Journal of Statistical Physics (2024) 191:24
https://doi.org/10.1007/s10955-024-03241-8

Optimal Total Variation Bounds for Stochastic Differential Delay Equations with Small Noises

Nguyen Tien Dung[1] · **Nguyen Thu Hang**[2] · **Tran Manh Cuong**[1]

Received: 20 October 2023 / Accepted: 23 January 2024

Abstract

In this paper, we study the central limit theorem for the solutions of stochastic differential delay equations with small noises. Our aim is to provide explicit estimates for the rate of convergence in total variation distance. We also show that the convergence rate is of optimal order.

Keywords Central limit theorem · Stochastic differential delay equation · Malliavin calculus

Mathematics Subject Classification 60F05 · 60H07 · 60H10

1 Introduction and Main Results

It is well known that stochastic dynamical systems with small noise have useful applications in several fields including physics, chemistry, and biology [3], filtering problems [16] and mathematical finance [7, 20], etc. Since the appearance of seminal work [8], various properties of such dynamical systems have been intensively studied. Among others, we cite [4] and references therein for large deviation results, [11] for averaging principle, [13] for moderate deviation results, [9, 12] for parameter estimators and [1] for abrupt convergence.

In the last years, the central limit theorem for stochastic dynamical systems with small noise has been gained much attention, see e.g. [6, 10, 15, 18, 19]. However, in this research line, most of the existing results are qualitative. We only find in the literature a recent preprint [2] in which a quantitative Wasserstein bound was obtained for multi-scale diffusion systems.

Communicated by Gregory Schehr.

✉ Nguyen Tien Dung
dung@hus.edu.vn

1 Department of Mathematics, VNU University of Science, Vietnam National University Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi 084, Vietnam

2 Department of Mathematics, Hanoi University of Mining and Geology, 18 Pho Vien, Bac Tu Liem, Hanoi 084, Vietnam

Published online: 21 February 2024

In this paper, for any $\varepsilon \in (0,1)$, we consider the stochastic dynamical system governed by stochastic differential delay equations with small noise of the form

$$\begin{cases} X_{\varepsilon,t} = \varphi(0) + \int_0^t b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})ds + \varepsilon \int_0^t \sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})dB_s, \ t \in [0,T] \\ X_{\varepsilon,t} = \varphi(t), \ t \in [-\tau, 0], \end{cases} \tag{1.1}$$

where the initial data $\varphi : [-\tau, 0] \to \mathbb{R}$ is a bounded deterministic function, $(B_t)_{t\in[0,T]}$ is a standard Brownian motion and b, σ are deterministic functions on $\mathbb{R}^2$.

Intuitively, as ε tends to 0, $X_{\varepsilon,t}$ converges to x_t, which solves the following deterministic differential delay equation

$$\begin{cases} x_t = \varphi(0) + \int_0^t b(x_s, x_{s-\tau})ds, \quad t \in [0,T] \\ x_t = \varphi(t), \ t \in [-\tau, 0]. \end{cases} \tag{1.2}$$

Define

$$\tilde{X}_{\varepsilon,t} := \frac{X_{\varepsilon,t} - x_t}{\varepsilon}, \quad t \in [-\tau, T]. \tag{1.3}$$

It is known from [18] that $\tilde{X}_{\varepsilon,t}$ converges to Y_t in $L^2(\Omega)$ as $\varepsilon \to 0$, where $(Y_t)_{t\geq 0}$ is unique solution to the following linear stochastic differential equation

$$\begin{cases} Y_t = \int_0^t \left(b_1'(x_s, x_{s-\tau})Y_s + b_2'(x_s, x_{s-\tau})Y_{s-\tau}\right) ds + \int_0^t \sigma(x_s, x_{s-\tau})dB_s, \quad t \in [0,T] \\ Y_t = 0, \quad t \in [-\tau, 0]. \end{cases} \tag{1.4}$$

We observe that Y_t is a normal random variable for each $t \in [0,T]$, see Remark 3.1 below. Thus the sequence $(\tilde{X}_{\varepsilon,t})_{\varepsilon\in(0,1)}$ satisfies the central limit theorem as $\varepsilon \to 0$, and hence, an important problem arising here is to investigate the rate of convergence via certain distances. There are three distances commonly used in the literature.

(i) The Wasserstein distance between the laws of $\tilde{X}_{\varepsilon,t}$ and Y_t :

$$d_{\mathrm{W}}(\tilde{X}_{\varepsilon,t}, Y_t) := \sup_{|g(x)-g(y)|\leq|x-y|} |Eg(\tilde{X}_{\varepsilon,t}) - Eg(Y_t)|.$$

(ii) The Kolmogorov distance between the laws of $\tilde{X}_{\varepsilon,t}$ and Y_t :

$$d_{\mathrm{K}}(\tilde{X}_{\varepsilon,t}, Y_t) := \sup_{x\in\mathbb{R}} |P(\tilde{X}_{\varepsilon,t} \leq x) - P(Y_t \leq x)|.$$

(iii) The total variation distance between the laws of $\tilde{X}_{\varepsilon,t}$ and Y_t :

$$\begin{aligned} d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t) &:= \sup_{A\in\mathcal{B}(\mathbb{R})} |P(\tilde{X}_{\varepsilon,t} \in A) - P(Y_t \in A)| \\ &= \frac{1}{2} \sup_{\|g\|_\infty\leq 1} |Eg(\tilde{X}_{\varepsilon,t}) - Eg(Y_t)|, \end{aligned}$$

where $\mathcal{B}(\mathbb{R})$ is Borel σ-algebra on $\mathbb{R}$ and $\|g\|_\infty := \sup_{x\in\mathbb{R}} |g(x)|$. The Wasserstein distance is easy to bound. Indeed, Theorem 1 in [18] (see also Proposition 3.4 below) gives us

$$d_{\mathrm{W}}(\tilde{X}_{\varepsilon,t}, Y_t) \leq E|\tilde{X}_{\varepsilon,t} - Y_t| \leq C\varepsilon, \ 0 \leq t \leq T,$$

where C is a positive constant not depending on t and ε. On the other hand, we always have the following relationship

$$d_{\mathrm{K}}(\tilde{X}_{\varepsilon,t}, Y_t) \leq d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t).$$

Thus our present work focuses on bounding the total variation distance $d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t)$. For this purpose, we make the use the following assumption.

Assumption 1.1 $b, \sigma : \mathbb{R}^2 \to \mathbb{R}$ are twice differentiable functions with the partial derivatives bounded by L.

Notice that our Assumption 1.1 is slightly stronger than the conditions required in [18]. By employing a general result established in our recent paper [5] by means of Malliavin calculus (see Lemma 2.1 below), we obtain the following quantitative estimate for the total variation distance.

Theorem 1.1 *Let Assumption 1.1 hold. Consider the stochastic processes $(\tilde{X}_{\varepsilon,t})_{-\tau\le t\le T}$ and $(Y_t)_{-\tau\le t\le T}$ defined by (1.3) and (1.4), respectively. Then, we have*

$$d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t) \le \frac{Ct\varepsilon}{\sqrt{\mathrm{Var}(Y_t)}}\ \forall\, \varepsilon \in (0,1), 0 < t \le T, \tag{1.5}$$

where C is a positive constant not depending on t and ε.

Our next theorem points out that rate of convergence $O(\varepsilon)$ is of optimal order as $\varepsilon \to 0$.

Theorem 1.2 *Let Assumption 1.1 hold. We additionally assume that the second-order partial derivatives of b are continuous. Then, for any continuous and bounded function g, we have*

$$\lim_{\varepsilon\to 0} \frac{Eg(\tilde{X}_{\varepsilon,t}) - Eg(Y_t)}{\varepsilon} = \frac{1}{2\mathrm{Var}(Y_t)} E\left[g(Y_t)\delta\left(Z_t DY_t\right)\right],\ \ 0 < t \le T, \tag{1.6}$$

where $Z_t = 0$ for $t \in [-\tau, 0]$ and

$$\begin{aligned} Z_t = &\int_0^t b_1'(x_s, x_{s-\tau}) Z_s ds + \int_0^t b_2'(x_s, x_{s-\tau}) Z_{s-\tau} ds \\ &+ \int_0^t \left(b_{11}''(x_s, x_{s-\tau}) Y_s^2 + b_{12}''(x_s, x_{s-\tau}) Y_s Y_{s-\tau} + b_{22}''(x_s, x_{s-\tau}) Y_{s-\tau}^2\right) ds \\ &+ 2\int_0^t \left(\sigma_1'(x_s, x_{s-\tau}) Y_s + \sigma_2'(x_s, x_{s-\tau}) Y_{s-\tau}\right) dB_s,\quad t \in [0, T]. \end{aligned} \tag{1.7}$$

In particular, we have

$$\lim_{\varepsilon\to 0} \frac{d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t)}{\varepsilon} \ge \frac{1}{2\mathrm{Var}(Y_t)} E|E\left[\delta(Z_t DY_t)|Y_t\right]|,\ \ 0 < t \le T. \tag{1.8}$$

In the statement of the above theorem, D denotes Malliavin derivative operator and δ denotes the divergence operator (or Skorohod integral). The definition of these operators will be recalled in Sect. 2 below. We also use the notation: Given a differentiable function h of 2 variables, we denote

$$h_i'(x_1, x_2) := \frac{\partial h}{\partial x_i}(x_1, x_2),\ \ h_{ij}''(x_1, x_2) = \frac{\partial h}{\partial x_i \partial x_j}(x_1, x_2),\ \ 1 \le i, j \le 2.$$

The rest of this article is organized as follows. In Sect. 2, we recall some fundamental concepts of Malliavin calculus and a general estimate for the total variation distance between two Malliavin differentiable random variables. In Sect. 3, we prove Theorems 1.1 and 1.2. The conclusion is given in Sect. 4.

