Available online at www.sciencedirect.com

ScienceDirect

Journal of Differential Equations 323 (2022) 1–37

Journal of Differential Equations

www.elsevier.com/locate/jde

Fast-slow-coupled stochastic functional differential equations ☆

Fuke Wu [a], George Yin [b,*]

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

[b] *Department of Mathematics, University of Connecticut, Storrs, CT 06269-1009, USA*

Received 22 November 2021; revised 5 March 2022; accepted 20 March 2022
Available online 29 March 2022

Abstract

This paper focuses on two-time-scale coupled stochastic functional differential equations (SFDEs). The system under consideration has a slow component and a fast component. Both components depend on the segment process (an infinite dimension process) of the slow component. To overcome the difficulty due to the past dependence and the coupling of the segment process, such properties as the Hölder continuity and tightness on a space of continuous functions are investigated first for the segment process. In addition, it is also shown that the solution of a fixed-x equation depends continuously on the parameters. Then using the martingale problem formulation, an average principle is established by a direct averaging.

MSC: 34K50; 60G44; 60H10; 60J60; 93C70

Keywords: SFDE; Two-time scale; Tightness; Martingale method; Averaging principle

☆ The research of F. Wu was supported in part by the National Natural Science Foundation of China (Grant No. 61873320). The research of G. Yin was supported in part by the National Science Foundation under grant DMS-2114649.

* Corresponding author.
E-mail addresses: wufuke@hust.edu.cn (F. Wu), gyin@uconn.edu (G. Yin).

https://doi.org/10.1016/j.jde.2022.03.030
0022-0396/

1. Introduction and motivation

This work focuses on coupled stochastic functional differential equations (SFDEs) with two-time scales. The system under consideration has a slow component and a fast component. Both the slow and fast components depend on the segment process (to be specified) of the slow component. Our effort is devoted to obtaining a limit process.

Uncertainty and time delays are ubiquitous and pervasive, which are often encountered in our life. As a result, stochastic systems with delays have received much attention in systems and control, physics, biomedical sciences, epidemic modeling, communication networks, population dynamics, and related fields [5,9,10,18,21]. Taking random disturbances and delays into consideration, much effort has been devoted to the study of stochastic delay or functional differential equations (SDDEs or SFDEs for short) [21,23,24,35]. Because solutions of SDDEs and SFDEs are non-Markovian due to the dependence of history, methods based on Markovian setup for the solutions are no longer applicable; see for example, [2,24,37] and references therein.

From another perspective, many complex systems involve "fast" and "slow" motions. For example, learning processes in the brain involve two-time scales, from fast neuronal activity (a few milliseconds) to slow synaptic plasticity (minutes/hours) [8]. Combined with uncertainty, these systems are often modeled as SDEs with fast and slow time scales; see [3,7,28,33,34,39,40] and references therein. Assuming $\varepsilon > 0$ to be a small parameter, in [14], Khasminskii and Yin examined the following systems of stochastic differential equations with two-time scales

$$\begin{cases} dX_1^\varepsilon(t) = h_1(X_1^\varepsilon(t), X_2^\varepsilon(t))dt + \varsigma_1(X_1^\varepsilon(t), X_2^\varepsilon(t))dw_1(t), \\ dX_2^\varepsilon(t) = \dfrac{1}{\varepsilon} h_2(X_1^\varepsilon(t), X_2^\varepsilon(t))dt + \dfrac{1}{\sqrt{\varepsilon}} \varsigma_2(X_1^\varepsilon(t), X_2^\varepsilon(t))dw_2(t), \end{cases} \tag{1.1}$$

in which the component $X_2^\varepsilon(\cdot)$ is rapidly varying and $X_1^\varepsilon(\cdot)$ is slowly changing. In [8], Galtier and Wainrib studied a generic learning neural network model

$$\begin{cases} \dot{\Xi}^\varepsilon(t) = G(\Xi^\varepsilon(t), v^\varepsilon(t)), \\ dv^\varepsilon(t) = \dfrac{1}{\varepsilon}[f(\Xi^\varepsilon(t), v^\varepsilon(t)) + u(t)]dt + \dfrac{1}{\sqrt{\varepsilon}} \varsigma(\Xi^\varepsilon(t), v^\varepsilon(t))dw(t), \end{cases} \tag{1.2}$$

where $v^\varepsilon \in \mathbb{R}^n$ represents the fast activity of the individual elements in n neurons, $\Xi^\varepsilon(t) \in \mathbb{R}^{n\times n}$ is the connectivity matrix that varies slowly due to plasticity, and u represents the external input. One of the main features is: The original systems are complex and difficult to deal with, but the associated limit dynamic systems as $\varepsilon \to 0$ are considerably simpler; see [3,7,12,14,13,15, 16,27,28,32,33,39,40] and references therein. Several methods are commonly used to treat the corresponding asymptotic properties. One of them is based on analytic techniques [14,13,15] by means of asymptotic expansions of the associated transition densities through Kolmogorov-Fokker-Planck equations. For example, using the asymptotic expansion methods for (1.1), an averaging principle was established in [14]; it was shown that as $\varepsilon \to 0$, the fast component is averaged out, and the slow component $X_1^\varepsilon(\cdot)$ converges weakly to a limit $X(\cdot)$ satisfying

$$dX(t) = \bar{h}_1(X(t))dt + \bar{\varsigma}_1(X(t))dw(t), \tag{1.3}$$

where

$$\bar{h}_1(x_1) = \int h_1(x_1, x_2)\mu^{x_1}(dx_2), \quad \bar{\varsigma}_1^2(x_1) = \int \varsigma_1^2(x_1, x_2)\mu^{x_1}(dx_2)$$

is an average with respect to $\mu^{x_1}(\cdot)$ the invariant measure of the following fixed-x_1 equation

$$dX_2^{x_1}(t) = h_2(x_1, X_2^{x_1}(t))dt + \varsigma_2(x_1, X_2^{x_1}(t))dw_2(t).$$

Another method is probabilistic (a stochastic averaging method), for example, Khasminskii [12], Kushner [19,20], and Pardoux and Veretennikov [27]. Note that in the last reference above, partial differential equations were also used as a bridge for the averaging, whereas in [12,19,20,30,31, 38] probabilistic method was used as a primary tool.

The two-time-scale systems mentioned above can be extended to systems involving delays. When time delays have influence on two-time-scale diffusion systems, SDDEs and SFDEs with two-time scales have to be considered. Due to the lack of the Markovian property, techniques in the literature for treating Markov processes are not applicable. For example, the weak convergence methods developed by Kushner in [19,20] cannot be applied directly. For instance, when the perturbed test function method is used, it is necessary to consider the differential of the delay term. This implies that delay or functional differential is needed. By extending the functional Itô formula initiated by Dupire [6], we established in [36] a stochastic averaging principle for the two-time-scale functional diffusion system of the following form

$$\begin{cases} dX^\varepsilon(t) = b(X_t^\varepsilon, Y^\varepsilon(t))dt + \psi(X_t^\varepsilon, Y^\varepsilon(t))dw_1(t), \\ dY^\varepsilon(t) = \dfrac{1}{\varepsilon}h(Y^\varepsilon(t))dt + \dfrac{1}{\sqrt{\varepsilon}}\phi(Y^\varepsilon(t))dw_2(t), \end{cases} \tag{1.4}$$

where $x_t^\varepsilon := \{x^\varepsilon(u \wedge t) : 0 \le u \le T\}$ represents the delay from 0 to the current time t. Note that in the above, although the slow process depends on the fast process, the fast component involves no delays. However, the analysis is already rather complex. This paper treats an even more complex situation, where the fast process is also past dependent and involves an infinite-dimensional process. When the fast component depends on the segment process of the slow one, the slow component will appear in the stationary distribution of the corresponding fixed-x equation, making it difficult to apply the functional Itô formula. Thus, when fully coupled SDDEs or SFDEs with two-times scales are examined, different approaches need to be taken.

To substantially extend the results of [36], this paper examines coupled functional diffusion processes with two-time scales given by the following SFDE

$$dx^\varepsilon(t) = b(x_t^\varepsilon, \xi^\varepsilon(t))dt + \psi(x_t^\varepsilon, \xi^\varepsilon(t))dw_1(t), \tag{1.5a}$$

$$d\xi^\varepsilon(t) = \frac{1}{\varepsilon}h(x_t^\varepsilon, \xi^\varepsilon(t))dt + \frac{1}{\sqrt{\varepsilon}}\phi(x_t^\varepsilon, \xi^\varepsilon(t))dw_2(t), \tag{1.5b}$$

with initial data $\xi(0) \in \mathbb{R}^m$ and $x_0 \in C([-\tau, 0]; \mathbb{R}^n)$. In the above, $x_t^\varepsilon := \{x^\varepsilon(t+\theta) : -\tau \le \theta \le 0\}$ is termed a segment process or solution map process, $b = (b_1, b_2, \dots, b_n)' : C([-\tau, 0]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^n$, $\psi = [\psi_{ij}]_{n\times l_1} : C([-\tau, 0]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^{n\times l_1}$, $h = (h_1, h_2, \dots, h_m)' : C([-\tau, 0]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^m$, z' denotes the transpose of z, $\phi = [\phi_{ij}]_{m\times l_2} : C([-\tau, 0]; \mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^{m\times l_2}$, and $w_1(t)$ and $w_2(t)$ are two independent standard Brownian motions taking values in $\mathbb{R}^{l_1}$ and $\mathbb{R}^{l_2}$, respectively. Note that the fast and slow components are fully coupled through the segment process x_t^ε.

Since the functional term in $C([-\tau,0];\mathbb{R}^n)$ and coupled systems are considered, it is difficult to use the perturbed test function method together with the functional Itô formula. This paper aims to establish the averaging principle for system (1.5) by a direct averaging approach. Along this line, it is also necessary to examine such properties as continuity and tightness on the space $C([0,T];C([-\tau,0];\mathbb{R}^n))$ of the segment process $x^\varepsilon_\cdot$. In addition, we need to examine the fixed-x equation (or frozen x equation) and establish the continuous dependence on the fixed parameter x of the system.

The rest of the paper is arranged as follows. Section 2 provides notation, assumptions, and some preliminary results. Section 3 establishes existence and uniqueness of the global solutions for the coupled functional stochastic differential equations with two-time scales and the moment boundedness of the solutions. This section also presents continuity of the slow component and the corresponding segment process. Section 4 examines the invariant measure and the exponential ergodicity of the fixed-x equation and proves continuous dependence of the solution on the parameter x. These results are interesting in their own right. Section 5 establishes the averaging principle by examining the weak convergence of the slow-varying process $x^\varepsilon(\cdot)$ as $\varepsilon\to 0$. To establish this result, we prove the tightness of the segment process $x^\varepsilon_\cdot$ on the space $C([-\tau,0];C([-\tau,0];\mathbb{R}^n))$. Treating two classes of SFDEs, Section 6 derives the averaging principles of stochastic integro-differential equations (SIDEs) and SDDEs with two-time scales by using weak convergence methods. Some final thoughts are presented at the end of the paper.

2. Notation, assumptions, and preliminaries

Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space with the Euclidean norm $|\cdot|$. For a vector or matrix Ψ, denote its transpose by Ψ'; for a matrix Ψ, denote its trace norm by $|\Psi|=\sqrt{\mathrm{Tr}(\Psi'\Psi)}$. For a set or event A, $\overline{A}$ represents its complement. For $a,b\in\mathbb{R}^n$, $\langle a,b\rangle = a'b$ represents the inner product of a and b. Throughout the paper, K denotes a generic positive constant, whose value may change for different usage, so $K+K=K$ and $KK=K$ are understood in an appropriate sense. Similarly, K_α denotes a generic positive constant depending on parameter α. We use $\varepsilon>0$ to represent a small parameter.

In this paper, if $x(t)$ is a stochastic process, denote by $\mathcal{F}^x_t=\sigma\{x(s): s\le t\}$ the filtration generated by $\{x(s): s\le t\}$, and $\mathbb{E}^x_t$ the corresponding conditional expectation. For stochastic processes $\xi^\varepsilon(\cdot)$ and $x^\varepsilon(\cdot)$ depending on ε, we define $\mathcal{F}^\varepsilon_t$ as the σ-algebra generated by $\{\xi^\varepsilon(s), x^\varepsilon(s): s\le t\}$, and $\mathbb{E}^\varepsilon_t$ the conditional expectation on $\mathcal{F}^\varepsilon_t$.

For $\tau>0$, denote by $C([-\tau,0];\mathbb{R}^n)$ the family of continuous functions $\varphi(\cdot)$ from $[-\tau,0]$ to $\mathbb{R}^n$ with the norm $\|\varphi\|=\sup_{-\tau\le\theta\le 0}|\varphi(\theta)|$ and $C([0,T];\mathbb{R}^n)$ be the family of continuous function $x(\cdot)$ from $[0,T]$ to $\mathbb{R}^n$. For $p>0$, $L^p_{\mathcal{F}_t}(\Omega,\mathbb{R}^n)$ and $L^p_{\mathcal{F}_t}(\Omega, C([-\tau,0];\mathbb{R}^n))$ represent the families of $\mathbb{R}^n$ and $C([-\tau,0];\mathbb{R}^n)$-valued $\mathcal{F}_t$-measurable random variables with $\mathbb{E}|\cdot|^p<\infty$ and $\mathbb{E}\|\cdot\|^p<\infty$, respectively. Denote by $C^\infty_0(\mathbb{R}^n;\mathbb{R})$ the family of C^∞ functions on $\mathbb{R}^n$ with compact support. Also define $C([0,T];C([-\tau,0];\mathbb{R}^n))$ the family of continuous functions $x_\cdot$ from $[0,T]$ to $C([-\tau,0];\mathbb{R}^n)$, which has the following property.

Lemma 2.1. *The space $C([0,T];C([-\tau,0];\mathbb{R}^n))$ is complete and separable.*

Proof. Note that $C([0,T]\times[-\tau,0];\mathbb{R}^n)$ is complete and separable, and

$$C([0,T];C([-\tau,0];\mathbb{R}^n)) = C([0,T]\times[-\tau,0];\mathbb{R}^n).$$

The desired result follows. □

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 [19,22], let us recall the definitions of the p-lim and the infinitesimal operator $\hat{\mathcal{L}}^{\varepsilon}$ as follows.

Definition 2.1. Let $f, f^{\delta} \in \overline{\mathcal{M}}^{\varepsilon}$ for each $\delta > 0$. We say $f = \text{p-}\lim_{\delta} f^{\delta}$ 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 $\text{p-}\lim_{\delta} f^{\delta} = 0$ if $f(\cdot) = 0$ almost surely, where $f^{\delta} \in \overline{\mathcal{M}}^{\varepsilon}$ for each $\delta > 0$.

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 by Kurtz [22].

Lemma 2.2. *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

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

We need the following assumptions.

(A1) (Lipschitz condition) For any integer R, there exists a positive constant L_R such that for any $\varphi_1, \varphi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ and $\xi_1, \xi_2 \in \mathbb{R}^m$ with $\|\varphi_1\| \vee \|\varphi_2\| \vee |\xi_1| \vee |\xi_2| \le R$

$$|h(\varphi_1, \xi_1) - h(\varphi_2, \xi_2)|^2 \le L_R (\|\varphi_1 - \varphi_2\|^2 + |\xi_1 - \xi_2|^2), \tag{2.2}$$

and

$$|b(\varphi_1,\xi_1)-b(\varphi_2,\xi_2)|^2\vee|\psi(\varphi_1,\xi_1)-\psi(\varphi_2,\xi_2)|^2\le L_R\|\varphi_1-\varphi_2\|^2+L|\xi_1-\xi_2|^2. \tag{2.3}$$

In (2.3), $\xi_1, \xi_2 \in \mathbb{R}^m$ and L is a positive constant.

