Least squares estimation for path-distribution dependent stochastic differential equations

Panpan Ren and Jiang-Lun Wu

Department of Mathematics, Swansea University, Bay Campus, Swansea SA1 8EN, UK

E-mails: 673788@swansea.ac.uk j.l.wu@swansea.ac.uk

May 13, 2021

Abstract

We study a least squares estimator for an unknown parameter in the drift coefficient of a path-distribution dependent stochastic differential equation involving a small dispersion parameter $\varepsilon > 0$. The estimator, based on n (where $n \in \mathbb{N}$) discrete time observations of the stochastic differential equation, is shown to be convergent weakly to the true value as $\varepsilon \to 0$ and $n \to \infty$. This indicates that the least squares estimator obtained is consistent with the true value. Moreover, we obtain the rate of convergence and derive the asymptotic distribution of least squares estimator.

AMS subject Classification: 62F12, 62M05, 60G52, 60J75

Keywords: Path-distribution dependent stochastic differential equation, least squares estimator, consistency, asymptotic distribution.

1 Introduction

Nowadays, stochastic differential equations (SDEs) are widely used in modelling time evolution of dynamical systems influenced by random noise, see, the monographs [6, 12, 29, 32] (and references therein). Usually, there exist unknown parameters in such modelled systems, such as those stochastic models with comparably easier structured stochastic differential equations involving unknown quantities (see, [2, P.2-4]). Fundamental issues are to estimate certain parameters (i.e., deterministic quantities) appearing in the stochastic models by certain observations (or by experimental data). Viewing the drift part of the SDEs as the averaging evolution of the systems, estimating the drift parameter of SDEs is hence an important topic. To approach the true value of the unknown parameter, the asymptotic approach to statistical estimation is frequently taken an advantage due to its general applicability and relative simplicity (cf. [2]). As we know, the estimations upon the unknown quantities are generally based on continuous-time or discrete-time observations. Whereas, the parameter estimation relied on continuous-time observations is a mathematical idealisation although there is a vast literature concerned with such topic. On the other hand, no measuring device can follow continuously the sample paths of the diffusion processes involved, which are indeed rather tricky. Hence, in practice the investigation on the parameter estimations with the help of discrete-time observations has been received much more attention recently. Most importantly, the parameter estimation by the aid of discrete-time observations can be implemented conveniently with a powerful theory of simulation schemes and numerical analysis of diffusion processes.

So far, there are numerous methods to investigate the parameter estimations on the unknown parameters in the drift coeffecents; see, [17, 23, 30, 33] by maximum likelihood estimator (MLE for short), [4, 14, 17, 30] via least squares estimator (LSE for abbreviation), and [27] through trajectory-fitting estimator, to name a few. Diffusion processes with small noises have been applied considerably in mathematical finance, see, [16, 18, 35, 42] and references within. In particular, Kutoyants [18] investigated the issues upon parametric and nonparametric identification. Moreover, the asymptotic behavior of parametric estimators (e.g. the maximum likelihood, the Bayes and the minimum distance estimators) and the nonparametric estimators (e.g. the kernel-type estimators) was discussed.

In the past forty years, the asymptotic theory on parameter estimations for diffusion processes with small noises has also been developed very well, see, for instance, [7, 21, 22, 34, 36, 37] for SDEs driven by Lévy processes with arbitrary moments, and [8, 24, 25] for SDEs driven by α-stable Lévy noises which enjoy heavy tail properties.

On the other hand, from the stochastic modelling perspective and diverse demanding in practical problems, there has been increasing interest on studying stochastic differential equations with path-distribution coefficients, see e.g. [9, 10, 40] (and references therein). The distribution-dependent SDEs are also named as McKean-Vlasov SDEs or mean-filed SDEs, which have been studied intensively in the literature, see e.g. [5, 20] and references therein. Such kind of SDEs has been applied successfully in stochastic differential games and stochastic optimal optimisation, see, [19] and references within. Although McKean-Vlasov SDEs have been applied diffusively in different research areas, so far there is little work on parameter estimations except the existing literature [41], to the best of our knowledge. Recently in [31] we carried out least squares estimation for path-distribution dependent SDEs with monotone condition via discrete-time observation. In the present paper, we are concerned with the LSE problem for the path-distribution stochastic differential equations with small dispersion noise and involving unknown parameter in the drift. Our key start point is the associated discrete-time observations of path-distribution dependent SDEs (see (2.1) below). We then investigate parameter estimation for McKean-Vlasov SDEs which are not only path-dependent but also dependent on the law of the path. Since the state space of the window process is infinite dimensional, some new procedures need to be put forward. We succeeded the task by interpolating the discrete-time observations (see (2.4) below for more details). Moreover, our proposed estimator has wide applications in e.g. derivative pricing, future data prediction, stochastic filters, stochastic mean-field games which will appear in our forthcoming paper. For more applications of parameter estimations, see [1, 16, 17, 18].

Before closing the introduction part, we would like to explicate a bit more of the relation of the current paper with our previous paper [31]. The aim of the present paper is to construct a classical (or explicit) Euler-Maruyama approximation scheme to path-distribution dependent stochastic differential equations and then to derive the least squares estimation for the drift parameter of the path-distribution dependent stochastic differential equations. For stochastic differential equations with certain complexity like, for instance, the McKean-Vlasov stochastic differential equations, it is comparably easier to derive modified approximating schemes, such as the tamed Euler-Maruyama scheme for the McKean-Vlasov stochastic differential equations carried out in our previous paper [31]. Establishing convergent Euler-Maruyama approximations are difficulty for path-distribution dependent stochastic differential equations, as explored for various SDEs in the literature (see for instance [26] and references therein). It is even harder and tedious to obtain the convergence of the classical Euler-Maruyama approximations for our path-distribution dependent stochastic differential equations, due to the measure dependence coefficients. On the other hand, due to succinct and comparably easier formulation, classical Euler-Maruyama approximation schemes to stochastic differential equations, including our path-distribution dependent stochastic differential equations, are important and efficiently useful in practice. This motivates us to establish a classical Euler-Maruyama approximation scheme and to obtain more explicit contrast functions for the least squares estimation. While in our previous paper [31], we constructed a tamed Euler-Maruyama scheme for the McKean-Vlasov stochastic differential equations with monotone coefficients, and the associated contrast functions look more complicated for simulation. In the present paper, we establish classical Euler-Maruyama approximations for path-distribution dependent stochastic differential equations with Lipschitz coefficients and we take the advantage of the obtained classical Euler-Maruyama approximation scheme and the associated succinct contrast functions, several numerical simulations were implemented to support our theoretical result. We have done the numerical simulation to support our theoretical result. In summary, our present paper focuses more on applicable aspects by establishing a classical Euler-Maruyama approximation scheme for stochastic equations with (typical) Lipschitz coefficients, while the previous paper [31] dealt with a tamed Euler-Maruyama approximation scheme for stochastic equations with monotone coefficients (i.e., slightly more general coefficients).

The rest of the paper is arranged as follows. In Section 2, we introduce some notations, present the framework of our paper, and construct the LSE; Section 3 is devoted to the consistency of LSE. Section 4 focus on the asymptotic distribution of LSE. In Section5 we provide an example(named as Example 5.1) to illustrate our main results (i.e., theorems 3.1 and 4.1). Moreover we implement numerical simulations to support our theoretical analysis.

Throughout this paper, we use $c > 0$ for a generic constant which may change from line to line.

2 Preliminaries

sec0

We start with some notation and terminology which will be used later. For $d, m \in \mathbb{N}$, the set of all positive integers, let $(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|)$ be the d-dimensional Euclinean space with the inner product $\langle\cdot,\cdot\rangle$ inducing the norm $|\cdot|$ and $\mathbb{R}^d \otimes \mathbb{R}^m$ the collection of all $d \times m$ matrixes with real entries, which is endowed with the Hilbert-Schmidt norm $\|\cdot\|$. $\mathbf{0} \in \mathbb{R}^d$ denotes the zero vector. For a matrix A, A^* denotes the transpose of A. Concerning a square matrix A, A^{-1} means the inverse of A provided that $\det A \neq 0$. For $p \in \mathbb{N}$, let Θ be an open bounded convex subset of $\mathbb{R}^p$, and $\overline{\Theta}$ the closure of Θ. For $r > 0$ and $x \in \mathbb{R}^p$, $B_r(x)$ represents the closed ball centered at x with the radius r. For $z \in \mathbb{R}^d$, δ_z denotes Dirac's delta measure or unit mass at the point z. For a real number $a > 0$, $\lfloor a \rfloor$ stands for the integer part of a. For a random variable ξ, $\mathscr{L}_\xi$ denotes its law. For a fixed finite number $r_0 > 0$, $\mathscr{C} := C([-r_0, 0]; \mathbb{R}^d)$ means the family of all continuous functions $f : [-r_0, 0] \to \mathbb{R}^d$, which is a Polish (i.e., separable, complete metric) space under the uniform norm $\|f\|_\infty := \sup_{-r_0 \le \theta \le 0} |f(\theta)|$. Generally speaking, $r_0 > 0$ is named as the length of memory. For a continuous map $f : [-r_0, \infty) \to \mathbb{R}^d$ and $t \ge 0$, let $f_t \in \mathscr{C}$ be such that $f_t(\theta) = f(t + \theta)$ for $\theta \in [-r_0, 0]$. In general, $(f_t)_{t \ge 0}$ is called the window (or segment) process of $(f(t))_{t \ge -r_0}$. $\mathcal{P}_2(\mathscr{C})$ stands for the space of all probability measures on $\mathscr{C}$ with the finite second-order moment, i.e.,

$\mu(\|\cdot\|_\infty^2) := \int_{\mathscr{C}} \|\zeta\|_\infty^2 \mu(\mathrm{d}\zeta) < \infty$ for $\mu \in \mathcal{P}_2(\mathscr{C})$. Define the Wasserstein distance $\mathbb{W}_2$ on $\mathcal{P}_2(\mathscr{C})$ by

$$\mathbb{W}_2(\mu,\nu) = \inf_{\pi \in \mathcal{C}(\mu,\nu)} \Big(\int_{\mathscr{C}} \int_{\mathscr{C}} \|\zeta_1 - \zeta_2\|_\infty^2 \pi(\mathrm{d}\zeta_1, \mathrm{d}\zeta_2) \Big)^{1/2}, \quad \mu,\nu \in \mathcal{P}_2(\mathscr{C}),$$

where $\mathcal{C}(\mu,\nu)$ signifies the collection of all probability measures on $\mathscr{C} \times \mathscr{C}$ with marginals μ and ν (i.e., $\pi \in \mathcal{C}(\mu,\nu)$ such that $\pi(\cdot,\mathscr{C}) = \mu(\cdot)$ and $\pi(\mathscr{C},\cdot) = \nu(\cdot)$), respectively. Under the distance $\mathbb{W}_2$, $\mathcal{P}_2(\mathscr{C})$ is a Polish space; see, [3, Lemma 5.3 & Theorem 5.4]. Let $(B(t))_{t\geq 0}$ be an m-dimensional Brownian motion defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with the filtration $(\mathscr{F}_t)_{t\geq 0}$ satisfying the usual condition (i.e., $\mathscr{F}_0$ contains all $\mathbb{P}$-null sets and $\mathscr{F}_t = \mathscr{F}_{t+} := \bigcap_{s>t} \mathscr{F}_s$).

Throughout the paper, we fix the time horizon $T > 0$. For the scale parameter $\varepsilon \in (0,1)$, we consider a path-distribution dependent SDE on $(\mathbb{R}^d, \langle \cdot, \cdot \rangle, |\cdot|)$ in the form

eq1

$$\text{(2.1)} \qquad \mathrm{d}X^\varepsilon(t) = b(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}, \theta)\mathrm{d}t + \varepsilon\, \sigma(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon})\mathrm{d}B(t), \quad t \in (0,T], \qquad X_0^\varepsilon = \xi \in \mathscr{C},$$

where $b : \mathscr{C} \times \mathcal{P}_2(\mathscr{C}) \times \Theta \to \mathbb{R}^d$ and $\sigma : \mathscr{C} \times \mathcal{P}_2(\mathscr{C}) \to \mathbb{R}^d \times \mathbb{R}^m$. In literature, the drift coefficient b is also called the trend coefficient or damping coefficient or translation coefficient, and the diffusion coefficient is also named as volatility coefficient. In (2.1), we assume that the drift b and the diffusion σ are known apart from the parameter $\theta \in \Theta$ and we stipulate that $\theta_0 \in \Theta$ is the true value of $\theta \in \Theta$.

For any $\zeta_1, \zeta_2 \in \mathscr{C}$ and $\mu, \nu \in \mathcal{P}_2(\mathscr{C})$, we assume that

(A1) There exist $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$ such that

$$\sup_{\theta \in \overline{\Theta}} |b(\zeta_1, \mu, \theta) - b(\zeta_2, \nu, \theta)|^2 \leq \alpha_1 \|\zeta_1 - \zeta_2\|_\infty^2 + \alpha_2 \mathbb{W}_2(\mu,\nu)^2,$$

and

$$\|\sigma(\zeta_1, \mu) - \sigma(\zeta_2, \nu)\|^2 \leq \beta_1 \|\zeta_1 - \zeta_2\|_\infty^2 + \beta_2 \mathbb{W}_2(\mu,\nu)^2;$$

(A2) $(\sigma\sigma^*)(\zeta_1, \mu)$ is invertible, and there exists an $L_1 > 0$ such that

$$\|(\sigma\sigma^*)^{-1}(\zeta_1, \mu) - (\sigma\sigma^*)^{-1}(\zeta_2, \nu)\| \leq L_1\{\|\zeta_1 - \zeta_2\|_\infty + \mathbb{W}_2(\mu,\nu)\};$$

(A3) For the initial value $X_0^\varepsilon = \xi$, there exists an $L_2 > 0$ such that

$$|\xi(t) - \xi(s)| \leq L_2 |t - s|, \quad t, s \in [-r_0, 0].$$

(A4) There exist constants $K_0, p_0 > 0$ such that

$$|b(\zeta, \mu, \theta_1) - b(\zeta, \mu, \theta_2)| \leq K_0(1 + \|\zeta\|_\infty + \mathbb{W}_2(\mu, \delta_{\zeta_0}))^{p_0} |\theta_1 - \theta_2|, \quad \theta_1, \theta_2 \in \overline{\Theta},$$

where $\zeta_0(s) \equiv \mathbf{0} \in \mathbb{R}^d$ for any $s \in [-r_0, 0]$.

