence. Do so and still index the convergent subsequence by ε and denote the limit as $x^N(\cdot)$. By the Skorohod representation, with a slight abuse of notation (without changing notation), we may assume that $x^{\varepsilon,N}(\cdot)$ converges to $x^N(\cdot)$ in the sense of w.p.1. We proceed to characterize the limit process $x^N(\cdot)$ by use of the martingale problem formulation. In what follows, we characterize the weak limit by applying the following lemma [39,46].

Lemma 5.4. *Let $X^{\varepsilon}(\cdot)$ be an $\mathbb{R}^n$-valued process defined on $[0,T]$, with $X^{\varepsilon}(0)=X(0)$ being deterministic and independent of ε. Let $\{X(\cdot)\}$ be tight on $D([0,T];\mathbb{R}^n)$. Suppose* (A4) *holds and L is the corresponding operator defined by* (5.7). *For each $f(\cdot)\in C^4_0(\mathbb{R}^n;\mathbb{R})$* (*or any dense subset of it*)*, each $T<\infty$, there exists $f^{\varepsilon}(\cdot)\in\mathcal{D}(\hat{\mathcal{L}}^{\varepsilon})$ such that*

$$\text{p-}\lim_{\varepsilon\to 0}[f^{\varepsilon}(\cdot)-f(X^{\varepsilon}(\cdot))]=0, \tag{5.14}$$

and

$$\text{p-}\lim_{\varepsilon\to 0}[\hat{\mathcal{L}}^{\varepsilon}f^{\varepsilon}(\cdot)-L(\cdot,X^{\varepsilon}_{\cdot})f(X^{\varepsilon}(\cdot))]=0. \tag{5.15}$$

Then, $X^{\varepsilon}(\cdot)\Rightarrow x(\cdot)$, where $x(\cdot)$ is the solution of the stochastic differential equation (5.3).

According to the definition of p-lim, to prove (5.14) for $x^{\varepsilon,N}(t)$, for any $f(\cdot)\in C^4_0(\mathbb{R}^n;\mathbb{R})$, we need to look for a function $f^{\varepsilon,N}(\cdot)\in\mathcal{D}(\hat{\mathcal{L}}^{\varepsilon,N})$ and verify the corresponding conditions

$$\begin{cases}\sup_{t,\varepsilon}\mathbb{E}|f^{\varepsilon,N}(t)-f(x^{\varepsilon,N}(t))|<\infty,\\ \lim_{\varepsilon\to 0}\mathbb{E}|f^{\varepsilon,N}(t)-f(x^{\varepsilon,N}(t))|=0\ \text{ for each } t.\end{cases} \tag{5.16}$$

Similarly, to prove (5.15) for the above $x^{\varepsilon,N}(t)$ and $f(\cdot)$, we need to verify the conditions

$$\begin{cases}\sup_{t,\varepsilon}\mathbb{E}|\hat{\mathcal{L}}^{\varepsilon,N}f^{\varepsilon,N}(t)-L^N(t,x^{\varepsilon,N}_t)f(x^{\varepsilon,N}(t))|<\infty,\\ \lim_{\varepsilon\to 0}\mathbb{E}|\hat{\mathcal{L}}^{\varepsilon,N}f^{\varepsilon,N}(t)-L^N(t,x^{\varepsilon,N}_t)f(x^{\varepsilon,N}(t))|=0\ \text{ for each } t.\end{cases} \tag{5.17}$$

Remark 5.5. In the process of the averaging, the fast-changing variable $\xi^{\varepsilon}(t)$ is treated as noise and is averaged out. Noting the underlying system is a functional differential equation, we use the perturbed test functional method to examine the weak convergence. Introducing the perturbed test functionals allows us to eliminate the noise terms $\xi^{\varepsilon}(t)$ through averaging, and obtain the desired terms in the limit.

Proof of Theorem 5.1. For any $f(\cdot)\in C^4_0(\mathbb{R}^n;\mathbb{R})$, to use the perturbed test functional method, for any $t<T$, we choose

$$f_1^{\varepsilon,N}(t) := V_1(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) = \int_t^T f_y'(x^{\varepsilon,N}(t))\mathbb{E}_t^{\xi^\varepsilon}[b^N(x_t^{\varepsilon,N}, \xi^\varepsilon(u)) - \bar{b}^N(x_t^{\varepsilon,N})]du, \quad (5.18)$$

$$f_2^{\varepsilon,N}(t) := V_2(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) = \sum_{i,j=1}^n \int_t^T f_{y_i y_j}(x^{\varepsilon,N}(t))\mathbb{E}_t^{\xi^\varepsilon}[a_{ij}^N(x_t^{\varepsilon,N}, \xi^\varepsilon(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})]du. \quad (5.19)$$

In building the perturbations, the slow-changing variable $x^{\varepsilon,N}(t)$ and the corresponding delay term $x_t^{\varepsilon,N}$ are considered as parameters. Making change of variable u/ε to u yields that

$$f_1^{\varepsilon,N}(t) = \varepsilon \int_{t/\varepsilon}^{T/\varepsilon} f_y'(x^{\varepsilon,N}(t))\mathbb{E}_{t/\varepsilon}^{\xi}[b^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{b}^N(x_t^{\varepsilon,N})]du, \quad (5.20)$$

$$f_2^{\varepsilon,N}(t) = \varepsilon \sum_{i,j=1}^n \int_{t/\varepsilon}^{T/\varepsilon} f_{y_i y_j}(x^{\varepsilon,N}(t))\mathbb{E}_{t/\varepsilon}^{\xi}[a_{ij}^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})]du. \quad (5.21)$$

Note that $\mathbb{E}_t^\xi$ is the conditional expectation with respect to the σ-algebra $\mathcal{F}_t^\xi$ generated by $\xi(t)$ in (3.2). Define

$$f^{\varepsilon,N}(t) = f(x^{\varepsilon,N}(t)) + f_1^{\varepsilon,N}(t) + \frac{1}{2} f_2^{\varepsilon,N}(t). \quad (5.22)$$

Assumption (A2) shows that $b(x,\xi)$ is boundedness preserving on x and $b(x,\cdot)$ is continuous and bounded, which implies that $b^N(x,\cdot)$ is a bounded and continuous function for any $x \in D([0,T]; S_N)$, i.e., $b^N(x,\cdot) \in C_b(\mathbb{R}^m; \mathbb{R}^n)$. Note that the invariant measure μ is ergodic exponentially. Applying (A3), (A.17) in the appendix, and homogeneity of $\xi(\cdot)$ gives

$$\begin{aligned}
\sup_{t\le T} |f_1^{\varepsilon,N}(t)| &= \varepsilon \sup_{t\le T} \Big| \int_{t/\varepsilon}^{T/\varepsilon} f_y(x^{\varepsilon,N}(t))\mathbb{E}_{t/\varepsilon}^{\xi}[b^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{b}^N(x_t^{\varepsilon,N})]du \Big|, \\
&\le \varepsilon |f_y(x^{\varepsilon,N}(t))| \sup_{t\le T} \int_{t/\varepsilon}^{T/\varepsilon} |\mathcal{P}_{t/\varepsilon,u} b^N(x_t^{\varepsilon,N}, \xi(t/\varepsilon)) - \mu b^N(x_t^{\varepsilon,N}, \xi)|du \\
&= \varepsilon |f_y(x^{\varepsilon,N}(t))| \sup_{t\le T} \int_{t/\varepsilon}^{T/\varepsilon} |\mathcal{P}_{u-t/\varepsilon} b^N(x_t^{\varepsilon,N}, \xi(t/\varepsilon)) - \mu b^N(x_t^{\varepsilon,N}, \xi)|du \\
&\le \varepsilon K \sup_{t\le T} \int_{t/\varepsilon}^{T/\varepsilon} e^{-\frac{\tilde{\xi}}{2}(u-t/\varepsilon)} du,
\end{aligned}$$

$$
\begin{aligned}
&= \varepsilon K(1 - e^{-\frac{\tilde{\xi}(T-t)}{2\varepsilon}}) \\
&= O(\varepsilon),
\end{aligned} \tag{5.23}
$$

which implies

$$|f_1^{\varepsilon,N}(t)| \to 0 \text{ w.p.1 as } \varepsilon \to 0,$$

since $f(\cdot) \in C_0^4(\mathbb{R}^n; \mathbb{R})$. Assumption (A2) also reveals that $\psi(x,\xi)$ is boundedness preserving on x and $\psi(x,\cdot)$ is bounded and continuous, which implies that $\psi^N(x,\cdot) \in C_b(\mathbb{R}^m; \mathbb{R}^{n\times l_2})$ for any $x \in D([0,T]; S_N)$. This, together with the definition of $A(x,\xi)$ in (A3), implies that $a_{ij}(x,\cdot) \in C_b(\mathbb{R}^m; \mathbb{R})$ for any $x \in D([0,T]; S_N)$. Applying the technique similar to (5.23) gives

$$
\begin{aligned}
\sup_{t\le T} |f_2^{\varepsilon,N}(t)| &= \varepsilon \sum_{i,j=1}^{n} \sup_{t\le T} \Big| \int_{t/\varepsilon}^{T/\varepsilon} f_{y_i y_j}(x^{\varepsilon,N}(t)) \mathbb{E}^{\xi}_{t/\varepsilon}[a_{ij}^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})]du \Big| \\
&\le \varepsilon \sum_{i,j=1}^{n} |f_{y_i y_j}(x^{\varepsilon,N}(t))| \sup_{t\le T} \Big| \int_{t/\varepsilon}^{T/\varepsilon} \mathbb{E}^{\xi}_{t/\varepsilon}[a_{ij}^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})]du \Big| \\
&= \varepsilon \sum_{i,j=1}^{n} |f_{y_i y_j}(x^{\varepsilon,N}(t))| \sup_{t\le T} \int_{t/\varepsilon}^{T/\varepsilon} |\mathcal{P}^x_{u-t/\varepsilon} a_{ij}^N(x_t^{\varepsilon,N}, \xi(t/\varepsilon)) - \mu a_{ij}^N(x_t^{\varepsilon,N}, \xi)| du \\
&\le \varepsilon K \sup_{t\le T} \int_{t/\varepsilon}^{T/\varepsilon} e^{-\frac{\tilde{\xi}}{2}(u-t/\varepsilon)} du, \\
&= \varepsilon K(1 - e^{-\frac{\tilde{\xi}(T-t)}{2\varepsilon}}) \\
&= O(\varepsilon),
\end{aligned} \tag{5.24}
$$

where $\mu a_{ij}^N(x_t^{\varepsilon,N}, \xi) = \mathbb{E}_\mu a_{ij}^N(x_t^{\varepsilon,N}, \xi)$, which implies $|f_2^{\varepsilon,N}(t)| \to 0$ w.p.1 as $\varepsilon \to 0$ since $f(\cdot) \in C_0^4(\mathbb{R}^n; \mathbb{R})$. Note that (5.23) and (5.24) also imply that there exist $\varepsilon_0 > 0$ and η,

$$\sup_{t\ge 0, \varepsilon\in(0,\varepsilon_0)} \mathbb{E}|f^{\varepsilon,N}(t) - f(x^{\varepsilon,N}(t))| \le \sup_{t\ge 0, \varepsilon\in(0,\varepsilon_0)} [\mathbb{E}|f_1^{\varepsilon,N}(t)| + \mathbb{E}|f_2^{\varepsilon,N}(t)|] < \eta,$$

and

$$\lim_{\varepsilon\to 0} \mathbb{E}|f^{\varepsilon,N}(t) - f(x^{\varepsilon,N}(t))| = 0 \text{ for each } t. \tag{5.25}$$