(A2) (Dissipative condition) For any $\varphi_1, \varphi_2 \in C([-\tau, 0]; \mathbb{R}^n)$, there exist λ_1, λ_2, and L such that for any $\xi_1, \xi_2 \in \mathbb{R}^m$,

$$\langle \xi_1-\xi_2, h(\varphi_1,\xi_1)-h(\varphi_2,\xi_2)\rangle \le -\lambda_1|\xi_1-\xi_2|^2+L\|\varphi_1-\varphi_2\|^2$$

and

$$|\phi(\varphi_1,\xi_1)-\phi(\varphi_2,\xi_2)|^2\le\lambda_2(|\xi_1-\xi_2|^2+\|\varphi_1-\varphi_2\|^2).$$

(A3) (Linear growth condition) There exists a constant $L > 0$ such that

$$|b(\varphi,0)|^2\vee|\psi(\varphi,0)|^2\vee|h(\varphi,0)|^2\le L(1+\|\varphi\|^2), \tag{2.4}$$

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

(A4) The initial data $\xi(0) \in L^p_{\mathcal{F}_0}(\Omega, \mathbb{R}^m)$ and $x_0 \in L^p_{\mathcal{F}_0}(\Omega, C([-\tau, 0]; \mathbb{R}^n))$ for some $p > 2$, and $x(\theta)$ for $\theta \in [-\tau, 0]$ is Hölder-continuous with exponent $\gamma_0 > 0$.

3. Coupled functional diffusions with two-time scales

The following theorem establishes the existence and uniqueness of the strong solution for Eq. (1.5) together with the moment bounds of the solution, as well as continuity of the slow-varying component $x^\varepsilon(\cdot)$.

Theorem 3.1. *Under Assumptions* (A1)–(A4)*, for any* $\varepsilon > 0$*, SFDE* (1.5) *has a unique global strong solution* $(x^{\varepsilon\prime}(t), \xi^{\varepsilon\prime}(t))'$*. Moreover, if* $2\lambda_1 > \lambda_2$*, for any* $T > 0$*, there exist positive constants* $p > 2$ *and* $K_{p,T}$ *depending on* p, T*, and initial data* x_0 *and* $\xi(0)$ *independent of* ε *such that*

$$\mathbb{E}\Big[\sup_{0\le t\le T}|x^\varepsilon(t)|^p\Big]\le K_{p,T} \tag{3.1}$$

and

$$\sup_{0\le t\le T}\mathbb{E}|\xi^\varepsilon(t)|^p\le K_{p,T}, \tag{3.2}$$

and

$$\mathbb{E}|x^\varepsilon(t)-x^\varepsilon(s)|^p\le K_{p,T}(t-s)^{p/2},\ \textit{for all } 0\le s<t\le T. \tag{3.3}$$

Proof. We divide the proof into the following three steps.

Step 1: Existence and uniqueness of the global solution. The SFDE (1.5) can be rewritten as

$$
\begin{aligned}
d\begin{pmatrix} x^{\varepsilon}(t) \\ \xi^{\varepsilon}(t) \end{pmatrix} &= \begin{pmatrix} b(x_t^{\varepsilon}, \xi^{\varepsilon}(t)) \\ \frac{1}{\varepsilon} h(x_t^{\varepsilon}, \xi^{\varepsilon}(t)) \end{pmatrix} dt \\
&\quad + \begin{pmatrix} \psi(x_t^{\varepsilon}(t), \xi^{\varepsilon}(t)) & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \phi(x_t^{\varepsilon}(t), \xi^{\varepsilon}(t)) \end{pmatrix} d\begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix}.
\end{aligned}
\tag{3.4}
$$

According to Assumptions (A1) and (A2), for any $\varphi_1, \varphi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ and $\xi_1, \xi_2 \in \mathbb{R}^m$ with $\|\varphi_1\| \vee \|\varphi_2\| \vee |\xi_1| \vee |\xi_2| \le R$,

$$
\begin{aligned}
&\left|\begin{pmatrix} b(\varphi_1, \xi_1) \\ \frac{1}{\varepsilon} h(\varphi_1, \xi_1) \end{pmatrix} - \begin{pmatrix} b(\varphi_2, \xi_2) \\ \frac{1}{\varepsilon} h(\varphi_2, \xi_2) \end{pmatrix}\right|^2 = \left|\begin{pmatrix} b(\varphi_1, \xi_1) - b(\varphi_2, \xi_2) \\ \frac{1}{\varepsilon}[h(\varphi_1, \xi_1) - h(\varphi_2, \xi_2)] \end{pmatrix}\right|^2 \\
&= |b(\varphi_1, \xi_1) - b(\varphi_2, \xi_2)|^2 + \frac{1}{\varepsilon^2}|h(\varphi_1, \xi_1) - h(\varphi_2, \xi_2)|^2 \\
&\le \Big(1 + \frac{1}{\varepsilon^2}\Big) L_R(\|\varphi_1 - \varphi_2\|^2 + \|\xi_1 - \xi_2\|^2)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\begin{pmatrix} \psi(\varphi_1, \xi_1) & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \phi(\varphi_1, \xi_1) \end{pmatrix} - \begin{pmatrix} \psi(\varphi_2, \xi_2) & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \phi(\varphi_2, \xi_2) \end{pmatrix}\right|^2 \\
&= \left|\begin{pmatrix} \psi(\varphi_1, \xi_1) - \psi(\varphi_2, \xi_2) & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}}[\phi(\varphi_1, \xi_1) - \phi(\varphi_2, \xi_2)] \end{pmatrix}\right|^2 \\
&= |\psi(\varphi_1, \xi_1) - \psi(\varphi_2, \xi_2)|^2 + \frac{1}{\varepsilon}|\phi(\varphi_1, \xi_1) - \phi(\varphi_2, \xi_2)|^2 \\
&\le \Big(L_R + \frac{\lambda_2}{\varepsilon}\Big)(\|\varphi_1 - \varphi_2\|^2 + \|\xi_1 - \xi_2\|^2),
\end{aligned}
$$

which shows that Eq. (3.4) (or Eq. (1.5)) satisfies the local Lipschitz condition. Since this equation is autonomous, the local Lipschitz condition also implies that the local linear growth condition, yielding that Eq. (3.4) (or Eq. (1.5)) has a local solution (see [23, Theorem 2.8, P154]).

By Assumptions (A1), (A2), and (A3),

$$
\begin{aligned}
\varphi'(0) b(\varphi, \xi) &\le \frac{1}{2}(|\varphi(0)|^2 + |b(\varphi, \xi)|^2) \\
&= \frac{1}{2}|\varphi(0)|^2 + \frac{1}{2}|b(\varphi, \xi) - b(\varphi, 0) + b(\varphi, 0)|^2 \\
&\le \frac{1}{2}|\varphi(0)|^2 + |b(\varphi, \xi) - b(\varphi, 0)|^2 + |b(\varphi, 0)|^2
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{1}{2}\|\varphi\|^2 + L|\xi|^2 + L(1+\|\varphi\|^2)\\
&\le K(\|\varphi\|^2 + |\xi|^2 + 1), \qquad (3.5)
\end{aligned}
$$

$$
\begin{aligned}
\xi' h(\varphi,\xi) &\le -\lambda_1|\xi|^2 + \xi' h(\varphi,0)\\
&\le -\lambda_1|\xi|^2 + \frac{|\xi|^2}{2} + \frac{|h(\varphi,0)|^2}{2}\\
&\le \Big(-\lambda_1 + \frac{1}{2}\Big)|\xi|^2 + \frac{L}{2}\|\varphi\|^2 + \frac{L}{2}\\
&\le K(\|\varphi\|^2 + |\xi|^2 + 1), \qquad (3.6)
\end{aligned}
$$

$$
\begin{aligned}
|\psi(\varphi,\xi)|^2 &= |\psi(\varphi,\xi) - \psi(\varphi,0) + \psi(\varphi,0)|^2\\
&\le 2|\psi(\varphi,\xi) - \psi(\varphi,0)|^2 + 2|\psi(\varphi,0)|^2\\
&\le 2L|\xi|^2 + 2L(1+\|\varphi\|^2)\\
&\le K(\|\varphi\|^2 + |\xi|^2 + 1) \qquad (3.7)
\end{aligned}
$$

and

$$
\begin{aligned}
|\phi(\varphi,\xi)|^2 &\le |\phi(\varphi,\xi) - \phi(\mathbf{0},0) + \phi(\mathbf{0},0)|^2\\
&\le 2|(\phi(\varphi,\xi) - \phi(\mathbf{0},0)|^2 + 2|\phi(\mathbf{0},0)|^2\\
&\le 2\lambda_2(\|\varphi\|^2 + |\xi|^2) + 2|\phi(\mathbf{0},0)|^2\\
&\le K(\|\varphi\|^2 + |\xi|^2 + 1), \qquad (3.8)
\end{aligned}
$$

where $\mathbf{0} \in C([-\tau,0];\mathbb{R}^n)$ represents zero segment process. These yield that for any $\varphi \in C([-\tau,0];\mathbb{R}^n)$ and $\xi \in \mathbb{R}^m$, from (A3)

$$
\begin{aligned}
&(\varphi'(0),\xi')\begin{pmatrix} b(\varphi,\xi)\\ \frac{1}{\varepsilon}h(\varphi,\xi)\end{pmatrix} + \frac{1}{2}\left|\begin{pmatrix}\psi(\varphi,\xi) & 0\\ 0 & \frac{1}{\sqrt{\varepsilon}}\phi(\varphi,\xi)\end{pmatrix}\right|^2\\
&\le \varphi'(0)b(\varphi,\xi) + \frac{1}{\varepsilon}\xi' h(\varphi,\xi) + \frac{1}{2}\Big[|\psi(\varphi,\xi)|^2 + \frac{1}{\varepsilon}|\phi(\varphi,\xi)|^2\Big]\\
&\le \Big(\frac{3}{2}K + \frac{3}{2\varepsilon}K\Big)(\|\varphi\|^2 + |\xi|^2 + 1)\\
&=: K_\varepsilon(\|\varphi\|^2 + |\xi|^2 + 1), \qquad (3.9)
\end{aligned}
$$

which shows that the coefficients of Eq. (3.4) satisfy the monotone condition.

Choosing sufficiently large integer R satisfying $|\xi(0)| \vee |x(0)| < R$, define the stopping time

$$\tau_R = \inf\{t \ge 0, |x^\varepsilon(t)| \vee |\xi^\varepsilon(t)| > R\} \wedge T.$$

Note that τ_R is increasing with respect to R. Define $\tau_\infty = \lim_{R\to\infty}\tau_R$ and let $z^\varepsilon(t) = ([x^\varepsilon(t)]', [\xi^\varepsilon(t)]')'$. Since Eq. (1.5) holds a local solution, for any integer $R > 0$ and any

$t \in [0, \tau_R]$, $z^\varepsilon(t)$ is well-posed. By the monotone condition, applying the Itô formula to $|z^\varepsilon(t)|^2$ yields that for any $t \in [0, T]$,

$$\begin{aligned}\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}|z^\varepsilon(s)|^2\Big] &= \mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}(|x^\varepsilon(s)|^2+|\xi^\varepsilon(s)|^2)\Big]\\ &\le \mathbb{E}[|x(0)|^2+|\xi(0)|^2]+K_\varepsilon\mathbb{E}\int_0^{t\wedge\tau_R}[\|x_u^\varepsilon\|^2+|\xi^\varepsilon(u)|^2+1]du\\ &\quad+\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}\int_0^s 2[z^\varepsilon(u)]'\Sigma^\varepsilon(u)dw(u)\Big],\end{aligned}\tag{3.10}$$

where

$$\Sigma^\varepsilon(u)=\begin{pmatrix}\psi(x_u^\varepsilon,\xi^\varepsilon(u)) & 0\\ 0 & \dfrac{1}{\sqrt{\varepsilon}}\phi(x_u^\varepsilon,\xi^\varepsilon(u))\end{pmatrix}\quad\text{and}\quad w(u)=\begin{pmatrix}w_1(u)\\ w_2(u)\end{pmatrix}.$$

By (A3), (3.7), and (3.8), applying the Burkholder-Davis-Gundy inequality gives

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}\int_0^s 2[z^\varepsilon(u)]'\Sigma^\varepsilon(u)dw(u)\Big]\\ &\le 12\mathbb{E}\Big|\int_0^{t\wedge\tau_R}|z^\varepsilon(u)|^2|\Sigma^\varepsilon(u)|^2du\Big|^{1/2}\\ &\le \frac{1}{2}\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}|z^\varepsilon(s)|^2\Big]+K\int_0^{t\wedge\tau_R}\Big[|\psi(x_u^\varepsilon,\xi^\varepsilon(u))|^2+\frac{1}{\varepsilon}|\phi(x_u^\varepsilon,\xi^\varepsilon(u))|^2\Big]du\\ &\le \frac{1}{2}\mathbb{E}\Big[\sup_{0\le s\le t}|z^\varepsilon(s\wedge\tau_R)|^2\Big]+K_\varepsilon\int_0^{t\wedge\tau_R}[\|x_u^\varepsilon\|^2+|\xi^\varepsilon(u)|^2+1]du.\end{aligned}\tag{3.11}$$

Substituting (3.11) into (3.10) yields

$$\begin{aligned}&\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}|z^\varepsilon(s)|^2\Big]\\ &\le \mathbb{E}[|x(0)|^2+|\xi(0)|^2]+K_\varepsilon\mathbb{E}\int_0^{t\wedge\tau_R}[\|x_u^\varepsilon\|^2+|\xi^\varepsilon(u)|^2+1]du\\ &\le \mathbb{E}[|x(0)|^2+|\xi(0)|^2]+K_\varepsilon\int_0^t[1+\mathbb{E}\|x_0\|^2+\mathbb{E}[\sup_{0\le v\le u\wedge\tau_R}|z^\varepsilon(v)|^2]]du.\end{aligned}\tag{3.12}$$

The Gronwall inequality gives

$$\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}|z^\varepsilon(s)|^2\Big]\le K_\varepsilon\mathbb{E}[\|x_0\|^2+|\xi(0)|^2+1]e^{K_\varepsilon T}:=K_{\varepsilon,T}.$$

Noting t's arbitrariness and letting $R\to\infty$ give

$$\mathbb{E}\Big[\sup_{0\le s\le T}|z^\varepsilon(s)|^2\Big]\le K_\varepsilon\mathbb{E}[\|x_0\|^2+|\xi(0)|^2+1]e^{K_\varepsilon T}=K_{\varepsilon,T}.$$

Recalling the definition of τ_R, this implies that for any $T>0$,

$$\begin{aligned}R^2\mathbb{P}(\tau_R\le T)&\le\mathbb{E}[|x^\varepsilon(\tau_R\wedge T)|^2\mathbf{1}_{\{\tau_R\le T\}}\vee|\xi^\varepsilon(\tau_R\wedge T)|^2\mathbf{1}_{\{\tau_R\le T\}}]\\&\le\mathbb{E}[|x^\varepsilon(\tau_R\wedge T)|^2+|\xi^\varepsilon(\tau_R\wedge T)|^2]\mathbf{1}_{\{\tau_R\le T\}}\\&\le\mathbb{E}\Big[\sup_{0\le s\le T}|z^\varepsilon(s)|^2\Big]\le K_{\varepsilon,T},\end{aligned}$$

which implies that

$$\mathbb{P}(\tau_R\le T)\le\frac{K_{\varepsilon,T}}{R^2}.$$

Hence it follows that

$$\sum_{R=1}^{\infty}\mathbb{P}(\tau_R\le T)<\infty.$$

The Borel–Cantelli lemma gives that for any $T>0$,

$$\mathbb{P}(\tau_\infty\le T)=0.$$

Due to the arbitrariness of T, $\mathbb{P}(\tau_\infty=\infty)=1$, which shows that the local solution is actually global. This gives the existence and uniqueness of the global solution of Eq. (1.5).