2 Preliminaries

This paper is strongly based on techniques of Malliavin calculus. For the reader's convenience, let us recall some elements of Malliavin calculus (for more details see [14]). We suppose that$(B_t)_{t\in[0,T]}$ is defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is a natural filtration generated by the Brownian motion B. For $h \in L^2[0, T]$, we denote by $B(h)$ the Wiener integral

$$B(h) = \int_0^T h(t)dB_t.$$

Let $\mathcal{S}$ denote a dense subset of $L^2(\Omega, \mathcal{F}, P)$ that consists of smooth random variables of the form

$$F = f(B(h_1), B(h_2), ..., B(h_n)), \tag{2.1}$$

where $n \in \mathbb{N}$, $f \in C_0^\infty(\mathbb{R}^n)$, $h_1, h_2, ..., h_n \in L^2[0, T]$. If F has the form (2.1), we define its Malliavin derivative as the process $DF := D_t F$, $t \in [0, T]$ given by

$$D_t F = \sum_{k=1}^n \frac{\partial f}{\partial x_k}(B(h_1), B(h_2), ..., B(h_n))h_k(t).$$

More generally, for each $k \geq 1$, we can define the iterated derivative operator on a cylindrical random variable by setting

$$D^k_{t_1,...,t_k} F = D_{t_1} ... D_{t_k} F.$$

For any $1 \leq p, k < \infty$, we denote by $\mathbb{D}^{k,p}$ the closure of $\mathcal{S}$ with respect to the norm

$$\begin{aligned} ||F||^p_{k,p} \;&:= E|F|^p + E\left[\left(\int_0^T |D_u F|^2 du\right)^{\frac{p}{2}}\right] \\ &+ ... + E\left[\left(\int_0^T ... \int_0^T |D^k_{t_1,...,t_k} F|^2 dt_1 ... dt_k\right)^{\frac{p}{2}}\right]. \end{aligned}$$

A random variable F is said to be Malliavin differentiable if it belongs to $\mathbb{D}^{1,2}$. The derivative operator D satisfies the chain rule, i.e, $D\phi(F) = \phi'(F)DF$ for any differentiable function ϕ with bounded derivative. Furthermore, we have the following relations between Malliavin derivative and the integrals

$$D_r\left(\int_0^T u_s ds\right) = \int_r^T D_r u_s ds$$

and

$$D_r\left(\int_0^T u_s dB_s\right) = u_r + \int_r^T D_r u_s dB_s, 0 \leq r \leq T,$$

for all $0 \leq r \leq T$, where $(u_t)_{t\in[0,T]}$ is an $\mathbb{F}$-adapted and Malliavin differentable stochastic process.

An important operator in the Malliavin calculus theory is the divergence operator δ. It is the adjoint of derivative operator D. The domain of δ is the set of all functions $u \in L^2(\Omega \times [0, T])$ such that

$$E|\langle DF, u\rangle_{L^2[0,T]}| \leq C(u)\|F\|_{L^2(\Omega)},$$

where $C(u)$ is some positive constant depending on u. In particular, if $u \in Dom\delta$, then $\delta(u)$ is characterized by following duality relationships

$$\delta(uF) = F\delta(u) - \langle DF, u\rangle_{L^2[0,T]} \tag{2.2}$$

$$E[\langle DF, u\rangle_{L^2[0,T]}] = E[F\delta(u)] \text{ for any } F \in \mathbb{D}^{1,2}. \tag{2.3}$$

We have the following general result.

Lemma 2.1 *Let $F_1 \in \mathbb{D}^{2,4}$ be such that $\|DF_1\|_{L^2[0,T]} > 0$ a.s. Then, for any random variable $F_2 \in \mathbb{D}^{1,2}$ and any measurable function g with $\|g\|_\infty = \sup_{x\in\mathbb{R}} |g(x)| \leq 1$, we have*

$$\begin{aligned} &|Eg(F_1) - Eg(F_2)| \\ &\leq C\left(E\|DF_1\|^{-8}_{L^2[0,T]} E\left(\int_0^T \int_0^T |D_\theta D_r F_1|^2 d\theta dr\right)^2 \right. \\ &\left. +(E\|DF_1\|^{-2}_{L^2[0,T]})^2\right)^{\frac{1}{4}} \|F_1 - F_2\|_{1,2}, \end{aligned} \tag{2.4}$$

provided that the expectations exist, where C is an absolute constant.

Proof This lemma is Theorem 3.1 in our recent paper [5]. Here we note that the inequality (2.4) follows from the relation

$$\begin{aligned} Eg(F_1) - Eg(F_2) = E\left[\int_{F_2}^{F_1} g(z)dz\delta\left(\frac{DF_1}{\|DF_1\|^2_{L^2[0,T]}}\right)\right] \\ -E\left[\frac{g(F_2)\langle DF_1 - DF_2, DF_1\rangle_{L^2[0,T]}}{\|DF_1\|^2_{L^2[0,T]}}\right]. \end{aligned} \tag{2.5}$$

We also have $E\left[\left(\delta\left(\frac{DF_1}{\|DF_1\|^2_{L^2[0,T]}}\right)\right)^2\right] < \infty.$ □

3 Proofs of Main Results

Hereafter, we denote by C a generic constant which may vary at each appearance. For any $a, b \in \mathbb{R}$, we denote $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. In our proofs, we frequently use the fundamental inequality

$$(a_1 + ... + a_n)^p \leq n^{p-1}(a_1^p + ... + a_n^p),$$

for all $a_1, ..., a_n \geq 0$ and $p \geq 2$.

3.1 Some Fundamental Estimates

In this subsection, we collect some fundamental properties of the solution to (1.1). We first note that, under Assumption 1.1, the functions b and σ are Lipschitz continuous and have linear growth. Indeed, we have

$$|b(x_1, y_1) - b(x_2, y_2)| + |\sigma(x_1, y_1) - \sigma(x_2, y_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|) \tag{3.1}$$

for all $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and

$$|b(x,y)| + |\sigma(x,y)| \le |b(0,0)| + |\sigma(0,0)| + |b(x,y) - b(0,0)| + |\sigma(x,y) - \sigma(0,0)|$$
$$\le |b(0,0)| + |\sigma(0,0)| + L(|x| + |y|)\ \forall x, y \in \mathbb{R}. \tag{3.2}$$

Lemma 3.1 *Let Assumption 1.1 hold. Consider the solution $(X_{\varepsilon,t})_{t\in[-\tau,T]}$ to the equation (1.1). Then, for every $p \ge 2$, we have*

$$\sup_{0\le t\le T} E|X_{\varepsilon,t}|^p \le C\ \forall \varepsilon \in (0,1). \tag{3.3}$$

where C is a positive constant not depending on ε.

Proof See Lemma 1 in [18]. □

Proposition 3.1 *Let Assumption 1.1 hold. Consider the stochastic process $(\tilde{X}_{\varepsilon,t})_{t\in[0,T]}$ defined by (1.3). Then, for all $p \ge 2$ we have*

$$E|X_{\varepsilon,t} - x_t|^p \le Ct^{\frac{p}{2}}\varepsilon^p\ \forall \varepsilon \in (0,1), 0 \le t \le T, \tag{3.4}$$

where C is a positive constant not depending on t and ε.