Thus (5.16) holds. According to Definition 2.1,

$$\text{p-}\lim_{\varepsilon\to 0} [f^{\varepsilon,N}(\cdot) - f(x^{\varepsilon,N}(\cdot))] = 0,$$

that is, (5.14) holds.

To prove (5.15), for $(t, y, x) \in [0, T] \times \mathbb{R}^n \times D([0, T]; \mathbb{R}^n)$, let us define

$$V(t, y, x) = V_1(t, y, x) + \frac{1}{2} V_2(t, y, x).$$

According to the definitions of $f^{\varepsilon,N}(\cdot)$, $\mathcal{L}^{\varepsilon,N}$, L^N, $\mathbb{L}^N$, applying (4.7) and (4.11) gives

$$\begin{aligned}
\mathcal{L}^{\varepsilon,N} f^{\varepsilon,N}(t) &= \text{p-}\lim_{\delta\to 0} \frac{\mathbb{E}_t^{\varepsilon,N} f^{\varepsilon,N}(t+\delta) - f^{\varepsilon,N}(t)}{\delta} \\
&= \text{p-}\lim_{\delta\to 0} \frac{\mathbb{E}_t^{\varepsilon,N} f(x^{\varepsilon,N}(t+\delta)) - f(x^{\varepsilon,N}(t))}{\delta} \\
&\quad + \text{p-}\lim_{\delta\to 0} \frac{\mathbb{E}_t^{\varepsilon,N} V(t+\delta, x^{\varepsilon,N}(t+\delta), x_{t+\delta}^{\varepsilon,N}) - V(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N})}{\delta} \\
&= L^N(t, x_t^{\varepsilon,N}) f(x^{\varepsilon,N}(t)) + \mathbb{L}^N(t, x_t^{\varepsilon,N}) V(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}),
\end{aligned} \tag{5.26}$$

where $\mathbb{E}_t^{\varepsilon,N}$ is the conditional expectation with respect to the σ-algebra $\mathcal{F}_t^{\varepsilon,N}$, and $\mathbb{L}^N V(t, y, x) = \mathbb{L}^N(t, x_t^{\varepsilon,N}) V_1(t, y, x) + \mathbb{L}^N(t, x_t^{\varepsilon,N}) V_2(t, y, x)/2$.

Let us first examine $\mathbb{L}^N(t, x_t^{\varepsilon,N}) V_1(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N})$. Note that

$$\begin{aligned}
&\mathbb{L}^N(t, x_t^{\varepsilon,N}) V_1(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) \\
&\quad = \varepsilon \sum_{i=1}^n \mathbb{L}^N(t, x_t^{\varepsilon,N}) \Big[\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t)) \mathbb{E}_{t/\varepsilon}^{\xi} [b_i^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{b}_i^N(x_t^{\varepsilon,N})] du \Big].
\end{aligned} \tag{5.27}$$

Note that $\mathcal{F}_t^{\xi^\varepsilon} \subset \mathcal{F}_t^{\varepsilon,N}$. This implies that when we apply $\mathbb{E}_t^{\varepsilon,N}$ to $\xi^\varepsilon(\cdot)$, it is equivalent to use $\mathbb{E}_t^{\xi^\varepsilon}$. This, together with the definition of $\mathbb{L}^N$ yields

$$\begin{aligned}
&\varepsilon \mathbb{L}^N(t, x_t^{\varepsilon,N}) \Big[\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t)) \mathbb{E}_{t/\varepsilon}^{\xi} (b_i^N(x_t^{\varepsilon,N}, \xi(u)) - \bar{b}_i^N(x_t^{\varepsilon,N})) du \Big] \\
&= I_1^\varepsilon(x^{\varepsilon,N}(t), x_t^{\varepsilon,N}, \xi^\varepsilon(t)) + \varepsilon \sum_{j=1}^n I_{2j}^\varepsilon(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) + \frac{\varepsilon}{2} \sum_{k,j=1}^n I_{3kj}^\varepsilon(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) \\
&\quad + \varepsilon \sum_{j=1}^n I_{4j}^\varepsilon(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) + \varepsilon \sum_{k,j=1}^n I_{5kj}^\varepsilon(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}) + \frac{\varepsilon}{2} \sum_{k,j=1}^n I_{6kj}^\varepsilon(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N}),
\end{aligned} \tag{5.28}$$

where

$$\begin{aligned}
I_1^\varepsilon(y, x, \xi^\varepsilon(t)) &= -f_{y_i}(y)[b_i^N(x, \xi^\varepsilon(t)) - \bar{b}_i^N(x)], \\
I_{2j}^\varepsilon(t, y, x) &= \int_{t/\varepsilon}^{T/\varepsilon} f_{y_i y_j}(y) \mathbb{E}_{t/\varepsilon}^{\xi} [b_i^N(x, \xi(u)) - \bar{b}_i^N(x)] du \, b_j^N(x, \xi^\varepsilon(t)),
\end{aligned}$$

$$I^{\varepsilon}_{3kj}(t,y,x)=\int_{t/\varepsilon}^{T/\varepsilon} f_{y_iy_ky_j}(y)\mathbb{E}^{\xi}_{t/\varepsilon}[b^N_i(x,\xi(u))-\bar{b}^N_i(x)]du a^N_{kj}(x,\xi^{\varepsilon}(t)),$$

$$I^{\varepsilon}_{4j}(t,y,x)=\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(y)\partial_j[\mathbb{E}^{\xi}_{t/\varepsilon}(b^N_i(x,\xi(u))-\bar{b}^N_i(x))]du b^N_j(x,\xi^{\varepsilon}(t)),$$

$$I^{\varepsilon}_{5kj}(t,y,x)=\int_{t/\varepsilon}^{T/\varepsilon} f_{y_iy_k}(y)\partial_j[\mathbb{E}^{\xi}_{t/\varepsilon}(b^N_i(x,\xi(u))-\bar{b}^N_i(x))]du a^N_{kj}(x,\xi^{\varepsilon}(t)),$$

$$I^{\varepsilon}_{6kj}(t,y,x)=\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(y)\partial^2_{kj}[\mathbb{E}^{\xi}_{t/\varepsilon}(b^N_i(x,\xi(u))-\bar{b}^N_i(x))]du a^N_{kj}(x,\xi^{\varepsilon}(t)).$$

It can be seen that I^{ε}_1 is from the differential on t in this integral, I^{ε}_{2j} are from the differentials on $x^{\varepsilon,N}(t)$, I^{ε}_{3j} are from the second-order differentials on $x^{\varepsilon,N}(t)$, I^{ε}_{4j} are from the functional differentials on $x^{\varepsilon,N}_t$, I^{ε}_{5kj} and I^{ε}_{6kj} are from the mixed partial derivatives on $x^{\varepsilon,N}(t)$ and $x^{\varepsilon,N}_t$. Recall that $b^N(x,\cdot)\in C_b(\mathbb{R}^m;\mathbb{R}^n)$. This, together with the boundedness preserving property of $b(x,\xi)$ on x, leads to that $b^N(x,\xi)$ is bounded for any $x\in D([0,T];S_N)$ and $\xi\in\mathbb{R}^m$. Since and $f(\cdot)\in C^4_0(\mathbb{R}^n;\mathbb{R})$, similar to (5.23),

$$\begin{aligned}
&\sup_{t\le T}|I^{\varepsilon}_{2j}(t,x^{\varepsilon,N}(t),x^{\varepsilon,N}_t)|\\
&=\sup_{t\le T}\Big|\int_{t/\varepsilon}^{T/\varepsilon} f_{y_iy_j}(x^{\varepsilon,N}(t))\mathbb{E}^{\xi}_{t/\varepsilon}[b^N_i(x^{\varepsilon,N}_t,\xi(u))-\bar{b}^N_i(x_t)]du b^N_j(x^{\varepsilon,N}_t,\xi^{\varepsilon}(t))\Big|\\
&\le K_N\sup_{t\le T}\Big|\int_{t/\varepsilon}^{T/\varepsilon}|f_{y_iy_j}(x^{\varepsilon,N}(t))|e^{-\frac{\tilde{\varsigma}}{2}(u-t/\varepsilon)}du\Big|\\
&=\frac{2}{\tilde{\varsigma}}K_N(1-e^{-\frac{\tilde{\varsigma}(T-t)}{2\varepsilon}})<\infty,
\end{aligned}$$

which implies

$$\varepsilon\sup_{t\le T}\Big|\sum_{j=1}^n I^{\varepsilon}_{2j}(t,x^{\varepsilon,N}(t),x^{\varepsilon,N}_t)\Big|\le\varepsilon\sup_{t\le T}\sum_{j=1}^n|I^{\varepsilon}_{2j}(t,x^{\varepsilon,N}(t),x^{\varepsilon,N}_t)|=O(\varepsilon).$$