Step 2: Proof of (3.1) and (3.2). For $p>2$ and $\lambda>0$, applying the Itô formula yields that for any $s>0$,

$$\begin{aligned}e^{\frac{\lambda}{\varepsilon}s}[1+|\xi^\varepsilon(s)|^2]^{\frac{p}{2}}&=[1+|\xi(0)|^2]^{\frac{p}{2}}+\frac{\lambda}{\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p}{2}}du\\&\quad+\frac{p}{\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}[\xi^\varepsilon(u)]'h(x^\varepsilon_u,\xi^\varepsilon(u))du\\&\quad+\frac{p}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}|\phi(x^\varepsilon_u,\xi^\varepsilon(u))|^2du\end{aligned}$$

$$
\begin{aligned}
&+\frac{p(p-2)}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p-4}{2}}|[\xi^\varepsilon(u)]'\phi(x_u^\varepsilon,\xi^\varepsilon(u))|^2du\\
&+M^\varepsilon(s)\\
\le{}&[1+|\xi(0)|^2]^{\frac{p}{2}}+\frac{\lambda}{\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p}{2}}du\\
&+\frac{p}{\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}[\xi^\varepsilon(u)]'h(x_u^\varepsilon,\xi^\varepsilon(u))du\\
&+\frac{p(p-1)}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}|\phi(x_u^\varepsilon,\xi^\varepsilon(u))|^2du\\
&+M^\varepsilon(s),
\end{aligned}
\tag{3.13}
$$

where $M^\varepsilon(s)$ is a local martingale with $\mathbb{E}M^\varepsilon(s)=0$. By Assumptions (A2), (A3), and the Young inequality, for any ε_1,

$$
\begin{aligned}
[\xi^\varepsilon(u)]'h(x_u^\varepsilon,\xi^\varepsilon(u))&\le-\lambda_1|\xi^\varepsilon(u)|^2+[\xi^\varepsilon(u)]'h(x_u^\varepsilon,0)\\
&\le-\lambda_1|\xi^\varepsilon(u)|^2+\frac{\varepsilon_1|\xi^\varepsilon(u)|^2}{2}+\frac{|h(x_u^\varepsilon,0)|^2}{2\varepsilon_1}\\
&\le\Big(-\lambda_1+\frac{\varepsilon_1}{2}\Big)|\xi^\varepsilon(u)|^2+\frac{L}{2\varepsilon_1}\|x_u^\varepsilon\|^2+\frac{L}{2\varepsilon_1}.
\end{aligned}
\tag{3.14}
$$

By Assumption (A2) and the Young inequality, for any ε_2,

$$
\begin{aligned}
&|\phi(x_u^\varepsilon,\xi^\varepsilon(u))|^2\\
&\quad\le|(\phi(x_u^\varepsilon,\xi^\varepsilon(u))-\phi(\mathbf{0},0)+\phi(\mathbf{0},0)|^2\\
&\quad\le|(\phi(x_u^\varepsilon,\xi^\varepsilon(u))-\phi(\mathbf{0},0)|^2+|\phi(\mathbf{0},0)|^2+2(\phi(x_u^\varepsilon,\xi^\varepsilon(u))-\phi(\mathbf{0},0))'\phi(\mathbf{0},0)\\
&\quad\le\lambda_2(\|x_u^\varepsilon\|^2+|\xi^\varepsilon(u)|^2)+|\phi(\mathbf{0},0)|^2+\varepsilon_2|(\phi(x_u^\varepsilon,\xi^\varepsilon(u))-\phi(\mathbf{0},0)|^2+\frac{1}{\varepsilon_2}|\phi(\mathbf{0},0)|^2\\
&\quad\le\lambda_2(1+\varepsilon_2)|\xi^\varepsilon(u)|^2+\lambda_2(1+\varepsilon_2)\|x_u^\varepsilon\|^2+\Big(1+\frac{1}{\varepsilon_2}\Big)|\phi(\mathbf{0},0)|^2.
\end{aligned}
\tag{3.15}
$$

Substituting (3.14) and (3.15) into (3.13) yields

$$
\begin{aligned}
&e^{\frac{\lambda}{\varepsilon}s}[1+|\xi^\varepsilon(s)|^2]^{\frac{p}{2}}\\
&\quad\le[1+|\xi(0)|^2]^{\frac{p}{2}}\\
&\qquad+\frac{p}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}\Big\{\Big[-2\lambda_1+(p-1)\lambda_2(1+\varepsilon_2)+\frac{2\lambda}{p}+\varepsilon_1\Big][1+|\xi^\varepsilon(u)|^2]^{\frac{p}{2}}
\end{aligned}
$$

$$+\Big[\frac{L}{\varepsilon_1}+(p-1)\lambda_2(1+\varepsilon_2)\Big][1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}\|x_u^\varepsilon\|^2+A[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}\Big\}du$$
$$+M^\varepsilon(s), \tag{3.16}$$

where

$$A=\frac{L}{\varepsilon_1}+(p-1)\Big(1+\frac{1}{\varepsilon_2}\Big)|\phi(\mathbf{0},0)|^2-\Big[-2\lambda_1+(p-2)\lambda_2(1+\varepsilon_2)+\frac{2\lambda}{p}+\varepsilon_1\Big]$$

is a constant. Applying the Young inequality yields that for any ε_3 such that

$$[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}\|x_u^\varepsilon\|^2\le\frac{(p-2)\varepsilon_1\varepsilon_3}{p}[1+|\xi^\varepsilon(u)|^2]^{\frac{p}{2}}+\frac{2}{p(\varepsilon_1\varepsilon_3)^{\frac{p-2}{2}}}\|x_u^\varepsilon\|^p,$$

which shows that

$$\begin{aligned}&e^{\frac{\lambda}{\varepsilon}s}[1+|\xi^\varepsilon(s)|^2]^{\frac{p}{2}}\\&\le[1+|\xi(0)|^2]^{\frac{p}{2}}+\frac{p}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}\Big\{-B[1+|\xi^\varepsilon(u)|^2]^{\frac{p}{2}}+A[1+|\xi^\varepsilon(u)|^2]^{\frac{p-2}{2}}\\&\quad+\Big[\frac{L}{\varepsilon_1}+(p-1)\lambda_2(1+\varepsilon_2)\Big]\frac{2}{p(\varepsilon_1\varepsilon_3)^{\frac{p-2}{2}}}\|x_u^\varepsilon\|^p\Big\}du+M^\varepsilon(s),\end{aligned}\tag{3.17}$$

where

$$B=2\lambda_1-(p-1)\lambda_2(1+\varepsilon_2)-\frac{2\lambda}{p}-\varepsilon_1-\frac{(p-2)L\varepsilon_3}{p}-\frac{p-2}{p}(p-1)\lambda_2(1+\varepsilon_2)\varepsilon_1\varepsilon_3.$$

Noting that $2\lambda_1>\lambda_2$, we can choose λ, ε_1, ε_2 and ε_3 sufficiently small, and $p>2$ but sufficiently close to 2 such that $B>0$. This also implies that there exists a constant K such that

$$-B|z|^{\frac{p}{2}}+A|z|^{\frac{p-2}{2}}<K$$

according to the boundedness of the polynomial function. It follows from this result that there exists a constant $K>0$ such that

$$\begin{aligned}e^{\frac{\lambda}{\varepsilon}s}|\xi^\varepsilon(s)|^p&\le e^{\frac{\lambda}{\varepsilon}s}[1+|\xi^\varepsilon(s)|^2]^{\frac{p}{2}}\\&\le[1+|\xi(0)|^2]^{\frac{p}{2}}+\frac{p}{2\varepsilon}\int_0^s e^{\frac{\lambda}{\varepsilon}u}[K+K\|x_u^\varepsilon\|^p]du+M^\varepsilon(s)\\&\le[1+|\xi(0)|^2]^{\frac{p}{2}}+\frac{pK}{2\lambda}\Big[1+\sup_{0\le u\le s}\|x_u^\varepsilon\|^p\Big][e^{\frac{\lambda}{\varepsilon}s}-1]+M^\varepsilon(s),\end{aligned}$$

which shows that

$$
\begin{aligned}
\mathbb{E}|\xi^\varepsilon(s)|^p &\le \mathbb{E}[1+|\xi(0)|^2]^{\frac{p}{2}} + \frac{pK}{2\lambda}\Big[1+\mathbb{E}\Big(\sup_{0\le u\le s}\|x_u^\varepsilon\|^p\Big)\Big] \\
&\le \mathbb{E}[1+|\xi(0)|^2]^{\frac{p}{2}} + \frac{pK}{2\lambda}\Big[1+\mathbb{E}\|x_0\|^p+\mathbb{E}\Big(\sup_{0\le u\le s}|x^\varepsilon(u)|^p\Big)\Big] \\
&\le K_p(1+\mathbb{E}|\xi(0)|^p+\mathbb{E}\|x_0\|^p) + K_p\mathbb{E}\Big(\sup_{0\le u\le s}|x^\varepsilon(u)|^p\Big)
\end{aligned} \tag{3.18}
$$

since

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{0\le u\le s}\|x_u^\varepsilon\|^p\Big] &= \mathbb{E}\Big[\sup_{0\le u\le s}\sup_{-\tau\le\theta\le 0}|x^\varepsilon(u+\theta)|^p\Big] \\
&\le \mathbb{E}\Big[\sup_{-\tau\le u\le s}|x^\varepsilon(u)|^p\Big] \\
&\le \mathbb{E}\|x_0\|^p + \mathbb{E}\Big[\sup_{0\le u\le s}|x^\varepsilon(u)|^p\Big].
\end{aligned} \tag{3.19}
$$

For any $p > 2$, it is easy to observe that

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{0\le s\le t}|x^\varepsilon(s)|^p\Big] \le 3^{p-1}\Big\{&\mathbb{E}|x(0)|^p + \mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s b(x_u^\varepsilon,\xi^\varepsilon(u))du\Big|^p\Big] \\
&+\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s \psi(x_u^\varepsilon,\xi^\varepsilon(u))dw_1(u)\Big|^p\Big]\Big\}.
\end{aligned} \tag{3.20}
$$

Computation of (3.5) implies $|b(\varphi,\xi)|^2 \le K(\|\varphi\|^2+|\xi|^2+1)$. Applying the Hölder inequality gives

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s b(x_u^\varepsilon,\xi^\varepsilon(u))du\Big|^p\Big] &\le t^{p-1}\int_0^t \mathbb{E}|b(x_u^\varepsilon,\xi^\varepsilon(u))|^p du \\
&\le t^{p-1}K_p\int_0^t \mathbb{E}(\|x_u^\varepsilon\|^2+|\xi^\varepsilon(u)|^2+1)^{\frac{p}{2}}du \\
&\le K_{p,t}\int_0^t [\mathbb{E}\|x_u^\varepsilon\|^p+\mathbb{E}|\xi^\varepsilon(u)|^p+1]du.
\end{aligned} \tag{3.21}
$$

By (3.7), applying the Burkholder-Davis-Gundy inequality and the Hölder inequality yield

$$
\mathbb{E}\Big[\sup_{0\le s\le t}\Big|\int_0^s \psi(x_u^\varepsilon,\xi^\varepsilon(u))dw_1(u)\Big|^p\Big]
$$

$$\le K_p\mathbb{E}\Big|\int_0^t |\psi(x_u^\varepsilon,\xi^\varepsilon(u))|^2 du\Big|^{\frac{p}{2}}$$

$$\le K_p t^{\frac{p-2}{2}}\int_0^t \mathbb{E}|\psi(x_u^\varepsilon,\xi^\varepsilon(u))|^p du$$

$$\le K_{p,t}\int_0^t [\mathbb{E}\|x_u^\varepsilon\|^p+\mathbb{E}|\xi^\varepsilon(u)|^p+1]du. \tag{3.22}$$

Substituting (3.20) and (3.21) into (3.22), together with (3.19), gives

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le s\le t}|x^\varepsilon(s)|^p\Big]\\
&\le K_p\mathbb{E}|x(0)|^p+K_{p,t}\int_0^t[\mathbb{E}\|x_u^\varepsilon\|^p+\mathbb{E}|\xi^\varepsilon(u)|^p+1]du\\
&\le K_p\mathbb{E}|x(0)|^p+K_{p,t}\int_0^t\Big[\mathbb{E}\|x_0\|^p+\mathbb{E}\Big(\sup_{0\le v\le u}|x^\varepsilon(v)|^p\Big)\\
&\quad+K_p(1+\mathbb{E}|\xi(0)|^p+\mathbb{E}\|x_0\|^p)+K_p\mathbb{E}\Big(\sup_{0\le v\le u}|x^\varepsilon(v)|^p\Big)+1\Big]du\\
&\le K_p\mathbb{E}|x(0)|^p+K_{p,t}\int_0^t\Big[\mathbb{E}\|x_0\|^p+\mathbb{E}|\xi(0)|^p+\mathbb{E}\Big(\sup_{0\le v\le u}|x^\varepsilon(v)|^p\Big)+1\Big]du.
\end{aligned}$$

The Gronwall inequality gives that there exists a $K_{p,T}$ such that

$$\mathbb{E}\Big[\sup_{0\le s\le T}|x^\varepsilon(s)|^p\Big]\le K_{p,T},$$

which leads to the desired (3.1). This, together with (3.18) yields

$$\sup_{0\le s\le T}\mathbb{E}|\xi^\varepsilon(s)|^p\le K_{p,T}.$$

Thus (3.2) holds.