Proof For every $\varepsilon \in (0,1)$, we have

$$X_{\varepsilon,t} - x_t = \int_0^t \big(b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b(x_s, x_{s-\tau})\big)\,ds + \varepsilon \int_0^t \sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})dB_s,\ 0 \le t \le T.$$

Consequently,

$$E|X_{\varepsilon,t} - x_t|^p \le 2^{p-1}E\left|\int_0^t \big(b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b(x_s, x_{s-\tau})\big)\,ds\right|^p$$
$$+ 2^{p-1}\varepsilon^p E\left|\int_0^t \sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})dB_s\right|^p,\ 0 \le t \le T.$$

By the Hölder and Burkholder–Davis–Gundy inequalities, for all $p \ge 2$, we deduce

$$E|X_{\varepsilon,t} - x_t|^p \le Ct^{p-1}\int_0^t E\left|b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b(x_s, x_{s-\tau})\right|^p ds$$
$$+ C\varepsilon^p t^{\frac{p}{2}-1}\int_0^t E|\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})|^p ds, \quad 0 \le t \le T,$$

where C is a positive constant depending only on p. Recalling (3.1), (3.2) and (3.3), we get

$$E|X_{\varepsilon,t} - x_t|^p \le Ct^{p-1}\int_0^t E\left(|X_{\varepsilon,s} - x_s| + |X_{\varepsilon,s-\tau} - x_{s-\tau}|\right)^p ds$$
$$+ C\varepsilon^p t^{\frac{p}{2}-1}\int_0^t E(1 + |X_{\varepsilon,s}| + |X_{\varepsilon,s-\tau}|)^p ds$$
$$\le Ct^{p-1}\int_0^t \left(E|X_{\varepsilon,s} - x_s|^p + E|X_{\varepsilon,s-\tau} - x_{s-\tau}|^p\right) ds + C\varepsilon^p t^{\frac{p}{2}},$$

where C is a positive constant depending only on L and p. Since $X_{\varepsilon,s} = x_s = \varphi(s), s \in [-\tau, 0]$, it holds that

$$\int_0^t E|X_{\varepsilon,s-\tau} - x_{s-\tau}|^p ds = 0 \text{ if } t \le \tau,$$

and

$$\int_0^t E|X_{\varepsilon,s-\tau}-x_{s-\tau}|^p ds = \int_0^\tau E|X_{\varepsilon,s-\tau}-x_{s-\tau}|^p ds + \int_\tau^t E|X_{\varepsilon,s-\tau}-x_{s-\tau}|^p ds$$
$$= \int_\tau^t E|X_{\varepsilon,s-\tau}-x_{s-\tau}|^p ds = \int_0^{t-\tau} E|X_{\varepsilon,s}-x_s|^p ds$$
$$\leq \int_0^t E|X_{\varepsilon,s}-x_s|^p ds, 0\leq\tau\leq t\leq T.$$

Consequently,

$$E|X_{\varepsilon,t}-x_t|^p \leq Ct^{p-1}\int_0^t E|X_{\varepsilon,s}-x_s|^p ds + C\varepsilon^p t^{\frac{p}{2}}$$
$$\leq C\int_0^t E|X_{\varepsilon,s}-x_s|^p ds + C\varepsilon^p t^{\frac{p}{2}},\ 0\leq t\leq T,$$

where C is a positive constant not depending on t and ε. Then, by Gronwall's lemma, we get the desired conclusion (3.4). Indeed,

$$E|X_{\varepsilon,t}-x_t|^p \leq C\varepsilon^p t^{\frac{p}{2}} e^{Ct} \leq C\varepsilon^p t^{\frac{p}{2}} e^{CT} \leq C\varepsilon^p t^{\frac{p}{2}},\quad 0\leq t\leq T.$$

So the proof of the proposition is complete. □

Proposition 3.2 *Let Assumption 1.1 hold. Consider the solution $(X_{\varepsilon,t})_{t\in[-\tau,T]}$ to the Eq. (1.1). Then, for each $0\leq t\leq T$, the random variable $X_{\varepsilon,t}$ is Malliavin differentiable. Moreover, the derivative $D_\theta X_{\varepsilon,t}$ satisfies*
(i) When $t\in[-\tau,0]$, $D_\theta X_{\varepsilon,t}=0$ for all $0\leq\theta\leq T$,
(ii) When $t\in(0,T]$, $D_\theta X_{\varepsilon,t}=0$ for $\theta>t$ and

$$\begin{aligned} D_\theta X_{\varepsilon,t} &= \varepsilon\sigma(X_{\varepsilon,\theta},X_{\varepsilon,\theta-\tau}) + \int_\theta^t b_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}ds \\ &\quad + \int_{\theta+\tau}^t b_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s-\tau}ds \\ &\quad + \varepsilon\int_\theta^t \sigma_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}dB_s \\ &\quad + \varepsilon\int_{\theta+\tau}^t \sigma_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s-\tau}dB_s,\ 0\leq\theta\leq t-\tau, \end{aligned} \tag{3.5}$$

$$\begin{aligned} D_\theta X_{\varepsilon,t} &= \varepsilon\sigma(X_{\varepsilon,\theta},X_{\varepsilon,\theta-\tau}) + \int_\theta^t b_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}ds \\ &\quad + \varepsilon\int_\theta^t \sigma_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}dB_s,\ (t-\tau)\vee 0<\theta\leq t, \end{aligned} \tag{3.6}$$

Here, we use the convention $[0,t-\tau]=\emptyset$ if $t<\tau$.

Proof See Proposition 1 and Remark 2 in [17]. □

Proposition 3.3 *Let Assumption 1.1 hold. Consider the solution $(X_{\varepsilon,t})_{t\in[-\tau,T]}$ to the equation (1.1). Then, for each $p\geq 2$, we have*

$$\sup_{0\leq\theta\leq t\leq T} E|D_\theta X_{\varepsilon,t}|^p \leq C\varepsilon^p\ \forall\,\varepsilon\in(0,1), \tag{3.7}$$

where C is a positive constant not depending on ε.

Proof For every $\varepsilon \in (0, 1)$, it follows from equations (3.5) and (3.6) that the Malliavin derivative $D_\theta X_{\varepsilon,t}$ satisfies

$$D_\theta X_{\varepsilon,t} = \varepsilon\sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau}) + \int_\theta^t b_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s} ds + \int_\theta^t b_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s-\tau} \mathbb{1}_{[\theta+\tau,t]}(s) ds + \varepsilon \int_\theta^t \sigma_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s} dB_s + \varepsilon \int_\theta^t \sigma_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s-\tau} \mathbb{1}_{[\theta+\tau,t]}(s) dB_s \tag{3.8}$$

for $0 \leq \theta \leq t \leq T$. We therefore get

$$E|D_\theta X_{\varepsilon,t}|^p \leq 5^{p-1}\Big(\varepsilon^p E|\sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau})|^p + E\left|\int_\theta^t b_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,t} ds\right|^p + E\left|\int_\theta^t b_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s-\tau} \mathbb{1}_{[\theta+\tau,t]}(s) ds\right|^p + \varepsilon^p E\left|\int_\theta^t \sigma_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s} dB_s\right|^p + \varepsilon^p E\left|\int_\theta^t \sigma_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s-\tau} \mathbb{1}_{[\theta+\tau,t]}(s) dB_s\right|^p\Big), \quad 0 \leq \theta \leq t \leq T.$$

By the Hölder and Burkholder–Davis–Gundy inequalities and the boundedness of the partial derivatives of b and σ, we deduce

$$E|D_\theta X_{\varepsilon,t}|^p \leq C\varepsilon^p E|\sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau})|^p + C(1+\varepsilon^p)\int_\theta^t E|D_\theta X_{\varepsilon,t}|^p ds + C(1+\varepsilon^p)\int_\theta^t E|D_\theta X_{\varepsilon,s-\tau}|^p ds$$
$$\leq C\varepsilon^p E|\sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau})|^p + C\int_\theta^t E|D_\theta X_{\varepsilon,t}|^p ds, \quad 0 \leq \theta \leq t \leq T,$$

where C is a positive constant not depending on ε. Furthermore, in view of the estimates (3.2) and (3.3), we have

$$E|\sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau})|^p \leq C \ \forall 0 \leq \theta \leq T. \tag{3.9}$$

So it holds that

$$E|D_\theta X_{\varepsilon,t}|^p \leq C\varepsilon^p + C\int_\theta^t E|D_\theta X_{\varepsilon,t}|^p ds, \quad 0 \leq \theta \leq t \leq T$$

Then, by using Gronwall's lemma, we obtain

$$E|D_\theta X_{\varepsilon,t}|^p \leq C\varepsilon^p, \ 0 \leq \theta \leq t \leq T.$$

The proof of the proposition is complete. □

We end this subsection by giving some remarks on the stochastic processes $(Y_t)_{-\tau\leq t\leq T}$ and $(Z_t)_{-\tau\leq t\leq T}$ defined by (1.4) and (1.7), respectively.