Likewise, we have

$$\frac{\varepsilon}{2}\sup_{t\le T}\Big|\sum_{k,j=1}^n I^{\varepsilon}_{3kj}(t,x^{\varepsilon,N}(t),x^{\varepsilon,N}_t)\Big|\le\frac{\varepsilon}{2}\sup_{t\le T}\sum_{k,j=1}^n|I^{\varepsilon}_{3kj}(t,x^{\varepsilon,N}(t),x^{\varepsilon,N}_t)|=O(\varepsilon)$$

since $A^N(x,\xi)$ is bounded for any $x \in D([0,T];S_N)$ and $\xi \in \mathbb{R}^m$ according to $a_{ij}(x,\cdot) \in C_b(\mathbb{R}^m;\mathbb{R})$ and the boundedness preserving property of $\psi(x,\xi)$ for x. By (A2), $\nabla_x b^N(x,\xi)$ is boundedness preserving for any $x \in D([0,T];\mathbb{R}^n)$, and continuous and bounded with respect to $\xi \in \mathbb{R}^m$. Therefore, for any $i,j = 1,2,\dots,n$, $\partial_j b_i^N(x,\cdot) \in C_b(\mathbb{R}^m;\mathbb{R})$ for any $x \in D([0,T];S_N)$. Hence applying similar technique to the estimation of I_{2j}^ε and recalling the boundedness of $b^N(x,\xi)$ for any $x \in D([0,T];S_N)$ and $\xi \in \mathbb{R}^m$ give that

$$
\begin{aligned}
&\sup_{t\le T}|I_{4j}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})|\\
&=\sup_{t\le T}\Big|\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t))\partial_j[\mathbb{E}_{t/\varepsilon}^\xi(b_i^N(x_t^{\varepsilon,N},\xi(u))-\bar b_i^N(x_t^{\varepsilon,N}))]du\, b_j^N(x_t^{\varepsilon,N},\xi^\varepsilon(t))\Big|\\
&=\sup_{t\le T}\Big|\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t))[\mathbb{E}_{t/\varepsilon}^\xi\partial_j b_i^N(x_t^{\varepsilon,N},\xi(t))-\partial_j\bar b_i^N(x_t^{\varepsilon,N})]du\, b_j^N(x_t^{\varepsilon,N},\xi^\varepsilon(t))\Big|\\
&\le K_N\sup_{t\le T}\int_{t/\varepsilon}^{T/\varepsilon}|f_{y_i}(x^{\varepsilon,N}(t))|e^{-\frac{\tilde\varsigma}{2}(u-t/\varepsilon)}du\\
&=\frac{2}{\tilde\varsigma}K_N(1-e^{-\frac{\tilde\varsigma(T-t)}{2\varepsilon}})<\infty
\end{aligned}
$$

with probability 1, which implies that

$$
\varepsilon\sup_{t\le T}\Big|\sum_{j=1}^n I_{4j}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big|\le\varepsilon\sup_{t\le T}\sum_{j=1}^n|I_{4j}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})|=O(\varepsilon).
$$

The same technique as the estimation of I_{3j}^ε gives

$$
\frac{\varepsilon}{2}\sup_{t\le T}\Big|\sum_{k,j=1}^n I_{5kj}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big|=O(\varepsilon).
$$

Similarly, (A2) also leads to that for any $i,j=1,2,\dots,n$, $\partial_{ij}^2 b^N(x,\cdot)\in C_b(\mathbb{R}^m;\mathbb{R})$ for any $x\in D([0,T];S_N)$. This, together with the boundedness of $a^N(x,\xi)$ for any $x\in D([0,T];S_N)$ and $\xi\in\mathbb{R}^m$, yields

$$
\begin{aligned}
&\sup_{t\le T}|I_{6kj}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})|\\
&\le\sup_{t\le T}\Big|\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t))\partial_{kj}^2[\mathbb{E}_{t/\varepsilon}^\xi(b_i^N(x_t^{\varepsilon,N},\xi(u))-\bar b_i^N(x_t^{\varepsilon,N}))]du\, a_{kj}^N(x_t^{\varepsilon,N},\xi^\varepsilon(t))\Big|
\end{aligned}
$$

$$\leq K_N \sup_{t\leq T}\Big|\int_{t/\varepsilon}^{T/\varepsilon} f_{y_i}(x^{\varepsilon,N}(t))[\mathbb{E}^{\xi}_{t/\varepsilon}(\partial^2_{kj} b_i^N(x_t^{\varepsilon,N},\xi(u)) - \partial^2_{kj}\bar{b}_i^N(x_t^{\varepsilon,N}))]du\Big|$$

$$\leq K_N \sup_{t\leq T}\int_{t/\varepsilon}^{T/\varepsilon} |f_{y_i}(x^{\varepsilon,N}(t))| e^{-\frac{\tilde{\varsigma}}{2}(u-t/\varepsilon)}du$$

$$\leq \frac{2K_N}{\tilde{\varsigma}}(1-e^{-\frac{\tilde{\varsigma}(T-t)}{2\varepsilon}}) < \infty,$$

which leads to that

$$\frac{\varepsilon}{2}\sum_{k,j=1}^{n} I^{\varepsilon}_{6kj}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}) \leq \frac{\varepsilon}{2}\sum_{k,j=1}^{n} |I^{\varepsilon}_{6kj}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})| = O(\varepsilon).$$

Substituting these estimates into (5.28) gives

$$\begin{aligned}\mathbb{L}^N(t,x_t^{\varepsilon,N})V_1(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}) &= O(\varepsilon) - \sum_{i=1}^{n}[f_{y_i}(x^{\varepsilon,N}(t))(b_i^N(x_t^{\varepsilon,N},\xi^{\xi}(t)) - \bar{b}_i^N(x_t^{\varepsilon,N}))]\\ &= O(\varepsilon) - f_y(x^{\varepsilon,N}(t))[b^N(x_t^{\varepsilon,N},\xi^{\varepsilon}(t)) - \bar{b}^N(x_t^{\varepsilon,N})] \quad (5.29)\end{aligned}$$

with probability 1. To proceed, let us estimate $\mathbb{L}^N(t,x_t^{\varepsilon,N})V_2(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})$. Note that

$$\begin{aligned}&\mathbb{L}^N(t,x_t^{\varepsilon,N})V_2(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\\ &= \varepsilon\sum_{i,j=1}^{n}\mathbb{L}^N(t,x_t^{\varepsilon,N})\Big[\int_{t/\varepsilon}^{T/\varepsilon} f_{y_iy_j}(x^{\varepsilon,N}(t))\mathbb{E}^{\xi}_{t/\varepsilon}(a_{ij}^N(x_t^{\varepsilon,N},\xi(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N}))du\Big].\end{aligned} \quad (5.30)$$

Similar to (5.28), we have

$$\begin{aligned}&\varepsilon\mathbb{L}^N(t,x_t^{\varepsilon,N})\Big[\int_{t/\varepsilon}^{T/\varepsilon} f_{y_iy_j}(x^{\varepsilon,N}(t))\mathbb{E}^{\xi}_{t/\varepsilon}(a_{ij}^N(x_t^{\varepsilon,N},\xi(u)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N}))du\Big]\\ &= J_1^{\varepsilon}(x^{\varepsilon,N}(t),x_t^{\varepsilon,N},\xi^{\varepsilon}(t)) + \varepsilon\sum_{k=1}^{n} J^{\varepsilon}_{2k}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}) + \varepsilon\sum_{k=1}^{n} J^{\varepsilon}_{3k}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\\ &\quad + \frac{\varepsilon}{2}\sum_{k,l=1}^{n} J^{\varepsilon}_{4kl}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}) + \frac{\varepsilon}{2}\sum_{k,l=1}^{n} J^{\varepsilon}_{5kl}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\\ &\quad + \frac{\varepsilon}{2}\sum_{k,l=1}^{n} J^{\varepsilon}_{6kl}(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}),\end{aligned} \quad (5.31)$$

where

$$J_1^\varepsilon(y,x,\xi^\varepsilon(t)) = -f_{y_i y_j}(y)[a_{ij}^N(x,\xi^\varepsilon(t)) - \bar{a}_{ij}^N(x)],$$

$$J_{2k}^\varepsilon(t,y,x) = \int\limits_{t/\varepsilon}^{T/\varepsilon} f_{y_i y_j y_k}(y)\mathbb{E}_{t/\varepsilon}^{\xi}[a_{ij}^N(x,\xi(u)) - \bar{a}_{ij}^N(x)]du b_k^N(x,\xi^\varepsilon(t)),$$

$$J_{3k}^\varepsilon(t,y,x) = \int\limits_{t/\varepsilon}^{T/\varepsilon} f_{y_i y_j}(y)\partial_k[\mathbb{E}_{t/\varepsilon}^{\xi}(a_{ij}^N(x,\xi(u)) - \bar{a}_{ij}^N(x))]du b_k^N(x,\xi^\varepsilon(t)),$$

$$J_{4kl}^\varepsilon(t,y,x) = \int\limits_{\varepsilon/t}^{T/\varepsilon} f_{y_i y_j y_k y_l}(y)\mathbb{E}_{t/\varepsilon}^{\xi}[a_{ij}^N(x,\xi(u)) - \bar{a}_{ij}^N(x)]du a_{kl}^N(x,\xi^\varepsilon(t)),$$

$$J_{5kl}^\varepsilon(t,y,x) = \int\limits_{\varepsilon/t}^{T/\varepsilon} f_{y_i y_j y_k}(y)\partial_l[\mathbb{E}_{t/\varepsilon}^{\xi}(a_{ij}^N(x,\xi(u)) - \bar{a}_{ij}^N(x))]du a_{kl}^N(x,\xi^\varepsilon(t)),$$

$$J_{6kl}^\varepsilon(t,y,x) = \int\limits_{\varepsilon/t}^{T/\varepsilon} f_{y_i y_j}(y)\partial_{kl}^2[\mathbb{E}_{t/\varepsilon}^{\xi}(a_{ij}^N(x,\xi(u)) - \bar{a}_{ij}^N(x))]du a_{kl}^N(x,\xi(t)).$$