Step 3: Proof of (3.3). For $p>2$ determined by Step 2, it is easy to observe that

$$\mathbb{E}|x^\varepsilon(t)-x^\varepsilon(s)|^p\le 2^{p-1}\Big[\mathbb{E}\Big|\int_s^t b(x_u^\varepsilon,\xi^\varepsilon(u))du\Big|^p+\mathbb{E}\Big|\int_s^t \psi(x_u^\varepsilon,\xi^\varepsilon(u))dw_1(u)\Big|^p\Big]. \tag{3.23}$$

By the Hölder inequality, (A3), (3.1), and (3.2),

$$
\begin{aligned}
&\mathbb{E}\Big|\int_s^t b(x_u^\varepsilon, \xi^\varepsilon(u))du\Big|^p \\
&\le (t-s)^{p-1}\int_s^t \mathbb{E}|b(x_u^\varepsilon, \xi^\varepsilon(u))|^p du \\
&\le 3^{p-1}(t-s)^{p-1}\int_s^t \mathbb{E}[\|x_u^\varepsilon\|^p + \mathbb{E}|\xi^\varepsilon(u)|^p + 1]du \\
&\le 3^{p-1}(t-s)^{p-1}\int_s^t \Big[\mathbb{E}\|x_0\|^p + \mathbb{E}\Big(\sup_{0\le u\le T}|x^\varepsilon(u)|^p\Big) + \sup_{0\le u\le T}\mathbb{E}|\xi^\varepsilon(u)|^p + 1\Big]du \\
&\le K_{p,T}(t-s)^p. \qquad (3.24)
\end{aligned}
$$

By (A3), (3.1), and (3.2), applying the same technique as [23, Theorem 7.1, P39] gives

$$
\begin{aligned}
\mathbb{E}\Big|\int_s^t \psi(x_u^\varepsilon, \xi^\varepsilon(u))dw_1(u)\Big|^p &\le \Big[\frac{p(p-1)}{2}\Big]^{\frac{p}{2}}(t-s)^{\frac{p-2}{2}}\mathbb{E}\int_s^t |\psi(x_u^\varepsilon, \xi^\varepsilon(u))|^p du \\
&\le K_{p,T}(t-s)^{\frac{p}{2}}. \qquad (3.25)
\end{aligned}
$$

Substituting (3.24) and (3.25) into (3.23) leads to

$$
\mathbb{E}|x^\varepsilon(t) - x^\varepsilon(s)|^p \le K_{p,T}(t-s)^{\frac{p}{2}},
$$

which is the desired assertion (3.3). This completes this proof. □

Remark 3.2. As mentioned before, the solution process $(x^{\varepsilon\prime}(t), \xi^{\varepsilon\prime}(t))'$ is not a Markov process. We cannot express some convergence conditions in the sense of transition probability as in Kushner [19]. Even if we consider the segment process x_t^ε (noting that $(x_t^{\varepsilon\prime}, \xi^{\varepsilon\prime}(t))'$ is an $\mathcal{F}_t^\varepsilon$-adapted Markov process [2,24]), Kushner's method still needs to be modified because convergence of the pair $(x^\varepsilon(t), x_t^\varepsilon)$ needs to be considered altogether.

Remark 3.3. It is worth noting that the boundedness of pth moment and continuity are uniform w.r.t. ε. This implies that $x^\varepsilon(\cdot) \to x(\cdot)$ with probability 1, since for any $t \in [0, T]$, $\mathbb{E}[\sup_{0\le s\le T}|x^\varepsilon(s)|^p] \le K_{p,T}$. The Vitali convergence theorem shows $\mathbb{E}[\sup_{0\le s\le T}|x^\varepsilon(t)|^p] \to \mathbb{E}[\sup_{0\le s\le T}|x(t)|^p] \le K_{p,T}$ for any $t \in [0, T]$. Similarly, $\mathbb{E}|x^\varepsilon(t) - x^\varepsilon(s)|^p \to \mathbb{E}|x(t) - x(s)|^p \le K_{p,T}(t-s)^{\frac{p}{2}}$. This also shows that $x(t)$ is continuous in the sense of pth moment.

Applying [11, Theorem 2.8, p.53] gives the following corollary:

Corollary 3.4. *Under the conditions of Theorem 3.1, the solution process $x^\varepsilon(\cdot)$ is locally Hölder-continuous with exponent γ with probability 1 for any $\gamma \in (0, 1/2 - 1/p)$. That is,*

$$\mathbb{P}\left(\omega : \sup_{\substack{s,t\in[0,T],\\ 0<t-s<h(\omega)}} \frac{|x^\varepsilon(\omega,t)-x^\varepsilon(\omega,s)|}{|t-s|^\gamma} < \kappa\right) = 1,$$

where p is defined in Theorem 3.1, $h(\omega)$ is an almost surely positive random variable, and $\kappa > 0$ is an appropriate constant.

Remark 3.5. In fact, if $x^\varepsilon(\cdot) \to x(\cdot)$ with probability 1, Remark 3.3 and [11, Theorem 2.8, p.53] also give that $x(t)$ is locally Hölder-continuous with exponent γ with probability 1 for any $\gamma \in (0, 1/2 - 1/p)$.

By the above results, we can establish continuity of x_t^ε in the almost sure sense and the pth moment.

Corollary 3.6. *Under the conditions of Theorem 3.1, $x_\cdot^\varepsilon$ is locally Hölder-continuous with exponent $\gamma \wedge \gamma_0$ with probability 1, and for p defined in Theorem 3.1 and γ from Corollary 3.4,*

$$\lim_{\delta\to 0} \mathbb{E}\left(\sup_{\substack{t,s\in[0,T]\\ |t-s|\le\delta}} \|x_t^\varepsilon - x_s^\varepsilon\|^p \right) = 0. \tag{3.26}$$

Proof. From Corollary 3.4, there exists a $\tilde{\Omega}$ with $\mathbb{P}(\tilde{\Omega}) = 1$ and $h(\omega) > 0$ for any $\omega \in \tilde{\Omega}$ such that for any $t, s \in [0, T]$ and $0 < |t-s| < h(\omega)$,

$$|x^\varepsilon(\omega,t) - x^\varepsilon(\omega,s)| \le \kappa |t-s|^\gamma \tag{3.27}$$

with probability 1, where $\gamma \in (0, 1/2 - 1/p)$. This implies that for any $\theta \in [-\tau, 0]$, if $(t \wedge s) + \theta \ge 0$,

$$|x^\varepsilon(\omega, t+\theta) - x^\varepsilon(\omega, s+\theta)| \le \kappa |t-s|^\gamma.$$

If $(t \vee s) + \theta \le 0$, (A4) yields that

$$|x^\varepsilon(\omega, t+\theta) - x^\varepsilon(\omega, s+\theta)| \le K |t-s|^{\gamma_0}.$$

If $(t \wedge s) + \theta \le 0$ and $(t \vee s) + \theta \ge 0$, (A4) and (3.27) lead to

$$|x^\varepsilon(\omega, t+\theta) - x^\varepsilon(\omega, s+\theta)| \le K |t-s|^{\gamma\wedge\gamma_0}.$$

These three cases imply that for any $t, s \in [0, T]$ and $\theta \in [-\tau, 0]$,

$$\sup_{-\tau\le\theta\le 0} |x^\varepsilon(\omega, t+\theta) - x^\varepsilon(\omega, s+\theta)| \le K |t-s|^{\gamma\wedge\gamma_0}. \tag{3.28}$$

That is,

$$\|x_t^\varepsilon(\omega) - x_s^\varepsilon(\omega)\| \le K |t-s|^{\gamma\wedge\gamma_0}$$

for any $\omega \in \tilde{\Omega}$ and $|t-s| \in (0, h(\omega))$. This implies that

$$\mathbb{P}\left(\omega : \sup_{\substack{t,s\in[0,T],\\ 0<|t-s|<h(\omega)}} \frac{\|x_t^\varepsilon - x_s^\varepsilon\|}{|t-s|^{\gamma\wedge\gamma_0}} \le K\right) = 1, \tag{3.29}$$

which shows that $x_\cdot^\varepsilon$ is almost surely Hölder-continuous with exponent $\gamma \vee \gamma_0$.

For p defined in Theorem 3.1, (3.29) also indicates that as $\delta \to 0$,

$$\sup_{\substack{t,s\in[0,T],\\ |t-s|<\delta}} \|x_t^\varepsilon - x_s^\varepsilon\|^p \to 0, \quad \text{a.s.}$$

Note that from (3.19), $\mathbb{E}[\sup_{0\le t\le T} \|x_t^\varepsilon\|^p] \le \mathbb{E}\|x_0\|^p + K_{p,T}$. The Lebesgue dominated convergence theorem gives

$$\lim_{\delta\to 0} \mathbb{E}\left(\sup_{\substack{t,s\in[0,T],\\ |t-s|<\delta}} \|x_t^\varepsilon - x_s^\varepsilon\|^p\right) = 0,$$

as desired. □

Remark 3.7. According to [24, Chapter II, Lemma 2.1], continuity of the solution process $x^\varepsilon(\cdot)$ on $[-\tau, T]$ implies that the corresponding segment process $x_\cdot^\varepsilon$ is also continuous on $[0, T]$. But this result cannot give the Hölder continuity.

4. Fast-varying process and fixed-x process

To obtain the weak convergence of the slow-varying $x^\varepsilon(t)$ as $\varepsilon \to 0$, certain ergodicity is crucial. In (1.5), $\xi^\varepsilon(t)$ is rapidly varying in contrast to $x^\varepsilon(t)$. Define $\tilde{\xi}^\varepsilon(t) = \xi^\varepsilon(\varepsilon t)$. Then

$$d\tilde{\xi}^\varepsilon(t) = h(x_{\varepsilon t}^\varepsilon, \tilde{\xi}^\varepsilon(t))dt + \phi(x_{\varepsilon t}^\varepsilon, \tilde{\xi}^\varepsilon(t))d\tilde{w}_2(t), \tag{4.1}$$

where $\tilde{w}_2(t) = w_2(\varepsilon t)/\sqrt{\varepsilon}$ is a standard Brownian motion. Let us consider the following fixed-x equation

$$d\xi^x(t) = h(x, \xi^x(t))dt + \phi(x, \xi^x(t))d\tilde{w}_2(t). \tag{4.2}$$

Theorem 4.1. *Let Assumptions* (A1)*,* (A2)*, and* (A4) *hold. Then the fixed-x equation* (4.2) *has a unique strong global solution $\xi^x(t)$ for any $t \in [0, T]$, which is $\mathcal{F}_{\varepsilon t}^{w_2}$-adapted and satisfies the following properties:*

- (i) *the solution is a homogeneous strong Markov process;*
- (ii) $\mathbb{E}(\sup_{t\in[0,T]} |\xi^x(t)|^2) \le K_{x,T}$*, where $K_{x,T}$ is a constant depending on x and T;*
- (iii) *when $2\lambda_1 > \lambda_2$, there exists a unique invariant measure $\mu^x(\cdot)$, which is exponentially ergodic;*
- (iv) *for any $x_1, x_2 \in C([-\tau, 0]; \mathbb{R}^n)$, $\sup_{t\ge 0} \mathbb{E}|\xi^{x_1}(t) - \xi^{x_2}(t)|^2 \le K\|x_1 - x_2\|^2$, where K is a constant.*

Proof. Existence and uniqueness of the strong solution of (4.2) follows from the existing result [23, Theorem 3.6, p. 58] by (2.2), (A2), and (3.6).

(i) Since (4.2) is autonomous when x is fixed, by the same statement as [23, Theorem 9.5, p. 90] (we replace the uniform Lipschitz condition by the local Lipschitz condition and the Dissipative condition (A2)), we can prove the solution of (4.2) is a homogeneous strong Markov process.

(ii) Applying the Itô formula gives that for any $t > 0$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{s\in[0,t]} |\xi^\varepsilon(s)|^2\Big] \\
&\le \mathbb{E}|\xi(0)|^2 + \mathbb{E}\Big[\sup_{s\in[0,t]} \int_0^s (2[\xi^\varepsilon(u)]' h(x,\xi^\varepsilon(u)) + |\phi(x,\xi^\varepsilon(u))|^2) du\Big] \\
&+2\mathbb{E}\Big[\sup_{s\in[0,t]} \int_0^s [\xi^\varepsilon(u)]'\phi(x,\xi^\varepsilon(u)) d\tilde{w}_2(u)\Big].
\end{aligned} \tag{4.3}$$

By (3.6) and (3.8),

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{s\in[0,t]} \int_0^s (2[\xi^\varepsilon(u)]' h(x,\xi^\varepsilon(u)) + |\phi(x,\xi^\varepsilon(u))|^2) ds \\
&\le K\mathbb{E}\int_0^s [\|x\|^2 + |\xi^\varepsilon(u)|^2 + 1] du
\end{aligned} \tag{4.4}$$

and by the Burkholder-Davis-Gundy inequality, the Young inequality and (3.8),

$$\begin{aligned}
&2\mathbb{E}\Big[\sup_{s\in[0,t]} \int_0^s \xi^\varepsilon(u)]'\phi(x,\xi^\varepsilon(u)) d\tilde{w}_2(u)\Big] \\
&\le K\mathbb{E}\Big|\int_0^t |\xi^\varepsilon(u)]'\phi(x,\xi^\varepsilon(u))|^2 du\Big|^{\frac{1}{2}} \\
&\le \frac{1}{2}\mathbb{E}\Big[\sup_{s\in[0,t]} |\xi^\varepsilon(s)|^2\Big] + K\mathbb{E}\int_0^t |\phi(x,\xi^\varepsilon(u))|^2 du \\
&\le \frac{1}{2}\mathbb{E}\Big[\sup_{s\in[0,t]} |\xi^\varepsilon(s)|^2\Big] + K\mathbb{E}\int_0^s [\|x\|^2 + |\xi^\varepsilon(u)|^2 + 1] du.
\end{aligned} \tag{4.5}$$

Substituting (4.4) and (4.5) into (4.3) gives

$$\mathbb{E}\Big[\sup_{s\in[0,t]}|\xi^{\varepsilon}(s)|^2\Big] \le 2\mathbb{E}|\xi(0)|^2 + K\int_0^s \Big[\|x\|^2 + \mathbb{E}\Big(\sup_{v\in[0,u]}|\xi^{\varepsilon}(v)|^2\Big) + 1\Big]du.$$

The Gronwall inequality yields the desired assertion.

(iii) This result is from the existing result; see [2, Theorem 1.5, p. 5].

(iv) Applying the Itô formula gives

$$\begin{aligned}
&e^{(2\lambda_1-\lambda_2)t}|\xi^{x_1}(t)-\xi^{x_2}(t)|^2\\
&= (2\lambda_1-\lambda_2)\int_0^t e^{(2\lambda_1-\lambda_2)s}|\xi^{x_1}(s)-\xi^{x_2}(s)|^2 ds\\
&\quad + \int_0^t [e^{(2\lambda_1-\lambda_2)s}2\langle \xi^{x_1}(s)-\xi^{x_2}(s), h(x_1,\xi^{x_1}(s))-h(x_2,\xi^{x_2}(s))\rangle\\
&\quad + \|\phi(x_1,\xi^{x_1}(s))-\phi(x_2,\xi^{x_2}(s))\|^2]ds\\
&\quad +2\int_0^t e^{(2\lambda_1-\lambda_2)s}2\langle \xi^{x_1}(s)-\xi^{x_2}(s), \phi(x_1,\xi^{x_1}(s))-\phi(x_2,\xi^{x_2}(s))\rangle d\tilde{w}_1(s)\\
&\le (2L+\lambda_2)\int_0^t e^{-(2\lambda_1-\lambda_2)(t-s)}ds\|x_1-x_2\|^2\\
&\quad +2\int_0^t e^{(2\lambda_1-\lambda_2)s}2\langle \xi^{x_1}(s)-\xi^{x_2}(s), \phi(x_1,\xi^{x_1}(s))-\phi(x_2,\xi^{x_2}(s))\rangle d\tilde{w}_1(s).
\end{aligned}$$

By Assumption (A1), (2.2), and (A2), taking the expectation gives

$$\begin{aligned}
\mathbb{E}|\xi^{x_1}(t)-\xi^{x_2}(t)|^2 &\le (2L+\lambda_2)\int_0^t e^{-(2\lambda_1-\lambda_2)(t-s)}ds\|x_1-x_2\|^2\\
&\le \frac{2L+\lambda_2}{2\lambda_1-\lambda_2}\|x_1-x_2\|^2. \qquad (4.6)
\end{aligned}$$

This completes the proof. □

Remark 4.2. The conditions in Theorem 4.1 yield the exponential ergodicity of the invariant measure $\mu^x(\cdot)$. It can also be seen that (iv) implies that as $x_1 \to x_2$, $\xi^{x_1}(t)-\xi^{x_2}(t) \xrightarrow{\mathbb{P}} 0$.