We further assume that

(B1) There exists $K_1 > 0$ such that

$$\sup_{\theta \in \overline{\Theta}} \|(\nabla_\theta b)(\zeta_1, \mu, \theta) - (\nabla_\theta b)(\zeta_2, \nu, \theta)\| \leq K_1\{\|\zeta_1 - \zeta_2\|_\infty + \mathbb{W}_2(\mu,\nu)\},$$

where $(\nabla_\theta b)$ means the gradient operator w.r.t. the third spatial variable.

(B2) There exists $K_2 > 0$ such that

$$\sup_{\theta\in\overline{\Theta}} \|(\nabla_\theta(\nabla_\theta b^*))(\zeta_1,\mu,\theta) - (\nabla_\theta(\nabla_\theta b^*))(\zeta_2,\nu,\theta)\| \le K_2\{\|\zeta_1-\zeta_2\|_\infty + \mathbb{W}_2(\mu,\nu)\}.$$

Definition 2.1. A continuous adapted process $(X_t^\varepsilon)_{t\ge0}$ on $\mathscr{C}$ is called strong solution of (2.1) if

$$\mathbb{E}\|X_t^\varepsilon\|_\infty^2 + \int_0^t \mathbb{E}\{|b(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon},\theta)| + \|\sigma(X_t^\varepsilon,\mathscr{L}_{X_t^\varepsilon})\|^2)\mathrm{d}s\} < \infty, \quad t\ge 0,$$

and $X^\varepsilon(t) = X_t^\varepsilon(0)$ satisfies $\mathbb{P}-a.s$

$$X^\varepsilon(t) = X_t^\varepsilon(0) + \int_0^t b(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon},\theta)\mathrm{d}s + \int_0^t \sigma(X_t^\varepsilon,\mathscr{L}_{X_t^\varepsilon})\mathrm{d}B(s), \quad t\ge 0.$$

Before we move forward, let's give some remarks. Under (**A1**), (2.1) admits a unique strong solution $(X^\varepsilon(t))_{t\in[-r_0,T]}$; see, for instance, [10, Theorem 3.1]. For more details on existence and uniqueness of strong solutions to distribution-dependent SDEs, we would like to refer to [5, 28, 40] and references within. As far as existence and uniqueness of weak solutions, please consult [13, 20, 39] for reference. (**B1**) and (**B2**) are imposed merely to discuss the asymptotic distribution of LSE constructed below; see Theorem 4.1. (**A3**) is put to analyze continuity of the window process associated with (2.2); see Lemma 3.3 for more details. Obviously, (**A2**) holds provided that $\sigma(\cdot,\cdot)\equiv\sigma\in\mathbb{R}^d\otimes\mathbb{R}^m$, a constant matrix, such that $\sigma\sigma^*$ is invertible. Moreover, for the scalar setting of (2.1), (**A2**) is also true in case of $\sigma(x,\mu) = 1+|x|$ for any $x\in\mathbb{R}$ and $\mu\in\mathcal{P}_2(\mathbb{R})$.

Without loss of generality, given the stepsize $\delta>0$, let define $M=\frac{r_0}{\delta}$, $n=\frac{T}{\delta}$ for each integers $n,M\in\mathbb{N}$ sufficiently large. Suppose that the solution process $(X^\varepsilon(t))_{t\in[-r_0,T]}$ is observed at regularly spaced time points $t_k = k\delta$ for $k=0,1,\cdots,n$. In this paper, our goal is to investigate the LSE on the parameter $\theta\in\Theta$ based on the sampling data $(X^\varepsilon(t_k))_{k=0}^n$ with small dispersion ε and large sample size n (i.e., small step size δ).

Motivated by [24, 25, 34], for our present setting we construct the following contrast function

eq2 (2.2)
$$\Psi_{n,\varepsilon}(\theta) = \varepsilon^{-2}\delta^{-1}\sum_{k=1}^n P_k^*(\theta)\Lambda_{k-1}^{-1}P_k(\theta).$$

Herein, for $k=1,\cdots,n$,

w1 (2.3)
$$P_k(\theta) := X^\varepsilon(t_k) - X^\varepsilon(t_{k-1}) - b(\widehat{X}_{t_{k-1}}^\varepsilon,\mathscr{L}_{\widehat{X}_{t_{k-1}}^\varepsilon},\theta)\delta \quad \text{and} \quad \Lambda_k := (\sigma\sigma^*)(\widehat{X}_{t_k}^\varepsilon,\mathscr{L}_{\widehat{X}_{t_k}^\varepsilon}),$$

where $\widehat{X}_{k\delta}^\varepsilon = \{\widehat{X}_{k\delta}^\varepsilon(s) : -r_0\le s\le 0\}$ is a $\mathscr{C}$-valued random variable defined as follows: for any $s\in[-(i+1)\delta,-i\delta]$, $i=1,\cdots,M-1$,

w2 (2.4)
$$\widehat{X}_{k\delta}^\varepsilon(s) = X^\varepsilon((k-i)\delta) + \frac{s+i\delta}{\delta}\{X^\varepsilon((k-i)\delta) - X^\varepsilon((k-i-1)\delta)\},$$

i.e., $\widehat{X}_{k\delta}^\varepsilon$ is the linear interpolation of $X^\varepsilon((k-M)\delta),\cdots,X^\varepsilon(k\delta)$. To achieve the LSE of $\theta\in\Theta$, it suffices to choose an argument $\widehat{\theta}_{n,\varepsilon}\in\Theta$ such that

eq3 (2.5)
$$\Psi_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon}) = \min_{\theta\in\Theta}\Psi_{n,\varepsilon}(\theta).$$

Next, we write $\widehat{\theta}_{n,\varepsilon} \in \Theta$ satisfying (2.5) by

$$\widehat{\theta}_{n,\varepsilon} = \arg\min_{\theta\in\Theta} \Psi_{n,\varepsilon}(\theta).$$

Set

$$\Phi_{n,\varepsilon}(\theta) := \varepsilon^2(\Psi_{n,\varepsilon}(\theta) - \Psi_{n,\varepsilon}(\theta_0)).$$

It follows from (2.5) that

eq4 $$\Phi_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon}) = \min_{\theta\in\Theta} \Phi_{n,\varepsilon}(\theta). \tag{2.6}$$

Likewise, we reformulate $\widehat{\theta}_{n,\varepsilon} \in \Theta$ ensuring (2.6) to hold true as

eq5 $$\widehat{\theta}_{n,\varepsilon} = \arg\min_{\theta\in\Theta} \Phi_{n,\varepsilon}(\theta). \tag{2.7}$$

Through the whole paper, $\widehat{\theta}_{n,\varepsilon}$ such that (2.7) holds is named as the LSE of $\theta \in \Theta$.

Before we end this section, we give some remarks.

Remark 2.1. For an invertible $\sigma(\cdot,\cdot) \in \mathbb{R}^d \otimes \mathbb{R}^d$, observe that

$$\frac{\triangle B_k}{\sqrt{\delta}} \approx \frac{1}{\varepsilon^{-1}\sqrt{\delta}}\sigma^{-1}(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}})P_k(\theta)$$

provided that the stepsize $\delta \in (0,1)$ is sufficiently small. Then, we can design the contrast function $\Psi_{n,\varepsilon}(\cdot)$ as

$$\begin{aligned}\Psi_{n,\varepsilon}(\theta) &= \varepsilon^{-2}\delta^{-1}|\sigma^{-1}(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}})P_k(\theta)|^2 \\ &= \varepsilon^{-2}\delta^{-1}P_k^*(\theta)((\sigma^{-1})^*\sigma^{-1})(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}})P_k(\theta) \\ &= \varepsilon^{-2}\delta^{-1}P_k^*(\theta)\Lambda_{k-1}^{-1}P_k(\theta).\end{aligned}$$

Motivated by the invertible setup above, we establish the contrast function for the setting that the diffusion $\sigma(\cdot,\cdot)$ need not to be invertible; see (2.2) for further details. On the other hand, if $b(\cdot,\cdot,\theta)$ is explicit w.r.t. the parameter θ, then the LSE $\widehat{\theta}_{n,\varepsilon}$ can indeed be obtained by Fermat's theorem.

Remark 2.2. Formally, the contrast function $\Psi_{n,\varepsilon}$ can be defined as in (2.2) with $\widehat{X}^\varepsilon_{t_k}$ replaced by $X^\varepsilon_{t_k}$. Nevertheless, $X^\varepsilon_{t_k}$ cannot be available provided that $(X^\varepsilon(t))_{t\in[0,T]}$ is observed only at the points $t = k\delta$. So, in our paper, we approximate the window process $X^\varepsilon_{t_k}$ via the linear interpolation. In detail, please see (2.4).

Remark 2.3. The contrast function can indeed be simulated as follows. In the first place, we introduce the following stochastic interacting particle systems: for each $i \in \mathcal{S}_N := \{1, \cdots, N\}$

$$\mathrm{d}X^{\varepsilon,i}(t) = b(X^{\varepsilon,i}_t, \mu^{\varepsilon,N}_t, \theta)\mathrm{d}t + \sigma(X^{\varepsilon,i}_t, \mu^{\varepsilon,N}_t)\mathrm{d}W^i(t),\ t \ge 0, X^{\varepsilon,i}_0 = X^\varepsilon_0$$

where $(W^i(t), X^{\varepsilon,i}_0), i \in \mathcal{S}_N$, are independent copies of $(W(t), X_0)$, and $\mu^{\varepsilon,N}_t := \frac{1}{N}\sum_{j=1}^d \delta_{\widehat{X}^{\varepsilon,j}_t}$ with δ_x being the Dirac measure centered at the point x and

$$\widehat{X}^{\varepsilon,i}_{k\delta}(s) := X^{\varepsilon,i}((k-i)\delta) + \frac{s+i\delta}{\delta}\{X^{\varepsilon,i}((k-i)\delta) - X^{\varepsilon,i}((k-i-1)\delta)\}$$

for $s \in [-(i+1)\delta, -i\delta]$. Set

$$P_k^{i,N}(\theta) := X^{\varepsilon,i}(t_k) - X^{\varepsilon,i}(t_{k-1}) - b(\widehat{X}^{\varepsilon,i}_{t_{k-1}}, \mu^{\varepsilon,N}_{k\delta}, \theta)\delta.$$

With $P_k^{i,N}(\theta)$ in hand, we then define

$$\Psi^{i,N}_{n,\varepsilon}(\theta) = \varepsilon^{-2}\delta^{-1}\sum_{k=1}^{n}(P_k^{i,N})^*(\theta)(\theta)(\Lambda^{i,N}_{k-1})^{-1}P_k^{i,N}(\theta)$$

with $\Lambda_k^{i,N} := (\sigma\sigma^*)(\widehat{X}^{\varepsilon,i}_{t_k}, \mu^{\varepsilon,N}_{t_k})$, whence, the construction of estimator goes back to the classical set up(see e.g. [8, 24, 25]). By the law of large number, we have $\Psi^{i,N}_{n,\varepsilon}(\theta) \to \Psi_{n,\varepsilon}(\theta)$ so that $\Psi_{n,\varepsilon}$ can be simulated numerically.

3 The consistency of LSE

First of all, let's consider the following deterministic ordinary differential equation

$$\text{(3.1)} \qquad \mathrm{d}X^0(t) = b(X^0_t, \mathscr{L}_{X^0_t}, \theta_0)\mathrm{d}t, \quad t > 0, \quad X^0_0 = \xi \in \mathscr{C}.$$

Under (**A1**), (3.1) possesses a unique solution $(X^0(t))_{t \ge -r_0}$. Herein, it is worth pointing out that (2.1) and (3.1) share the same initial datum. For the sake of notation brevity, for a random variable $\zeta \in \mathscr{C}$ with $\mathscr{L}_\zeta \in \mathcal{P}_2(\mathscr{C})$, let

$$\text{(3.2)} \qquad \Lambda(\zeta, \theta, \theta_0) = b(\zeta, \mathscr{L}_\zeta, \theta_0) - b(\zeta, \mathscr{L}_\zeta, \theta) \quad \text{and} \quad \widehat{\sigma}(\zeta) = (\sigma\sigma^*)^{-1}(\zeta, \mathscr{L}_\zeta).$$

Set

$$\text{(3.3)} \qquad \Xi(\theta) := \int_0^T \Lambda^*(X^0_t, \theta, \theta_0)\widehat{\sigma}(X^0_t)\Lambda(X^0_t, \theta, \theta_0)\mathrm{d}t, \qquad \theta \in \Theta,$$

where (X^0_t) is the segment process generated by the solution $(X^0(t))$ to (3.1).

Our first main result, which is concerned with the consistency of the LSE of $\theta \in \Theta$, is stated as below.

Theorem 3.1. Let (**A1**)-(**A4**) hold and assume further $\Xi(\theta) > 0$ for any $\theta \in \overline{\Theta}$. Then

$$\widehat{\theta}_{n,\varepsilon} \to \theta_0 \qquad \text{in probability as } \varepsilon \to 0 \quad \text{and} \quad n \to \infty.$$

The proof of Theorem 3.1 is based on several auxiliary lemmas below.