Recall that $a_{ij}(x,\xi) = \sum_{l=1}^{l_2} \psi_{il}(x,\xi)\psi_{lj}(x,\xi)$. Condition (A2) implies that $\psi(x,\xi)$, $\nabla_x \psi_{ij}(x,\xi)$, and $\nabla_x^2 \psi_{ij}(x,\xi)$, $i,j = 1,2,\dots,n$ are boundedness preserving with respect to $x \in D([0,T];\mathbb{R}^n)$, and continuous and bounded with respect to $\xi \in \mathbb{R}^m$ for $x \in D([0,T];S_N)$. These imply that for any $x \in D([0,T];S_N)$, $A(x,\cdot) \in C_b(\mathbb{R}^m;\mathbb{R}^{n\times n})$, $\partial_k a_{ij}(x,\cdot)$ and $\partial_{kl}^2 a_{ij}(x,\cdot) \in C_b(\mathbb{R}^m;\mathbb{R})$ for $i,j = 1,2,\dots,n$. Applying the same technique as the estimates of $I_{2j}^\varepsilon(t,x(t),x_t)$ yields

$$\varepsilon \sup_{t\le T}\Big|\sum_{k=1}^n J_{2k}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big| = O(\varepsilon) \text{ and } \frac{\varepsilon}{2}\sup_{t\le T}\Big|\sum_{k,l=1}^n J_{4kl}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big| = O(\varepsilon).$$

Similar to the estimates of I_{4j}^ε, I_{5kj}^ε, and I_{6kj}^ε, we obtain

$$\varepsilon \sup_{t\le T}\Big|\sum_{k=1}^n J_{3k}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big| = O(\varepsilon) \quad \text{and}$$
$$\frac{\varepsilon}{2}\sup_{t\le T}\Big|\sum_{k,l=1}^n J_{vkl}^\varepsilon(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N})\Big| = O(\varepsilon) \quad (v=5,6).$$

These estimates lead to

$$\mathbb{L}^N(t,x_t^{\varepsilon,N})V_2(t,x^{\varepsilon,N}(t),x_t^{\varepsilon,N}) = O(\varepsilon) - \sum_{i,j=1}^n f_{y_i y_j}(x^{\varepsilon,N}(t))[a_{ij}^N(x_t^{\varepsilon,N},\xi^\varepsilon(t)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})] \tag{5.32}$$

with probability 1. The estimates of $\mathbb{L}^N(t, x_t^{\varepsilon,N})V_1(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N})$ and $\mathbb{L}^N(t, x_t^{\varepsilon,N})V_2(t, x^{\varepsilon,N}(t), x_t^{\varepsilon,N})$, together with (5.26), yield

$$\hat{\mathcal{L}}^{\varepsilon,N} f^{\varepsilon,N}(t) = O(\varepsilon) + L^N(t, x_t^{\varepsilon,N}) f(x^{\varepsilon,N}(t)) - f_y(x^{\varepsilon,N}(t))[b^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t)) - \bar{b}^N(x_t^{\varepsilon,N})] \\ -\frac{1}{2}\sum_{i,j=1}^n f_{y_i y_j}(x^{\varepsilon,N}(t))[a_{ij}^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t)) - \bar{a}_{ij}^N(x_t^{\varepsilon,N})] \tag{5.33}$$

with probability 1. Applying the generator L^N defined by (4.7) to the solution process $x^{\varepsilon,N}(t)$ in the stochastic functional differential equation (5.2) gives

$$L^N(t, x_t^{\varepsilon,N}) f(x^{\varepsilon,N}(t)) = f_y(x^{\varepsilon,N}(t)) b^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t)) + \frac{1}{2}\sum_{i,j=1}^n f_{y_i y_j}(x^{\varepsilon,N}(t)) a_{ij}^N(x_t^{\varepsilon,N}, \xi^\varepsilon(t)).$$

Substituting this into (5.33) yields

$$\begin{aligned}\hat{\mathcal{L}}^{\varepsilon,N} f^{\varepsilon,N}(t) &= O(\varepsilon) + f_y(x^{\varepsilon,N}(t))\bar{b}^N(x_t^{\varepsilon,N}) + \frac{1}{2}\sum_{i,j=1}^n f_{y_i y_j}(x^{\varepsilon,N}(t))\bar{a}_{ij}^N(x_t^{\varepsilon,N}) \\ &= O(\varepsilon) + L^N(t, x_t^{\varepsilon,N}) f(x^{\varepsilon,N}(t))\end{aligned} \tag{5.34}$$

with probability 1, which implies that (5.17) holds since $f(\cdot) \in C_0^4(\mathbb{R}^n; \mathbb{R}^n)$ and the definition of $f^\varepsilon(\cdot)$. Thus (5.15) holds. This, together with (5.25) yields $x^{\varepsilon,N}(\cdot) \Rightarrow x^N(\cdot)$ as $\varepsilon \to 0$ by virtue of Lemma 5.4, where $x^N(\cdot)$ solves the martingale problem with operator L^N.

Next, we move from the weak convergence of the truncated process to that of untruncated processes. The argument is similar to that of [27, p.46]. For any continuous deterministic initial value $x(0)$, let $\mathbb{P}(\cdot)$ and $\mathbb{P}^N(\cdot)$ denote the probabilities induced by $x(\cdot)$ and $x^N(\cdot)$, respectively, on the Borel sets of $D([0,T]; \mathbb{R}^n)$. By (A4), the martingale problem has a unique solution for each $x(0)$, so $\mathbb{P}(\cdot)$ is unique. For each $T < \infty$, the uniqueness implies that $\mathbb{P}(\cdot)$ agrees with $\mathbb{P}^N(\cdot)$ on all Borel sets of the set of paths in $D([0,T]; S_N)$ for each $t \le T$. However, $\mathbb{P}\{\sup_{t\le T} |x(t)| \le N\} \to 1$ as $N \to \infty$. This together with the weak convergence of $x^{\varepsilon,N}(\cdot)$ implies that $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$. Moreover, the uniqueness implies that the limit does not depend on the chosen subsequences. The proof is thus complete. □

6. Extension and examples

In assumption (A2), some requirements may be weakened. It is important to do this since in a wide range of applications for control systems, biology, and communications systems, the b and ψ etc. may not be boundedness preserving, but only possess such properties in a local sense. In this section, we relax assumption (A2).

(A2') $b(x,\xi)$, $\psi(x,\xi)$, $\nabla_x b_i(x,\xi)$, $\nabla_x \psi_{ij}(x,\xi)$, $\nabla_x^2 b_i(x,\xi)$ and $\nabla_x^2 \psi_{ij}(x,\xi)$, $i, j = 1, 2, \dots, n$ are continuous with respect to $\xi \in \mathbb{R}^m$ for $x \in D([0,T]; G)$, where $G \subset \mathbb{R}^n$ is a compact set; Moreover, there exist the boundedness preserving functionals $K_1(x)$ and $K_2(x)$ such that

$$|b(x,\xi)| \vee |\nabla_x b_i(x,\xi)| \vee |\nabla^2_{xx} b_i(x,\xi)| \le K_1(x)(1+|\xi|^2), \tag{6.1}$$

$$|\psi(x,\xi)| \vee |\nabla_x \psi_{ij}(x,\xi)| \vee |\nabla^2_{xx} \psi(x,\xi)| \le K_2(x)(1+|\xi|^2). \tag{6.2}$$

Remark 6.1. In this assumption, since $K_1(x)$ and $K_2(x)$ are both boundedness preserving functionals, for any $x \in D([0,T], S_N)$, both $K_1(x)$ and $K_2(x)$ are bounded, that is, there exist $K_N > 0$ such that $K_1(x) \vee K_2(x) \le K_N$. This implies that $\mathbb{E}|b^N(x,\xi)| \vee \mathbb{E}|\psi^N(x,\xi)| \le K_N(1+\mathbb{E}|\xi|^2) < \infty$ if Assumption (A1) holds. According to the existing results, we cannot obtain the existence of the solution of (5.2). However, if both $b^N(\cdot,\xi)$ and $\psi^N(\cdot,\xi)$ are continuous functionals for any $\xi \in \mathbb{R}^m$, this condition can guarantee the existence of the weak solution for (5.2); see [19].

Theorem 6.2. *Assume the conditions Theorem 5.1 with the modification of replacing* (A2) *by* (A2'). *The results in Theorem 5.1 continue to hold.*

To prove this theorem, let us present the following lemma.

Lemma 6.3. *Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a continuous function such that*

$$|F(y)| \le K(1+|y|^2).$$

Then under conditions in Theorem 3.2, for any $t < T$,

$$\int_t^T [\mathbb{E}_t^{\xi^\varepsilon} F(\xi^\varepsilon(u)) - \mu F]du = O(\varepsilon).$$

Proof. For any $\varsigma > 0$, following (A.6) in the appendix,

$$\mathbb{E}|\xi(t)|^2 \le e^{-\varsigma t}\mathbb{E}|\xi(0)|^2 + \varsigma^{-1}\Big[\kappa^{-1}|h(0)|^2 + \frac{\kappa_2^{-1}-1}{1-\kappa_2}|\phi(0)|^2\Big](1-e^{-\varsigma t}). \tag{6.3}$$

This implies that for any deterministic initial value $\xi(0) = y_0$,

$$\lim_{t\to\infty} \mathbb{E}|\xi(t)|^2 \le \frac{1}{\varsigma}\Big[\kappa^{-1}|h(0)|^2 + \frac{\kappa_2^{-1}-1}{1-\kappa_2}|\phi(0)|^2\Big] =: B$$

and

$$\mathbb{E}|\xi(t)|^2 \le |y_0|^2 + B.$$

Moreover, noting that the initial value is deterministic and $\xi(t)$ is a Markov process, we have

$$\begin{aligned}\mathbb{E}_\mu|\xi|^2 &= \int_{\mathbb{R}^m} |\xi|^2 \mu(d\xi) \\ &= \lim_{t\to\infty} \int_{\mathbb{R}^m} |\xi|^2 p(z,0;d\xi,t)\end{aligned}$$

$$
\begin{aligned}
&= \lim_{t\to\infty} \mathbb{E}(|\xi(t)|^2|\mathcal{F}_0) \\
&= \lim_{t\to\infty} \mathbb{E}|\xi(t)|^2 \le B,
\end{aligned}
$$

where $p(z, 0; d\xi, t)$ denotes the transition probability as in (5.4).