Using the invariant measure μ^x, define

$$\bar{b}(x) = \int_{\mathbb{R}^m} b(x,\xi)\mu^x(d\xi) \quad \text{and} \quad \bar{\Sigma}(x) = \int_{\mathbb{R}^m} \psi(x,\xi)\psi'(x,\xi)\mu^x(d\xi)$$

Let us introduce the following assumption.

(A5) The following equation

$$dx(t) = \bar{b}(x_t)dt + \bar{\psi}(x_t)dB(t) \tag{4.7}$$

has a solution that is unique in the weak sense (i.e., uniqueness in the sense of in distribution) on $[0, T]$ for the same initial data $x_0 \in C([-\tau, 0]; \mathbb{R}^n)$ as Eq. (1.5a), where $B(t)$ is a standard Brownian motion, $\bar{\psi}(\cdot)\bar{\psi}'(\cdot) = \bar{\Sigma}(\cdot)$.

Remark 4.3. In general, the existence and uniqueness of the solution for Eq. (4.7) cannot be obtained by the conditions imposed on b and ψ. If $x^\varepsilon(\cdot) \to x(\cdot)$ as $\varepsilon \to 0$, by Remark 3.3, we know $\mathbb{E}[\sup_{0\le t\le T} |x(t)|^p] \le K_{p,T}$. Hence (4.7) has a global solution if $\bar{b}$ and $\bar{\psi}$ satisfy the local Lipschitz condition.

For any $V \in C_0^\infty(\mathbb{R}^d; \mathbb{R})$, applying the Itô formula to $V(x(t))$ for Eq. (4.7) yields that

$$M_V(t) := V(x(t)) - V(x(0)) - \int_0^t \mathcal{L}(x_s)V(x(s))ds = \int_0^t V_x(x(s))\bar{\psi}(x_s)dB(s) \tag{4.8}$$

is a martingale, where

$$\mathcal{L}(x_s)V(x(s)) = V_x(x(s))\bar{b}(x_s) + \frac{1}{2}\sum_{i,j=1}^n \bar{\psi}_i(x_s)\bar{\psi}_j(x_s)V_{x_i x_j}(x(s)).$$

It is well-known that the existence and uniqueness of the weak solution for Eq. (4.7) is equivalent to the existence and uniqueness of solution for the martingale problem (4.8).

5. Weak convergence and asymptotic approximation

This section shows that the sequence $\{x^\varepsilon(\cdot)\}_{\varepsilon\in(0,1]}$ converges weakly to a stochastic process that is the solution of an appropriate SFDE. In order to obtain the desired weak convergence, we need to prove tightness of $\{x^\varepsilon(\cdot)\}_{\varepsilon\in(0,1]}$ and $\{x^\varepsilon_\cdot\}_{\varepsilon\in(0,1]}$. In fact, for $C([-\tau, 0]; \mathbb{R}^n)$, sup norm is used, if $\{x^\varepsilon_\cdot\}_{\varepsilon\in(0,1]}$ is tight, $\{x^\varepsilon(\cdot)\}_{\varepsilon\in(0,1]}$ is naturally tight. Hence, we need only establish tightness of the segment process $x^\varepsilon_\cdot \in C([0, T]; C([-\tau, 0]; \mathbb{R}^n))$. We state the following lemma.

Lemma 5.1. *Let $X^\varepsilon_\cdot \in C([0, T]; C([-\tau, 0]; \mathbb{R}^n))$ be a sequence of $\mathcal{F}^\varepsilon_t$-measurable $C([-\tau, 0]; \mathbb{R}^n)$-valued processes satisfying that*

(i) *for any $\eta > 0$, there exists a constant M satisfying*

$$\sup_{\varepsilon\in(0,1]} \mathbb{P}\Big(\sup_{0\le t\le T} |X^\varepsilon_t(0)| > M\Big) \le \frac{\eta}{2}; \tag{5.1}$$

(ii) *for any $k \in \mathbb{N}$ and $\eta > 0$, there exists a constant δ_k satisfying*

$$\sup_{\varepsilon\in(0,1]} \mathbb{P}\Bigg(\sup_{t\in[0,T]} \max_{\substack{\theta_1,\theta_2\in[-\tau,0]\\ |\theta_1-\theta_2|\le\delta_k}} |X_t^\varepsilon(\theta_1) - X_t^\varepsilon(\theta_2)| > \frac{1}{k}\Bigg) \le \frac{\eta}{2^{T+k+2}}; \tag{5.2}$$

(iii) *for any $k \in \mathbb{N}$ and η, there exists a constant δ_k satisfying*

$$\sup_{\varepsilon\in(0,1]} \mathbb{P}\Bigg(\max_{\substack{t,s\in[0,T]\\ |t-s|\le\delta_k}} \|X_t^\varepsilon - X_s^\varepsilon\| > \frac{1}{k}\Bigg) \le \frac{\eta}{2^{T+k+2}}. \tag{5.3}$$

Then $X_\cdot^\varepsilon$ is tight in $C([0,T]; C([-\tau,0];\mathbb{R}^n))$.

Proof. Define

$$B_1^\varepsilon = \Big\{\omega : \sup_{0\le t\le T} |X_t^\varepsilon(\omega,0)| \le M\Big\},$$

$$B_{2,k}^\varepsilon = \Bigg\{\omega : \sup_{t\in[0,T]} \max_{\substack{\theta_1,\theta_2\in[-\tau,0]\\ |\theta_1-\theta_2|\le\delta_k}} |X_t^\varepsilon(\omega,\theta_1) - X_t^\varepsilon(\omega,\theta_2)| \le \frac{1}{k}\Bigg\},$$

$$B_{3,k}^\varepsilon = \Bigg\{\omega : \max_{\substack{t,s\in[0,T]\\ |t-s|\le\delta_k}} \|X_t^\varepsilon(\omega) - X_s^\varepsilon(\omega)\| \le \frac{1}{k}\Bigg\},$$

and let

$$B_2^\varepsilon = \bigcap_{k=1}^\infty B_{2,k}^\varepsilon, \quad B_3^\varepsilon = \bigcap_{k=1}^\infty B_{3,k}^\varepsilon \quad \text{and} \quad B^\varepsilon = B_1^\varepsilon \cap B_2^\varepsilon \cap B_3^\varepsilon.$$

By Lemma 2.1, $C([0,T]; C([-\tau,0];\mathbb{R}^n))$ is separable and complete. This shows that a single measure $\mathbb{P}$ is tightness [4, p.8, Theorem 1.3]. Note that for any $\varepsilon \in (0,1]$ and any $\eta > 0$, by conditions (5.1), there exists a $M > 0$ sufficiently large such that

$$\mathbb{P}(\overline{B_1^\varepsilon}) = \mathbb{P}\Big(\sup_{0\le t\le T} |X_t^\varepsilon(0)| > M\Big) \le \frac{\eta}{2}.$$

By (5.2),

$$\mathbb{P}(\overline{B_2^\varepsilon}) = \mathbb{P}\Big(\bigcup_{k=1}^\infty \overline{B_{2,k}^\varepsilon}\Big) \le \sum_{k=1}^\infty \mathbb{P}(\overline{B_{2,k}^\varepsilon}) \le \sum_{k=1}^\infty \frac{\eta}{2^{T+k+2}} \le \frac{\eta}{2^{T+2}}.$$

By (5.3),

$$\mathbb{P}(\overline{B_3^\varepsilon}) = \mathbb{P}\Big(\bigcup_{k=1}^\infty \overline{B_{3,k}^\varepsilon}\Big) \le \sum_{k=1}^\infty \mathbb{P}(\overline{B_{3,k}^\varepsilon}) \le \sum_{k=1}^\infty \frac{\eta}{2^{T+k+2}} \le \frac{\eta}{2^{T+2}}.$$

These three estimates imply that for any $T > 0$,

$$\begin{aligned}\mathbb{P}(B^\varepsilon) &= 1-\mathbb{P}(\overline{B^\varepsilon})\\ &\geq 1-\mathbb{P}(\overline{B_1^\varepsilon})-\mathbb{P}(\overline{B_2^\varepsilon})-\mathbb{P}(\overline{B_3^\varepsilon})\\ &> 1-\eta.\end{aligned}$$

Similarly, define

$$A_1^\varepsilon = \Big\{X_\cdot^\varepsilon \in C([0,T];C([-\tau,0];\mathbb{R}^n)) : \sup_{0\leq t\leq T}|X_t^\varepsilon(0)|\leq M\Big\},$$

$$A_{2,k}^\varepsilon = \Big\{X_\cdot^\varepsilon \in C([0,T];C([-\tau,0];\mathbb{R}^n)) : \sup_{t\in[0,T]}\max_{\substack{\theta_1,\theta_2\in[-\tau,0]\\ |\theta_1-\theta_2|\leq\delta_k}}|X_t^\varepsilon(\theta_1)-X_t^\varepsilon(\theta_2)|\leq \frac{1}{k}\Big\},$$

$$A_{3,k}^\varepsilon = \Big\{X_\cdot^\varepsilon \in C([0,T];C([-\tau,0];\mathbb{R}^n)) : \max_{\substack{t,s\in[0,T]\\ |t-s|\leq\delta_k}}\|X_t^\varepsilon-X_s^\varepsilon\|\leq \frac{1}{k}\Big\},$$

and let

$$A_2^\varepsilon = \bigcap_{k=1}^\infty A_{2,k}^\varepsilon \text{ and } A_3^\varepsilon = \bigcap_{k=1}^\infty A_{3,k}^\varepsilon.$$

According to [25, Theorem 47.1 (Ascoli's Theorem), p.290] or [29] with small modifications,

$$A^\varepsilon = A_1^\varepsilon \cap A_2^\varepsilon \cap A_3^\varepsilon$$

is relatively compact.

Let $\mathbb{P}_{X_\cdot^\varepsilon}(\cdot)$ be the probability measure induced by $X_\cdot^\varepsilon \in C([0,T];C([-\tau,0];\mathbb{R}^n))$. Then for any $i=1,2,3$,

$$\mathbb{P}_{X_\cdot^\varepsilon}(A_i^\varepsilon) = \mathbb{P}(B_i^\varepsilon).$$

It follows that $\mathbb{P}_{X_\cdot^\varepsilon}(A^\varepsilon) > 1-\eta$. This demonstrates that $X_\cdot^\varepsilon$ is tight. □

Following this lemma, we can examine the tightness of $\{x_\cdot^\varepsilon\}_{0<\varepsilon\leq 1}$.

Theorem 5.2. *Assuming* (A1)-(A4), $\{x_\cdot^\varepsilon\}$ *for* $0<\varepsilon\leq 1$ *is tight on* $C([0,T];C([-\tau,0];\mathbb{R}^n))$.

Proof. To prove the tightness, let us verify (5.1)–(5.3) for $\{x_\cdot^\varepsilon\}_{0<\varepsilon\leq 1}$. For any $\eta>0$, from (3.1) and the Chebyshev inequality, there exists $M > (2K_{p,T})^{1/p}/\eta^{1/p}$ such that

$$\mathbb{P}\Big(\sup_{0\leq t\leq T}|x_t^\varepsilon(0)|>M\Big) \leq \frac{\mathbb{E}\Big[\sup_{0\leq t\leq T}|x_t^\varepsilon(0)|^p\Big]}{M^p} = \frac{\mathbb{E}\Big[\sup_{0\leq t\leq T}|x^\varepsilon(t)|^p\Big]}{M^p} \leq \frac{K_{p,T}}{M^p} < \frac{\eta}{2}. \tag{5.4}$$

This shows that (5.1) holds.

Note that (3.28) also shows that

$$\mathbb{P}\left(\lim_{\delta\to 0}\sup_{\substack{t,s\in[-\tau,T]\\|t-s|\le\delta}}|x^{\varepsilon}(t)-x^{\varepsilon}(s)|=0\right)=1,$$

which implies that for any $k>0$,

$$\lim_{\delta\to 0}\mathbb{P}\left(\sup_{\substack{t,s\in[-\tau,T]\\|t-s|<\delta}}|x^{\varepsilon}(t)-x^{\varepsilon}(s)|>\frac{1}{k}\right)=0.$$

This further leads to that for any $\eta>0$, there exists a δ_k such that

$$\mathbb{P}\left(\sup_{\substack{t,s\in[-\tau,T]\\|t-s|<\delta_k}}|x^{\varepsilon}(t)-x^{\varepsilon}(s)|>\frac{1}{k}\right)<\frac{\eta}{2^{T+k+2}}.$$

This, together with $x_t^{\varepsilon}(\theta)=x^{\varepsilon}(t+\theta)$ for any $\theta\in[-\tau,0]$, yields

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,T]}\max_{\substack{\theta_1,\theta_2\in[-\tau,0]\\|\theta_1-\theta_2|\le\delta_k}}|x_t^{\varepsilon}(\theta_1)-x_t^{\varepsilon}(\theta_2)|>\frac{1}{k}\right)\\
&=\mathbb{P}\left(\sup_{t\in[0,T]}\max_{\substack{\theta_1,\theta_2\in[-\tau,0]\\|\theta_1-\theta_2|\le\delta_k}}|x^{\varepsilon}(t+\theta_1)-x^{\varepsilon}(t+\theta_2)|>\frac{1}{k}\right)\\
&\le\mathbb{P}\left(\sup_{\substack{t,s\in[-\tau,T]\\|t-s|\le\delta_k}}|x^{\varepsilon}(t)-x^{\varepsilon}(s)|>\frac{1}{k}\right)<\frac{\eta}{2^{T+k+2}},
\end{aligned}$$

which indicates that (5.2) holds.

Eq. (3.29) shows that

$$\mathbb{P}\left(\lim_{\delta\to 0}\sup_{\substack{t,s\in[0,T],\\0<|t-s|<\delta}}\|x_t^{\varepsilon}-x_s^{\varepsilon}\|=0\right)=1,$$

which implies that for any $k>0$,

$$\lim_{\delta\to 0}\mathbb{P}\left(\sup_{\substack{t,s\in[0,T],\\0<|t-s|<\delta}}\|x_t^{\varepsilon}-x_s^{\varepsilon}\|>\frac{1}{k}\right)=0.$$

This further implies that for any $\eta>0$, there exists $\delta_k>0$ such that

$$\mathbb{P}\left(\max_{\substack{t,s\in[0,T]\\|t-s|\le\delta_k}}\|x_t^{\varepsilon}-x_s^{\varepsilon}\|>\frac{1}{k}\right)\le\frac{\eta}{2^{T+k+2}}.$$

Hence (5.3) follows. Then the tightness of $x^\varepsilon_\cdot$ in $C([0,T];C([-\tau,0];\mathbb{R}^n))$ follows from Lemma 5.1 as desired. □

To proceed, let us state the main theorem of this paper.

Theorem 5.3. *Under Assumptions* (A1)-(A4)*, $x^\varepsilon(\cdot)$ converges weakly to $x(\cdot)$ that is the solution of* (4.7).