Lemma 3.2. Under (**A1**), for any $p > 0$, there exists a constant $C_{p,T} > 0$ such that

$$\text{(3.4)} \qquad \mathbb{E}\Big(\sup_{-r_0 \le t \le T}|X^\varepsilon(t)|^p\Big) \le C_{p,T}(1 + \|\xi\|^p_\infty)$$

and

$$\text{(3.5)} \qquad \sup_{0 \le t \le T}\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|^p_\infty \le C_{p,T}(1 + \|\xi\|^p_\infty).$$

Proof. For any $t \in [0,T]$, a direct calculation shows from (2.4) that

r5

$$
\begin{aligned}
&\|\widehat{X}^{\varepsilon}_{\lfloor t/\delta\rfloor\delta}\|_\infty\\
&= \sup_{-r_0\le v\le 0}|\widehat{X}^{\varepsilon}_{\lfloor t/\delta\rfloor\delta}(v)|\\
&= \max_{k=0,\cdots,M-1}\sup_{-(k+1)\delta\le v\le -k\delta}\Big|\frac{1}{\delta}((k+1)\delta+v)X^{\varepsilon}((\lfloor t/\delta\rfloor-k)\delta)\\
&\quad -\frac{1}{\delta}(k\delta+v)X^{\varepsilon}((\lfloor t/\delta\rfloor-k-1)\delta)\Big|\\
&\le \max_{k=0,\cdots,M-1}\sup_{-(k+1)\delta\le v\le -k\delta}(|X^{\varepsilon}(\lfloor t/\delta\rfloor\delta-k\delta)|+|X^{\varepsilon}(\lfloor t/\delta\rfloor\delta-(k+1)\delta)|)\\
&\le 2\sup_{-r_0\le s\le t}|X^{\varepsilon}(s)|,
\end{aligned}
\tag{3.6}
$$

where in the first inequality we have used the fact that for any $v\in[-(k+1)\delta,-k\delta]$,

$$
\frac{1}{\delta}((k+1)\delta+v)\in[0,1]\quad\text{ and }\quad \frac{1}{\delta}(k\delta+v)\in[0,1].
$$

Once (3.4) is available, (3.5) can be obtained from (3.6). So, in what follows, it remains to show that (3.4) holds true.

By Hölder's inequality, it is sufficient to show that (3.4) holds for any $p\ge 2$. From (**A1**), one has, for any $\zeta\in\mathscr{C}$ and $\mu\in\mathcal{P}_2(\mathscr{C})$,

r2

$$
|b(\zeta,\mu,\theta)|^2\le 2\,\{\alpha_1\|\zeta\|_\infty^2+\alpha_2\mathbb{W}_2(\mu,\delta_{\zeta_0})^2+|b(\zeta_0,\delta_{\zeta_0},\theta)|^2\},
\tag{3.7}
$$

and

r3

$$
\|\sigma(\zeta,\mu)\|^2\le 2\,\{\beta_1\|\zeta\|_\infty^2+\beta_2\mathbb{W}_2(\mu,\delta_{\zeta_0})^2+\|\sigma(\zeta_0,\delta_{\zeta_0})\|^2\},
\tag{3.8}
$$

where $\zeta_0(s)=\mathbf{0}\in\mathbb{R}^d$ for any $s\in[-r_0,0]$. For $k\ge 0$, define the stopping time $\tau_k:=inf\{t\ge 0:\|X_t^{\varepsilon}\|_\infty\ge k\}$ and for any $p\ge 2$, by Hölder's inequality and Burkhold-Davis-Gundy's (BDG's for brevity) inequality (see, [26, Theorem 7.3, P.40]), we deduce from (3.7) and (3.8) that

c3.9

$$
\begin{aligned}
&1+\mathbb{E}\Big(\sup_{-r_0\le s\le t\wedge\tau_k}|X^{\varepsilon}(s)|^p\Big)\\
&\le 1+c\,\|\xi\|_\infty^p+c\,t^{p-1}\mathbb{E}\int_0^{t\wedge\tau_k}|b(X_s^{\varepsilon},\mathscr{L}_{X_s^{\varepsilon}},\theta)|^p\mathrm{d}s+c\,\mathbb{E}\Big(\int_0^{t\wedge\tau_k}\|\sigma(X_s^{\varepsilon},\mathscr{L}_{X_s^{\varepsilon}})\|^2\mathrm{d}s\Big)^{p/2}\\
&\le 1+c\,\|\xi\|_\infty^p+c(t^{p-1}+t^{\frac{p-2}{2}})\mathbb{E}\int_0^{t\wedge\tau_k}\{|b(X_s^{\varepsilon},\mathscr{L}_{X_s^{\varepsilon}},\theta)|^p+\|\sigma(X_s^{\varepsilon},\mathscr{L}_{X_s^{\varepsilon}})\|^p\}\mathrm{d}s\\
&\le 1+c\,\|\xi\|_\infty^p+c(t^{p-1}+t^{\frac{p-2}{2}})\int_0^t\{1+\mathbb{E}\|X_{s\wedge\tau_k}^{\varepsilon}\|_\infty^p+\mathbb{W}_2(\mathscr{L}_{X_s^{\varepsilon}},\delta_{\zeta_0})^p\}\mathrm{d}s\\
&\le 1+c\,\|\xi\|_\infty^p+c(t^{p-1}+t^{\frac{p-2}{2}})\int_0^t\{1+\mathbb{E}\|X_{s\wedge\tau_k}^{\varepsilon}\|_\infty^p\}\mathrm{d}s.
\end{aligned}
\tag{3.9}
$$

where $c>0$ is a generic constant, whose value may change from line to line and we also used Definition 2.1 in the last display. (3.9), together with (3.6), leads to

$$
1+\mathbb{E}\Big(\sup_{-r_0\le s\le t\wedge\tau_k}|X^{\varepsilon}(s)|^p\Big)\le 1+c\,\|\xi\|_\infty^p+c(t^{p-1}+t^{\frac{p-2}{2}})1+\int_0^t\mathbb{E}\Big(\sup_{-r_0\le r\le s\wedge\tau_k}|X^{\varepsilon}(r)|^p\Big)\mathrm{d}s,
$$

Then, the desired assertion (3.4) follows from Gronwall's inequality, followed by Fatou's lemma. □

le5 **Lemma 3.3.** Let (**A1**) be satisfied. Then, there is a constant $C_T > 0$ such that

r6 (3.10)
$$\sup_{0\le t\le T} \mathbb{E}\|X_t^\varepsilon - X_t^0\|_\infty^2 \le C_T\,\varepsilon^2.$$

Proof. Note that

$$\mathbb{E}\|X_t^\varepsilon - X_t^0\|_\infty^2 \le \mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s) - X^0(s)|^2\Big) =: A(t,\varepsilon), \quad t\in[0,T],$$

where we have used $X_0^\varepsilon = X_0^0 = \xi$. By Hölder's inequality, Doob's submartingale inequality as well as Itô's isometry, we obtain from (**A1**), (3.4) and (3.8) that

$$\begin{aligned}
A(t,\varepsilon) &\le 2\,t\int_0^t \mathbb{E}|b(X_s^\varepsilon, \mathscr{L}_{X_s^\varepsilon},\theta_0) - b(X_s^0,\mathscr{L}_{X_s^0},\theta_0)|^2\mathrm{d}s + 2\,\varepsilon^2\,\mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^s \sigma(X_u^\varepsilon,\mathscr{L}_{X_u^\varepsilon})\mathrm{d}B(u)\Big|^2\Big)\\
&\le 2\,t\int_0^t \mathbb{E}|b(X_s^\varepsilon, \mathscr{L}_{X_s^\varepsilon},\theta_0) - b(X_s^0,\mathscr{L}_{X_s^0},\theta_0)|^2\mathrm{d}s + 8\,\varepsilon^2\int_0^t \mathbb{E}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})\|^2\mathrm{d}s\\
&\le 2\,t\int_0^t\{\alpha_1\mathbb{E}\|X_s^\varepsilon - X_s^0\|_\infty^2 + \alpha_2\mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon},\mathscr{L}_{X_s^0})^2\}\mathrm{d}s\\
&\quad + c\,\varepsilon^2\int_0^t\{1+\mathbb{E}\|X_s^\varepsilon\|_\infty^2 + \mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon},\delta_{\zeta_0})^2\}\mathrm{d}s\\
&\le c\,t\int_0^t \mathbb{E}\|X_s^\varepsilon - X_s^0\|_\infty^2\mathrm{d}s + c\,\varepsilon^2\int_0^t\{1+\mathbb{E}\|X_s^\varepsilon\|_\infty^2\}\mathrm{d}s\\
&\le c\,t\int_0^t A(s,\varepsilon)\mathrm{d}s + c(1+C_{2,T})\,\varepsilon^2 t, \qquad t\in[0,T],
\end{aligned}$$

for a generic constant $c>0$. As a result, (3.10) follows by Gronwall's inequality. □

Lemma 3.4. Assume that (**A1**) and (**A3**) hold. Then, for any $\beta\in(0,1)$, there exist constants $c, c_\beta > 0$ such that

y2 (3.11)
$$\sup_{0\le t\le T}\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta} - X_t^0\|_\infty^2 \le c\,(\varepsilon^2 + c_\beta\delta^\beta).$$

Proof. Due to (3.10), for any $t\in[0,T]$,

r7 (3.12)
$$\begin{aligned}
\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta} - X_t^0\|_\infty^2 &\le 2\{\mathbb{E}\|X_t^\varepsilon - \widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2 + \mathbb{E}\|X_t^\varepsilon - X_t^0\|_\infty^2\}\\
&\le c\{\varepsilon^2 + \mathbb{E}\|X_t^\varepsilon - \widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2\}.
\end{aligned}$$

So, to get (3.11), we only need to show that, for any $\beta\in(0,1)$, there exists $c_\beta>0$ such that

y4 (3.13)
$$\sup_{t\in[0,T]}\mathbb{E}\|X_t^\varepsilon - \widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2 \le c_\beta\delta^\beta.$$

For any $t\in[0,T)$, there exists an integer $k_0\in[0,n-1]$ such that $t\in[k_0\delta,(k_0+1)\delta)$ so that $\lfloor t/\delta\rfloor = k_0$. By Hölder's inequality, for any $\beta\in(0,1)$,

$$\begin{aligned}
\mathbb{E}\|X_t^\varepsilon - \widehat{X}^\varepsilon_{k_0\delta}\|_\infty^2 &= \mathbb{E}\Big(\sup_{-r_0\le v\le 0}|X^\varepsilon(t+v) - \widehat{X}^\varepsilon_{k_0\delta}(v)|^2\Big)\\
&\le \Big(\mathbb{E}\Big(\sup_{-r_0\le v\le 0}|X^\varepsilon(t+v) - \widehat{X}^\varepsilon_{k_0\delta}(v)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\
&\le M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{-(k+1)\delta\le v\le -k\delta}|X^\varepsilon(t+v) - \widehat{X}^\varepsilon_{k_0\delta}(v)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta},
\end{aligned}$$

where $M>0$ is an integer such that $r_0=M\delta$. For $v\in[-(k+1)\delta,-k\delta]$, it follows from (2.4) that

$$\begin{aligned}X^\varepsilon(t+v)-\widehat{X}^\varepsilon_{k_0\delta}(v)=&\,\frac{(k+1)\delta+v}{\delta}\Big(X^\varepsilon(t+v)-X^\varepsilon((k_0-k)\delta)\Big)\\&-\frac{k\delta+v}{\delta}\Big(X^\varepsilon(t+v)-X^\varepsilon((k_0-k-1)\delta)\Big).\end{aligned}$$

As a consequence, we deduce that

y5

$$\begin{aligned}&\mathbb{E}\|X^\varepsilon_t-\widehat{X}^\varepsilon_{k_0\delta}\|^2_\infty\\&\le c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0-k+1)\delta}|X^\varepsilon(s)-X^\varepsilon((k_0-k)\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\quad+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0-k+1)\delta}|X^\varepsilon(s)-X^\varepsilon((k_0-k-1)\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&=:A_1(\varepsilon,\delta)+A_2(\varepsilon,\delta).\end{aligned}\tag{3.14}$$

For any $t\in[l\delta,(l+1)\delta]$ with $l=0,1,\cdots,n-1$, we deduce from Hölder's inequality and BDG's inequality that

y3

$$\begin{aligned}&\mathbb{E}\Big(\sup_{l\delta\le s\le t}|X^\varepsilon(s)-X^\varepsilon(l\delta)|^{\frac{2}{1-\beta}}\Big)\\&\le c\Big\{\delta^{\frac{1+\beta}{1-\beta}}\int_{l\delta}^t\mathbb{E}|b(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s},\theta)|^{\frac{2}{1-\beta}}\mathrm{d}s+\mathbb{E}\Big(\int_{l\delta}^t\|\sigma(\widehat{X}^\varepsilon_s,\mathscr{L}_{\widehat{X}^\varepsilon_s})\|^2\mathrm{d}s\Big)^{\frac{1}{1-\beta}}\Big\}\\&\le c\delta^{\frac{\beta}{1-\beta}}\int_{l\delta}^t\{\mathbb{E}|b(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s},\theta)|^{\frac{2}{1-\beta}}+\mathbb{E}\|\sigma(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s},\theta)\|^{\frac{2}{1-\beta}}\}\mathrm{d}s.\end{aligned}\tag{3.15}$$

This, combining (3.7) with (3.8) and (3.15), yields that, for any $t\in[l\delta,(l+1)\delta]$,

y1

$$\begin{aligned}\mathbb{E}\Big(\sup_{l\delta\le s\le t}|X^\varepsilon(s)-X^\varepsilon(l\delta)|^{\frac{2}{1-\beta}}\Big)&\le c\,\delta^{\frac{\beta}{1-\beta}}\int_{l\delta}^t\{1+\mathbb{E}\|X^\varepsilon_s\|_\infty^{\frac{2}{1-\beta}}+\mathbb{W}_2(\mathscr{L}_{X^\varepsilon_s},\delta_{\zeta_0})^{\frac{2}{1-\beta}}\}\mathrm{d}s\\&\le c\,\delta^{\frac{1}{1-\beta}}\Big\{1+\sup_{l\delta\le s\le t}\mathbb{E}\|X^\varepsilon_s\|_\infty^{\frac{2}{1-\beta}}\Big\}\\&\le c\,\delta^{\frac{1}{1-\beta}},\end{aligned}\tag{3.16}$$

where in the last procedure we have exploited (3.5).

In the sequel, we divide three cases to show the estimates on $A_1(\varepsilon,\delta)$ and $A_2(\varepsilon,\delta)$.