Remark 3.1 (i) Since $(x_t)_{-\tau\leq t\leq T}$ is deterministic and bounded, $\int_0^t \sigma(x_s, x_{s-\tau}) dB_s$ is a centered normal random variable with finite variance for each $0 \leq t \leq T$. Hence, it is easy to see that the linear integral equation (1.4) admits a unique solution $(Y_t)_{-\tau\leq t\leq T}$ satisfying

$$\sup_{0\leq t\leq T} E|Y_t|^p \leq C < \infty, \ \ p \geq 2.$$

Moreover, for each $t \in [0, T]$, the random variable Y_t is Malliavin differentiable and its derivative is given by $D_\theta Y_t = 0$ for $\theta > t$ and

$$D_\theta Y_t = \sigma(x_\theta, x_{\theta-\tau}) + \int_\theta^t b_1'(x_s, x_{s-\tau}) D_\theta Y_s ds \\ + \int_\theta^t b_2'(x_s, x_{s-\tau}) D_\theta Y_{s-\tau} ds, \quad 0 \leq \theta \leq t - \tau, \tag{3.10}$$

$$D_\theta Y_t = \sigma(x_\theta, x_{\theta-\tau}) + \int_\theta^t b_1'(x_s, x_{s-\tau}) D_\theta Y_s ds, \; t - \tau \vee 0 < \theta \leq t. \tag{3.11}$$

(ii) We have

$$EY_t = \int_0^t \left(b_1'(x_s, x_{s-\tau}) EY_s + b_2'(x_s, x_{s-\tau}) EY_{s-\tau} \right) ds, \quad 0 \leq t \leq T.$$

This is a linear equation with the initial data $EY_t|_{t=0} = 0$, and hence, $EY_t = 0$ for all $0 \leq t \leq T$.
(iii) Note that $D_\theta Y_t$ is deterministic for all $0 \leq \theta \leq t \leq T$. Hence, by Clark-Ocone formula (see Proposition 1.3.14 in [14]), we have

$$Y_t = \int_0^t E[D_\theta Y_t | \mathcal{F}_\theta] dB_\theta = \int_0^t D_\theta Y_t dB_\theta$$

This representation formula shows that Y_t is a normal random variable for each $t \in (0, T]$. Moreover, we have $\mathrm{Var}(Y_t) = \|DY_t\|^2_{L^2[0,T]}$.

Remark 3.2 Denote by $h(t)$ the sum of the last two addends in the right hand side of (1.7). We can verify that

$$\sup_{0 \leq t \leq T} E|h(t)|^p \leq C < \infty, \;\; p \geq 2.$$

Hence, the linear integral equation (1.7) admits a unique solution $(Z_t)_{-\tau \leq t \leq T}$ satisfying

$$\sup_{0 \leq t \leq T} E|Z_t|^p \leq C < \infty, \;\; p \geq 2.$$

3.2 Proof of Theorem 1.1

The proof of Theorem 1.1 will be given at the end of this subsection. In order to be able to apply Lemma 2.1, we first need to prepare some technical results.

Proposition 3.4 *Suppose Assumption 1.1. Let $(\tilde{X}_{\varepsilon,t})_{-\tau \leq t \leq T}$ and $(Y_t)_{-\tau \leq t \leq T}$ be as in Theorem 1.1. Then, for every $p \geq 2$, we have*

$$E|\tilde{X}_{\varepsilon,t} - Y_t|^p \leq C t^p \varepsilon^p \;\forall \varepsilon \in (0, 1), 0 \leq t \leq T,$$

where C is a positive constant not depending on t and ε.

Proof For every $\varepsilon \in (0, 1)$, recalling (1.3) and (1.4), we have

$$\tilde{X}_{\varepsilon,t} - Y_t = \frac{1}{\varepsilon} \int_0^t \left(b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b(x_s, x_{s-\tau}) \right) ds \\ - \int_0^t \left(b_1'(x_s, x_{s-\tau}) Y_s + b_2'(x_s, x_{s-\tau}) Y_{s-\tau} \right) ds$$

$$+\int_0^t \left(\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - \sigma(x_s, x_{s-\tau})\right) dB_s,\ 0 \le t \le T. \tag{3.12}$$

For each $s \in [0, T]$, using Taylor's expansion, we get

$$b(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b(x_s, x_{s-\tau}) = b_1'(x_s, x_{s-\tau})\left(X_{\varepsilon,s} - x_s\right) + b_2'(x_s, x_{s-\tau})\left(X_{\varepsilon,s-\tau} - x_{s-\tau}\right) + \frac{1}{2}R_{\varepsilon,s},$$

where the remainder term $R_{\varepsilon,s}$ is given by

$$\begin{aligned} R_{\varepsilon,s} &= b_{11}''\left(x_s + \xi_1(X_{\varepsilon,s} - x_s), x_{s-\tau} + \xi_2(X_{\varepsilon,s-\tau} - x_{s-\tau})\right)(X_{\varepsilon,s} - x_s)^2 \\ &\quad + 2b_{12}''\left(x_s + \xi_1(X_{\varepsilon,s} - x_s), x_{s-\tau} + \xi_2(X_{\varepsilon,s-\tau} - x_{s-\tau})\right)(X_{\varepsilon,s} - x_s)(X_{\varepsilon,s-\tau} - x_{s-\tau}) \\ &\quad + b_{22}''\left(x_s + \xi_1(X_{\varepsilon,s} - x_s), x_{s-\tau} + \xi_2(X_{\varepsilon,s-\tau} - x_{s-\tau})\right)\left(X_{\varepsilon,s-\tau} - x_{s-\tau}\right)^2, \end{aligned} \tag{3.13}$$

where ξ_1, ξ_2 are random variables lying between 0 and 1. We now can rewrite (3.12) as follows

$$\begin{aligned} \tilde{X}_{\varepsilon,t} - Y_t &= \int_0^t b_1'(x_s, x_{s-\tau})(\tilde{X}_{\varepsilon,s} - Y_s)ds + \int_0^t b_2'(x_s, x_{s-\tau})(\tilde{X}_{\varepsilon,s-\tau} - Y_{s-\tau})ds \\ &\quad + \frac{1}{2\varepsilon}\int_0^t R_{\varepsilon,s}ds + \int_0^t \left(\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau})\right. \\ &\quad \left.-\sigma(x_s, x_{s-\tau})\right) dB_s,\ 0 \le t \le T. \end{aligned} \tag{3.14}$$

Hence, for every $p \ge 2$, we obtain

$$\begin{aligned} E|\tilde{X}_{\varepsilon,t} - Y_t|^p &\le 4^{p-1}\Bigg(E\left|\int_0^t b_1'(x_s, x_{s-\tau})(\tilde{X}_{\varepsilon,s} - Y_s)ds\right|^p \\ &\quad + E\left|\int_0^t b_2'(x_s, x_{s-\tau})(\tilde{X}_{\varepsilon,s-\tau} - Y_{s-\tau})ds\right|^p \\ &\quad + \frac{1}{2^p\varepsilon^p}E\left|\int_0^t R_{\varepsilon,s}ds\right|^p + E\left|\int_0^t \left(\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - \sigma(x_s, x_{s-\tau})\right) dB_s\right|^p\Bigg). \end{aligned}$$

Then, by using the Hölder and Burkholder-Davis-Gundy inequalities, we deduce

$$\begin{aligned} E|\tilde{X}_{\varepsilon,t} - Y_t|^p &\le 3^{p-1}\Bigg(L^p t^{p-1}\int_0^t E|\tilde{X}_{\varepsilon,s} - Y_s|^p ds + L^p t^{p-1}\int_0^t E|\tilde{X}_{\varepsilon,s-\tau} - Y_{s-\tau}|^p ds \\ &\quad + \frac{t^{p-1}}{2^p\varepsilon^p}\int_0^t E|R_{\varepsilon,s}|^p ds + Ct^{\frac{p}{2}-1}\int_0^t E|\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - \sigma(x_s, x_{s-\tau})|^p ds\Bigg), \end{aligned}$$

where $C > 0$ depending only on p. By the boundedness of the partial derivatives of b, we have

$$|R_{\varepsilon,s}| \le 2L(X_{\varepsilon,s} - x_s)^2 + 2L\left(X_{\varepsilon,s-\tau} - x_{s-\tau}\right)^2,\ 0 \le s \le T.$$

This, combined with the estimate (3.4) and the fact $X_{\varepsilon,s-\tau} = x_{s-\tau}$ for $s \le \tau$, gives us

$$E|R_{\varepsilon,s}|^p \le Cs^p\varepsilon^{2p},\ 0 \le s \le T \tag{3.15}$$

for some $C > 0$ not depending on s and ε. Similarly, recalling the estimate (3.1), we also have

$$E|\sigma(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - \sigma(x_s, x_{s-\tau})|^p \le Cs^{\frac{p}{2}}\varepsilon^p,\ 0 \le s \le T.$$

Consequently, we get

$$E|\tilde{X}_{\varepsilon,t}-Y_t|^p \le Ct^p\varepsilon^p + C\int_0^t E|\tilde{X}_{\varepsilon,s}-Y_s|^p ds + \int_0^t E|\tilde{X}_{\varepsilon,s-\tau}-Y_{s-\tau}|^p ds$$
$$\le Ct^p\varepsilon^p + C\int_0^t E|\tilde{X}_{\varepsilon,s}-Y_s|^p ds,\ 0\le t\le T,$$

where C is a positive constant not depending on t and ε. By using Gronwall's lemma, we obtain

$$E|\tilde{X}_{\varepsilon,t}-Y_t|^p \le Ct^p\varepsilon^p,\ 0\le t\le T.$$

The proof of the proposition is complete. □

Proposition 3.5 *Suppose Assumption 1.1. Let* $(\tilde{X}_{\varepsilon,t})_{-\tau\le t\le T}$ *and* $(Y_t)_{-\tau\le t\le T}$ *be as in Theorem 1.1. Then we have*

$$E\|D\tilde{X}_{\varepsilon,t}-DY_t\|^2_{L^2[0,T]} \le Ct^2\varepsilon^2\ \forall\,\varepsilon\in(0,1), 0\le t\le T, \tag{3.16}$$

where C *is a positive constant not depending on* t *and* ε.