To prove this theorem, let us introduce the following the truncated SFDE. For each $N>0$, define $b^N(\varphi,\xi)=b(\varphi,\xi)q^N(\varphi(0))$, $\psi^N(\varphi,\xi)=\psi(\varphi,\xi)q^N(\varphi(0))$, and

$$q^N(y)=\begin{cases}1, & \text{when } \ y\in S_N,\\ 0, & \text{when } \ y\in\mathbb{R}^n-S_{N+1},\\ \text{smooth}, & \text{otherwise},\end{cases}$$

$S_N=\{y:|y|\le N\}$. For sufficiently large such that $|x(0)|\le N$, let us consider

$$dx^{\varepsilon,N}(t)=b^N(x_t^{\varepsilon,N},\xi^{\varepsilon,N}(t))dt+\psi^N(x_t^{\varepsilon,N},\xi^{\varepsilon,N}(t))dw_1(t), \tag{5.5}$$

where $x_t^{\varepsilon,N}=\{x^{\varepsilon,N}(t+\theta):-\tau\le\theta\le 0\}$ and $\xi^{\varepsilon,N}(t)$ is the solution of Eq. (1.5b) when x_t^ε is replaced by $x_t^{\varepsilon,N}$. It can be observed that $x^{\varepsilon,N}(t)=x^\varepsilon(t)$ up until the first exit from S_N. Then $x^{\varepsilon,N}(t)$ is said to be an N-truncation of $x^\varepsilon(t)$. According to the definition of x_t, it is easily seen that $x_t^{\varepsilon,N}\in C([0,T];S_{N+1})$ if $x^{\varepsilon,N}(t)\in S_{N+1}$ since $\|x_t^{\varepsilon,N}\|=\sup_{-\tau\le\theta\le 0}|x^{\varepsilon,N}(t+\theta)|\le N+1$. Let $\mathcal{F}_t^{\varepsilon,N}=\sigma\{\xi^{\varepsilon,N}(s),x^{\varepsilon,N}(s):s\le t\}$. We can likewise define the corresponding $\overline{\mathcal{M}}^{\varepsilon,N}$ and $\hat{\mathcal{L}}^{\varepsilon,N}$ accordingly.

Remark 5.4. Note that $b^N(\cdot)$ and $\psi^N(\cdot)$ have better properties. Because $\|x_t^{\varepsilon,N}\|\le N+1$, in (A1), $b^N(\cdot)$ and $\psi^N(\cdot)$ are, in fact, globally Lipschitz continuous. The $x^{\varepsilon,N}(t)$ and the corresponding segment process $x_t^{\varepsilon,N}$ inherit all properties of $x^\varepsilon(t)$ and x_t^ε, for example, Theorem 3.1 holds and $x_t^{\varepsilon,N}\in C([0,T];C([-\tau,0];S_{N+1}))$ is tight and continuous.

Proof of Theorem 5.3. Let us first prove $x^{\varepsilon,N}(\cdot)$ converges weakly to $x^N(\cdot)$, where $x^N(\cdot)$ is the solution of Eq. (4.7) with truncated coefficients

$$dx^N(t)=\bar{b}^N(x_t^N)dt+\bar{\psi}^N(x_t^N)dB(t) \tag{5.6}$$

and $\bar{b}^N(\varphi)=\bar{b}(\varphi)q^N(\varphi(0))$ and $\bar{\psi}^N(\varphi)=\bar{\psi}(\varphi)q^N(\varphi(0))$. Let us give the definitions of $\bar{\Sigma}^N(\cdot)=\bar{\psi}^N(\cdot)[\bar{\psi}^N(\cdot)]'$ and

$$\mathcal{L}^N(x_s^N)V(x^N(s))=V_x(x^N(s))b^N(x_s^N)+\frac{1}{2}\sum_{i,j=1}^n\psi_i^N(x_s^N)\psi_j^N(x_s^N)V_{x_ix_j}(x^N(s)).$$

Since $\{x_\cdot^{\varepsilon,N}\}$ is tight, by the Prohorov theorem, it is sequentially compact. There exist $x^N\in C([0,T];C([-\tau,0];S_{N+1}))$ and a subsequence $\{\varepsilon_n\}_{n\ge 1}$ with $\varepsilon_n\to 0$ as $n\to\infty$ such that $x^{\varepsilon_n,N}\Rightarrow x^N$. Since the distribution of $x^{\varepsilon_n,N}(t)$ is the marginal distribution of $x_t^{\varepsilon_n,N}$,

$x^{\varepsilon_n,N}(t) = x_t^{\varepsilon_n,N}(0) \Rightarrow x_t^N(0) = x^N(t)$. By the Skorohod representation theorem, we may assume that $x_\cdot^{\varepsilon_n,N}$ converges to $x_\cdot^N$ in the sense of w.p.1. Because of $x_t^{\varepsilon,N} \in C([-\tau,0]; S_{N+1})$ with the uniform norm $\|\cdot\|$, $x^{\varepsilon_n,N}(\cdot)$ also converges to $x^N(\cdot)$ with probability 1.

We proceed to characterize the limit process $x^N(\cdot)$ by use of the martingale problem formulation. For any $V \in C_0^\infty(\mathbb{R}^n;\mathbb{R})$, applying the Itô formula to $V(x^{\varepsilon_n,N}(t))$ for Eq. (1.5a) yields

$$\begin{aligned} M_V^{\varepsilon_n,N}(t) &:= V(x^{\varepsilon_n,N}(t)) - V(x(0)) - \int_0^t \mathcal{L}^{\varepsilon_n,N}(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))V(x^{\varepsilon_n,N}(s))ds \\ &= \int_0^t V_x(x^{\varepsilon_n,N}(s))\psi^N(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))dw_2(s) \end{aligned} \tag{5.7}$$

is a martingale, where

$$\begin{aligned} &\mathcal{L}^{\varepsilon_n,N}(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))V(x^{\varepsilon_n,N}(s)) \\ &\quad = V_x(x^{\varepsilon_n,N}(s))b^N(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s)) \\ &\qquad + \frac{1}{2}\sum_{i,j=1}^n \psi_i^N(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))\psi_j^N(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))V_{x_i x_j}(x^{\varepsilon_n,N}(s)). \end{aligned}$$

This is equivalent to that

$$\begin{aligned} &\mathbb{E}\Big\{h(x^{\varepsilon_n,N}(s_j), j \le k)\Big[V(x^{\varepsilon_n,N}(t)) - V(x^{\varepsilon_n,N}(s)) \\ &\quad - \int_s^t \mathcal{L}^{\varepsilon_n,N}(x_u^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(u))V(x^{\varepsilon_n,N}(u))du\Big]\Big\} = 0 \end{aligned} \tag{5.8}$$

for arbitrary k, t and s with $s_1 < s_2 < \cdots < s_k < s < t$, and any bounded and continuous function $h(\cdot)$. Since $x^{\varepsilon_n,N}(\cdot)$ converges to $x^N(\cdot)$ w.p.1 as $n \to \infty$, by the Lebesgue dominated convergence theorem,

$$\mathbb{E}[h(x^{\varepsilon_n,N}(s_j), j \le k)V(x^{\varepsilon_n,N}(s))] \to \mathbb{E}[h(x^N(s_j), j \le k)V(x^N(s))] \tag{5.9}$$

for all $0 < s \le t$. Then let us consider convergence of

$$\mathbb{E}[h(x^{\varepsilon_n,N}(u_j), j \le k)\int_s^t \mathcal{L}^{\varepsilon_n,N}(x_u^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(u))V(x^{\varepsilon_n,N}(u))du].$$

Choose Δ sufficiently small. Then

$$
\begin{aligned}
&\mathbb{E}\Big[h(x^{\varepsilon_n,N}(s_j), j\le k)\int_s^t \mathcal{L}^{\varepsilon_n,N}(x_u^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u))V(x^{\varepsilon_n,N}(u))du\Big]\\
&\quad=\mathbb{E}\Big[h(x^{\varepsilon_n,N}(s_j), j\le k)\\
&\qquad\times\sum_{l\Delta=s}^{t}\int_{l\Delta}^{(l+1)\Delta}\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))V(x^{\varepsilon_n,N}(l\Delta))du\Big]\\
&\quad+\mathbb{E}\Big[h(x^{\varepsilon_n,N}(s_j), j\le k)\sum_{l\Delta=s}^{t}\int_{l\Delta}^{(l+1)\Delta}\mathcal{L}^{\varepsilon_n,N}(x_u^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u))V(x^{\varepsilon_n,N}(u))\\
&\qquad\qquad-\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))V(x^{\varepsilon_n,N}(l\Delta))du\Big]\\
&\quad=:I_1^{\varepsilon_n,N}+I_2^{\varepsilon_n,N},
\end{aligned}
\tag{5.10}
$$

where

$$
\begin{aligned}
\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)=\xi^{\varepsilon_n,N}(l\Delta)&+\frac{1}{\varepsilon_n}\int_{l\Delta}^{u}h(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(v))dv\\
&+\frac{1}{\sqrt{\varepsilon_n}}\int_{l\Delta}^{u}\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(v))dw_2(v).
\end{aligned}
\tag{5.11}
$$

Note that $x_{l\Delta}^{\varepsilon_n,N}$ in (5.11) is at the beginning of the interval so it is a fixed constant. Making change of variable u to $\varepsilon_n u$ gives

$$
\begin{aligned}
\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(\varepsilon_n u)=\xi^{\varepsilon_n,N}(l\Delta)&+\frac{1}{\varepsilon_n}\int_{l\Delta}^{\varepsilon_n u}h(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(v))dv\\
&+\frac{1}{\sqrt{\varepsilon_n}}\int_{l\Delta}^{\varepsilon_n u}\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(v))dw_2(v)\\
=\xi^{\varepsilon_n,N}(l\Delta)&+\int_{\frac{l\Delta}{\varepsilon_n}}^{u}h(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(\varepsilon_n v))dv\\
&+\int_{\frac{l\Delta}{\varepsilon_n}}^{u}\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(\varepsilon_n v))d\tilde{w}_2(v).
\end{aligned}
\tag{5.12}
$$

Compared with Eq. (4.2), it is obvious that for any $u\in[l\Delta,(l+1)\Delta]$,

$$\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)=\xi^{x_{l\Delta}^{\varepsilon_n,N}}\Big(\frac{u}{\varepsilon_n}\Big), \tag{5.13}$$

where $\xi^{x_{l\Delta}^{\varepsilon_n,N}}(\cdot)$ is solution the fixed-x equation (4.2) with $x=x_{l\Delta}^{\varepsilon_n,N}$. According to expression of $\mathcal{L}^{\varepsilon_n,N}(\cdot)V(\cdot)$,

$$\begin{aligned}
&\int_{l\Delta}^{(l+1)\Delta}\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))V(x^{\varepsilon_n,N}(l\Delta))du\\
&=\int_{l\Delta}^{(l+1)\Delta}\Big[V_x(x^{\varepsilon_n,N}(l\Delta))b^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))\\
&\quad+\frac{1}{2}\sum_{i,j=1}^{n}\psi_i^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))\psi_j^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))V_{x_ix_j}(x^{\varepsilon_n,N}(l\Delta))\Big]du\\
&=V_x(x^{\varepsilon_n,N}(l\Delta))I_{11}^{\varepsilon_n,N}+\frac{1}{2}\sum_{i,j=1}^{n}I_{12ij}^{\varepsilon_n,N}V_{x_ix_j}(x^{\varepsilon_n,N}(l\Delta)),
\end{aligned} \tag{5.14}$$

where

$$I_{11}^{\varepsilon_n,N}=\int_{l\Delta}^{(l+1)\Delta}b^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))du$$

and

$$I_{12ij}^{\varepsilon_n,N}=\int_{l\Delta}^{(l+1)\Delta}\psi_i^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))\psi_j^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u))du.$$

Making change of variable u to $\varepsilon_n u$ and using (5.13) yield that

$$\begin{aligned}
I_{11}^{\varepsilon_n,N}&=\varepsilon_n\int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}}b^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))du\\
&=\varepsilon_n\int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}}b^N(x_{l\Delta}^{N},\xi^{x_{l\Delta}^{N}}(u))du\\
&\quad+\varepsilon_n\int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}}[b^N(x_{l\Delta}^{\varepsilon_n,N},\xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))-b^N(x_{l\Delta}^{N},\xi^{x_{l\Delta}^{N}}(u))]du
\end{aligned}$$

$$=: I_{11,1}^{\varepsilon_n,N} + I_{11,2}^{\varepsilon_n,N}.$$

According to ergodicity of $\xi^x(\cdot)$ and the definition of $\bar{b}^N$, as $n \to \infty$,

$$I_{11,1}^{\varepsilon_n,N} = \Delta \frac{1}{\Delta/\varepsilon_n} \int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}} b^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))du \to \bar{b}^N(x_{l\Delta}^N)\Delta, \quad \text{a.s.} \tag{5.15}$$

Recall that $x_{l\Delta}^{\varepsilon_n,N} \to x_{l\Delta}^N$ with probability 1 as $n \to \infty$. Applying (iv) of Theorem 4.1 gives that

$$\xi^{x_{l\Delta}^{\varepsilon_n,N}}(u) - \xi^{x_{l\Delta}^N}(u) \xrightarrow{\mathbb{P}} 0.$$

By the integral mean value theorem and the local Lipschitz condition, since b^N satisfies the global Lipschitz condition, there exists a $u^* \in \mathbb{R}$ such that

$$\begin{aligned}|I_{11,2}^{\varepsilon_n,N}| &= |b^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u^*)) - b^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u^*))|\Delta \\ &\le K_N\Delta(\|x_{l\Delta}^{\varepsilon_n,N} - x_{l\Delta}^N\| + |\xi^{x_{l\Delta}^{\varepsilon_n,N}}(u^*) - \xi^{x_{l\Delta}^N}(u^*))|) \xrightarrow{\mathbb{P}} 0.\end{aligned} \tag{5.16}$$

Note that $x_{l\Delta}^{\varepsilon_n,N} \to x_{l\Delta}^N$ with probability 1. Combining (5.15) with (5.16) gives

$$V_x(x^{\varepsilon_n,N}(l\Delta))I_{11}^{\varepsilon_n,N} - V_x(x^N(l\Delta))\bar{b}^N(x_{l\Delta}^N)\Delta \xrightarrow{\mathbb{P}} 0, \quad \text{a.s..} \tag{5.17}$$

Let us estimate $I_{12ij}^{\varepsilon_n,N}$. By making change of variable u to εu gives that

$$\begin{aligned}I_{12ij}^{\varepsilon_n,N} &= \varepsilon_n \int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}} \psi_i^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))\psi_j^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))du \\ &= \varepsilon_n \int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}} \psi_i^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))\psi_j^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))du \\ &\quad + \varepsilon_n \int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}} [\psi_i^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))\psi_j^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u)) \\ &\quad - \psi_i^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))\psi_j^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))]du \\ &=: I_{12ij,1}^{\varepsilon_n,N} + I_{12ij,2}^{\varepsilon_n,N}.\end{aligned} \tag{5.18}$$

Similar to (5.15), the ergodicity of $\xi^x(\cdot)$ and the definition of $\bar{\Sigma}^N$ yield that as $n \to \infty$,

$$I_{12ij,1}^{\varepsilon_n,N} = \Delta \frac{1}{\Delta/\varepsilon_n} \int_{\frac{l\Delta}{\varepsilon_n}}^{\frac{(l+1)\Delta}{\varepsilon_n}} \psi_i^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))\psi_j^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))du \to \bar{\Sigma}_{ij}^N(x_{l\Delta}^N)\Delta, \text{ a.s.} \tag{5.19}$$