Case 1: $k\ge k_0+1$. With regard to such case, $(k_0+1-k)\delta\in[-r_0,0]$. We infer from (**A1**) and (3.14), in addition to $M\delta=r_0$, that

$$A_1(\varepsilon,\delta)+A_2(\varepsilon,\delta)\le cM^{1-\beta}\delta=c\,r_0^{1-\beta}\delta^\beta.$$

Case 2: $k_0=k$. For this case, $t\in[k\delta,(k+1)\delta)$. Again, one gets from (3.14) that

$$\begin{aligned}&A_1(\varepsilon,\delta)+A_2(\varepsilon,\delta)\\&\le c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{-\delta\le s\le\delta}|X^\varepsilon(s)-X^\varepsilon(0)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\quad+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{-\delta\le s\le\delta}|X^\varepsilon(s)-X^\varepsilon(-\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}.\end{aligned}$$

This, besides (**A3**) and (3.16), implies that

$$\begin{aligned}A_1(\varepsilon,\delta)+A_2(\varepsilon,\delta)&\le c\,\delta^\beta+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{-\delta\le s\le\delta}|X^\varepsilon(s)-X^\varepsilon(0)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\le c\,\delta^\beta+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{0\le s\le\delta}|X^\varepsilon(s)-X^\varepsilon(0)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\le c\,\delta^\beta+c\,M^{1-\beta}\delta\\&\le c\,\delta^\beta.\end{aligned}$$

Case 3: $k\le k_0-1$. Also, by making use of (3.16), it follows that

$$\begin{aligned}&A_1(\varepsilon,\delta)+A_2(\varepsilon,\delta)\\&\le c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0-k+1)\delta}|X^\varepsilon(s)-X^\varepsilon((k_0-k-1)\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\quad+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}|X^\varepsilon((k_0-k-1)\delta)-X^\varepsilon((k_0-k)\delta)|^{\frac{2}{1-\beta}}\Big)^{1-\beta}\\&\le c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0-k)\delta}|X^\varepsilon(s)-X^\varepsilon((k_0-k-1)\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\quad+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}\Big(\sup_{(k_0-k)\delta\le s\le(k_0-k+1)\delta}|X^\varepsilon(s)-X^\varepsilon((k_0-k)\delta)|^{\frac{2}{1-\beta}}\Big)\Big)^{1-\beta}\\&\quad+c\,M^{1-\beta}\max_{k=0,\cdots,M-1}\Big(\mathbb{E}|X^\varepsilon((k_0-k-1)\delta)-X^\varepsilon((k_0-k)\delta)|^{\frac{2}{1-\beta}}\Big)^{1-\beta}\\&\le c\delta^\beta.\end{aligned}$$

By summing up the three cases above, (3.13) holds true. □

lem2 **Lemma 3.5.** Let (**A1**)-(**A3**) hold. Then,

a7 (3.17)
$$\Phi^{(1)}_{n,\varepsilon}(\theta):=\delta\sum_{k=1}^n\Lambda^*(\widehat X^\varepsilon_{t_{k-1}},\theta,\theta_0)\widehat\sigma(\widehat X^\varepsilon_{t_{k-1}})\Lambda(\widehat X^\varepsilon_{t_{k-1}},\theta,\theta_0)\to\Xi(\theta)$$

in L^1 uniformly w.r.t. θ as $\varepsilon\to0$ and $\delta\to0$ (i.e., $n\to\infty$), in which $\Xi(\cdot)$ is introduced in (3.3).

Proof. It is straightforward to see that

$$\begin{aligned}&\delta\sum_{k=1}^n\Lambda^*(\widehat X^\varepsilon_{t_{k-1}},\theta,\theta_0)\widehat\sigma(\widehat X^\varepsilon_{t_{k-1}})\Lambda(\widehat X^\varepsilon_{t_{k-1}},\theta,\theta_0)-\int_0^T\Lambda^*(X^0_s,\theta,\theta_0)\widehat\sigma(X^0_s)\Lambda(X^0_s,\theta,\theta_0)\mathrm{d}s\\&=\int_0^T\Big\{\Lambda^*(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\widehat\sigma(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta})\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)-\Lambda^*(X^0_s,\theta,\theta_0)\widehat\sigma(X^0_s)\Lambda(X^0_s,\theta,\theta_0)\Big\}\mathrm{d}s\\&=\int_0^T\Big(\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)-\Lambda(X^0_s,\theta,\theta_0)\Big)^*\widehat\sigma(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta})\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\mathrm{d}s\\&\quad+\int_0^T\Lambda^*(X^0_s,\theta,\theta_0)\Big(\widehat\sigma(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta})-\widehat\sigma(X^0_s)\Big)\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\mathrm{d}s\\&\quad+\int_0^T\Lambda^*(X^0_s,\theta,\theta_0)\widehat\sigma(X^0_s)\Big(\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)-\Lambda(X^0_s,\theta,\theta_0)\Big)\mathrm{d}s\\&=:J_1(\varepsilon,\delta)+J_2(\varepsilon,\delta)+J_3(\varepsilon,\delta).\end{aligned}$$

Next, for any random variables $\zeta_1, \zeta_2 \in \mathscr{C}$ with $\mathscr{L}_{\zeta_1}, \mathscr{L}_{\zeta_2} \in \mathcal{P}_2(\mathscr{C})$, observe from (**A1**) that

a1
$$
\begin{aligned}
|\Lambda(\zeta_1,\theta,\theta_0)-\Lambda(\zeta_2,\theta,\theta_0)| &\le |b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta_0)|+|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta)|\\
&\le c\{\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mathscr{L}_{\zeta_1},\mathscr{L}_{\zeta_2})\}.
\end{aligned}
\tag{3.18}
$$

For a random variable $\zeta \in \mathscr{C}$ with $\mathscr{L}_\zeta \in \mathcal{P}_2(\mathscr{C})$, employing (**A2**) gives that

a6
$$
\|\widehat{\sigma}(\zeta)\| \le \|\widehat{\sigma}(\zeta)-\widehat{\sigma}(\zeta_0)\|+\|\widehat{\sigma}(\zeta_0)\| \le c\,\{1+\|\zeta\|_\infty+\mathbb{W}_2(\mathscr{L}_\zeta,\delta_{\zeta_0})\}.
\tag{3.19}
$$

Consequently, combining (3.7) with (3.18) and (3.19), we deduce that

$$
\begin{aligned}
&|J_1(\varepsilon,\delta)|+|J_3(\varepsilon,\delta)|\\
&\le c\int_0^T\{\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\mathscr{L}_{X^0_s})\}\\
&\quad\times\{1+\|X^0_s\|_\infty+\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\delta_{\zeta_0})\}^2\mathrm{d}s\\
&\le c\int_0^T\Big\{\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty+\sqrt{\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty^2}\Big\}\\
&\quad\times\{1+\|X^0_s\|_\infty^2+\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty^2+\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty^2\}\mathrm{d}s.
\end{aligned}
$$

This, together with (3.5) and (3.11) as well as Hölder's inequality, implies that

a8
$$
\begin{aligned}
&\mathbb{E}|J_1(\varepsilon,\delta)|+\mathbb{E}|J_3(\varepsilon,\delta)|\\
&\le c\int_0^T\sqrt{\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty^2}\{1+\|X^0_s\|_\infty^4+\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty^4\}\mathrm{d}s\\
&\to 0
\end{aligned}
\tag{3.20}
$$

as $\varepsilon \to 0$ and $\delta \to 0$. Next, making use of (**A2**) and (3.7), we derive that

$$
\begin{aligned}
|J_2(\varepsilon,\delta)| \le c\int_0^T&(1+\|X^0_s\|_\infty)(1+\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\delta_{\zeta_0}))\\
&\times\Big(\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty+\sqrt{\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty^2}\Big)\mathrm{d}s.
\end{aligned}
$$

Again, using (3.5) and (3.11) and utilizing Hölder's inequality gives that

a9
$$
\begin{aligned}
\mathbb{E}|J_2(\varepsilon,\delta)| &\le c\int_0^T\sqrt{\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}-X^0_s\|_\infty^2}\{1+\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|_\infty^2\}\mathrm{d}s\\
&\to 0
\end{aligned}
\tag{3.21}
$$

whenever $\varepsilon \to 0$ and $\delta \to 0$. Hence, (3.17) follows immediately from (3.20) and (3.21). □

lem4 **Lemma 3.6.** Let (**A1**)-(**A4**) hold. Then,

s5
$$
\Phi^{(2)}_{n,\varepsilon}(\theta) := \sum_{k=1}^n \Lambda^*(\widehat{X}^\varepsilon_{t_{k-1}},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})P_k(\theta_0) \longrightarrow 0
\tag{3.22}
$$

in probability uniformly w.r.t. θ as $\varepsilon \to 0$ and $\delta \to 0$.

Proof. Note that

$$
\begin{aligned}
\Phi^{(2)}_{n,\varepsilon}(\theta) &= \int_0^T \Lambda^*(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta})(b(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s},\theta_0)-b(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta_0))\mathrm{d}s\\
&\quad+\varepsilon\int_0^T \Lambda^*(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta})\sigma(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s})\mathrm{d}B(s)\\
&=:\Upsilon_1(\varepsilon,\delta,\theta)+\Upsilon_2(\varepsilon,\delta,\theta).
\end{aligned}
$$

By means of (3.7), together with (3.5), it follows from Hölder's inequality that, for some constants $C_1, C_2 > 0$,

s3

$$
\begin{aligned}
\mathbb{E}|\Upsilon_1(\varepsilon,\delta,\theta)| &\le C_1\int_0^T \mathbb{E}\{(1+\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}},\delta_{\zeta_0})^2)\\
&\qquad\times\|X^\varepsilon_t-\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}},\mathscr{L}_{X^\varepsilon_t})\}\mathrm{d}t\\
&\le C_2\int_0^T\{1+\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^4\}(\mathbb{E}\|X^\varepsilon_t-\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2)^{1/2}\mathrm{d}t\\
&\to 0
\end{aligned}\tag{3.23}
$$

uniformly w.r.t. θ as $\varepsilon\to 0$ and $\delta\to 0$, where the last procedure is due to (3.13). Next, by BDG's inequality and Hölder's inequality, we derive from (3.7), (3.8), and (3.19) followed by (3.4) and (3.5) that

s4

$$
\begin{aligned}
\mathbb{E}|\Upsilon_2(\varepsilon,\delta,\theta)|^p &\le c\,\varepsilon^p\,\mathbb{E}\Big(\int_0^T|\Lambda^*(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta})\sigma(X^\varepsilon_t,\mathscr{L}_{X^\varepsilon_t})|^2\mathrm{d}t\Big)^{p/2}\\
&\le c\,\varepsilon^p\int_0^T\mathbb{E}\{|\Lambda(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta},\theta,\theta_0)|^2\cdot\|\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta})\|^2\cdot\|\sigma(X^\varepsilon_t,\mathscr{L}_{X^\varepsilon_t})\|^2\}^{p/2}\mathrm{d}t\\
&\le c\,\varepsilon^p\int_0^T\mathbb{E}\Big\{\Big(1+\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}},\delta_{\zeta_0})^2\Big)^{p/2}\\
&\qquad\times\Big(1+\|X^\varepsilon_t\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{X^\varepsilon_t},\delta_{\zeta_0})^2\Big)^{p/2}\Big\}\mathrm{d}t\\
&\le c\,\varepsilon^p\int_0^T\{1+\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}\|_\infty^{2p}+\mathbb{E}\|X_t\|_\infty^{2p}\}\mathrm{d}t\\
&\le c\,\varepsilon^p,\quad p>2
\end{aligned}\tag{3.24}
$$

where $c>0$ is a generic constant. On the other hand, for any $\theta_1,\theta_2\in\overline{\Theta}$, by using the BDG inequality and the Hölder inequality, it follows from (**A4**), (3.8), and (3.19) that

s66

$$
\begin{aligned}
&\mathbb{E}|\Upsilon_2(\varepsilon,\delta,\theta_1)-\Upsilon_2(\varepsilon,\delta,\theta_2)|^p\\
&\le c\,\varepsilon^p\int_0^T\mathbb{E}\{|b(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta_1)-b(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta_2)|^2\\
&\quad\times\cdot\|\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}\|^2\cdot\|\sigma(X^\varepsilon_s,\mathscr{L}_{X^\varepsilon_s}\}^{p/2}\mathrm{d}t\\
&\le c\,\varepsilon^p|\theta_1-\theta_2|^p,\quad p>2.
\end{aligned}\tag{3.25}
$$

Consequently, combining (3.23), (3.24) with (3.25), we obtain (3.22) from [11, Theorem 20, p378]. $\square$

To make the content self-contained, we reformulate [38, Theorem 5.9] as the following lemma.

lem3 **Lemma 3.7.** Let $(M_n)_{n\geq 1}$ be random functions and M a fixed function of θ such that, for any $\varepsilon > 0$,

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \longrightarrow 0 \qquad \text{in probability}$$

and $\sup_{|\theta-\theta_0|\geq\varepsilon} M(\theta) < M(\theta_0)$. Then, any sequence of estimators $\widehat{\theta}_n$ with $M_n(\widehat{\theta}_n) \geq M_n(\theta_0)$ converges in probability to θ_0.

With Lemmas 3.5-3.7 in hand, we are in the position to complete the proof of Theorem 3.1.