Proof We consider two cases separately.
Case 1. $0\le\theta\le t-\tau$. From the equations (3.5) and (3.10), we have

$$\begin{aligned}D_\theta\tilde{X}_{\varepsilon,t}-D_\theta Y_t &= \sigma(X_{\varepsilon,\theta},X_{\varepsilon,\theta-\tau})-\sigma(x_\theta,x(\theta-\tau))\\ &\quad+\int_\theta^t b_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta\tilde{X}_{\varepsilon,s}ds-\int_\theta^t b_1'(x_s,x_{s-\tau})D_\theta Y_s ds\\ &\quad+\int_{\theta+\tau}^t b_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta\tilde{X}_{\varepsilon,s-\tau}ds-\int_{\theta+\tau}^t b_2'(x_s,x_{s-\tau})D_\theta Y_{s-\tau}ds\\ &\quad+\int_\theta^t \sigma_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}dB_s+\int_{\theta+\tau}^t \sigma_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s-\tau}dB_s.\end{aligned}$$

We therefore deduce

$$\begin{aligned}E|D_\theta\tilde{X}_{\varepsilon,t}-D_\theta Y_t|^2 &\le 5E\left|\sigma(X_{\varepsilon,\theta},X_{\varepsilon,\theta-\tau})-\sigma(x_\theta,x(\theta-\tau))\right|^2\\ &+5E\left|\int_\theta^t\Big(b_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta\tilde{X}_{\varepsilon,s}-b_1'(x_s,x_{s-\tau})D_\theta Y_s\Big)ds\right|^2\\ &+5E\left|\int_{\theta+\tau}^t\Big(b_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta\tilde{X}_{\varepsilon,s-\tau}-b_2'(x_s,x_{s-\tau})D_\theta Y_{s-\tau}\Big)ds\right|^2\\ &+5E\left|\int_\theta^t\sigma_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s}dB_s\right|^2+5E\left|\int_{\theta+\tau}^t\sigma_2'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta X_{\varepsilon,s-\tau}dB_s\right|^2\\ &:=\sum_{k=1}^5 I_k.\end{aligned}\tag{3.17}$$

Using the estimates (3.1) and (3.4), we have

$$I_1\le 5L^2\left(E|X_{\varepsilon,\theta}-x_\theta|^2+E|X_{\varepsilon,\theta-\tau}-x_{\theta-\tau}|^2\right)\le C\theta\varepsilon^2.$$

For I_2, we use the Cauchy-Schwarz inequality to get

$$I_2=5E\left|\int_\theta^t\Big(b_1'(X_{\varepsilon,s},X_{\varepsilon,s-\tau})D_\theta\tilde{X}_{\varepsilon,s}-b_1'(x_s,x_{s-\tau})D_\theta Y_s\Big)ds\right|^2$$

$$\leq 10E\left|\int_\theta^t \left(b_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b_1'(x_s, x_{s-\tau})\right) D_\theta \tilde{X}_{\varepsilon,s} ds\right|^2$$
$$+ 10E\left|\int_\theta^t b_1'(x_s, x_{s-\tau})\big(D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s\big) ds\right|^2$$
$$\leq 10(t-\theta)\int_\theta^t \sqrt{E|b_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) - b_1'(x_s, x_{s-\tau})|^4}\sqrt{E|D_\theta \tilde{X}_{\varepsilon,s}|^4} ds$$
$$+ 10(t-\theta)\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds.$$

So, by using Lipschitz property of b_1' and the estimates (3.4) and (3.7), we obtain

$$I_2 \leq 10L^2(t-\theta)\int_\theta^t \sqrt{E|X_{\varepsilon,s} - x_s|^4 + E|X_{\varepsilon,s-\tau} - x_{s-\tau}|^4}\sqrt{E|D_\theta \tilde{X}_{\varepsilon,s}|^4} ds$$
$$+ 10(t-\theta)\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds$$
$$\leq C(t-\theta)^3\varepsilon^2 + C(t-\theta)\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds$$
$$\leq Ct^3\varepsilon^2 + C\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds,$$

where C is a positive constant not depending on t and ε. For the term I_3, by using the same arguments as in the estimate of I_2, we also have

$$I_3 = 5E\left|\int_{\theta+\tau}^t \left(b_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta \tilde{X}_{\varepsilon,s-\tau} - b_2'(x_s, x_{s-\tau}) D_\theta Y_{s-\tau}\right) ds\right|^2$$
$$\leq Ct^3\varepsilon^2 + C\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds.$$

On the other hand, by the Itô isometry and the estimate (3.7), we deduce

$$I_4 + I_5 = 5\int_\theta^t E\left|\sigma_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s}\right|^2 ds + 5\int_{\theta+\tau}^t E\left|\sigma_2'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s-\tau}\right|^2 ds$$
$$\leq 10L^2\int_\theta^t E|D_\theta X_{\varepsilon,s}|^2 ds \leq Ct\varepsilon^2.$$

We now insert the estimates of $I_k, k = 1, ..., 5$ into (3.17) and we obtain

$$E|D_\theta \tilde{X}_{\varepsilon,t} - D_\theta Y_t|^2 \leq Ct\varepsilon^2 + C\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds, \quad 0 \leq t \leq T.$$

Case 2. $(t-\tau)\vee 0 < \theta \leq t$. From the equations (3.6) and (3.11), we have

$$D_\theta \tilde{X}_{\varepsilon,t} - D_\theta Y_t = \sigma(X_{\varepsilon,\theta}, X_{\varepsilon,\theta-\tau}) - \sigma(x_\theta, x(\theta-\tau))$$
$$+ \int_\theta^t b_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta \tilde{X}_{\varepsilon,s} ds - \int_\theta^t b_1'(x_s, x_{s-\tau}) D_\theta Y_s ds$$
$$+ \int_\theta^t \sigma_1'(X_{\varepsilon,s}, X_{\varepsilon,s-\tau}) D_\theta X_{\varepsilon,s} dB_s, \quad 0 \leq t \leq T,$$

and hence,

$$E|D_\theta \tilde{X}_{\varepsilon,t} - D_\theta Y_t|^2$$

$$\leq I_1 + I_2 + I_4$$
$$\leq Ct\varepsilon^2 + C\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds,\ \ 0 \leq t \leq T.$$

Thus we always have

$$E|D_\theta \tilde{X}_{\varepsilon,t} - D_\theta Y_t|^2 \leq Ct\varepsilon^2 + C\int_\theta^t E|D_\theta \tilde{X}_{\varepsilon,s} - D_\theta Y_s|^2 ds,\ \ 0 \leq \theta \leq t \leq T,$$

where C is a positive constant not depending on t and ε. As a consequence,

$$E\|D\tilde{X}_{\varepsilon,t} - DY_t\|^2_{L^2[0,T]} = \int_0^t E|D_\theta \tilde{X}_{\varepsilon,t} - D_\theta Y_t|^2 d\theta$$
$$\leq Ct^2\varepsilon^2 + C\int_0^t E\|D\tilde{X}_{\varepsilon,s} - DY_s\|^2_{L^2[0,T]} ds,\ \ 0 \leq t \leq T.$$

An application of Gronwall's lemma gives us

$$E\|D\tilde{X}_{\varepsilon,t} - DY_t\|^2_{L^2[0,T]} \leq Ct^2\varepsilon^2,\ \ 0 \leq t \leq T.$$

This completes the proof of the proposition. □

Proof of Theorem 1.1 Fixed $\varepsilon \in (0, 1)$ and $t \in (0, T]$, we consider the random variables $F_1 = Y_t$ and $F_2 = \tilde{X}_{\varepsilon,t}$. Thanks to Propositions 3.4 and 3.5 we have

$$\|F_1 - F_2\|_{1,2} = \|\tilde{X}_{\varepsilon,t} - Y_t\|_{1,2}$$
$$= \left(E|\tilde{X}_{\varepsilon,t} - Y_t|^2 + E\|D\tilde{X}_{\varepsilon,t} - DY_t\|^2_{L^2[0,T]}\right)^{\frac{1}{2}}$$
$$\leq \left(Ct^2\varepsilon^2 + Ct^2\varepsilon^2\right)^{\frac{1}{2}} \leq Ct\varepsilon.$$

We recall that $D_\theta Y_t$ are deterministic for all $0 \leq \theta \leq t \leq T$. Hence, $D_r D_\theta F_1 = D_r D_\theta Y_t = 0,\ 0 \leq r, \theta \leq t \leq T$. On the other hand, we have $\|DF_1\|^2_{L^2[0,T]} = \|DY_t\|^2_{L^2[0,T]} = \mathrm{Var}(Y_t)$.