Note that for any $\varphi \in C([-\tau, 0]; S_{N+1})$, $\psi_i^N(\varphi, \xi)$ is bounded and satisfies the global Lipschitz condition. This implies that for any u,

$$\begin{aligned}
&|\psi_i^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u))\psi_j^N(x_{l\Delta}^{\varepsilon_n,N}, \xi^{x_{l\Delta}^{\varepsilon_n,N}}(u)) - \psi_i^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))\psi_j^N(x_{l\Delta}^N, \xi^{x_{l\Delta}^N}(u))| \\
&\le K_N(|x_{l\Delta}^{\varepsilon_n,N} - x_{l\Delta}^N| + |\xi^{x_{l\Delta}^{\varepsilon_n,N}}(u) - \xi^{x_{l\Delta}^N}(u))|).
\end{aligned}$$

Similar to the estimate of $I_{11,2}^{\varepsilon_n,N}$ in (5.16), as $n \to \infty$,

$$I_{12ij,2}^{\varepsilon_n,N} \xrightarrow{\mathbb{P}} 0.$$

This, together with (5.19) and $x^{\varepsilon_n,N}(l\Delta) \to x^N(l\Delta)$ with probability 1, gives

$$\frac{1}{2}\sum_{i,j=1}^{n} I_{12ij}^{\varepsilon_n,N} V_{x_i x_j}(x^{\varepsilon_n,N}(l\Delta)) - \frac{1}{2}\sum_{i,j=1}^{n} \bar{\Sigma}_{ij}^N(x_{l\Delta}^N) V_{x_i x_j}(x^N(l\Delta))\Delta \xrightarrow{\mathbb{P}} 0. \tag{5.20}$$

By the Lebesgue dominated convergence theorem, (5.17) and (5.20) give

$$I_1^{\varepsilon_n,N} \to \mathbb{E}\Big[h(x^N(s_j), j \le k) \sum_{l\Delta=s}^{t} \mathcal{L}^N(x_{l\Delta}^N) V(x^N(l\Delta))\Delta\Big]. \tag{5.21}$$

Now let us consider $I_2^{\varepsilon_n,N}$. By the integral mean value theorem, there exists $u^* \in [l\Delta, (l+1)\Delta]$ such that

$$\begin{aligned}
I_2^{\varepsilon_n,N} = \mathbb{E}\Big\{h(x^{\varepsilon_n,N}(s_j), j \le k) \sum_{l\Delta=s}^{t} \Big[&\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*)) \\
&-\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N}, \xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))\Big]\Delta\Big\}.
\end{aligned} \tag{5.22}$$

Note that for $u \in [l\Delta, (l+1)\Delta]$,

$$\begin{aligned}
\xi^{\varepsilon_n,N}(u) = \xi^{\varepsilon_n,N}(l\Delta) &+ \frac{1}{\varepsilon_n}\int_{l\Delta}^{u} h(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))ds \\
&+ \frac{1}{\sqrt{\varepsilon_n}}\int_{l\Delta}^{u} \phi(x_s^{\varepsilon_n,N}, \xi^{\varepsilon_n,N}(s))dw_1(s),
\end{aligned}$$

and that $\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)$ satisfies Eq. (5.11). It follows that

$$
\begin{aligned}
&\xi^{\varepsilon_n,N}(u)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)\\
&=\frac{1}{\varepsilon_n}\int_{l\Delta}^{u}[h(x_s^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(s))-h(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s))]ds\\
&\quad+\frac{1}{\sqrt{\varepsilon_n}}\int_{l\Delta}^{u}[\phi(x_s^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(s))-\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s))]dw_1(s).
\end{aligned}
$$

Applying the Itô formula gives

$$
\begin{aligned}
&e^{\frac{2\lambda_1-\lambda_2}{\varepsilon_n}u}|\xi^{\varepsilon_n,N}(u)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)|^2\\
&=(2\lambda_1-\lambda_2)\frac{1}{\varepsilon_n}\int_{l\Delta}^{u}e^{\frac{2\lambda_1-\lambda_2}{\varepsilon}s}|\xi^{\varepsilon_n,N}(s)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s)|^2ds\\
&\quad+\frac{1}{\varepsilon_n}\int_{l\Delta}^{u}e^{\frac{2\lambda_1-\lambda_2}{\varepsilon_n}s}[2\langle\xi^{\varepsilon_n,N}(s)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s),\\
&\qquad\qquad h(x_s^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(s))-h(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s))\rangle\\
&\qquad+\|\phi(x_s^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(s))-\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s))\|^2]ds\\
&\quad+\frac{2}{\sqrt{\varepsilon_n}}\int_{l\Delta}^{u}e^{\frac{2\lambda_1-\lambda_2}{\varepsilon_n}s}2\langle\xi^{\varepsilon_n,N}(s)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s),\phi(x_s^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(s))\\
&\qquad-\phi(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(s))\rangle dw_1(s).
\end{aligned}
$$

By Assumption (A1) and (A2), taking the expectation on both sides gives

$$
\begin{aligned}
&\mathbb{E}|\xi^{\varepsilon_n,N}(u)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u)|^2\\
&\le(2L+\lambda_2)\frac{1}{\varepsilon_n}\int_{l\Delta}^{u}e^{-\frac{2\lambda_1-\lambda_2}{\varepsilon_n}(u-s)}\mathbb{E}\|x_s^{\varepsilon_n,N}-x_{l\Delta}^{\varepsilon_n,N}\|^2ds\\
&\le\frac{2L+\lambda_2}{2\lambda_1-\lambda_2}\Big[\sup_{s\in[l\Delta,(l+1)\Delta]}\mathbb{E}\|x_s^{\varepsilon_n,N}-x_{l\Delta}^{\varepsilon_n,N}\|^2\Big].
\end{aligned}
\tag{5.23}
$$

Let $\zeta_\Delta=\sup_{s\in[l\Delta,(l+1)\Delta]}\mathbb{E}\|x_s^{\varepsilon_n,N}-x_{l\Delta}^{\varepsilon_n,N}\|^2$. (3.26) shows $\zeta_\Delta\to 0$ as $\Delta\to 0$. Note that $V\in C_0^\infty(\mathbb{R}^d;\mathbb{R})$. This, combined with (A1), implies that $\mathcal{L}^{\varepsilon_n,N}(\cdot,\cdot)V(\cdot)$ satisfies the global Lipschitz condition, that is, there exists constant $K_N>0$ such that

$$
\begin{aligned}
&|\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*))\\
&\qquad-\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))|^2
\end{aligned}
$$

$$\le K_N[\|x_{u^*}^{\varepsilon_n,N}-x_{l\Delta}^{\varepsilon_n,N}\|^2+|x^{\varepsilon_n,N}(u^*)-x^{\varepsilon_n,N}(l\Delta)|^2 \\ +|\xi^{\varepsilon_n,N}(u^*)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*)|^2]. \tag{5.24}$$

By (3.3), (3.28), and (5.23), taking expectation yields

$$\begin{aligned}
&\mathbb{E}|\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*)) \\
&\qquad -\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))|^2 \\
&\le K_N[\mathbb{E}\|x_{u^*}^{\varepsilon_n,N}-x_{l\Delta}^{\varepsilon_n,N}\|^2+\mathbb{E}|x^{\varepsilon_n,N}(u^*)-x^{\varepsilon_n,N}(l\Delta)|^2 \\
&\qquad +\mathbb{E}|\xi^{\varepsilon_n,N}(u^*)-\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*)|^2] \\
&\le K_N\zeta_\Delta+K_{p,N,T}\Delta+\frac{\kappa^2(2L+\lambda_2)}{2\lambda_1-\lambda_2}\Delta^{2(\gamma\wedge\gamma_0)} \\
&\le K_N\zeta_\Delta+K_{p,N,T}\Delta^{2(\gamma\wedge\gamma_0)\wedge 1}.
\end{aligned}$$

This implies that

$$\begin{aligned}
&\mathbb{E}\Big|\sum_{l\Delta=s}^{t}\Big[\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*)) \\
&\qquad -\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))\Big]\Delta\Big| \\
&\le \sum_{l\Delta=s}^{t}\mathbb{E}\Big|\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*)) \\
&\qquad -\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))\Big|\Delta \\
&\le K_N\zeta_\Delta+K_{p,N,T}(t-s)\Delta^{(\gamma\wedge\gamma_0)\wedge 1/2}.
\end{aligned}$$

Noting that h is bounded, we therefore obtain

$$\begin{aligned}
I_2^{\varepsilon_n,N}&=\mathbb{E}\Big[h(x^{\varepsilon_n,N}(s_j),\, j\le k)\sum_{l\Delta=s}^{t}\Big[\mathcal{L}^{\varepsilon_n,N}(x_{u^*}^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u^*))V(x^{\varepsilon_n,N}(u^*)) \\
&\qquad -\mathcal{L}^{\varepsilon_n,N}(x_{l\Delta}^{\varepsilon_n,N},\xi^{\varepsilon_n,N,x_{l\Delta}^{\varepsilon_n,N}}(u^*))V(x^{\varepsilon_n,N}(l\Delta))\Big]\Delta \\
&=K_N\zeta_\Delta+K_{p,N,T}(t-s)\Delta^{(\gamma\wedge\gamma_0)\wedge 1/2}.
\end{aligned}$$

Substituting $I_1^{\varepsilon_n,N}$ and $I_2^{\varepsilon_n,N}$ into (5.10) yields

$$\mathbb{E}\Big[h(x^{\varepsilon_n,N}(s_j),\, j\le k)\int_s^t \mathcal{L}^{\varepsilon_n,N}(x_u^{\varepsilon_n,N},\xi^{\varepsilon_n,N}(u))V(x^{\varepsilon_n,N}(u))du\Big]$$

$$\to \mathbb{E}\Big[h(x^N(s_j), j \le k) \sum_{l\Delta=s}^{t} \mathcal{L}^N(x^N_{l\Delta}) V(x^N(l\Delta))\Delta\Big] + K_N \zeta_\Delta + O(\Delta^{(\gamma\wedge\gamma_0)\wedge 1/2}),$$

which, together with (5.9) gives

$$\begin{aligned} 0 = \mathbb{E}\Big\{h(x^{\varepsilon_n,N}(s_j), j \le k)\Big[V(x^{\varepsilon_n,N}(t)) - V(x^{\varepsilon_n,N}(s)) \\ - \int_s^t \mathcal{L}^{\varepsilon_n,N}(x^{\varepsilon_n,N}_u, \xi^{\varepsilon_n,N}(u)) V(x^{\varepsilon_n,N}(u))du\Big]\Big\} \\ \to \mathbb{E}\Big\{h(x^N(s_j), j \le k)\Big[V(x^N(t)) - V(x^N(s)) - \sum_{l\Delta=s}^{t} \mathcal{L}^N(x^N_{l\Delta}) V(x^N(l\Delta))\Delta\Big]\Big]\Big\} \\ + K_N\zeta_\Delta + O(\Delta^{(\gamma\wedge\gamma_0)\wedge 1/2}). \end{aligned}$$

Letting $\Delta \to 0$ gives

$$\mathbb{E}\Big\{h(x^N(s_j), j \le k)\Big[V(x^N(t)) - V(x^N(s)) - \int_s^t \mathcal{L}^N(x^N_u) V(x^N(u))du\Big]\Big\} = 0.$$

This shows that $x^{\varepsilon_n,N}(\cdot)$ converges weakly to $x^N(\cdot)$, where $x^N(\cdot)$ solves the martingale problem with operator $\mathcal{L}^N$. This also shows that $x^N(\cdot)$ is the weak solution of Eq. (5.6).

Next, we move from the weak convergence of the truncated process to that of untruncated processes. The argument is similar to that of [19, p.46]. For any continuous initial value $x_0 \in C([-\tau, 0]; \mathbb{R}^n)$ independent of ε, 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 $C([0, T]; \mathbb{R}^n)$. By (A5), 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 $\mathbb{P}(\cdot)$ determined by Eq. (4.7) agrees with $\mathbb{P}^N(\cdot)$ determined by Eq. (5.6) on all Borel sets of the set of paths in $C([0, T]; S_N)$ 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 $x^{\varepsilon_n,N}(\cdot)$ implies that $x^{\varepsilon_n}(\cdot) \Rightarrow x(\cdot)$. Moreover, the uniqueness implies that the limit does not depend on the chosen subsequences. This completes this proof. □

6. SIDEs and SDDEs with two-time scales

As a class of special SFDEs, SIDEs arise widely in biology, ecology, medicine and physics (see [1,17,26,35]). Let us consider the following two-time-scale SIDE:

$$dx^\varepsilon(t) = B\Big(\int_{-\tau}^{0} x^\varepsilon(t+\theta)\mu(d\theta), \xi^\varepsilon(t)\Big)dt + \Psi\Big(\int_{-\tau}^{0} x^\varepsilon(t+\theta)\mu(d\theta), \xi^\varepsilon(t)\Big)dw_1(t), \tag{6.1a}$$

$$d\xi^\varepsilon(t) = \frac{1}{\varepsilon}H\Big(\int_{-\tau}^{0} x^\varepsilon(t+\theta)\mu(d\theta), \xi^\varepsilon(t)\Big)dt + \frac{1}{\sqrt{\varepsilon}}\Phi\Big(\int_{-\tau}^{0} x^\varepsilon(t+\theta)\mu(d\theta), \xi^\varepsilon(t)\Big)dw_2(t), \tag{6.1b}$$

with initial data $\xi(0) \in \mathbb{R}^m$ and $x_0 \in C([-\tau, 0]; \mathbb{R}^n)$, where μ is a probability measure on $[-\tau, 0]$, $B : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n\times l_2}$, $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $\Phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m\times l_1}$. Let us impose the following assumptions on these coefficients.

(Â1) (Lipschitz condition) For any integer R, there exists positive constant L_R such that for any $X_1, X_2 \in \mathbb{R}^n$, $\xi_1, \xi_2 \in \mathbb{R}^m$ with $|X_1| \vee |X_2| \vee |\xi_1| \vee |\xi_2| \le R$,

$$|H(X_1, \xi_1) - H(X_2, \xi_2)|^2 \le L_R(|X_1 - X_2|^2 + |\xi_1 - \xi_2|^2), \tag{6.2}$$

and

$$|B(X_1, \xi_1) - B(X_2, \xi_2)|^2 \vee |\Psi(X_1, \xi_1) - \Psi(X_2, \xi_2)|^2 \le L_R|X_1 - X_2|^2 + L|\xi_1 - \xi_2|^2, \tag{6.3}$$

where L is some constant. In (6.3), $\xi_1, \xi_2 \in \mathbb{R}^m$ are arbitrary.

(Â2) (Dissipative condition) For any $X_1, X_2 \in \mathbb{R}^n$, there exist λ_1, λ_2 and L such that for any ξ_1, $\xi_2 \in \mathbb{R}^m$,

$$\langle \xi_1 - \xi_2, H(X_1, \xi_1) - H(X_2, \xi_2)\rangle \le -\lambda_1|\xi_1 - \xi_2|^2 + L|X_1 - X_2|^2$$

and

$$|\Phi(X_1, \xi_1) - \Phi(\varphi_2, \xi_2)|^2 \le \lambda_2(|\xi_1 - \xi_2|^2 + |X_1 - X_2|^2).$$

(Â3) (Linear growth condition) There exists a constant $L > 0$ such that

$$|B(X, 0)|^2 \vee |\Psi(X, 0)|^2 \vee |H(X, 0)|^2 \le L(1 + |X|^2), \tag{6.4}$$

for any $X \in \mathbb{R}^n$.