Proof of Theorem 3.1. From (2.2), we infer that

s6
$$\begin{aligned}
&\Phi_{n,\varepsilon}(\theta)\\
&= \delta^{-1}\sum_{k=1}^{n}\Big\{P_k^*(\theta)\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})P_k(\theta) - P_k^*(\theta_0)\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})P_k(\theta_0)\Big\}\\
&= \delta^{-1}\sum_{k=1}^{n}\Big\{\Big(P_k(\theta_0) + \Lambda(\widehat{X}^{\varepsilon}_{t_{k-1}},\theta,\theta_0)\delta\Big)^*\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})\Big(P_k(\theta_0) + \Lambda(\widehat{X}^{\varepsilon}_{t_{k-1}},\theta,\theta_0)\delta\Big)\\
&\quad - P_k^*(\theta_0)\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})P_k(\theta_0)\Big\}\\
&= \delta\sum_{k=1}^{n}\Lambda^*(\widehat{X}^{\varepsilon}_{t_{k-1}},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})\Lambda(\widehat{X}^{\varepsilon}_{t_{k-1}},\theta,\theta_0) + 2\sum_{k=1}^{n}\Lambda^*(\widehat{X}^{\varepsilon}_{t_{k-1}},\theta,\theta_0)\widehat{\sigma}(\widehat{X}^{\varepsilon}_{t_{k-1}})P_k(\theta_0)\\
&= \Phi^{(1)}_{n,\varepsilon}(\theta) + 2\Phi^{(2)}_{n,\varepsilon}(\theta),
\end{aligned}\tag{3.26}$$

where $\Phi^{(1)}_{n,\varepsilon}(\theta)$ and $\Phi^{(2)}_{n,\varepsilon}(\theta)$ are defined in (3.17) and (3.22), respectively. In terms of Lemmas 3.5 and 3.6, we deduce from Chebyshev's inequality that

$$\sup_{\theta\in\Theta} |-\Phi_{n,\varepsilon}(\theta) - (-\Xi(\theta))| \to 0 \qquad \text{in probability,}$$

where $\Xi(\cdot)$ is defined as in (3.17). On the other hand, for any $\kappa > 0$, notice that

$$\sup_{|\theta-\theta_0|\geq\kappa} (-\Xi(\theta)) < -\Xi(\theta_0) = 0$$

due to $\Xi(\cdot) > 0$. Moreover, according to the notion of $\widehat{\theta}_{n,\varepsilon}$, one has $-\Phi_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon}) \geq -\Phi_{n,\varepsilon}(\theta_0) = 0$. As far as our present model is concerned, all of the assumptions in Lemma 3.7 with $M_n(\cdot) = -\Phi_{n,\varepsilon}(\cdot)$ and $M(\cdot) = -\Xi(\cdot)$ are fulfilled. As a consequence, we conclude that $\widehat{\theta}_{n,\varepsilon} \to \theta_0$ in probability as $\varepsilon \to 0$ and $n \to \infty$, as required. □

4 The asymptotic distribution of LSE

sec3

In this section, to begin, we recall some materials on derivatives for matrix-valued functions and introduce some notation. For a differentiable mapping $V = (V_1, \cdots, V_d)^* : \mathbb{R}^p \to \mathbb{R}^d$, its gradient operator $(\nabla_x V)(x) \in \mathbb{R}^d \otimes \mathbb{R}^p$ w.r.t. the argument $x = (x_1, \cdots, x_p)^* \in \mathbb{R}^p$ is given by

A2
$$(\nabla_x V)(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} V_1(x) & \frac{\partial}{\partial x_2} V_1(x) & \cdots & \frac{\partial}{\partial x_p} V_1(x) \\ \frac{\partial}{\partial x_1} V_2(x) & \frac{\partial}{\partial x_2} V_2(x) & \cdots & \frac{\partial}{\partial x_p} V_2(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial}{\partial x_1} V_d(x) & \frac{\partial}{\partial x_2} V_d(x) & \cdots & \frac{\partial}{\partial x_p} V_d(x) \end{pmatrix}. \tag{4.1}$$

If $V = (V_1, \cdots, V_d) : \mathbb{R}^p \to (\mathbb{R}^d)^*$ (i.e., the d-dimensional raw vector) is differentiable, its gradient operator $(\nabla_x V)(x) \in \mathbb{R}^p \otimes \mathbb{R}^d$ w.r.t. the argument $x = (x_1, \cdots, x_p)^* \in \mathbb{R}^p$ reads as follows

A3
$$(\nabla_x V)(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} V_1(x) & \frac{\partial}{\partial x_1} V_2(x) & \cdots & \frac{\partial}{\partial x_1} V_d(x) \\ \frac{\partial}{\partial x_2} V_1(x) & \frac{\partial}{\partial x_2} V_2(x) & \cdots & \frac{\partial}{\partial x_2} V_d(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial}{\partial x_p} V_1(x) & \frac{\partial}{\partial x_p} V_2(x) & \cdots & \frac{\partial}{\partial x_p} V_d(x) \end{pmatrix}. \tag{4.2}$$

So, from (4.1) and (4.2), one has $\nabla_x V^*(x) = (\nabla_x V)^*(x)$ for a differentiable function $V : \mathbb{R}^p \to \mathbb{R}^d$. Let $V = (V_{ij})_{p\times d} : \mathbb{R} \to \mathbb{R}^p \otimes \mathbb{R}^d$ be differentiable. Then, the derivative $\frac{\partial}{\partial x} V(x) \in \mathbb{R}^p \otimes \mathbb{R}^d$ of the matrix-valued mapping V w.r.t. the scalar argument $x \in \mathbb{R}$ enjoys the form

A1
$$\frac{\partial}{\partial x} V(x) = \begin{pmatrix} \frac{\partial}{\partial x} V_{11}(x) & \frac{\partial}{\partial x} V_{12}(x) & \cdots & \frac{\partial}{\partial x} V_{1d}(x) \\ \frac{\partial}{\partial x} V_{21}(x) & \frac{\partial}{\partial x} V_{22}(x) & \cdots & \frac{\partial}{\partial x} V_{2d}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial}{\partial x} V_{p1}(x) & \frac{\partial}{\partial x} V_{p2}(x) & \cdots & \frac{\partial}{\partial x} V_{pd}(x) \end{pmatrix}. \tag{4.3}$$

For a differentiable function $V = (V_{ij})_{p\times d} : \mathbb{R}^p \to \mathbb{R}^p \otimes \mathbb{R}^d$, the gradient operator, denoted by $\nabla_x V(x) \in \mathbb{R}^p \otimes \mathbb{R}^{pd}$, of $V(x)$ w.r.t. the variable $x = (x_1, \cdots, x_p)^* \in \mathbb{R}^p$ is formulated as

$$(\nabla_x V)(x) = \Big(\frac{\partial}{\partial x_1} V(x), \frac{\partial}{\partial x_2} V(x), \cdots, \frac{\partial}{\partial x_p} V(x)\Big),$$

where $\frac{\partial}{\partial x_i} V(x)$ is defined as in (4.3). Moreover, for a differentiable function $V = (V_{ij})_{p\times d} : \mathbb{R}^p \to \mathbb{R}^d$, we have

z1
$$(\nabla_x^{(2)} V^*)(x) := (\nabla_x(\nabla_x V^*))(x) = (\nabla_x(\nabla_x V)^*)(x). \tag{4.4}$$

For $A = (A_1, A_2, \cdots, A_p) \in \mathbb{R}^p \otimes \mathbb{R}^{pd}$ with $A_k \in \mathbb{R}^p \otimes \mathbb{R}^d$, $k = 1, \cdots, p$, and $B \in \mathbb{R}^d$, let's define $A \circ B \in \mathbb{R}^p \otimes \mathbb{R}^p$ by

$$A \circ B = (A_1 B, A_2 B, \cdots, A_p B).$$

Set, for any $\theta \in \Theta$,

z3
$$I(\theta) := \int_0^T (\nabla_\theta b)^*(X_s^0, \mathscr{L}_{X_s^0}, \theta)\widehat{\sigma}(X_s^0)(\nabla_\theta b)(X_s^0, \mathscr{L}_{X_s^0}, \theta)\mathrm{d}s, \tag{4.5}$$

and, for any random variables $\zeta_1, \zeta_2 \in \mathscr{C}$ with $\mathscr{L}_{\zeta_1}, \mathscr{L}_{\zeta_2} \in \mathcal{P}_2(\mathscr{C})$,

s0
$$\Upsilon(\zeta_1, \zeta_2, \theta_0) := (\nabla_\theta b)^*(\zeta_1, \mathscr{L}_{\zeta_1}, \theta_0)\widehat{\sigma}(\zeta_1)\sigma(\zeta_2, \mathscr{L}_{\zeta_2}). \tag{4.6}$$

Furthermore, we set

z2
$$K(\theta) := -2\int_0^T \{(\nabla_\theta^{(2)} b^*)(X_s^0, \mathscr{L}_{X_s^0}, \theta) \circ (\widehat{\sigma}(X_s^0)\Lambda(X_s^0, \theta, \theta_0))\}\mathrm{d}s, \qquad \theta \in \Theta. \tag{4.7}$$

Another main result in this paper is presented as below, which reveals the asymptotic distribution of $\widehat{\theta}_{n,\varepsilon}$.

th2 **Theorem 4.1.** Let the assumptions of Theorem 3.1 hold and suppose further that (**A2**) and (**A3**) hold and that $I(\cdot)$ and $K(\cdot)$ defined in (4.5) and (4.7), respectively, are continuous. Then,

$$\varepsilon^{-1}(\widehat{\theta}_{n,\varepsilon} - \theta_0) \to I^{-1}(\theta_0)\int_0^T \Upsilon(X_s^0, X_s^0, \theta_0)\mathrm{d}B(s) \qquad \text{in probability}$$

as $\varepsilon \to 0$ and $n \to \infty$, where $I(\cdot)$ and $\Upsilon(\cdot)$ are given in (4.5) and (4.6), respectively.

Before we proceed to complete the proof of Theorem 4.1, let's prepare the lemmas below.

le1 **Lemma 4.2.** Assume that (**A1**)- (**A3**) and (**B1**)- (**B2**) hold. Then,

s1
$$\int_0^T \Upsilon(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, X_t^\varepsilon, \theta_0)\mathrm{d}B(t) \longrightarrow \int_0^T \Upsilon(X_t^0, X_t^0, \theta_0)\mathrm{d}B(t) \qquad \text{in probability} \tag{4.8}$$

as $\varepsilon \to 0$ and $\delta \to 0$. Moreover,

s2
$$\varepsilon^{-1}(\nabla_\theta \Phi_{n,\varepsilon})(\theta_0) \to -2\int_0^T \Upsilon(X_s^0, X_s^0, \theta_0)\mathrm{d}B(s) \qquad \text{in probability} \tag{4.9}$$

whenever $\varepsilon \to 0$ and $\delta \to 0$.

Proof. We first claim that

s7
$$\int_0^T \|\Upsilon(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, X_t^\varepsilon, \theta_0) - \Upsilon(X_t^0, X_t^0, \theta_0)\|^2\mathrm{d}t \to 0 \qquad \text{in probability} \tag{4.10}$$

as $\varepsilon \to 0$ and $\delta \to 0$. For any $\kappa > 0$ and $\rho > 0$, by the aid of (4.10) and by making use of [6, Theorem 2.6, P.63], we have

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\int_0^T (\Upsilon(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, X_t^\varepsilon, \theta_0) - \Upsilon(X_t^0, X_t^0, \theta_0))\mathrm{d}B(t)\Big| \ge \kappa\Big)\\
&\le \mathbb{P}\Big(\int_0^T \|\Upsilon(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, X_t^\varepsilon, \theta_0) - \Upsilon(X_t^0, X_t^0, \theta_0)\|^2\mathrm{d}t \ge \kappa^2\rho\Big) + \rho.
\end{aligned}$$

Thus, (4.8) follows from (4.10) and the arbitrariness of ρ. So, in what follows, it remains to show that (4.10) holds true. Observe that

$$\begin{aligned}
&\Upsilon(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, X_t^\varepsilon, \theta_0) - \Upsilon(X_t^0, X_t^0, \theta_0)\\
&= \{(\nabla_\theta b)^*(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}, \mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}}, \theta_0) - (\nabla_\theta b)^*(X_t^0, \mathscr{L}_{X_t^0}, \theta_0)\}\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta})\sigma(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon})\\
&\quad + (\nabla_\theta b)^*(X_t^0, \mathscr{L}_{X_t^0}, \theta_0)\{\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}) - \widehat{\sigma}(X_t^0)\}\sigma(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon})\\
&\quad + (\nabla_\theta b)^*(X_t^0, \mathscr{L}_{X_t^0}, \theta_0)\widehat{\sigma}(X_t^0)\{\sigma(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}) - \sigma(X_t^0, \mathscr{L}_{X_t^0})\}\\
&=: \Sigma_1(t,\varepsilon,\delta) + \Sigma_2(t,\varepsilon,\delta) + \Sigma_3(t,\varepsilon,\delta).
\end{aligned}$$

From (**B1**), (3.8), (3.13), and (3.19), it follows that

s8 (4.11)
$$\begin{aligned}&\int_0^T (\|\Sigma_1(t,\varepsilon,\delta)\|^2+\|\Sigma_2(t,\varepsilon,\delta)\|^2)\mathrm{d}t\\ &\le c\int_0^T (1+\|X_t^\varepsilon\|_\infty^4)\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t+\widehat{\Pi}(\varepsilon,\delta),\end{aligned}\tag{4.11}$$

where

$$\widehat{\Pi}(\varepsilon,\delta):=c\int_0^T (1+\|X_t^\varepsilon\|_\infty^4)\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t.$$

For any $\rho>0$, one gets from (4.11) that

$$\begin{aligned}&\mathbb{P}\Big(\int_0^T (\|\Sigma_1(t,\varepsilon,\delta)\|^2+\|\Sigma_2(t,\varepsilon,\delta)\|^2)\mathrm{d}t\ge\rho\Big)\\ &\le \mathbb{P}(\widehat{\Pi}(\varepsilon,\delta)\ge\rho/2)+\mathbb{P}\Big(c\int_0^T (1+\|X_t^\varepsilon\|_\infty^4)\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t\ge\frac{\rho}{2}\Big).\end{aligned}$$

By the Chebyshev inequality, in addition to (3.5) and (3.11),

$$\begin{aligned}\mathbb{P}(\widehat{\Pi}(\varepsilon,\delta)\ge\rho/2)&\le\frac{c}{\rho}\int_0^T (1+\mathbb{E}\|X_t^\varepsilon\|_\infty^4)\mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}s\\ &\longrightarrow 0\end{aligned}$$

as $\varepsilon\to0$ and $\delta\to0$. Also, for any $K>0$, by Chebyshev's inequality, besides (3.5),

$$\begin{aligned}&\mathbb{P}\Big(c\int_0^T (1+\|X_t^\varepsilon\|_\infty^4)\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t\ge\frac{\rho}{2}\Big)\\ &\le\mathbb{P}\Big(c(1+K^4)\int_0^T \|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t\ge\frac{\rho}{4}\Big)\\ &\quad+\mathbb{P}\Big(c\int_0^T (1+\sup_{-r_0\le s\le t}|X_s^\varepsilon|^8)\mathbf{1}_{\{\sup_{-r_0\le s\le t}|X^\varepsilon(t)|\ge K\}}\mathrm{d}t\ge\frac{\rho}{4}\Big)\\ &\le\frac{c(1+K^4)}{\rho}\int_0^T \mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t\\ &\quad+\frac{c}{\rho}\int_0^T \Big(1+\mathbb{E}\Big(\sup_{-r_0\le s\le t}|X_s^\varepsilon\|_\infty^{12}\Big)\Big)^{1/2}\Big(\mathbb{P}\Big(\sup_{-r_0\le s\le t}|X^\varepsilon(t)|\ge K\Big)\Big)^{1/2}\mathrm{d}t\\ &\le\frac{c(1+K^4)}{\rho}\int_0^T \mathbb{E}\|\widehat{X}^\varepsilon_{\lfloor t/\delta\rfloor\delta}-X_t^0\|_\infty^2\mathrm{d}t+\frac{c}{\rho K}\int_0^T \Big(1+\mathbb{E}\Big(\sup_{-r_0\le s\le t}|X_s^\varepsilon\|_\infty^2\Big)\Big)\mathrm{d}t.\end{aligned}$$

where $c>0$ is a generic constant. This, together with (3.11), leads to

a2 (4.12)
$$\int_0^T (\|\Sigma_1(t,\varepsilon,\delta)\|^2+\|\Sigma_2(t,\varepsilon,\delta)\|^2)\mathrm{d}t\longrightarrow0\quad\text{in probability}\tag{4.12}$$

by taking $\varepsilon\to0$ and $\delta\to0$ followed by taking $K\uparrow\infty$. Furthermore, (**A1**), (3.10), (3.19) as well as (**B1**) imply that

a3 (4.13)
$$\int_0^T \mathbb{E}\|\Sigma_3(t,\varepsilon,\delta)\|^2\mathrm{d}t\le c\int_0^T \mathbb{E}\|X_t^\varepsilon-X_t^0\|_\infty^2\mathrm{d}t\longrightarrow0\tag{4.13}$$

as $\varepsilon \to 0$. As a result, (4.10) follows from (4.12), (4.13) and Chebyshev's inequality.