Now, for any measurable function g with $\|g\|_\infty = \sup_{x\in\mathbb{R}} |g(x)| \leq 1$, we apply Lemma 2.1 to get the following

$$|Eg(\tilde{X}_{\varepsilon,t}) - Eg(Y_t)| = |Eg(F_1) - Eg(F_2)|$$
$$\leq C\left(E\|DF_1\|^{-8}_{L^2[0,T]} E\left(\int_0^t\int_0^t |D_\theta D_r F_1|^2 d\theta dr\right)^2\right.$$
$$\left. + \left(E\|DF_1\|^{-2}_{L^2[0,T]}\right)^2\right)^{\frac{1}{4}} \|F_1 - F_2\|_{1,2} \leq \frac{Ct\varepsilon}{\sqrt{\mathrm{Var}(Y_t)}}.$$

Taking the supremum over all measurable functions g bounded by 1 yields

$$d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t) \leq \frac{Ct\varepsilon}{\sqrt{\mathrm{Var}(Y_t)}}.$$

The proof of Theorem 1.1 is complete. □

3.3 Proof of Theorem 1.2

Theorem 1.2 will be proved by using the interpolation formula (2.5). We will carry out the proof in three steps.

Step 1. In this step, we show that for all $p \geq 2$,

$$\lim_{\varepsilon\to 0} E\left|\frac{\tilde{X}_{\varepsilon,t}-Y_t}{\varepsilon}-\frac{1}{2}Z_t\right|^p = 0, \quad 0\leq t\leq T. \tag{3.18}$$

It follows from (3.14) that, for each $0 \leq t \leq T$,

$$\begin{aligned}\tilde{X}_{\varepsilon,t}-Y_t &= \int_0^t b_1'(x_s,x_{s-\tau})(\tilde{X}_{\varepsilon,s}-Y_s)ds\\ &\quad+\int_0^t b_2'(x_s,x_{s-\tau})(\tilde{X}_{\varepsilon,s-\tau}-Y_{s-\tau})ds+\frac{1}{2\varepsilon}\int_0^t R_{\varepsilon,s}ds\\ &\quad+\int_0^t \sigma_1'(x_s,x_{s-\tau})\left(X_{\varepsilon,s}-x_s\right)dB_s\\ &\quad+\int_0^t \sigma_2'(x_s,x_{s-\tau})\left(X_{\varepsilon,s}-x_s\right)dB_s+\frac{1}{2}\int_0^t Q_{\varepsilon,s}dB_s,\end{aligned}$$

where $R_{\varepsilon,s}$ is defined by (3.13) and $Q_{\varepsilon,s}$ is given by

$$\begin{aligned}Q_{\varepsilon,s} &= \sigma_{11}''\big(x_s+\xi_3(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_4(X_{\varepsilon,s-\tau}-x_{s-\tau})\big)(X_{\varepsilon,s}-x_s)^2\\ &\quad+2\sigma_{12}''\big(x_s+\xi_3(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_4(X_{\varepsilon,s-\tau}-x_{s-\tau})\big)(X_{\varepsilon,s}-x_s)(X_{\varepsilon,s-\tau}-x_{s-\tau})\\ &\quad+\sigma_{22}''\big(x_s+\xi_3(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_4(X_{\varepsilon,s-\tau}-x_{s-\tau})\big)\left(X_{\varepsilon,s-\tau}-x_{s-\tau}\right)^2\end{aligned}$$

for some random variables ξ_3, ξ_4 lying between 0 and 1. Hence, recalling (1.7), we get

$$\begin{aligned}\frac{\tilde{X}_{\varepsilon,t}-Y_t}{\varepsilon}-\frac{1}{2}Z_t &= \int_0^t b_1'(x_s,x_{s-\tau})\left(\frac{\tilde{X}_{\varepsilon,s}-Y_s}{\varepsilon}-\frac{1}{2}Z_t\right)ds\\ &\quad+\int_0^t b_2'(x_s,x_{s-\tau})\left(\frac{\tilde{X}_{\varepsilon,s-\tau}-Y_{s-\tau}}{\varepsilon}-\frac{1}{2}Z_{s-\tau}\right)ds\\ &\quad+\frac{1}{2}\int_0^t b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}^2ds\\ &\quad-\frac{1}{2}\int_0^t b_{11}''(x_s,x_{s-\tau})Y_s^2ds+\frac{1}{2}\int_0^t b_{22}''\big((x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}\\ &\quad+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\big)\big)\tilde{X}_{\varepsilon,s-\tau}^2ds\\ &\quad-\frac{1}{2}\int_0^t b_{22}''(x_s,x_{s-\tau})Y_{s-\tau}^2ds+\int_0^t b_{12}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}\right.\\ &\quad\left.+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}\tilde{X}_{\varepsilon,s-\tau}ds\\ &\quad-\int_0^t b_{12}''(x_s,x_{s-\tau})Y_sY_{s-\tau}ds+\int_0^t \sigma_1'(x_s,x_{s-\tau})\tilde{X}_{\varepsilon,s}dB_s\\ &\quad-\int_0^t \sigma_1'(x_s,x_{s-\tau})Y_sdB_s+\int_0^t \sigma_2'(x_s,x_{s-\tau})\tilde{X}_{\varepsilon,s-\tau}dB_s\\ &\quad-\int_0^t \sigma_2'(x_s,x_{s-\tau})Y_{s-\tau}dB_s+\frac{1}{2\varepsilon}\int_0^t Q_{\varepsilon,s}dB_s,\ \ 0\leq t\leq T.\end{aligned}$$

As a consequence, by using the Hölder and Burkholder—Davis–Gundy inequalities, we deduce

$$E\left|\frac{\tilde{X}_{\varepsilon,t}-Y_t}{\varepsilon}-\frac{1}{2}Z_t\right|^p \le C\int_0^t E\left|\frac{\tilde{X}_{\varepsilon,s}-Y_s}{\varepsilon}-\frac{1}{2}Z_t\right|^p ds + C(K_{1,\varepsilon}+K_{2,\varepsilon}),\ \ 0\le t\le T \tag{3.19}$$

for some $C>0$ and for all $\varepsilon\in(0,1)$, where $K_{1,\varepsilon}$, $K_{2,\varepsilon}$ are given by

$$\begin{aligned}
K_{1,\varepsilon} :=& \int_0^T E\left|\sigma_1'(x_s,x_{s-\tau})(\tilde{X}_{\varepsilon,s}-Y_s)\right|^p ds\\
&+\int_0^T E\left|\sigma_2'(x_s,x_{s-\tau})(\tilde{X}_{\varepsilon,s-\tau}-Y_{s-\tau})\right|^p ds+\frac{C}{\varepsilon^p}\int_0^T E|Q_{\varepsilon,s}|^p ds,\\
K_{2,\varepsilon} :=& \int_0^T E\left|b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}\right.\right.\\
&\left.\left.+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}^2-b_{11}''(x_s,x_{s-\tau})Y_s^2\right|^p ds\\
&+\int_0^T E\left|b_{22}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}\right.\right.\\
&\left.\left.+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s-\tau}^2-b_{22}''(x_s,x_{s-\tau})Y_{s-\tau}^2\right|^p ds\\
&+\int_0^T E\left|b_{12}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}\tilde{X}_{\varepsilon,s-\tau}\right.\\
&\left.-b_{12}''(x_s,x_{s-\tau})Y_sY_{s-\tau}\right|^p ds.
\end{aligned}$$

Using the same arguments as in the proof of (3.15), we have $E|Q_{\varepsilon,s}|^p\le Cs^p\varepsilon^{2p},\ \ 0\le s\le T$. Hence, from Proposition 3.4, it is easy to see that $K_{1,\varepsilon}\to 0$ as $\varepsilon\to 0$. To estimate $K_{2,\varepsilon}$, we observe that

$$\begin{aligned}
&\int_0^T E\left|b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}^2-b_{11}''(x_s,x_{s-\tau})Y_s^2\right|^p ds\\
&\le 2^{p-1}\int_0^T E\left|b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)(\tilde{X}_{\varepsilon,s}^2-Y_s^2)\right|^p ds\\
&\quad+2^{p-1}\int_0^T E\left|\left(b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)-b_{11}''(x_s,x_{s-\tau})\right)Y_s^2\right|^p ds\\
&\le 2^{p-1}L^p\int_0^T\sqrt{E|\tilde{X}_{\varepsilon,s}-Y_s|^{2p}E|\tilde{X}_{\varepsilon,s}+Y_s|^{2p}}ds\\
&\quad+2^{p-1}\int_0^T\sqrt{E\left|\left(b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)-b_{11}''(x_s,x_{s-\tau})\right)\right|^{2p}E|Y_s|^{4p}}ds.
\end{aligned}$$