Let us define

$$b(\varphi, \xi) = B\Big(\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta), \xi\Big), \quad \psi(\varphi, \xi) = \Psi\Big(\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta), \xi\Big),$$

$$h(\varphi, \xi) = H\Big(\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta), \xi\Big), \quad \phi(\varphi, \xi) = \Phi\Big(\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta), \xi\Big).$$

Note that $\|\varphi\| \le R$ for any $\varphi \in C([-\tau, 0]; \mathbb{R}^n)$ implies $|\varphi(\theta)| \le R$ for any $\theta \in [-\tau, 0]$, and

$$\Big|\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta)\Big| \le \int_{-\tau}^{0} \|\varphi\|\mu(d\theta) = \|\varphi\|.$$

According to (Â1), for any $\varphi_1, \varphi_2 \in C([-\tau, 0]; \mathbb{R}^n)$ and $\xi_1, \xi_2 \in \mathbb{R}^m$ with $\|\varphi_1\| \vee \|\varphi_2\| \vee |\xi_1| \vee |\xi_2| \le R$,

$$
\begin{aligned}
|h(\varphi_1,\xi_1)-h(\varphi_2,\xi_2)|^2 &= \Big|H\Big(\int_{-\tau}^{0}\varphi_1(\theta)\mu(d\theta),\xi_1\Big)-H\Big(\int_{-\tau}^{0}\varphi_2(\theta)\mu(d\theta),\xi_2\Big)\Big|^2\\
&\le L_R\Big(\Big|\int_{-\tau}^{0}\varphi_1(\theta)\mu(d\theta)-\int_{-\tau}^{0}\varphi_2(\theta)\mu(d\theta)\Big|^2+|\xi_1-\xi_2|^2\Big)\\
&\le L_R\Big(\Big|\int_{-\tau}^{0}|\varphi_1(\theta)-\varphi_2(\theta)|^2\mu(d\theta)\Big|+|\xi_1-\xi_2|^2\Big)\\
&\le L_R(\|\varphi_1-\varphi_2\|^2+|\xi_1-\xi_2|^2).
\end{aligned}
$$

Similarly, for any $\xi_1,\xi_2\in\mathbb{R}^m$ and $\varphi_1,\varphi_2\in C([-\tau,0];\mathbb{R}^n)$ with $\|\varphi_1\|\vee\|\varphi_2\|\le R$,

$$
\begin{aligned}
|b(\varphi_1,\xi_1)-b(\varphi_2,\xi_2)|^2 &= \Big|B\Big(\int_{-\tau}^{0}\varphi_1(\theta)\mu(d\theta),\xi_1\Big)-B\Big(\int_{-\tau}^{0}\varphi_2(\theta)\mu(d\theta),\xi_2\Big)\Big|^2\\
&\le L_R\|\varphi_1-\varphi_2\|^2+L|\xi_1-\xi_2|^2
\end{aligned}
$$

and

$$|\psi(\varphi_1,\xi_1)-\psi(\varphi_2,\xi_2)|^2\le L_R\|\varphi_1-\varphi_2\|^2+L|\xi_1-\xi_2|^2.$$

These imply Assumption (A1) holds. Likewise, Assumptions (A2) and (A3) hold.

Let ξ^X be the solution of the fixed-X equation

$$d\xi(t)=H(X,\xi(t))dt+\Phi(X,\xi(t))d\tilde{w}_2(t). \tag{6.5}$$

Theorem 4.1 shows that this equation has a unique invariant measure μ^X. Let us define

$$\bar{B}(X)=\int_{\mathbb{R}^m}B(X,\xi)\mu^X(d\xi)\ \text{ and }\ \bar{\Sigma}(X)=\int_{\mathbb{R}^m}\Psi(X,\xi)\Psi'(X,\xi)\mu^X(d\xi). \tag{6.6}$$

Assume that there exists a unique weak solution for SIDE

$$dx(t)=\bar{B}\Big(\int_{-\tau}^{0}x(t+\theta)\mu(d\theta)\Big)dt+\bar{\Psi}\Big(\int_{-\tau}^{0}x(t+\theta)\mu(d\theta)\Big)dB(t), \tag{6.7}$$

with the initial data $x_0\in C([-\tau,0];\mathbb{R}^n)$, where $\bar{\Psi}(\cdot)\bar{\Psi}'(\cdot)=\bar{\Sigma}(\cdot)$. Then we have the following theorem.

Theorem 6.1. *Under Assumptions* ($\hat{\text{A}}$1)-($\hat{\text{A}}$3) *and* (A4)*, Eq.* (6.1) *has a unique global solution* $((x^\varepsilon(t))',(\xi^\varepsilon(t))')'$*. Moreover, if* $2\lambda_1>\lambda_2$*,* $x^\varepsilon(\cdot)$ *converges weakly to* $x(\cdot)$*, the solution of* (6.7)*.*

As an example, let us consider the following special 2-dimensional linear SIDEs with two-time scales:

$$dx^\varepsilon(t) = \Big[\alpha_1 \int_{t-\tau}^{t} x^\varepsilon(s)ds + \beta\xi^\varepsilon(t)\Big]dt + \chi\xi^\varepsilon(t)dw_1(t), \tag{6.8a}$$

$$d\xi^\varepsilon(t) = \frac{1}{\varepsilon}\Big[\alpha_2 \int_{t-\tau}^{t} x^\varepsilon(s)ds - \lambda\xi^\varepsilon(t)\Big]dt + \frac{\rho}{\sqrt{\varepsilon}} \int_{t-\tau}^{t} x^\varepsilon(s)ds dw_2(t) \tag{6.8b}$$

with initial data $\xi(0) \in \mathbb{R}$ and $x_0 \in C([-\tau, 0]; \mathbb{R})$, where $\lambda > 0$, α_1, α_2, β, χ, $\rho \in \mathbb{R}$. Choose $\mu(\cdot)$ to be the uniform distribution on $[-\tau, 0]$, i.e., $\mu(d\theta) = d\theta/\tau$. Then it can be observed that

$$\int_{t-\tau}^{t} x^\varepsilon(s)ds = \int_{-\tau}^{0} x^\varepsilon(t+\theta)d\theta = \tau \int_{-\tau}^{0} x^\varepsilon(t+\theta)\mu(d\theta).$$

This shows that Eq. (6.8) satisfies Assumptions ($\hat{\text{A}}$1)-($\hat{\text{A}}$3) and holds a unique global solution $(x^\varepsilon(t), \xi^\varepsilon(t))'$. Let us consider the following fixed-X equation

$$d\xi(t) = (\alpha_2 X - \lambda\xi(t))dt + \rho X dw_2(t).$$

This equation describes the mean reverting Ornstein–Uhlenbeck process with stationary normal distribution μ^X being $N(\alpha_2 X/\lambda, (\rho X)^2/(2\lambda))$, which is exponentially ergodic (see [23, p.306]). It is easy to observe that

$$\mathbb{E}_{\mu^X}\xi = \int_{\mathbb{R}} \xi\mu^X(d\xi) = \frac{\alpha_2 X}{\lambda}, \quad \mathbb{E}_{\mu^X}\xi^2 = \int_{\mathbb{R}} \xi^2\mu^X(d\xi) = \frac{2\alpha_2^2 + \lambda\rho^2}{2\lambda^2}X^2.$$

Let us define

$$dx(t) = \Big(\alpha_1 + \frac{\beta\alpha_2}{\lambda}\Big) \int_{t-\tau}^{t} x(s)dsdt + \frac{|\chi|}{\sqrt{2}\lambda}\sqrt{2\alpha_2^2 + \lambda\rho^2}\Big|\int_{t-\tau}^{t} x(s)ds\Big|dB(t), \tag{6.9}$$

where $B(t)$ is a scalar Brownian motion. Since $\lambda > 0$, if Eq. (6.9) has a unique global solution and Eqs. (6.9) and (6.8a) have the same initial data satisfying (A4), by Theorem 6.1, $x^\varepsilon(\cdot) \Rightarrow x(\cdot)$ determined by Eq. (6.9).

Choosing μ being the Dirac measure at $-\tau$, we have

$$\int_{-\tau}^{0} \varphi(\theta)\mu(d\theta) = \varphi(-\tau).$$

Then Eq. (6.1) may be rewritten as the following SDDE

$$dx^\varepsilon(t) = B(x^\varepsilon(t-\tau), \xi^\varepsilon(t))dt + \Psi(x^\varepsilon(t-\tau), \xi^\varepsilon(t))dw_1(t), \tag{6.10a}$$

$$d\xi^\varepsilon(t) = \frac{1}{\varepsilon}H(x^\varepsilon(t-\tau), \xi^\varepsilon(t))dt + \frac{1}{\sqrt{\varepsilon}}\Phi(x^\varepsilon(t-\tau), \xi^\varepsilon(t))dw_2(t). \tag{6.10b}$$

According to Theorem 6.1, under Assumptions ($\hat{A}1$)-($\hat{A}3$) and (A4), Eq. (6.10) has a unique global solution $((x^\varepsilon(t))', (\xi^\varepsilon(t))')'$. Moreover, if $2\lambda_1 > \lambda_2$, $x^\varepsilon(\cdot)$ converges weakly to $x(\cdot)$ determined by the following stochastic pure delay differential equation

$$dx(t) = \bar{B}(x(t-\tau))dt + \bar{\Psi}(x^\varepsilon(t-\tau)dB(t),$$

where $\bar{B}$ and $\bar{\Psi}\bar{\Psi}' = \bar{\Sigma}$ are determined by (6.6). There always exists a global solution for this stochastic pure delay differential equation (see [23, p.157]).

Remark 6.2. (Final remarks). Let us recapture the main advances of this paper. Considering two-time-scale stochastic functional differential equations, we treat coupled systems, which are more versatile but are far more difficult to deal with. To overcome the difficulty due to the past dependence and the coupled systems, the Hölder continuity and the tightness of certain processes are obtained together with continuous dependence of the parameters. Then a direct averaging is performed to obtain the desired limit stochastic functional equations. An immediate question is: Can we handle systems in which not only does the fast component depend on the segment process of the slow component, but also depends on the segment process of the fast-varying component? At this point, it seems that the current techniques cannot be used to treat the corresponding systems. More sophisticated methods are needed, which deserves further in depth investigation.

References

[1] J.A.D. Appleby, M. Riedle, Almost sure asymptotic stability of stochastic Volterra integro-differential equations with fading perturbations, Stoch. Anal. Appl. 24 (4) (2006) 813–826.

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

[3] 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.

[4] P. Billingsley, Convergence of Probability Measures, second edition, A Wiley-Interscience Publication, New York, 1999.

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

[6] B. Dupire, Functional Itô Calculus, Portfolio Research paper 2009-04, Bloomberg, 2009.

[7] M.I. Freidlin, A.D. Wentzell, Random Perturbations of Dynamical Systems, 3rd edition, Springer, Berlin Heidelberg, 2012.

[8] M. Galtier, G. Wainrib, Multiscale analysis of slow-fast neuronal learning models with noise, J. Math. Neurosci. 2 (2012) 13.

[9] K. Gopalsamy, Stability and Oscillations in Delay Difference Differential Equations of Population Dynamics, Kluwer Academic, Dordrecht, 1992.

[10] J.K. Hale, S.M.V. Lunel, Introduction to Functional Differential Equations, Springer, Berlin, 1993.

[11] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer, New York, 1988.

[12] R.Z. Khasminskii, On stochastic processes defined by differential equations with a small parameter, Theory Probab. Appl. 11 (1966) 211–228.

[13] R.Z. Khasminskii, G. Yin, Asymptotic series for singularly perturbed Kolmogorov-Fokker-Planck equations, SIAM J. Appl. Math. 56 (1996) 1766–1793.

[14] 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.

[15] R.Z. Khasminskii, G. Yin, Limit behavior of two-time-scale diffusions revisited, J. Differ. Equ. 212 (2005) 85–113.
[16] R.Z. Khasminskii, Stochastic Stability of Differential Equations, 2nd ed., Springer, Heidelberg, 2012.
[17] I. Kim, K-H. Kim, P. Kim, Parabolic Littlewood-Paley inequality for $\phi(-\Delta)$-type operators and applications to stochastic integro-differential equations, Adv. Math. 249 (2013) 161–203.
[18] Y. Kuang, Delay Differential Equations with Applications in Population Dynamics, Academic Press, Boston, 1993.
[19] H.J. Kushner, Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory, MIT Press, Cambridge, MA, 1984.
[20] H.J. Kushner, Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems, Birkhäuser, Boston, MA, 1990.
[21] H.J. Kushner, Numerical Methods for Controlled Stochastic Delay Systems, Birkhäuser Boston, Boston, MA, 2008.
[22] T. Kurtz, Semigroups of conditioned shifts and approximation of Markov processes, Ann. Probab. 3 (4) (1975) 618–642.
[23] X. Mao, Stochastic Differential Equations and Applications, second edition, Horwood, Chichester, 2007.
[24] S.-E.A. Mohammed, Stochastic Functional Differential Equations, Longman, Harlow/New York, 1986.
[25] J. Munkres, Topology, Prentice Hall, Upper Saddle River, 2000.
[26] W.J. Padgett, Chris P. Tsokos, On stochastic integro-differential equation of Volterra type, SIAM J. Appl. Math. 23 (1972) 499–512.
[27] E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation I, Ann. Probab. 29 (3) (2001) 1061–1085, II, Ann. Probab. 31 (3) (2003) 1166–1192, III, Ann. Probab. 33 (3) (2005) 1111–1133.
[28] G.A. Pavliotis, A.M. Stuart, Multiscale Methods: Averaging and Homogenization, Springer, Berlin, 2008.
[29] B. Przeradzki, The existence of bounded solutions for differential equations in Hilbert space, Ann. Pol. Math. LVI.2 (1992) 103–121.
[30] K.M. Ramachandran, A singularly perturbed stochastic delay system with small parameter, Stoch. Anal. Appl. 11 (1993) 209–230.
[31] K.M. Ramachandran, Stability of stochastic delay differential equation with a small parameter, Stoch. Anal. Appl. 26 (4) (2008) 710–723.
[32] M. Röckner, L. Xie, Diffusion approximation for fully coupled stochastic differential equations, Ann. Probab. 49 (3) (2021) 1205–1236.
[33] A.V. Skorokhod, F.C. Hoppensteadt, H.D. Salehi, Random Perturbation Methods with Applications in Science and Engineering, Springer, New York, 2002.
[34] 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.
[35] F. Wu, Y. Xu, Stochastic Lotka-Volterra population dynamics with infinite delay, SIAM J. Appl. Math. 70 (2009) 641–657.
[36] F. Wu, G. Yin, An averaging principle for two-time-scale stochastic functional differential equations, J. Differ. Equ. 269 (2020) 1037–1077.
[37] 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 (2017) 1226–1252.
[38] G. Yin, K.M. Ramachandran, A differential delay equation with wideband noise perturbations, Stoch. Process. Appl. 35 (1990) 231–249.
[39] G. Yin, H.Q. Zhang, Singularly perturbed Markov chains: limit results and applications, Ann. Appl. Probab. 17 (2007) 207–229.
[40] G. Yin, Q. Zhang, Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach, Springer, Now York, 2013.