According to the definition of the distance in variation (see the definition in Lemma A.1 or [16]), applying (A.17) gives

$$
\begin{aligned}
\mathrm{Var}(\mathcal{P}_t - \mu) &:= \sup_{|g|\le 1} (\mathcal{P}_t - \mu) g \\
&= \sup_{|g|\le 1} \int_{\mathbb{R}^m} g(w)(p(z,0; dw, t) - \mu(dw)) \\
&\le K e^{-\frac{\tilde{\xi}}{2} t}. \qquad (6.4)
\end{aligned}
$$

Noting $|F(y)| \le K(1 + |y|^2)$ and making a change of variable similar to (5.20), the Cauchy-Schwarz inequality, the Markov property, and the stationarity give that

$$
\begin{aligned}
&\int_t^T [\mathbb{E}_t^{\xi^\varepsilon} F(\xi^\varepsilon(u)) - \mu F] du \\
&= \varepsilon \int_{t/\varepsilon}^{T/\varepsilon} [\mathbb{E}_{t/\varepsilon}^{\xi} F(\xi(u)) - \mu F] du \\
&= \varepsilon \int_{t/\varepsilon}^{T/\varepsilon} [\mathcal{P}_{t/\varepsilon, u} F(\xi(u)) - \mu F] du \\
&= \varepsilon \int_{t/\varepsilon}^{T/\varepsilon} [\mathcal{P}_{u - t/\varepsilon} F(\xi(u)) - \mu F] du \\
&\le \varepsilon K \int_{t/\varepsilon}^{T/\varepsilon} \Big[\int_{\mathbb{R}^m} (1 + |w|^2)(p(z, 0; dw, u - t/\varepsilon) + \mu(dw))\Big]^{1/2} \\
&\quad \times \Big[\int_{\mathbb{R}^m} |p(z, 0; dw, u - t/\varepsilon) - \mu(dw)|\Big]^{1/2} du \\
&\le \varepsilon K \int_{t/\varepsilon}^{T/\varepsilon} \Big[\int_{\mathbb{R}^m} (1 + |w|^2)(p(z, 0; dw, u - t/\varepsilon) + \mu(dw))\Big]^{1/2} [\mathrm{var}(\mathcal{P}_{u-t/\varepsilon} - \mu)]^{1/2} du \\
&\le \varepsilon K \int_{t/\varepsilon}^{T/\varepsilon} (1 + \mathbb{E}|\xi(u - t/\varepsilon)|^2 + \mathbb{E}_\mu |\xi|^2)[\mathrm{var}(\mathcal{P}_{u-t/\varepsilon} - \mu)]^{1/2} du
\end{aligned}
$$

$$
\begin{aligned}
&\leq \varepsilon(1+|y_0|^2+2B)\int_{t/\varepsilon}^{T/\varepsilon}[\mathrm{var}(\mathcal{P}_{u-t/\varepsilon}-\mu)]^{1/2}du \\
&\leq \varepsilon K\int_{t/\varepsilon}^{T/\varepsilon} e^{-\frac{\tilde{\varsigma}}{4}(u-t/\varepsilon)}du \\
&\leq \varepsilon\frac{4K}{\tilde{\varsigma}}(1-e^{-\frac{\tilde{\varsigma}(T-t)}{4\varepsilon}})=O(\varepsilon).
\end{aligned}
\tag{6.5}
$$

The desired assertion thus follows. □

Proof of Theorem 6.2. We first prove that when (A2) is replaced by (A2'), $\{x^{\varepsilon,N}(\cdot)\}$ is tight in $D([0,T];\mathbb{R}^n)$. In reference to the proof of Theorem 5.2, we need to use $x^{\varepsilon,N}(t)$ and estimate (5.12) and (5.13) under Assumption (A2'). Note that $K_1(\cdot)$ and $K_2(\cdot)$ are boundedness preserving. This implies that there exists a constant K_N such that $|K_1(x)|\vee|K_2(x)|\leq K_N$ for $x\in D([0,T];S_N)$. To prove the probability boundedness of $x^{\varepsilon,N}(t)$, based on (5.10),

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t\leq T}|x^{\varepsilon,N}(t)|^2\Big] \\
&\leq 3\mathbb{E}|x(0)|^2+3T\int_0^T\mathbb{E}|b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))|^2du+12\mathbb{E}\int_0^T|\psi^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))|^2du \\
&\leq 3\mathbb{E}|x(0)|^2+3T\int_0^T\mathbb{E}K_1(x_u^{\varepsilon,N})(1+|\xi^\varepsilon(u)|^2)du+\int_0^T\mathbb{E}K_2(x_u^{\varepsilon,N})(1+|\xi^\varepsilon(u)|^2)du \\
&\leq 3\mathbb{E}|x(0)|^2+K_N\int_0^T(1+\mathbb{E}|\xi^\varepsilon(u)|^2)du+K_N\int_0^T(1+\mathbb{E}|\xi^\varepsilon(u)|^2)du \\
&\leq K_N,
\end{aligned}
\tag{6.6}
$$

which gives the desired boundedness in (5.8) by the Chebyshev inequality. Applying the Hölder inequality yields that there exists a random variable $\widetilde{K}_1^{\varepsilon,N}(\delta)$ such that

$$
\begin{aligned}
\mathbb{E}_t^\varepsilon\Big|\int_t^{t+\delta}b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))du\Big|^2 &\leq \delta\mathbb{E}_t^\varepsilon\int_t^{t+\delta}|b^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))|^2du \\
&\leq \delta\mathbb{E}_t^\varepsilon\int_t^{t+\delta}K_1(x_u^{\varepsilon,N})(1+|\xi^\varepsilon(u)|^2)du \\
&\leq \mathbb{E}_t^\varepsilon\widetilde{K}_1^{\varepsilon,N}(\delta)
\end{aligned}
$$

such that

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\mathbb{E}\widetilde{K}_1^{\varepsilon,N}(\delta)=\lim_{\delta\to 0}O(\delta^2)=0. \tag{6.7}$$

Under Assumption (A2'), applying the martingale isometry to the second term in (5.11), there is a random variable $\widetilde{K}_2^{\varepsilon,N}(\delta)$ such that

$$\begin{aligned}&\mathbb{E}_t^\varepsilon\Big|\int_t^{t+\delta}\sigma^N(x_u^{\varepsilon,N},\xi^\varepsilon(u))dw(u)\Big|^2\\ &=\mathbb{E}_t^\varepsilon\int_t^{t+\delta}K_2(x_u^{\varepsilon,N})(1+|\xi^\varepsilon(u)|^2)du\le\mathbb{E}_t^\varepsilon\widetilde{K}_2^{\varepsilon,N}(\delta)\end{aligned} \tag{6.8}$$

satisfying

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\mathbb{E}\widetilde{K}_2^{\varepsilon,N}(\delta)=\lim_{\delta\to 0}O(\delta)=0.$$

Taking $\widetilde{K}^{\varepsilon,N}(\delta)=\widetilde{K}_1^{\varepsilon,N}(\delta)+\widetilde{K}_2^{\varepsilon,N}(\delta)$, and using (6.7) and (6.8) yield that

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\widetilde{K}^{\varepsilon,N}(\delta)=\lim_{\delta\to 0}O(\delta)=0.$$

This, together with (6.6), gives that $x^{\varepsilon,N}(\cdot)$ is tight in $D([0,T];\mathbb{R}^n)$ by Lemma 5.3.

In the proof of Theorem 5.1, we need to re-estimate (5.23), (5.24), (5.28), and (5.31) under Assumption (A2'). Let us give the estimate (5.23). Since there exists the boundedness preserving functional $K_1(x)$ such that $|b(x,\xi)|\le K_1(x)(1+|\xi|^2)$, for any $x\in D([0,T];S_N)$, there must be constant K such that $|b^N(x,\xi)|\le K(1+|\xi|^2)$. Under assumption (A1), applying Lemma 6.3 directly gives

$$\sup_{t\le T}|f_1^{\varepsilon,N}(t)|=\sup_{t\le T}\Big|\int_t^T f_y(x^{\varepsilon,N}(t))\mathbb{E}_t^{\xi^\varepsilon}[b^N(x_t^{\varepsilon,N},\xi^\varepsilon(u))-\bar{b}^N(x_t^{\varepsilon,N})]du\Big|=O(\varepsilon). \tag{6.9}$$

Similarly, (5.24), (5.28), and (5.31) can be obtained under assumption (A2'). We have therefore proved the desired assertion. □

Let us use the above method developed to examine the following 2-dimensional two-time-scale functional diffusion system with the slow-varying component involving affine noise:

$$\begin{cases} d\xi^\varepsilon(t)=\dfrac{1}{\varepsilon}h(\xi^\varepsilon(t))dt+\dfrac{1}{\sqrt{\varepsilon}}\phi(\xi^\varepsilon(t))dw_1(t),\\[2ex] dx^\varepsilon(t)=b(x_t^\varepsilon)\xi^\varepsilon(t)dt+\psi(x_t^\varepsilon)\xi^\varepsilon(t)dw_2(t),\end{cases} \tag{6.10}$$

where $h(\cdot)$ and $\phi(\cdot)$ are appropriate scalar measurable functions such that assumption (A1) is satisfied, $b(\cdot)$ and $\psi(\cdot)$ $\nabla_x b_i(\cdot)$, $\nabla_x\psi_{ij}(\cdot)$, $\nabla_x^2 b_i(\cdot)$ and $\nabla_x^2\psi_{ij}(\cdot)$, $i,j=1,2,\dots,n$ are boundedness preserving functionals. Let μ be the invariant measure of the stochastic differential equation

$$d\xi(t) = h(\xi(t))dt + \phi(\xi(t))d\tilde{w}(t)$$

and

$$\bar{\xi} = \mathbb{E}_\mu \xi = \int \xi \mu(d\xi) \text{ and } \bar{\xi}_2 = \mathbb{E}_\mu \xi^2 = \int \xi^2 \mu(d\xi).$$

Applying Theorem 6.2 gives the following result.