We recall that $\sup_{0\le t\le T}E|Y_t|^p+\sup_{0\le t\le T}E|\tilde{X}_{\varepsilon,s}|^p<\infty$ for all $p>1$. Hence, by Proposition 3.4 and the dominated convergence theorem, we get

$$\begin{aligned}
&\int_0^T E\left|b_{11}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}\right.\right.\\
&\left.\left.-x_{s-\tau})\right)\tilde{X}_{\varepsilon,s}^2-b_{11}''(x_s,x_{s-\tau})Y_s^2\right|^p ds\to 0
\end{aligned}$$

as $\varepsilon\to 0$. Similarly, we also have

$$\int_0^T E\left|b_{22}''\left(x_s+\xi_1(X_{\varepsilon,s}-x_s),x_{s-\tau}+\xi_2(X_{\varepsilon,s-\tau}-x_{s-\tau})\right)\right.$$

$$\tilde{X}^2_{\varepsilon,s-\tau} - b''_{22}(x_s, x_{s-\tau})Y^2_{s-\tau}\Big|^p ds \to 0$$

and

$$\int_0^T E\left|b''_{12}\left(x_s + \xi_1(X_{\varepsilon,s} - x_s), x_{s-\tau} + \xi_2(X_{\varepsilon,s-\tau} - x_{s-\tau})\right)\right.$$
$$\left.\tilde{X}_{\varepsilon,s}\tilde{X}_{\varepsilon,s-\tau} - b''_{12}(x_s, x_{s-\tau})Y_s Y_{s-\tau}\right|^p ds \to 0.$$

Those imply that $K_{2,\varepsilon} \to 0$ as $\varepsilon \to 0$. From (3.19), an application of Gronwall's lemma gives us

$$E\left|\frac{\tilde{X}_{\varepsilon,t} - Y_t}{\varepsilon} - \frac{1}{2}Z_t\right|^p \le C(K_{1,\varepsilon} + K_{2,\varepsilon})e^{Ct},\ \ 0 \le t \le T.$$

This finishes the proof of Step 1.

Step 2. In this step, we prove (1.6). For simplicity, we write $\langle ., .\rangle$ instead of $\langle ., .\rangle_{L^2[0,T]}$ and $\|.\|$ instead of $\|.\|_{L^2[0,T]}$. Fix $t \in (0, T]$, by using the formula (2.5), we get

$$E[g(\tilde{X}_{\varepsilon,t})] - E[g(Y_t)] = E\left[\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz\delta\left(\frac{DY_t}{\|DY_t\|^2}\right)\right]$$
$$- E\left[\frac{g(\tilde{X}_{\varepsilon,t})\langle D\tilde{X}_{\varepsilon,t} - DY_t, DY_t\rangle}{\|DY_t\|^2}\right].$$

We recall that $\|DY_t\|^2 = \mathrm{Var}(Y_t) =: \beta_t^2$ and $\delta\left(\frac{DY_t}{\|DY_t\|^2}\right) = Y_t/\beta_t^2$. So we obtain

$$E[g(\tilde{X}_{\varepsilon,t})] - E[g(Y_t)] = \frac{1}{\beta_t^2}E\left[Y_t\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz\right] - \frac{1}{\beta_t^2}E\left[g(\tilde{X}_{\varepsilon,t})\langle D\tilde{X}_{\varepsilon,t} - DY_t, DY_t\rangle\right].$$

Then, for $\varepsilon \in (0, 1)$,

$$\begin{aligned}
&\frac{E[g(\tilde{X}_{\varepsilon,t})] - E[g(Y_t)]}{\varepsilon} - \frac{1}{2\beta_t^2}E\left[g(Y_t)Z_tY_t\right] + \frac{1}{2\beta_t^2}E\left[g(Y_t)\langle DZ_t, DY_t\rangle\right]\\
&= \frac{1}{\beta_t^2}E\left[\left(\frac{1}{\varepsilon}\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz - \frac{1}{2}g(Y_t)Z_t\right)Y_t\right]\\
&- \frac{1}{\beta_t^2}E\left[(g(\tilde{X}_{\varepsilon,t}) - g(Y_t))\left\langle\frac{D\tilde{X}_{\varepsilon,t} - DY_t}{\varepsilon}, DY_t\right\rangle\right]\\
&- \frac{1}{\beta_t^2}E\left[g(Y_t)\left\langle\frac{D\tilde{X}_{\varepsilon,t} - DY_t}{\varepsilon} - \frac{DZ_t}{2}, DY_t\right\rangle\right],\ \ 0 < t \le T.
\end{aligned} \tag{3.20}$$

We observe that

$$\begin{aligned}
\frac{1}{\varepsilon}\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz - \frac{1}{2}g(Y_t)Z_t &= \frac{\tilde{X}_{\varepsilon,t} - Y_t}{\varepsilon}\int_0^1 g(Y_t + z(\tilde{X}_{\varepsilon,t} - Y_t))dz - \frac{1}{2}g(Y_t)Z_t\\
&= \left(\frac{\tilde{X}_{\varepsilon,t} - Y_t}{\varepsilon} - \frac{Z_t}{2}\right)\int_0^1 g(Y_t + z(\tilde{X}_{\varepsilon,t} - Y_t))dz\\
&+ \frac{Z_t}{2}\int_0^1 (g(Y_t + z(\tilde{X}_{\varepsilon,t} - Y_t)) - g(Y_t))dz,
\end{aligned}$$

and hence,

$$E\left|\left(\frac{1}{\varepsilon}\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz - g(Y_t)Z_t\right)Y_t\right|$$
$$\leq \|g\|_\infty E\left|\left(\frac{\tilde{X}_{\varepsilon,t}-Y_t}{\varepsilon} - \frac{Z_t}{2}\right)Y_t\right|$$
$$+\frac{1}{2}E\left|Z_tY_t\int_0^1 (g(Y_t+z(\tilde{X}_{\varepsilon,t}-Y_t)) - g(Y_t))dz\right|.$$

Because the random variables Y_t and Z_t belong to $L^2(\Omega)$, we have

$$\lim_{\varepsilon\to 0} E\left|\left(\frac{\tilde{X}_{\varepsilon,t}-Y_t}{\varepsilon} - \frac{Z_t}{2}\right)Y_t\right| = 0 \quad \text{by the limit (3.18).}$$

By the dominated convergence theorem, we also have

$$\lim_{\varepsilon\to 0} E\left|Z_tY_t\int_0^1 (g(Y_t+z(\tilde{X}_{\varepsilon,t}-Y_t)) - g(Y_t))dz\right| = 0$$

So it holds that

$$\lim_{\varepsilon\to 0} E\left[\left(\frac{1}{\varepsilon}\int_{Y_t}^{\tilde{X}_{\varepsilon,t}} g(z)dz - \frac{1}{2}g(Y_t)Z_t\right)Y_t\right] = 0. \tag{3.21}$$

On the other hand, we have

$$E\left[(g(\tilde{X}_{\varepsilon,t}) - g(Y_t))\left\langle\frac{D\tilde{X}_{\varepsilon,t}-DY_t}{\varepsilon}, DY_t\right\rangle\right] \leq \frac{1}{\beta_t}E\left[\frac{|g(\tilde{X}_{\varepsilon,t}) - g(Y_t)|\|D\tilde{X}_{\varepsilon,t}-DY_t\|}{\varepsilon}\right]$$
$$\leq \frac{1}{\beta_t}(E|g(\tilde{X}_{\varepsilon,t}) - g(Y_t)|^2)^{\frac{1}{2}}\left(\frac{E\|D\tilde{X}_{\varepsilon,t}-DY_t\|^2}{\varepsilon^2}\right)^{\frac{1}{2}}.$$

Once again, by (3.16) and the dominated convergence theorem, we derive

$$\lim_{\varepsilon\to 0} E\left[(g(\tilde{X}_{\varepsilon,t}) - g(Y_t))\left\langle\frac{D\tilde{X}_{\varepsilon,t}-DY_t}{\varepsilon}, DY_t\right\rangle\right] = 0. \tag{3.22}$$

In view of Lemma 1.2.3 in [14], it follows from (3.16) and (3.18) that

$$\lim_{\varepsilon\to 0} E\left[g(Y_t)\left\langle\frac{D\tilde{X}_{\varepsilon,t}-DY_t}{\varepsilon} - \frac{DZ_t}{2}, DY_t\right\rangle\right] = 0. \tag{3.23}$$

Combining (3.20)–(3.23) yields

$$\lim_{\varepsilon\to 0}\frac{E[g(\tilde{X}_{\varepsilon,t})] - E[g(Y_t)]}{\varepsilon} = \frac{1}{2\beta_t^2}E\left[g(Y_t)Z_tY_t\right] - \frac{1}{2\beta_t^2}E\left[g(Y_t)\langle DZ_t, DY_t\rangle\right].$$

Then we obtain (1.6) by using the duality relationship (2.2).