For any $\theta \in \Theta$ and random variable $\zeta \in \mathscr{C}$ with $\mathcal{P}_2(\mathscr{C})$, note from (3.2) that

$$(\nabla_\theta \Lambda)(\zeta, \theta, \theta_0) = -(\nabla_\theta b)(\zeta, \mathscr{L}_\zeta, \theta).$$

A straightforward calculation shows that

$$\begin{aligned}(\nabla_\theta \Phi_{n,\varepsilon})(\theta) &= 2\sum_{k=1}^{n} (\nabla_\theta \Lambda)^*(\widehat{X}^\varepsilon_{t_{k-1}}, \theta, \theta_0)\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})\Big\{P_k(\theta_0) + \delta\Lambda(\widehat{X}^\varepsilon_{t_{k-1}}, \theta, \theta_0)\Big\}\\ &= -2\sum_{k=1}^{n} (\nabla_\theta b)^*(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta)\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})P_k(\theta).\end{aligned}$$

Therefore, one has

$$\begin{aligned}&\varepsilon^{-1}(\nabla_\theta \Phi_{n,\varepsilon})(\theta_0)\\ &= -\frac{2}{\varepsilon}\int_0^T (\nabla_\theta b)^*(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}, \mathscr{L}_{\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}}, \theta)\widehat{\sigma}(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta})(b(X^\varepsilon_s, \mathscr{L}_{X^\varepsilon_s}, \theta) - b(X^\varepsilon_{\lfloor s/\delta\rfloor\delta}, \mathscr{L}_{X^\varepsilon_{\lfloor s/\delta\rfloor\delta}}, \theta))\mathrm{d}s\\ &\quad - 2\int_0^T \Upsilon(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}, X^\varepsilon_s, \theta_0)\mathrm{d}B(s)\\ &= \Pi(\varepsilon, \delta) - 2\int_0^T \Upsilon(\widehat{X}^\varepsilon_{\lfloor s/\delta\rfloor\delta}, X^\varepsilon_s, \theta_0)\mathrm{d}B(s).\end{aligned}$$

Following the argument to obtain (3.20), we derive that

s00 (4.14)
$$\mathbb{E}|\Pi(\varepsilon, \delta)| \to 0$$

as $\varepsilon \to 0$ and $\delta \to 0$. Subsequently, (4.9) follows from (4.8) and (4.14) immediately. □

lem3.3 **Lemma 4.3.** Under the assumptions of Theorem 4.1,

a11 (4.15)
$$(\nabla^{(2)}_\theta \Phi_{n,\varepsilon})(\theta) \longrightarrow K_0(\theta) := K(\theta) + 2I(\theta) \quad \text{in probability}$$

as $\varepsilon \to 0, n \to \infty$, where $(\nabla^{(2)}_\theta \Phi_{n,\varepsilon}), I(\theta), K(\theta)$ are defined as in (4.4), (4.5), and (4.7), respectively.

Proof. By the chain rule, we infer from (4.4) that

$$\begin{aligned}(\nabla^{(2)}_\theta \Phi_{n,\varepsilon})(\theta) &= -2\sum_{k=1}^{n} (\nabla^{(2)}_\theta b^*)(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta) \circ \Big(\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})P_k(\theta)\Big)\\ &\quad - 2\sum_{k=1}^{n} (\nabla_\theta b)^*(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta)\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})(\nabla_\theta P_k)(\theta)\\ &= -2\sum_{k=1}^{n} (\nabla^{(2)}_\theta b^*)(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta) \circ \Big(\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})P_k(\theta_0)\Big)\\ &\quad - 2\,\delta\sum_{k=1}^{n} \Big\{(\nabla^{(2)}_\theta b^*)(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta) \circ \Big(\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})\Lambda(\widehat{X}^\varepsilon_{t_{k-1}}, \theta, \theta_0)\Big)\\ &\quad - (\nabla_\theta b)^*(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta)\widehat{\sigma}(\widehat{X}^\varepsilon_{t_{k-1}})(\nabla_\theta b)(\widehat{X}^\varepsilon_{t_{k-1}}, \mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}, \theta)\Big\}\\ &=: \Theta_1(\varepsilon, \delta) + \Theta_2(\varepsilon, \delta).\end{aligned}$$

Taking (**B2**) into consideration and mimicking the argument of Lemma 3.6, we obtain that

$$\Theta_1(\varepsilon,\delta)\to 0 \quad \text{ in probability as } \varepsilon\to 0,\ \ \delta\to 0.$$

Observe that

$$\begin{aligned}\Theta_2(\varepsilon,\delta) &= -2\int_0^T \Big\{(\nabla_\theta^{(2)} b^*)(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr L_{\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta)\circ\Big(\widehat\sigma(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta})\Lambda(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\theta,\theta_0)\Big)\mathrm{d}s\\ &\quad + 2\int_0^T \Big\{(\nabla_\theta b)^*(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr L_{\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta)\widehat\sigma(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta})(\nabla_\theta b)(\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta},\mathscr L_{\widehat X^\varepsilon_{\lfloor s/\delta\rfloor\delta}},\theta)\mathrm{d}s\\ &=: \Psi_1(\varepsilon,\delta)+\Psi_2(\varepsilon,\delta).\end{aligned}$$

Carrying out an analogous argument to derive Lemma 3.5, we infer that

z4 $$\Psi_1(\varepsilon,\delta)\to K(\theta) \quad \text{ in probability } \text{ as } \varepsilon\to 0,\delta\to 0 \tag{4.16}$$

by taking (**B2**) into account, and that

z5 $$\Psi_2(\varepsilon,\delta)\to 2I(\theta) \quad \text{ in probability } \text{ as } \varepsilon\to 0,\delta\to 0 \tag{4.17}$$

by using (**B1**). Thus, the desired assertion follows from (4.16) and (4.17) immediately. □

Now we start to finish the argument of Theorem 4.1 on the basis of the previous lemmas.

Proof of Theorem 4.1. The original idea on the proof of Theorem 4.1 is taken from [36]. To make the content self-contained, we herein provide a sketch of the proof. In terms of Theorem 3.1, there exists a sequence $\eta_{n,\varepsilon}\to 0$ as $\varepsilon\to 0$ and $n\to\infty$ such that $\widehat\theta_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0)\subset\Theta$, $\mathbb P$-a.s. By the Taylor expansion, one has

a4 $$(\nabla_\theta\Phi_{n,\varepsilon})(\widehat\theta_{n,\varepsilon}) = (\nabla_\theta\Phi_{n,\varepsilon})(\theta_0)+D_{n,\varepsilon}(\widehat\theta_{n,\varepsilon}-\theta_0),\quad \widehat\theta_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0) \tag{4.18}$$

with

$$D_{n,\varepsilon} := \int_0^1 (\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0+u(\widehat\theta_{n,\varepsilon}-\theta_0))\mathrm{d}u,\qquad \widehat\theta_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0).$$

Observe that, for $\widehat\theta_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0)$,

$$\begin{aligned}\|D_{n,\varepsilon}-K_0(\theta_0)\| &\le \|D_{n,\varepsilon}-(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)\|+\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)-K_0(\theta_0)\|\\ &\le \int_0^1 \|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0+u(\widehat\theta_{n,\varepsilon}-\theta_0))-(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)\|\mathrm{d}u\\ &\quad +\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)-K_0(\theta_0)\|\\ &\le \sup_{\theta\in B_{\eta_{n,\varepsilon}}(\theta_0)}\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta)-(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)\|+\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)-K_0(\theta_0)\|\\ &\le \sup_{\theta\in B_{\eta_{n,\varepsilon}}(\theta_0)}\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta)-K_0(\theta)\|+\sup_{\theta\in B_{\eta_{n,\varepsilon}}(\theta_0)}\|K_0(\theta)-K_0(\theta_0)\|\\ &\quad +2\|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta_0)-K_0(\theta_0)\|,\end{aligned}$$

in which $K_0(\cdot)$ is introduced in (4.15). This, together with Lemma 4.3 and continuity of $K_0(\cdot)$, gives that

a0
$$D_{n,\varepsilon} \to K_0(\theta_0) \qquad \text{in probability} \tag{4.19}$$

as $\varepsilon \to 0$ and $n \to \infty$. By following the exact line of [25, Theorem 2.2], we can deduce that $D_{n,\varepsilon}$ is invertible on the set

$$\Gamma_{n,\varepsilon} := \Big\{ \sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon})(\theta) - K_0(\theta_0)\| \le \frac{\alpha}{2}, \;\; \widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0) \Big\}$$

for some constant $\alpha > 0$. Let

$$\mathscr{D}_{n,\varepsilon} = \{D_{n,\varepsilon} \text{ is invertible }, \widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0)\}.$$

By virtue of Lemma 4.3, one has

n1
$$\lim_{\varepsilon \to 0, n \to \infty} \mathbb{P}\Big(\sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon})(\theta) - K_0(\theta_0)\| \le \frac{\alpha}{2} \Big) = 1. \tag{4.20}$$

On the other hand, recall that

n2
$$\lim_{\varepsilon \to 0, n \to \infty} \mathbb{P}\Big(\widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0) \Big) = 1. \tag{4.21}$$

By the fundamental fact: for any events A, B, $\mathbb{P}(AB) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$, we observe that

n3
$$\begin{aligned} 1 \ge \mathbb{P}(\Gamma_{n,\varepsilon}) &\ge \mathbb{P}\Big(\sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon})(\theta) - K_0(\theta_0)\| \le \frac{\alpha}{2} \Big) \\ &\quad + \mathbb{P}\Big(\widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0) \Big) - 1. \end{aligned} \tag{4.22}$$

Thus, taking advantage of (4.20), (4.21) as well as (4.22), we deduce from Sandwich theorem that

n4
$$\mathbb{P}(\mathscr{D}_{n,\varepsilon}) \ge \mathbb{P}(\Gamma_{n,\varepsilon}) \to 1 \tag{4.23}$$

as $\varepsilon \to 0$ and $n \to \infty$. Set

$$U_{n,\varepsilon} := D_{n,\varepsilon} \mathbf{1}_{\mathscr{D}_{n,\varepsilon}} + I_{p \times p} \mathbf{1}_{\mathscr{D}^c_{n,\varepsilon}},$$

where $I_{p\times p}$ is a $p \times p$ identity matrix. For $S_{n,\varepsilon} := \varepsilon^{-1}(\widehat{\theta}_{n,\varepsilon} - \theta_0)$, we deduce from (4.18) that

$$\begin{aligned} S_{n,\varepsilon} &= S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}_{n,\varepsilon}} + S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}^c_{n,\varepsilon}} \\ &= U^{-1}_{n,\varepsilon} D_{n,\varepsilon} S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}_{n,\varepsilon}} + S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}^c_{n,\varepsilon}} \\ &= \varepsilon^{-1} U^{-1}_{n,\varepsilon} \{(\nabla_\theta \Phi_{n,\varepsilon})(\widehat{\theta}_{n,\varepsilon}) - (\nabla_\theta \Phi_{n,\varepsilon})(\theta_0)\} \mathbf{1}_{\mathscr{D}_{n,\varepsilon}} + S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}^c_{n,\varepsilon}} \\ &= -\varepsilon^{-1} U^{-1}_{n,\varepsilon} (\nabla_\theta \Phi_{n,\varepsilon})(\theta_0) \mathbf{1}_{\mathscr{D}_{n,\varepsilon}} + S_{n,\varepsilon} \mathbf{1}_{\mathscr{D}^c_{n,\varepsilon}} \\ &\to I^{-1}(\theta_0) \int_0^T \Upsilon(X^0_s, \theta_0) \mathrm{d}B(s), \end{aligned}$$

as $\varepsilon \to 0$ and $n \to \infty$, where in the forth identity we dropped the term $(\nabla_\theta \Phi_{n,\varepsilon})(\widehat{\theta}_{n,\varepsilon})$ according to the notion of LSE and Fermat's lemma, and the last display follows from Lemma 4.2, (4.19) as well as (4.23) and by noting $K_0(\theta_0) = 2I(\theta_0)$. We therefore complete the proof. $\square$

5 An illustrative example

sec4

In this section, we intend to provide an example to demonstrate our main results. We first give the set-up of numerical example as following.