Corollary 6.4. *As $\varepsilon \to 0$, $x^\varepsilon(\cdot)$ the solution of (6.10) converges weakly to $x(\cdot)$ that is a solution of the stochastic functional differential equation*

$$dx(t) = \bar{\xi} b(x_t)dt + \sqrt{\bar{\xi}_2}|\psi(x_t)|d\widehat{w}(t),$$

where $\widehat{w}(t)$ is a scalar standard Brownian motion.

Let us consider the following two-dimensional stochastic integro-differential equation as an example:

$$\begin{cases} d\xi^\varepsilon(t) = -\dfrac{1}{\varepsilon}\alpha\xi^\varepsilon(t)dt + \dfrac{1}{\sqrt{\varepsilon}}\rho dw_1(t), \\ dx^\varepsilon(t) = \beta\Big(\displaystyle\int_0^t x^\varepsilon(s)ds\Big)\xi^\varepsilon(t)dt + \gamma\Big(\displaystyle\int_0^t x^\varepsilon(s)ds\Big)\xi^\varepsilon(t)dw_2(t) \end{cases} \tag{6.11}$$

with initial value $\xi(0)$ and $x(0)$ and $\alpha > 0, \rho \neq 0$, where $\beta(\cdot)$ and $\gamma(\cdot)$ are two continuous functions, $w_1(t)$ and $w_2(t)$ are two independent scalar standard Brownian motions.

It is well-known that

$$d\xi(t) = -\alpha\xi(t)dt + \rho d\tilde{w}(t) \tag{6.12}$$

defines a mean-reverting Ornstein-Uhlenbeck process with the normal distribution

$$\xi(t) \sim N\Big(\xi(0)e^{-\alpha t}, \frac{\rho^2(1-e^{-2\alpha t})}{2\alpha}\Big),$$

which implies that the invariant measure μ of (6.12) is exponentially ergodic with the normal distribution $N(0, \rho^2/(2\alpha))$. It is easily observed that

$$\mathbb{E}_\mu \xi = 0 \text{ and } \mathbb{E}_\mu \xi^2 = \int \xi^2 \mu(d\xi) = \frac{\rho^2}{2\alpha}.$$

According to Corollary 6.4, as $\varepsilon \to 0$, $x^\varepsilon(t)$ in (6.11) converges weakly to $x(t)$ satisfying the following stochastic integro-differential equation

$$dx(t) = \frac{\rho}{\sqrt{2\alpha}}\gamma\Big(\int_0^t x(s)ds\Big)d\widehat{w}(t),$$

where $\widehat{w}(t)$ is a standard Brownian motion.

Example 6.5. Now let us return to Example 1.2 in the introduction section. Although it is known that (3.3) has a unique global solution and this solution is pth moment bounded for any $p > 0$ and exponentially ergodic with the stationary distribution being the noncentral chi-square distribution, we only know that $k_r + \gamma_r \xi(t) \ge 0$. Moreover, if $\gamma_r \le 4k_r$, $k_r + \gamma_r \xi(t) > 0$. If $\gamma_r > 4k_r$, $k_r + \gamma_r \xi(t)$ can reach zero (see [11,34]). This implies that it is possible for the solution of the chemical Langevin equation to be negative. In fact, for the chemical Langevin equation (1.7), it is more difficult to prove $k_p \xi^\varepsilon(t) + \gamma_p x^\varepsilon(t) + \pi(x_t^\varepsilon) > 0$ since this term involves the delay and $\xi^\varepsilon(t)$ may be negative. Let us define

$$b^\varepsilon(x_t^\varepsilon, \xi^\varepsilon(t)) = k_p \xi^\varepsilon(t) - \gamma_p x^\varepsilon(t) - \pi(x_t^\varepsilon) \text{ and } \psi^\varepsilon(x_t^\varepsilon, \xi^\varepsilon(t)) = \sqrt{k_p \xi^\varepsilon(t) + \gamma_p x^\varepsilon(t) + \pi(x_t^\varepsilon)}.$$

Although it is possible to replace $b^\varepsilon(\cdot)$ and $\psi^\varepsilon(\cdot)$ by some positive functions or modify the processes by adding reflections, neither of these is common in the actual practice of chemical reactions. The most popular way of handling this is still to use the original model and simply assume the positiveness mentioned above.

To obtain the desired asymptotic results, we assume that (a) $\pi(x)$ is boundedness preserving and (b) $k_p \xi^\varepsilon(t) + \gamma_p x^\varepsilon(t) + \pi(x_t^\varepsilon) > 0$ for the solution $\xi^\varepsilon(t)$ and $x^\varepsilon(t)$ of (1.7). It follows that assumption (A2') holds. According to the mean-reverting property,

$$\mathbb{E}_\mu \xi = \lim_{t\to\infty} \mathbb{E}\xi(t) = \frac{k_r}{\gamma_r}.$$

Theorem 6.2 yields that there exists a standard Brownian motion $\widetilde{W}(t)$ such that $x^\varepsilon(t)$ in (1.7) converges weakly to $x(t)$ satisfying the following stochastic functional differential equation

$$dx(t) = \Big[\frac{k_p k_r}{\gamma_r} - \gamma_p x(t) - \pi(x_t)\Big]dt + \sqrt{\frac{k_p k_r}{\gamma_r} + \gamma_p x(t) + \pi(x_t)}d\widetilde{W}(t).$$

7. Concluding remarks

This paper develops averaging principles for functional diffusion systems with two-time scales, in which the functional term involves complete history, i.e., $x_t = \{x(u \wedge t) : 0 \le u \le T\}$. In this paper, the fast varying process is assumed to be a diffusion process. Using the ergodicity results in [3, Chapter 1], we can allow the fast process to satisfy a stochastic functional differential equation as long as the spectrum gap is small enough or the convergence to the invariant measure is fast enough (exponential ergodic), the approach presented in the paper still works.

We point out, however, in the two-time scale system discussed in this paper, the fast-varying process is independent of the slow-varying variable. When the slow-varying variable and the history appear in the fast-varying process, the system becomes more difficult to deal with. We will have to consider the transition probability involving the slow-varying variable as a parameter. This will be considered in our future work.

Appendix A. Proof of Theorem 3.2

We begin by recalling a lemma that was proved in [12, Lemma 7.1.5, P125].

Lemma A.1. *Let $\mathcal{P}_t, t \geq 0$ be a Markov semigroup on $B_b(\mathbb{R}^m; \mathbb{R})$ and let $K > 0$ and $t > 0$ be fixed. Then the following results are equivalent:*

(i) *for all $F \in C_b^2(\mathbb{R}^m; \mathbb{R})$ for all $x, y \in \mathbb{R}^m$, $|\mathcal{P}_t F(x) - \mathcal{P}_t F(y)| \leq K\|F\|_0|x - y|$, where $C_b^2(\mathbb{R}^m; \mathbb{R})$ is the set of bounded and continuous functions whose first and second partial derivatives are bounded;*

(ii) *for all $F \in B_b(\mathbb{R}^m; \mathbb{R})$ for all $x, y \in \mathbb{R}^m$, $|\mathcal{P}_t F(x) - \mathcal{P}_t F(y)| \leq K\|F\|_0|x - y|$, where $\|F\|_0 = \sup_{x\in\mathbb{R}^m} |F(x)|$;*

(iii) *for any $x, y \in \mathbb{R}^m$, the distance in variation $\mathrm{Var}(P_t(x, \cdot) - P_t(y, \cdot)) \leq K\|F\|_0|x - y|$ between $P_t(x, \cdot)$ and $P_t(y, \cdot)$, where the distance in variation between $P_t(x, \cdot)$ and $P_t(y, \cdot)$ is defined by*

$$\mathrm{Var}(P_t(x, \cdot) - P_t(y, \cdot)) = \sup_{|F(\cdot)|\leq 1} \int_{\mathbb{R}^m} d(P_t(x, dz) - P_t(y, dz))F(z).$$

Proof of Theorem 3.2. It is easily observed by conditions in this theorem,

$$2y'h(y) \leq -2\lambda_1|y|^2 + 2y'h(0) \leq -2\lambda_1|y|^2 + |y|^2 + |h(0)|^2, \tag{A.1}$$

$$|\phi(y)|^2 \leq 2|\phi(0)|^2 + 2\lambda_2|y|^2, \tag{A.2}$$

which implies that

$$2y'h(y) + |\phi(y)|^2 \leq (-2\lambda_1 + 2\lambda_2 + 1)|y|^2 + |h(0)|^2 + 2|\phi(0)|^2.$$

This reveals that the coefficients satisfy the monotone condition. This condition, together with the (local) Lipschitz conditions on $h(\cdot)$ and $\phi(\cdot)$, yields the existence and uniqueness of the global strong solution of system (3.2) with $\mathbb{E}(\sup_{t\in[0,T]} |\xi(t)|^2) \leq C_T$ and this solution is $\mathcal{F}_t^{\tilde{w}}$-adapted, where $\mathcal{F}_t^{\tilde{w}} = \sigma\{\tilde{w}(s); s \leq t\}$ (see [34, p.58 Theorem 3.5]). Note that $\tilde{w}(t) = w_1(\varepsilon t)/\sqrt{\varepsilon}$ implies that $\mathcal{F}_t^{\tilde{w}} = \mathcal{F}_{\varepsilon t}^{w_1}$. Hence the solution $x(t)$ is $\mathcal{F}_{\varepsilon t}^{w_1}$-adapted. Reference [34] also implies that this solution is homogeneous and has the strong Markov property.