Step 3. In this step, we verify (1.8). We consider the function $g(x) = \text{sign}\left(E\left[\delta\left(Z_tDY_t\right)\big|Y_t = x\right]\right)$ for $x \in \mathbb{R}$. Then, $\|g\|_\infty \leq 1$ and by the routine approximation argument, we can approximate g by a sequence $(g_n(x))_{n\geq 1}$ of continuous functions bounded by 1. Indeed, for example, we can use the following sequence

$$g_n(x) = \int_{-\infty}^{\infty} 1\!\!1_{\{|y|<n\}}g(y)\rho_n(x-y)dy, \ n \geq 1,$$

where ρ_n is the standard mollifier: $\rho_n(x) = n\rho(nx)$, where $\rho(x) = C\mathbb{1}_{\{|x|<1\}}e^{\frac{1}{x^2-1}}$ and C is a constant such that $\int_{-\infty}^{\infty}\rho(x)dx = 1$.

We obtain from (1.6) that, for all $n \geq 1$,

$$\lim_{\varepsilon\to 0}\frac{d_{\text{TV}}(\tilde{X}_{\varepsilon,t}, Y_t)}{\varepsilon} \geq \frac{1}{2\text{Var}(Y_t)}E\left[g_n(Y_t)\delta\left(Z_t DY_t\right)\right]$$
$$= \frac{1}{2\text{Var}(Y_t)}E\left[g_n(Y_t)E[\delta\left(Z_t DY_t\right)|Y_t]\right].$$

Letting $n \to \infty$ we obtain

$$\lim_{\varepsilon\to 0}\frac{d_{\text{TV}}(\tilde{X}_{\varepsilon,t}, Y_t)}{\varepsilon} \geq \frac{1}{2\text{Var}(Y_t)}E\left[g(Y_t)E[\delta\left(Z_t DY_t\right)|Y_t]\right]$$
$$= \frac{1}{2\text{Var}(Y_t)}E|E[\delta\left(Z_t DY_t\right)|Y_t]|.$$

The proof of Theorem 1.2 is complete. □

4 Conclusion and Example

The central limit theorem for stochastic dynamical systems with small noise has been extensively studied. However, most of the existing results are qualitative. In this paper, we used the techniques of Malliavin calculus to provide quantitative total variation estimates in the central limit theorem for stochastic differential delay equations with small noises. The significance of our results lie in the fact that we not only obtain explicit estimates for the rate of convergence, but also prove the optimality of these rates of convergence.

We also would like to emphasize that Lemma 2.1 is a key tool in the present paper. The proof of this lemma heavily relies on dimension one and hence, our results only hold true for one dimensional equations. The generalization to higher dimensions will be a difficult and interesting problem.

Example 4.1 Let us provide an explicit example to illustrate the theory. For any $\varepsilon \in (0, 1)$, we consider the following equation

$$\begin{cases} X_{\varepsilon,t} = 1 + \int_0^t X_{\varepsilon,s}ds + \varepsilon\int_0^t \left(2 + \sin(X_{\varepsilon,s} + X_{\varepsilon,s-\tau})\right)dB_s, & t \in [0, T] \\ X_{\varepsilon,t} = e^t,\ t \in [-\tau, 0], \end{cases} \tag{4.1}$$

It is easy to see that the functions $b(x, y) = x$ and $\sigma(x, y) = 2 + \sin(x + y)$ satisfy Assumption 1.1. Moreover, we have

$$\begin{cases} x_t = 1 + \int_0^t x_s ds,\ t \in [0, T] \\ x_t = e^t, \quad t \in [-\tau, 0]. \end{cases} \tag{4.2}$$

and

$$\begin{cases} Y_t = \int_0^t Y_s ds + \int_0^t (2 + \sin(x_s + x_{s-\tau}))\,dB_s, & t \in [0, T] \\ Y_t = 0, \quad t \in [-\tau, 0]. \end{cases} \tag{4.3}$$

Solving the equations (4.2) and (4.3) gives us $x_t = e^t$ for $t \in [-\tau, T]$ and

$$Y_t = \int_0^t e^{t-s}\left(2 + \sin(e^s + e^{s-\tau})\right)dB_s,\ \ t \in [0, T].$$

Furthermore, we have $\mathrm{Var}(Y_t) = \int_0^t e^{2(t-s)} \left(2 + \sin(e^s + e^{s-\tau})\right)^2 ds \geq t, \quad t \in [0, T]$. We now define $\tilde{X}_{\varepsilon,t} := \frac{X_{\varepsilon,t} - x_t}{\varepsilon} = \frac{X_{\varepsilon,t} - e^t}{\varepsilon}$, $t \in [-\tau, T]$. Then, thanks to Theorem 1.1, we conclude that

$$d_{\mathrm{TV}}(\tilde{X}_{\varepsilon,t}, Y_t) \leq \frac{Ct\varepsilon}{\sqrt{\mathrm{Var}(Y_t)}} \leq Ct^{\frac{1}{2}}\varepsilon \;\forall \varepsilon \in (0,1), 0 < t \leq T,$$

where C is a positive constant not depending on t and ε.

Acknowledgements The authors would like to thank the anonymous referees for valuable comments which led to the improvement of this work.

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Barrera, G., Jara, M.: Abrupt convergence for stochastic small perturbations of one dimensional dynamical systems. J. Stat. Phys. **163**(1), 113–138 (2016)
2. Bourguin, S., Spiliopoulos, K.: Quantitative fluctuation analysis of multiscale diffusion systems via Malliavin calculus. arXiv:2301.09005
3. Chauvière, A., Preziosi, L., Verdier, C.: Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling. CRC Press, Boca Raton (2010)
4. Chiarini, A., Fischer, M.: On large deviations for small noise Itô processes. Adv. Appl. Probab. **46**(4), 1126–1147 (2014)
5. Dung, N.T., Son, T.C.: Lipschitz continuity in the Hurst index of the solutions of fractional stochastic Volterra integro-differential equations. Stoch. Anal. Appl. **41**(4), 693–712 (2023)
6. Fatheddin, P.: Central limit theorem for a class of SPDEs. J. Appl. Probab. **52**(3), 786–796 (2015)
7. Fouque, J.P., Papanicolaou, G., Sircar, K.R.: Derivatives in Financial Markets with Stochastic Volatility. Cambridge University Press, Cambridge (2000)
8. Freidlin, M.I., Wentzell, A.D.: *Random perturbations of dynamical systems.* Translated from the Russian by Joseph Szücs. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 260. Springer, New York (1984)
9. Gloter, A., Sørensen, M.: Estimation for stochastic differential equations with a small diffusion coefficient. Stoch. Process. Appl. **119**(3), 679–699 (2009)
10. Hong, W., Li, S., Liu, W., Sun, X.: Central limit type theorem and large deviation principle for multi-scale McKean–Vlasov SDEs. Probab. Theory Relat. Fields **187**, 1–69 (2023)
11. Khasminskii, R.Z., Yin, G.: On averaging principles: an asymptotic expansion approach. SIAM J. Math. Anal. **35**(6), 1534–1560 (2004)
12. Kutoyants, Y.: Identification of Dynamical Systems with Small Noise. Mathematics and its Applications, 300. Kluwer Academic Publishers Group, Dordrecht (1994)
13. Budhiraja, A., Dupuis, P., Ganguly, A.: Moderate deviation principles for stochastic differential equations with jumps. Ann. Probab. **44**(3), 1723–1775 (2016)
14. Nualart, D.: The Malliavin Calculus and Related Topics. Probability and Its Applications, 2nd edn. Springer, Berlin (2006)
15. Pajor-Gyulai, Z., Salins, M.: On dynamical systems perturbed by a null-recurrent motion: the general case. Stoch. Process. Appl. **127**(6), 1960–1997 (2017)
16. Picard, J.: Efficiency of the extended Kalman filter for nonlinear systems with small noise. SIAM J. Appl. Math. **51**(3), 843–885 (1991)
17. Son, T.C., Dung, N.T., Tan, N.V., Cuong, T.M., Thao, H.T.P., Tung, P.D.: Weak convergence of delay SDEs with applications to Carathéodory approximation. Discret. Contin. Dyn. Syst. Ser. B **27**(9), 4725–4747 (2022)

18. Suo, Y., Tao, J., Zhang, W.: Moderate deviation and central limit theorem for stochastic differential delay equations with polynomial growth. Front. Math. China **13**(4), 913–933 (2018)
19. Spiliopoulos, K.: Fluctuation analysis and short time asymptotics for multiple scales diffusion processes. Stoch. Dyn. **14**(3), 1350026 (2014)
20. Uchida, M., Yoshida, N.: Asymptotic expansion for small diffusions applied to option pricing. Stat. Inference Stoch. Process. **7**(3), 189–223 (2004)