Example 5.1. Let $\theta = (\theta^{(1)}, \theta^{(2)})^* \in \Theta_0 := (c_1, c_2) \times (c_3, c_4) \subset \mathbb{R}^2$ for some $c_1 < c_2$ and $c_3 < c_4$. For any $\varepsilon \in (0, 1)$, consider the following scalar path-distribution dependent SDE

d1
$$\mathrm{d}X^\varepsilon(t) = \Big(\theta^{(1)} + \theta^{(2)} \int_{\mathscr{C}} b_0(X_t^\varepsilon, \zeta)\mathscr{L}_{X_t^\varepsilon}(\mathrm{d}\zeta)\Big)\mathrm{d}t + \varepsilon(1 + |X^\varepsilon(t)|)\,\mathrm{d}B(t), \quad t \in (0, T] \tag{5.1}$$

with the initial value $X_0^\varepsilon = \xi$, where $\theta \in \Theta_0$ is an unknown parameter with the true value $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}) \in \Theta_0$, and $b_0 : \mathscr{C} \times \mathscr{C} \to \mathbb{R}$ satisfy the global Lipschitz condition, i.e., there exists a constant $K > 0$ such that

d2
$$|b_0(\zeta_1, \zeta_2) - b(\zeta_1', \zeta_2')| \le K\{|\zeta_1 - \zeta_1'| + |\zeta_2 - \zeta_2'|\}, \qquad \zeta_1, \zeta_2, \zeta_1', \zeta_2' \in \mathscr{C}. \tag{5.2}$$

For any $\zeta \in \mathscr{C}$, $\mu \in \mathcal{P}_2(\mathscr{C})$ and $\theta = (\theta^{(1)}, \theta^{(2)})^*$, set

$$b(\zeta, \mu, \theta) := \theta^{(1)} + \theta^{(2)} \int_{\mathscr{C}} b_0(\zeta, \zeta')\mu(\mathrm{d}\zeta') \quad \text{and} \quad \sigma(\zeta, \mu) := 1 + |\zeta(0)|.$$

Then (5.1) can be reformulated as path distribution-dependent SDE (2.1).

5.1 Theoretical Result on Example 5.1

In this subsection, for (5.1) we aim to examine that all the assumptions imposed in Theorems 3.1 and 4.1 are applicable to the model (5.1). By a direct calculation, it follows from (5.2) that, for any $\mu, \nu \in \mathcal{P}_2(\mathscr{C})$ and $\zeta_1, \zeta_2 \in \mathscr{C}$,

d3
$$\begin{aligned} |b(\zeta_1, \mu, \theta) - b(\zeta_2, \nu, \theta)| &= |\theta^{(2)}| \cdot \Big| \int_{\mathscr{C}} b_0(\zeta_1, \zeta)\mu(\mathrm{d}\zeta) - \int_{\mathscr{C}} b_0(\zeta_2, \zeta')\nu(\mathrm{d}\zeta') \Big| \\ &\le |\theta^{(2)}| \int_{\mathscr{C}} \int_{\mathscr{C}} |b_0(\zeta_1, \zeta) - b_0(\zeta_2, \zeta')|\pi(\mathrm{d}\zeta, \mathrm{d}\zeta') \\ &\le K|\theta^{(2)}| \int_{\mathscr{C}} \int_{\mathscr{C}} \{|\zeta_1 - \zeta_2| + |\zeta - \zeta'|\}\pi(\mathrm{d}\zeta, \mathrm{d}\zeta') \\ &\le K(|c_3| \vee |c_4|)\{|\zeta_1 - \zeta_2| + \mathbb{W}_1(\mu, \nu)\} \\ &\le K(|c_3| \vee |c_4|)\{|\zeta_1 - \zeta_2| + \mathbb{W}_2(\mu, \nu)\}, \end{aligned} \tag{5.3}$$

in which $\pi \in \mathcal{C}(\mu, \nu)$. On the other hand, for any $x, y \in \mathbb{R}$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, one has

$$|\sigma(x, \mu) - \sigma(y, \nu)| \le |x - y|.$$

Hence, the assumption (**A1**) holds for (5.1). Next, for any $x, y \in \mathbb{R}$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, we have

$$|\sigma^{-2}(x, \mu) - \sigma^{-2}(y, \nu)| = \Big| \frac{1}{(1 + |x|)^2} - \frac{1}{(1 + |y|)^2} \Big| \le 4|x - y|.$$

So, (**A2**) is fulfilled. Furthermore, observe that

d4
$$(\nabla_\theta b)(\zeta, \mu, \theta) = \Big(1, \int_{\mathscr{C}} b_0(\zeta, \zeta')\mu(\mathrm{d}\zeta')\Big)^* \quad \text{and} \quad (\nabla_\theta(\nabla_\theta b))(\zeta, \mu, \theta) = \mathbf{0}_{2\times 2}, \tag{5.4}$$

where $\mathbf{0}_{2\times 2}$ stands for the 2×2-zero matrix. Thus, (5.3) yields that both (**B1**) and (**B2**) hold. We further assume that the initial value is global Lipschitz, i.e., there exists an $L>0$ such that

$$|\xi(t)-\xi(s)|\le L|t-s|,\qquad t,s\in[-r_0,0].$$

As a consequence, concerning (5.1), the assumptions (**A1**)-(**A3**) and (**B1**)-(**B2**) hold, respectively.

$$dX_t^\varepsilon=(\theta^{(1)}+\theta^{(2)}x_t^2)dt+\varepsilon(1+X^\varepsilon)dB(t)$$

According to (2.2), the contrast function admits the form below

$$\begin{aligned}\Psi_{n,\varepsilon}(\theta)=\varepsilon^{-2}\delta^{-1}\sum_{k=1}^n\frac{1}{(1+|X^\varepsilon(t_{k-1})|)^2}\\ \times\Big|X^\varepsilon(t_k)-X^\varepsilon(t_{k-1})-\Big(\theta^{(1)}+\theta^{(2)}\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)\Big)\delta\Big|^2.\end{aligned}$$

Observe that

$$\begin{aligned}\frac{\partial}{\partial\theta^{(1)}}\Psi_{n,\varepsilon}(\theta)=-2\,\varepsilon^{-2}\sum_{k=1}^n\frac{1}{(1+|X^\varepsilon(t_{k-1})|)^2}\Big\{X^\varepsilon(t_k)-X^\varepsilon(t_{k-1})\\ -\Big(\theta^{(1)}+\theta^{(2)}\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)\Big)\delta\Big\},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta^{(2)}}\Psi_{n,\varepsilon}(\theta)=-2\,\varepsilon^{-2}\sum_{k=1}^n\frac{1}{(1+|X^\varepsilon(t_{k-1})|)^2}\Big\{X^\varepsilon(t_k)-X^\varepsilon(t_{k-1})\\ -\Big(\theta^{(1)}+\theta^{(2)}\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)\Big)\delta\Big\}\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta).\end{aligned}$$

Subsequently, solving the equation below

$$\frac{\partial}{\partial\theta^{(1)}}\Psi_{n,\varepsilon}(\theta)=\frac{\partial}{\partial\theta^{(2)}}\Psi_{n,\varepsilon}(\theta)=0,$$

we obtain the LSE $\widehat{\theta}_{n,\varepsilon}=(\widehat{\theta}^{(1)}_{n,\varepsilon},\widehat{\theta}^{(2)}_{n,\varepsilon})^*$ of the unknown parameter $\theta=(\theta^{(1)},\theta^{(2)})^*\in\Theta_0$ possesses the formula

$$\widehat{\theta}^{(1)}_{n,\varepsilon}=\frac{A_2A_5-A_3A_4}{\delta(A_1A_5-A_4^2)}\qquad\text{and}\qquad\widehat{\theta}^{(2)}_{n,\varepsilon}=\frac{A_1A_3-A_2A_4}{\delta(A_1A_5-A_4^2)},$$

where

$$A_1:=\sum_{k=1}^n\frac{1}{(1+|X^\varepsilon(t_{k-1})|)^2},\qquad A_2:=\sum_{k=1}^n\frac{X^\varepsilon(t_k)-X^\varepsilon(t_{k-1})}{(1+|X^\varepsilon(t_{k-1})|)^2},$$

$$A_3:=\sum_{k=1}^n\frac{(Y^\varepsilon(t_k)-X^\varepsilon(t_{k-1}))\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)}{(1+|X^\varepsilon(t_{k-1})|)^2},\quad A_4:=\sum_{k=1}^n\frac{\int_{\mathscr{C}}b_0(\widehat{X}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{X}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)}{(1+|X^\varepsilon(t_{k-1})|)^2},$$

and

$$A_5 := \sum_{k=1}^{n} \frac{\Big(\int_{\mathscr{C}} b_0(\widehat{X}^{\varepsilon}_{t_{k-1}}, \zeta)\mathscr{L}_{\widehat{X}^{\varepsilon}_{t_{k-1}}}(\mathrm{d}\zeta)\Big)^2}{(1+|X^{\varepsilon}(t_{k-1})|)^2}.$$

In terms of Theorem 3.1, $\widehat{\theta}_{n,\varepsilon} \to \theta$ in probability as $\varepsilon \to 0$ and $n \to \infty$. Next, from (5.4), it follows that

$$I(\theta_0) = \int_0^T \frac{1}{(1+|X_s^0|)^2} \begin{pmatrix} 1 & b_0(X_s^0, X_s^0) \\ b_0(X_s^0, X_s^0) & b_0(X_s^0, X_s^0)^2 \end{pmatrix} \mathrm{d}s,$$

and, for $\zeta \in \mathscr{C}$,

$$\int_0^T \Upsilon(X_s^0, X_s^0, \theta_0)\mathrm{d}B(s) = \int_0^T \frac{1}{1+|X^0(s)|} \begin{pmatrix} 1 \\ b_0(X_s^0, X_s^0) \end{pmatrix} \mathrm{d}B(s).$$

At last, according to Theorem 4.1, we conclude that

$$\varepsilon^{-1}(\widehat{\theta}_{n,\varepsilon} - \theta_0) \to I^{-1}(\theta_0) \int_0^T \Upsilon(X_s^0, X_s^0, \theta_0)\mathrm{d}B(s) \qquad \text{in probability}$$

as $\varepsilon \to 0$ and $n \to \infty$ provided that $I(\cdot)$ is positive definite.

5.2 Numerical Result on Example 5.1

The numerical results are given in this subsection, aiming to illustrate the performance of proposed estimation on empirical problems. The simulating SDE corresponding to (5.1) is defined as follows:

$$dX^{\varepsilon}(t) = (\theta^{(1)} + \theta^{(2)}(X^{\varepsilon}(t) + X^{\varepsilon}(t-r_0) + \mathbb{E}X^{\varepsilon}(t)))dt + \varepsilon(1 + X^{\varepsilon}(t))dB(t)$$

where the true parameters are $\theta = (\theta^{(1)}, \theta^{(2)}) = (0, 1.7000e^{-4})$. By applying Euler-Maruyama scheme, we estimate θ where the perturbation scale ε is sufficiently small and sample size n is relatively large .

The data are sampled from the true function (given θ) with random noisy added as observation errors. Then the estimated parameters $\widehat{\theta}$ are computed according to our approach. We set particle size $n = 100$ and $n = 1000$ respectively with different levels of diffusion term.The experiments are repeated 1000 times, the mean and the stand deviation of $\widehat{\theta}$ are show in Table 1. In Case(a), we set $\varepsilon = 0.001$, the lowest relative error among three cases can be realised. (see Figure 1) . We gradually enlarge the level of diffusion part, then increasing of estimator error can be observed (see Figure 2 and 3), which complies with Theorem 3.1. The largest error (4.26%) is obtained while $\varepsilon = 0.1$, where the true function and the estimated function are oscillated due to large diffusion, while the error is still within the reasonable range. According to case (a)-(c) in Table 1, we observed that the larger number of particles leads to smaller standard deviation.

6 Conclusion

In the present paper, we established the LSE scheme for the stochastic parameter estimation of the path-distribution dependent SDEs. The numerical result is well coincide with theoretical result of Theorem 3.1. We would like to point out that the Maximum Likelihood estimation and the Bayesian estimation can also be used for stochastic parameter estimation with path-distribution

Table 1: Estimation Errors for the Example 5.1

True value $\theta = 1.7000e^{-4}$	Estimation Error(Avg)	Estimator Std
Case a: $\varepsilon = 0.001$		
n=1000	$(1.7272e^{-4})1.60\%$	$0.2615e^{-4}$
n=100	$(1.7392e^{-4})2.31\%$	$0.7525e^{-4}$
Case b: $\varepsilon = 0.01$		
n=1000	$(1.7331e^{-4})1.95\%$	$0.7704e^{-4}$
n=100	$(1.7349e^{-4})2.05\%$	$0.9929e^{-4}$
Case c: $\varepsilon = 0.1$		
n=1000	$(1.6275e^{-4})4.26\%$	$1.2343e^{-4}$
n=100	$(1.7596e^{-4})3.51\%$	$1.7317e^{-4}$

tab:PPer

dependent SDEs under real applications diversely and widely as in economics, finance, physics etc.. In our future work, we plan to study maximum likelihood estimation of certain parameters for the Black-Scholes or business cycle models.

References

ai [1] Ait-Sahalia, Y., Maximum likelihood estimation of discretely sampled diffusion: a closed-form approximation approach, *Econometrica*, **70** (2002), 223–262.

B08 [2] Bishwal, J. P. N., *Parameter Estimation in Stochastic Differential Equations*, Springer, Berlin, 2008.

Chen [3] Chen, M.-F., *From Markov Chains to Non-Equilibrium Particle Systems.*, Second Ed., World Scientific, Singapore, 2004.

D76 [4] Dorogovcev, A. Ja., The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. Math. Statist.*, **10** (1976), 73–82.

DST [5] Dos Reis, G., Salkeld, W., Tugaut, J., Freidlin-Wentzell LDPs in path space for McKean-Vlasov equations and the Functional Iterated Logarithm Law, *Ann. Appl. Probab.* **29** (2019) no. 3, 1487–1540

F98 [6] Friedman, A., *Stochastic Differential Equations and Applications*, Vol 1, Academic Press, INC, 1975.

GS [7] Gloter, A., Sørensen, M., Estimation for stochastic differential equations with a small diffusion coefficient, *Stochastic Process. Appl.*, **119** (2009), 679–699.

HL [8] Hu, Y., Long, H., Least squares estimator for Ornstein-Uhlenbeck processes driven by α-stable motions, *Stochastic Process. Appl.*, **119** (2009), 2465–2480.

Huang [9] Huang, X., Liu, C., Wang, F.-Y., Order preservation for path-distribution dependent, *Commun. Pure Appl. Anal.*, **17** (2018), no. 5, 2125–2133.

HRW [10] Huang, X., Röckner, M., Wang, F.-Y., Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, *Discrete Contin. Dyn. Syst.*, **39** (2019), no. 6, 3017–3035.

IH [11] Ibragimov, I. A., Khasminski, R. Z., *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York, Berlin, 1981.