Now let us examine the invariant measure. For any $\varsigma > 0$, applying the Itô formula to $e^{\varsigma t}|\xi(t)|^2$ yields that for any $t > s \geq 0$,

$$e^{\varsigma t}\mathbb{E}|\xi(t)|^2 = e^{\varsigma s}\mathbb{E}|\xi(s)|^2 + \mathbb{E}\int_s^t e^{\varsigma u}[\varsigma|\xi(u)|^2 + 2\xi'(u)h(\xi(u)) + |\phi(\xi(u))|^2]du. \tag{A.3}$$

According to the theorem's conditions, for any $\kappa_1 > 0$ we can reestablish the following inequality for (A.1):

$$2y'h(y) \leq -2\lambda_1|y|^2 + 2y'h(0) \leq -2\lambda_1|y|^2 + \kappa_1|y|^2 + \kappa_1^{-1}|h(0)|^2. \tag{A.4}$$

Similarly, for any $\kappa_2 \in (0, 1)$, we also reestablish (A.2) as follows:

$$|\phi(y)|^2 \leq \lambda_2|y|^2 + 2\phi'(y)\phi(0) - |\phi(0)|^2 \leq \lambda_2|y|^2 + \kappa_2|\phi(y)|^2 + \kappa_2^{-1}|\phi(0)|^2 - |\phi(0)|^2,$$

which implies that

$$|\phi(y)|^2 \le \frac{\lambda_2}{1-\kappa_2}|y|^2 + \frac{\kappa_2^{-1}-1}{1-\kappa_2}|\phi(0)|^2. \tag{A.5}$$

Substituting (A.4) and (A.5) into (A.3) yields

$$\begin{aligned}\mathbb{E}|\xi(t)|^2 &\le e^{-\varsigma(t-s)}\mathbb{E}|\xi(s)|^2 + e^{-\varsigma t}\mathbb{E}\int_s^t e^{\varsigma u}[\varsigma|\xi(u)|^2 + 2\xi'(u)h(\xi(u)) + |\phi(\xi(u))|^2]du \\ &\le e^{-\varsigma(t-s)}\mathbb{E}|\xi(s)|^2 - e^{-\varsigma t}\mathbb{E}\int_s^t e^{\varsigma u}\Big[2\lambda_1 - \frac{\lambda_2}{1-\kappa_2} - (\kappa_1+\varsigma)\Big]|\xi(u)|^2 du \\ &\quad + \Big[\kappa^{-1}|h(0)|^2 + \frac{\kappa_2^{-1}-1}{1-\kappa_2}|\phi(0)|^2\Big]e^{-\varsigma t}\int_u^t e^{\varsigma u}du.\end{aligned}$$

Note that $2\lambda_1 > \lambda_2$. Choose κ_1, κ_2 and ς sufficiently small such that

$$2\lambda_1 - \frac{\lambda_2}{1-\kappa_2} - (\kappa_1+\varsigma) > 0.$$

We therefore have

$$\mathbb{E}|\xi(t)|^2 \le e^{-\varsigma(t-s)}\mathbb{E}|\xi(s)|^2 + \varsigma^{-1}\Big[\kappa^{-1}|h(0)|^2 + \frac{\kappa_2^{-1}-1}{1-\kappa_2}|\phi(0)|^2\Big](1-e^{-\varsigma(t-s)}). \tag{A.6}$$

For the initial time s with different initial values $\xi_1(s) = y_1$ and $\xi_2(s) = y_2$, define $e(t) = \xi(t; y_1) - \xi(t; y_2)$. Then according to Eq. (3.1), we can write the equation about $e(t)$ as follows:

$$de(t) = \tilde{h}(t)dt + \tilde{\phi}(t)d\tilde{w}(t), \tag{A.7}$$

where $\tilde{h}(t) := h(\xi(t; y_1)) - h(\xi(t; y_2))$ and $\tilde{\phi}(t) := \phi(\xi(t; y_1)) - \phi(\xi(t; y_2))$. For any $\tilde{\varsigma} > 0$, applying the Itô formula to $e^{\tilde{\varsigma}t}|e(t)|^2$ gives

$$e^{\tilde{\varsigma}t}\mathbb{E}|e(t)|^2 = e^{\tilde{\varsigma}s}\mathbb{E}|e(s)|^2 + \mathbb{E}\int_s^t e^{\tilde{\varsigma}u}[\tilde{\varsigma}|e(u)|^2 + 2e'(u)\tilde{h}(u) + |\tilde{\phi}(u)|^2]du.$$

It is easily observed that

$$e'(t)\tilde{h}(t) \le -\lambda_1|e(t)|^2, \tag{A.8}$$

$$|\tilde{\phi}(t)|^2 \le \lambda_2|e(t)|^2, \tag{A.9}$$

which implies that

$$\mathbb{E}|e(t)|^2 \le e^{-\tilde{\varsigma}(t-s)}\mathbb{E}|e(s)|^2 - e^{-\tilde{\varsigma}t}\mathbb{E}\int_s^t e^{\tilde{\varsigma}u}(2\lambda_1 - \lambda_2 + \tilde{\varsigma})|e(u)|^2 du.$$

Note $2\lambda_1 > \lambda_2$ again. Let us choose $\tilde{\varsigma}$ sufficiently small such that $2\lambda_1 - \lambda_2 + \tilde{\varsigma} > 0$, which shows that

$$\mathbb{E}|e(t)|^2 \le e^{-\tilde{\varsigma}(t-s)}\mathbb{E}|e(s)|^2. \tag{A.10}$$

Since the solution $\xi(t)$ is a strong homogenous Markov process, we can adopt the remote start method to examine invariant measure. The idea can be described as follows. For the solution $\xi(t)$ with the initial value $\xi(s) = y$, in lieu of letting $t \to \infty$, we can keep t as fixed and letting the initial time $s \to -\infty$. To use this method, let us consider the following stochastic differential equation

$$d\xi(t) = h(x, \xi(t))dt + \phi(x, \xi(t))d\overleftrightarrow{W}(t), \tag{A.11}$$

with the initial data $y(s) = y \in \mathbb{R}^m$, where $\overleftrightarrow{W}(t)$ is a double-sided Wiener process defined by

$$\overleftrightarrow{W}(t) := \begin{cases} \tilde{w}(t), & t \ge 0, \\ \overline{w}(-t), & t < 0, \end{cases}$$

where $\overline{w}(t)$ is another Brownian motion independent of $\tilde{w}(t)$. Define the filtration

$$\bar{\mathcal{F}}_t := \bigcap_{r>t} \bar{\mathcal{F}}_r^0$$

and $\bar{\mathcal{F}}_r^0 := \sigma(\{\bar{w}(r_2) - \bar{w}(r_1) : -\infty < r_1 \le r_2 \le r\}, \mathcal{N})$, where $\mathcal{N}$ is the set of all $\mathbb{P}$-null sets. Eq. (A.11) is the same as Eq. (3.2) if the initial time may be any $s \in \mathbb{R}$. Hence the estimates of (A.6) and (A.10) still hold. By the Markov property of $\xi(t)$, for any $-\infty < s_2 < s_1 < t < \infty$, with t fixed, applying (A.10) gives

$$\begin{aligned} \mathbb{E}|\xi(t;\xi(s_2)) - \xi(t;\xi(s_1))|^2 &= \mathbb{E}|\xi(t, \xi(s_1;\xi(s_2))) - \xi(t;\xi(s_1))|^2 \\ &\le \mathbb{E}|\xi(s_1;\xi(s_2)) - \xi(s_1)|^2 e^{-\tilde{\varsigma}(t-s_1)} \\ &\le 2(\mathbb{E}|\xi(s_1;\xi(s_2))|^2 + \mathbb{E}|\xi(s_1)|^2)e^{-\tilde{\varsigma}(t-s_1)}. \end{aligned} \tag{A.12}$$

In (A.6), for any initial value $\xi(s) = y$ and $t \ge s$,

$$\mathbb{E}|\xi(t)|^2 \le \mathbb{E}|y|^2 + \varsigma^{-1}\Big[\kappa^{-1}|h(x,0)|^2 + \frac{\kappa_2^{-1} - 1}{1 - \kappa_2}|\phi(0)|^2\Big],$$

where implies that $\mathbb{E}|\xi(s_1;\xi(s_2))|^2$ and $\mathbb{E}|\xi(s_1)|^2$ are bounded for any bounded initial value y. This establishes that there exists constant K such that

$$\mathbb{E}|\xi(t;\xi(s_2)) - \xi(t;\xi(s_1))|^2 \le K e^{-\tilde{\varsigma}(t-s_1)} \to 0 \tag{A.13}$$

as $s_1 \to -\infty$. This implies that every sequence $\{\xi(-n)\}$ is a Cauchy sequence; all the sequences have the same limit ξ^* as $n \to \infty$. Let us denote its distribution by $\mu(\cdot)$. In addition, (A.13) reveals that ξ^* is independent of the initial value y. Noting that $\xi(t)$ is a homogeneous Markov process, for any $A \in \mathbb{R}^m$, we have

$$p(y, s; A, t) = p(y, 0; A, t - s).$$

The time homogeneity also implies that

$$p(y, -s; A, 0) \to \mu(A) \quad \text{as } s \to \infty.$$

Note that $\mathcal{P}_t F(y) = \mathbb{E} F(\xi(t; y))$ for any deterministic initial value y. For any $F \in B_b(\mathbb{R}^m; \mathbb{R})$, $\mathcal{P}_t F$ is also bounded. By the time homogeneity of $\xi(t)$, it follows from the Chapman-Kolmogorov equation of the transition probability that

$$\begin{aligned}
\mu \mathcal{P}_t F &= \int_{\mathbb{R}^m} \mathcal{P}_t F(z) \mu(dz) \\
&= \lim_{s\to\infty} \int_{\mathbb{R}^m} \mathcal{P}_t F(z) p(y, -s; dz, 0) \\
&= \lim_{s\to\infty} \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} F(w) p(z, 0; dw, t) p(y, -s; dz, 0) \\
&= \lim_{s\to\infty} \int_{\mathbb{R}^m} F(w) \int_{\mathbb{R}^m} p(z, 0; dw, t) p(y, -s; dz, 0) \\
&= \lim_{s\to\infty} \int_{\mathbb{R}^m} F(w) p(y, -s; dw, t) \\
&= \lim_{s\to\infty} \int_{\mathbb{R}^m} F(w) p(y, -(t+s); dw, 0) \\
&= \int_{\mathbb{R}^m} F(w) \mu(dw) \\
&= \mu F,
\end{aligned} \tag{A.14}$$

which implies that $\mu(\cdot)$ is an invariant measure of the solution $\xi(t)$. Choosing different initial values y_1 and y_2 at time zero and applying the second inequality in (A.12) give that for some constant K,

$$\mathbb{E}|\xi(t; y_1) - \xi(t; y_2)|^2 \le K|y_1 - y_2|^2 e^{-\tilde{\varsigma} t}. \tag{A.15}$$

Note that $F \in C_b^2(\mathbb{R}^m; \mathbb{R})$ implies that F satisfies the Lipschitz condition. This, together with (A.15) and the Lyapunov inequality, gives that for any $F \in C_b^2(\mathbb{R}^m; \mathbb{R})$,

$$
\begin{aligned}
|\mathcal{P}_t F(y_1) - \mathcal{P}_t F(y_2)| &= |\mathbb{E}F(\xi(t; y_1)) - \mathbb{E}F(\xi(t; y_2))| \\
&= |\mathbb{E}(F(\xi(t; y_1)) - F(\xi(t; y_2)))| \\
&\le K_1 \mathbb{E}|\xi(t; y_1) - \xi(t; y_2)| \\
&\le K_1 (\mathbb{E}|\xi(t; y_1) - \xi(t; y_2)|^2)^{\frac{1}{2}} \\
&\le K_1 \sqrt{K} |y_1 - y_2| e^{-\frac{\tilde{\xi}}{2} t},
\end{aligned}
$$

where $K_1 > 0$ is the Lipschitz constant. By Lemma A.1, for $F \in B_b(\mathbb{R}^m; \mathbb{R})$, there exists a constant $K > 0$ such that

$$
|\mathcal{P}_t F(y_1) - \mathcal{P}_t F(y_2)| \le K |y_1 - y_2| e^{-\frac{\tilde{\xi}}{2} t}. \tag{A.16}
$$

This, together with the definition of the invariant measure, yields

$$
|\mathcal{P}_t F(y) - \mu F| \le \int_{\mathbb{R}^m} |\mathcal{P}_t F(y) - \mathcal{P}_t F(z)| \mu(dz) \le K e^{-\frac{\tilde{\xi}}{2} t} \int_{\mathbb{R}^m} |y - z| \mu(dz). \tag{A.17}
$$

Then by [12, P40], $\mu(\cdot)$ is strongly exponential ergodic. If both $\mu^1(\cdot)$ and $\mu^2(\cdot)$ are invariant measures, it follows from (A.16) that

$$
\begin{aligned}
|\mu^1 F - \mu^2 F| &\le \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} |\mathcal{P}_t F(y) - \mathcal{P}_t F(z)| \mu^1(dy) \mu^2(dz) \\
&\le K e^{-\frac{\tilde{\xi}}{2} t} \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} |y - z| \mu^1(dy) \mu^2(dz) \to 0
\end{aligned} \tag{A.18}
$$

as $t \to \infty$, which implies the uniqueness of the invariant measure. □

References

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 4th ed., Garland, New York, 2002.

[2] V. Bally, L. Caramellino, R. Cont, Stochastic Integration by Parts and Functional Itô Calculus, Advanced Courses in Mathematics CRM Barcelona, Birkhäuser, Basel, 2016.

[3] J. Bao, G. Yin, C. Yuan, Asymptotic Analysis for Functional Stochastic Differential Equations, Springer Briefs in Mathematics, Springer, 2016.

[4] J. Bao, G. Yin, C. Yuan, Two-time-scale stochastic partial differential equations driven by α-stable noise: averaging principles, Bernoulli 23 (2017) 645–669.

[5] M. Bodnar, General model of a cascade of reactions with times: global stability analysis, J. Differ. Equ. 259 (2015) 777–795.

[6] D. Bratsun, D. Volfson, L.S. Tsimring, J. Hasty, Delay-induced stochastic oscillations in gene regulation, Proc. Natl. Acad. Sci. USA 102 (41) (2005) 14593–14598.

[7] T. Brett, T. Galla, Stochastic processes with distributed delays: chemical Langevin equation and linear-noise approximation, Phys. Rev. Lett. 110 (25) (2013) 250601.

[8] T. Brett, T. Galla, Gaussian approximations for stochastic systems with delay: chemical Langevin equation and application to a Brusselator system, J. Chem. Phys. 140 (12) (2014) 124112.

[9] C. Clayton, M. Shapira, Post-transcriptional regulation of gene expression in trypanosomes and leishmanias, Mol. Biochem. Parasitol. 156 (2) (2007) 93–101.

[10] R. Cont, D.A. Fournié, Functional Itô calculus and stochastic integral representation of martingales, Ann. Probab. 41 (1) (2013) 109–133.
[11] J.C. Cox, J.E. Ingersoll Jr., S.A. Ross, A theory of the term structure of interest rates, Econometrica 53 (1985) 385–408.
[12] G. Da Prato, J. Zabczyk, Ergodicity for Infinite-Dimensional Systems, Cambridge University Press, Cambridge, 1996.
[13] B. Dupire, Functional Itô Calculus, Portfolio Research paper 2009-04, 2009, Bloomberg.
[14] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (4) (1976) 403–434.
[15] K. Gopalsamy, Stability and Oscillations in Delay Difference Differential Equations of Population Dynamics, Kluwer Academic, Dordrecht, 1992.
[16] A. Guillin, C. Léonard, L. Wu, N. Yao, Transportation-information inequalities for Markov processes, Probab. Theory Relat. Fields 144 (2009) 669–695.
[17] C. Gupta, J.M. López, R. Azencott, M.R. Bennett, K. Josić, W. Ott, Modeling delay in genetic networks: from delay birth-death processes to delay stochastic differential equations, J. Chem. Phys. 140 (20) (2014) 204108.
[18] J.K. Hale, S.M.V. Lunel, Introduction to Functional Differential Equations, Springer, Berlin, 1993.
[19] K. Itô, M. Nisio, On stationary solution of a stochastic equation, Kyoto J. Math. 4 (1964) 1–75.
[20] R.Z. Khasminskii, On stochastic processes defined by differential equations with a small parameter, Theory Probab. Appl. 11 (1966) 211–228.
[21] R.Z. Khasminskii, G. Yin, Asymptotic series for singularly perturbed Kolmogorov-Fokker-Planck equations, SIAM J. Appl. Math. 56 (1996) 1766–1793.
[22] R.Z. Khasminskii, G. Yin, On transition densities of singularly perturbed diffusions with fast and slow components, SIAM J. Appl. Math. 56 (1996) 1794–1819.
[23] R.Z. Khasminskii, G. Yin, Limit behavior of two-time-scale diffusions revisited, J. Differ. Equ. 212 (2005) 85–113.
[24] R.Z. Khasminskii, Stochastic Stability of Differential Equations, 2nd ed, Springer, Heidelberg, 2012.
[25] V.B. Kolmanovskii, V.R. Nosov, Stability of Functional Differential Equations, Academic Press, New York, 1986.
[26] Y. Kuang, Delay Differential Equations with Applications in Population Dynamics, Academic Press, Boston, 1993.
[27] H.J. Kushner, Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory, MIT Press, Cambridge, MA, 1984.
[28] H.J. Kushner, Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems, Birkhäuser, Boston, MA, 1990.
[29] H.J. Kushner, Numerical Methods for Controlled Stochastic Delay Systems, Birkhäuser Boston, Boston, MA, 2008.
[30] T. Kurtz, Semigroups of conditioned shifts and approximation of Markov processes, Ann. Probab. 3 (4) (1975) 618–642.
[31] T. Kurtz, Approximation of Population Processes, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 36, SIAM, Philadelphia, 1981.
[32] B. Lewin, Genes IX, Jones & Bartlett Publishers, Burlington, MA, 2007.
[33] Y. Liu, J. Loros, J.C. Dunlap, Phosphorylation of the Neurospora clock protein frequency determines its degradation rate and strongly influences the period length of the circadian clock, Proc. Natl. Acad. Sci. 97 (1) (2000) 234–239.
[34] X. Mao, Stochastic Differential Equations and Applications, second edition, Horwood, Chichester, 2007.
[35] I. Melzner, M.A. Weniger, A.J. Bucur, S. Brüderlein, K. Dorsch, C. Hasel, F. Leithäuser, O. Ritz, M.J.S. Dyer, T.F.E. Barth, P. Möller, Biallelic deletion within 16p13. 13 including SOCS-1 in Karpas 1106P mediastinal B-cell lymphoma line is associated with delayed degradation of JAK2 protein, Int. J. Cancer 118 (8) (2006) 1941–1944.
[36] J. Miekisz, J. Poleszczuk, M. Bodnar, Stochastic models of gene expression with delayed degradation, Bull. Math. Biol. 73 (2011) 2231–2247.
[37] S.-E.A. Mohammed, Stochastic Functional Differential Equations, Longman, Harlow/New York, 1986.
[38] E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation I, Ann. Probab. 29 (3) (2001) 1061–1085;
E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation II, Ann. Probab. 31 (3) (2003) 1166–1192;
E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation III, Ann. Probab. 33 (3) (2005) 1111–1133.
[39] K.M. Ramachandran, A singularly perturbed stochastic delay system with small parameter, Stoch. Anal. Appl. 11 (1993) 209–230.
[40] K.M. Ramachandran, Stability of stochastic delay differential equation with a small parameter, Stoch. Anal. Appl. 26 (4) (2008) 710–723.

[41] P. Smolen, D.A. Baxter, J.H. Byrne, Modeling circadian oscillations with interlocking positive and negative feedback loops, J. Neurosci. 21 (17) (2001) 6644–6656.
[42] M. Thattai, A. van Oudenaarden, Intrinsic noise in gene regulatory networks, Proc. Natl. Acad. Sci. USA 98 (2001) 8614–8619.
[43] M. Turcotte, J. Garcia-Ojalvo, G.M. Süel, A genetic timer through noise-induced stabilization of an unstable state, Proc. Natl. Acad. Sci. USA 105 (41) (2008) 15732–15737.
[44] F. Wu, G. Yin, H. Mei, Stochastic functional differential equations with infinite delay: existence and uniqueness of solutions, solution maps, Markov properties and ergodicity, J. Differ. Equ. 262 (2012) 1226–1252.
[45] F. Wu, T. Tian, J.B. Rawlings, G. Yin, Approximate method for stochastic chemical kinetics with two-time scales by chemical Langevin equations, J. Chem. Phys. 144 (17) (2016) 174112.
[46] G. Yin, K.M. Ramachandran, A differential delay equation with wideband noise perturbations, Stoch. Process. Appl. 35 (1990) 231–249.
[47] G. Yin, H.Q. Zhang, Singularly perturbed Markov chains: limit results and applications, Ann. Appl. Probab. 17 (2007) 207–229.
[48] G. Yin, Q. Zhang, Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach, Springer, New York, 2013.