IkedaW [12] Ikeda, N. , Watanabe, S. *Stochastic Differential Equations and Diffusion Processes*, Second Ed., North-Holland Mathcmntilcal Library, 24, Amsterdam, 1989.

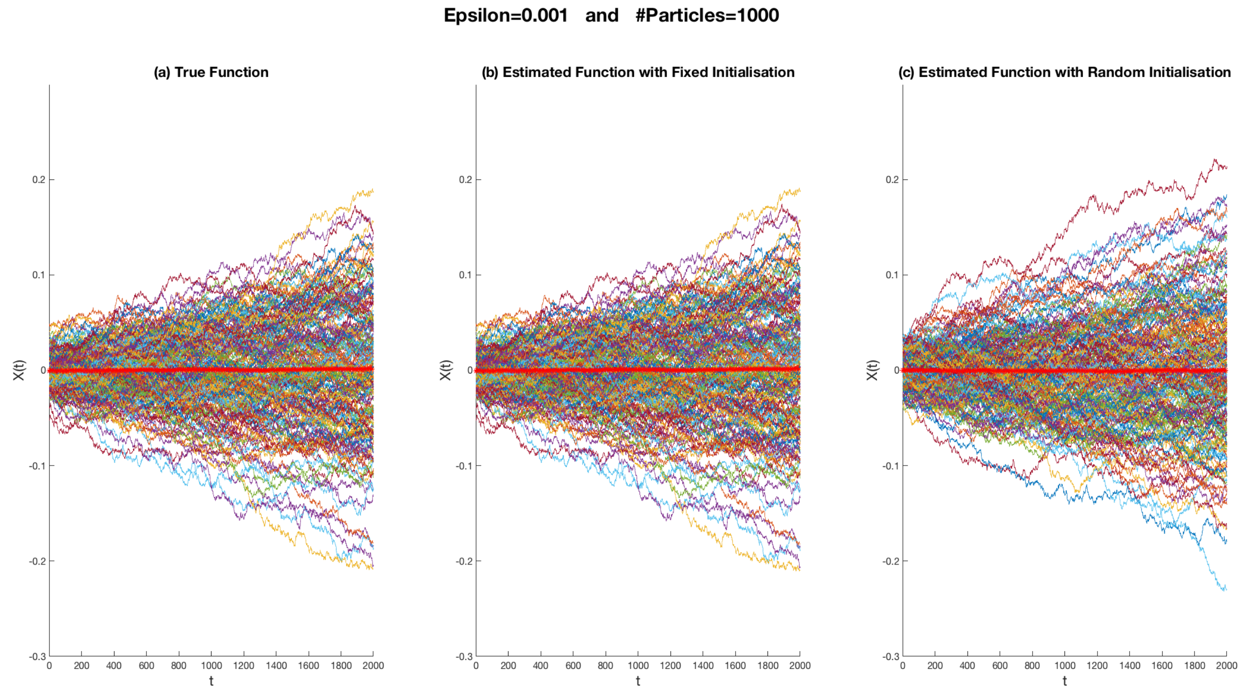

Figure 1: Simulation results with $\varepsilon = 0.001$ and $n = 1000$ which are corresponding to the Case (a). The (a) indicates the real data of the SDEs and the (b) indicates the estimated date with fixed initialisation and the (c) indicates the estimated data with the random initialisation and the red line indicates the mean value of the 1000 times simulation. We only draw 200 particles in these three figures (Figure 2 and Figure 3 below have similar behaviour with different ε).

JM [13] Jourdain, B., Méléard, S., Propagation of chaos and fluctuations for a moderate model with smooth initial data, *Ann. Inst. H. Poincaré Probab. Statist.*, **34** (1998), 727–766.

Ka88 [14] Kasonga, R. A., The consistency of a nonlinear least squares estimator for diffusion processes, *Stochastic Process. Appl.*, **30** (1988), 263–275.

K12 [15] Klebaner, F. C., *Introduction to Stochastic Calculus with Applications*, Third Ed., Imperial College Press, London, 2012.

Ku01 [16] Kunitomo, N., Takahashi, A., The asymptotic expansion approach to the valuation of interest rate contingent claims, *Math. Finance*, **11** (2001), 117–151.

Ku04 [17] Kutoyants, Yu. A., Statistical Inference for Ergodic Diffusion Processes, Springer-Verlag, London, Berlin, Heidelberg, 2004.

Ku94 [18] Kutoyants, Yu A., Identification of dynamical systems with small noise, Kluwer Academic Publishers Group, Dordrecht, 1994.

LM [19] Li, J., Min, H., Weak solutions of mean-field stochastic differential equations, *Stoch. Anal. Appl.*, **35** (2017), 542–568.

LMb [20] Li, J., Min, H., Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games, *SIAM J. Control Optim.*, **54** (2016), 1826–1858.

Li16 [21] Li, J., Wu, J.-L., On drift parameter estimation for mean-reversion type stochastic differential equations with discrete observations, *Adv. Difference Equ.*, 2016, Paper No. 90, 23 pages. (https://doi.org/10.1186/s13662-016-0819-1)

Li19 [22] Li, Jingjie; Wu, Jiang-Lun; Zhang, Guang, Estimation of intrinsic growth factors in a class of stochastic population model. (English summary) Stoch. Anal. Appl. 37 (2019), no. 4, 602–619.

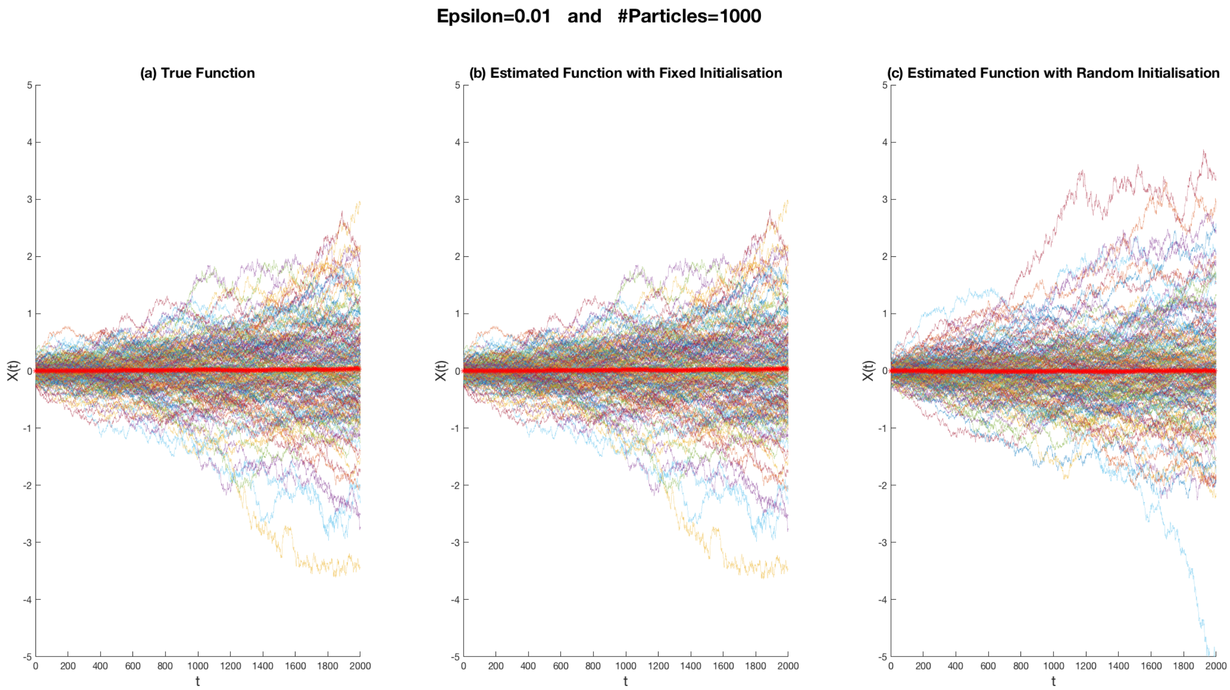

Figure 2: Simulation result with $\varepsilon = 0.01$ and $n = 1000$ which are corresponding to the Case (b).

LS01 [23] Liptser, R. S., Shiryaev, A. N., Statistics of Random Processes: II Applications, Second Edition , Springer-Verlag, Berlin, Heidelberg, New York, 2001.

LMS [24] Long, H., Ma, C., Shimizu, Y., Least squares estimators for stochastic differential equations driven by small Lévy noises, *Stochastic Process. Appl.*,**127** (2017), 1475–1495.

LSS [25] Long, H., Shimizu, Y., Sun, W., Least squares estimators for discretely observed stochastic processes driven by small Lévy noises, *J. Multivariate Anal.*, **116** (2013), 422–439.

M08 [26] Mao, X., *Stochastic Differential Equations and Applications*, Second Ed., Horwood Publishing Limited, Chichester, 2008.

M05 [27] Masuda, H., Simple estimators for non-linear Markovian trend from sampled data: I. ergodic cases. MHF Preprint Series 2005-7, Kyushu University, 2005.

MV [28] Mishura, Yu. S., Veretennikov, A. Yu., Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, arXiv:1603.02212v4.

Oksendal [29] Øksendal, B., *Stochastic Differential Equations. An Introduction with Applications*, Sixth Ed., Springer-Verlag, Berlin, Heidelberg, New York, 2003.

P199 [30] Prakasa Rao, B. L. S., *Statistical Inference for Diffusion Type Processes*, Kendall's Library of Statistics, 8. Edward Arnold, London, Oxford University Press, New York, 1999.

RW19 [31] Panpan Ren, Jiang-Lun Wu, *Least squares estimator for path-dependent Mckean-Vlasov SDEs via Discrete-time observations*, *Acata.Math. Scientia*, **39**, (2019) 691-716.

Protter [32] Protter, P., *Stochastic Integrations and Differential Equations*, Second edition. Applications of Mathematics (New York), 21. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2004.

SY06 [33] Shimizu, Y., Yoshida, N., Estimation of parameters for diffusion processes with jumps from discrete observations, *Stat. Inference Stoch. Process.*, **9** (2006), 227–277.

SM03 [34] Sørensen, M., Uchida, M., Small diffusion asymptotics for discretely sampled stochastic differential equations, *Bernoulli*, **9** (2003), 1051–1069.

Ta04 [35] Takahashi, A., Yoshida, N., An asymptotic expansion scheme for optimal investment problems, *Stat. Inference Stoch.Process.*, **7** (2004), 153–188.

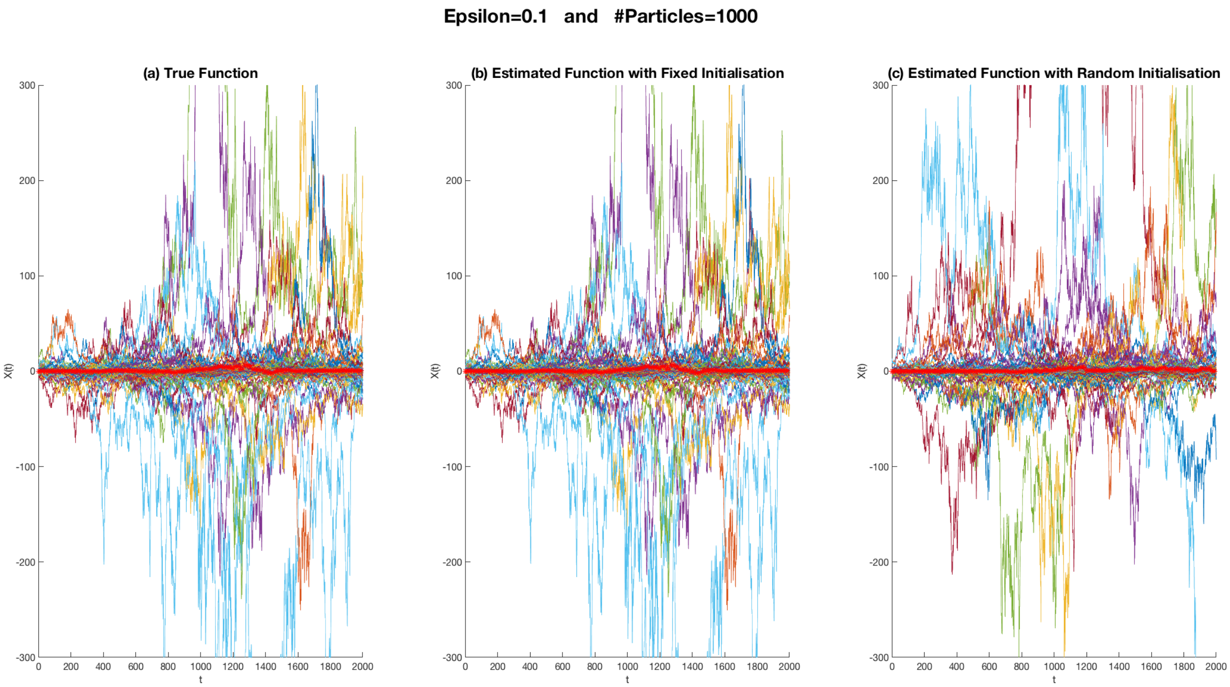

Figure 3: Simulation result with $\varepsilon = 0.1$ and $n = 1000$ which are corresponding to the Case (c).

U04 [36] Uchida, M., Estimation for discretely observed small diffusions based on approximate martingale estimating functions, *Scand. J. Statist.*, **31** (2004), 553–566.

U08 [37] Uchida, M., Approximate martingale estimating functions for stochastic differential equations with small noises, *Stochastic Process. Appl.*, **118** (2008), 1706–1721.

V98 [38] van der Vaart, A. W., *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge, 1998.

Ve [39] Veretennikov, A. Yu., On ergodic measures for McKean-Vlasov stochastic equations, In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, 471–486, Springer, Berlin, 2006.

RWW18 [40] Wang, F.-Y., Distribution-Dependent SDEs for Landau Type Equations, *Stoch.Proc. Appl.*, **128** (2018), 595–621.

Wen [41] Wen, J., Wang, X., Mao, S., Xiao, X., Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application, *Appl. Math. Comput.*, **274** (2016), 237–246.

Y1992b [42] Yoshida, N., Asymptotic expansion for statistics related to small diffusions, *J. Japan Statist. Soc.*, **22** (1992), 139–159.