Least squares estimator for path-dependent McKean-Vlasov SDEs via discrete-time observations

Panpan Ren and Jiang-Lun Wu
Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, UK
673788@swansea.ac.uk, J.L.Wu@swansea.ac.uk

October 9, 2018

Abstract

In this paper, we are interested in least squares estimator for a class of path-dependent McKean-Vlasov stochastic differential equations (SDEs). More precisely, we investigate the consistency and asymptotic distribution of the least squares estimator for the unknown parameters involved by establishing an appropriate contrast function. Comparing to the existing results in the literature, the innovations of our paper lie in three aspects: (i) We adopt a tamed Euler-Maruyama algorithm to establish the contrast function under the monotone condition, under which the Euler-Maruyama scheme no longer works; (ii) We take the advantage of linear interpolation with respect to the discrete-time observations to approximate the functional solution; (iii) Our model is more applicable and practice as we are dealing with SDEs with irregular coefficients (e.g., Hölder continuous) and path-distribution dependent.

AMS subject Classification: 62F12, 62M05, 60G52, 60J75
Keywords: McKean-Vlasov stochastic differential equation, tamed Euler-Maruyama scheme, weak monotonicity, least squares estimator, consistency, asymptotic distribution.

1 Introduction and main results

We start with some notation and terminology. Let $(\mathbb{R}^d, \langle\cdot,\cdot\rangle, |\cdot|)$ be the d-dimensional Euclidean space, and $\mathbb{R}^d\otimes\mathbb{R}^m$ the collection of all $d\times m$ matrices endowed with the Hilbert-Schmidt norm $\|\cdot\|$. For fixed $r_0>0$, $\mathscr{C}:=C([-r_0,0];\mathbb{R}^d)$ stands for the family of all continuous functions $f:[-r_0,0]\to\mathbb{R}^d$ which is a Banach space with the uniform norm $\|f\|_\infty:=\sup_{-r_0\le v\le 0}|f(v)|$. Given any integer $p\ge 1$, we use Θ to denote a bounded, open and convex subset of $\mathbb{R}^p$ whose closure is written as $\overline{\Theta}$. Let $\mathcal{P}(\mathscr{C})$ be the totality of all probability measures on $\mathscr{C}$. Set $\mathcal{P}_2(\mathscr{C}):=\{\mu\in\mathcal{P}(\mathscr{C}):\mu(\|\cdot\|_\infty^2):=\int_{\mathscr{C}}\|\xi\|_\infty^2\mu(\mathrm{d}\xi)<\infty\}$. $(\mathcal{P}_2(\mathscr{C}),\mathbb{W}_2)$ is a Polish space under the Warsserstein distance $\mathbb{W}_2$ on $\mathcal{P}_2(\mathscr{C})$ defined by

$$\mathbb{W}_2(\mu,\nu):=\inf_{\pi\in\mathscr{C}(\mu,\nu)}\left(\int_{\mathscr{C}\times\mathscr{C}}\|\xi-\eta\|_\infty^2\pi(\mathrm{d}\xi,\mathrm{d}\eta)\right)^{\frac{1}{2}},\quad \mu,\nu\in\mathcal{P}_2(\mathscr{C}),$$

where $\mathscr{C}(\mu,\nu)$ is the set of couplings for μ and ν. As usual, we use $\lfloor a\rfloor$ to denote the integer part of $a\ge 0$.

The time evolution for most of stochastic dynamical systems depends not only on the present state but also on the past path. So, path-dependent (i.e., functional) SDEs are much more desirable;

see, e.g., the monograph [27]. Since the pioneer work [14] due to Itô and Nisio, path-dependent SDEs have been investigated considerably owing to their theoretical and practical importance; see, e.g., Hairer et al. [9], Wang [37] and the references within.

McKean-Vlasov SDEs, which are SDEs with coefficients dependent on the law, were initiated by [25] inspired by Kac's programme in Kinetic theory. An excellent and thorough account of the general theory of McKean-Vlasov SDEs and their particle approximations can be found in [32]. McKean-Vlasov SDEs are alternatively referred to as mean-field SDEs in the literature, which have wide applications in interacting particle systems, optimal control problems, differential games, just to mention but a few. Recently, McKean-Vlasov SDEs have been extensively investigated on, e.g., wellposedness of strong/weak solutions (cf. [5, 8, 16, 26, 38]), Freidlin-Wentzell large deviation principles (cf. [5]), ergodicity (cf. [3, 4, 36]), links with nonlinear partial differential equations (cf. [2, 12, 12]), and distribution properties (cf.[11, 37]).

On the other hand, from stochastic and/or statistical aspects, there exist unknown parameters in various type SDEs arising in mathematical modeling (cf. [1]). Hence, there are vast of investigations paying attention to parameter estimations for SDEs via maximum likelihood estimator, least squares estimator (LSE for short), trajectory-fitting estimator, among others. See, for instance, [15, 17, 24, 28, 30]. In the same vein, the parameter estimations for SDEs (without path-dependence) with small noises have been developed very well; see, e.g., [7, 10, 19, 18, 20, 21, 23, 31, 33, 34], and references therein.

From above discussion, it is very natural to consider SDEs together with all four features of path dependence, distribution dependence, small noises and unknown parameter. So, in the present work, we focus on the following path-distribution SDE

$$\mathrm{d}X^\varepsilon(t)=b(X_t^\varepsilon,\mathscr{L}_{X_t^\varepsilon},\theta)\mathrm{d}t+\varepsilon\,\sigma(X_t^\varepsilon,\mathscr{L}_{X_t^\varepsilon})\mathrm{d}B(t),\quad t>0,\qquad X_0^\varepsilon=\xi\in\mathscr{C}. \tag{1.1}$$

Herein, $\varepsilon\in(0,1)$ is the scale parameter; for fixed t, $X_t^\varepsilon(v):=X^\varepsilon(t+v),v\in[-r_0,0]$, is called the segment (or window) process generated by $X^\varepsilon(t)$; $\mathscr{L}_{X_t^\varepsilon}$ stands for the distribution of X_t^ε; $b:\mathscr{C}\times\mathcal{P}_2(\mathscr{C})\times\Theta\to\mathbb{R}^d$ and $\sigma:\mathscr{C}\times\mathcal{P}_2(\mathscr{C})\to\mathbb{R}^d\otimes\mathbb{R}^m$ are continuous w.r.t. the first variable and the second variable; $\Theta\ni\theta$ is an unknown parameter whose true value is written as $\theta_0\in\Theta$; and $(B(t))_{t\ge0}$ is an m-dimensional Brownian motion on a filtered probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbb{P})$ satisfying the usual conditions, that is, $\mathscr{F}_t$ is non-decreasing (i.e., $\mathscr{F}_s\subseteq\mathscr{F}_t, s\le t$), $\mathscr{F}_0$ contains all $\mathbb{P}$-null sets and $\mathscr{F}_t$ is right continuous (i.e., $\mathscr{F}_t=\mathscr{F}_{t+}:=\bigcap_{s\uparrow t}\mathscr{F}_s$).

To guarantee the existence and uniqueness of solutions to (1.1), we assume that, for any $\zeta_1,\zeta_2\in\mathscr{C}$, $\mu,\nu\in\mathcal{P}_2(\mathscr{C})$, and $\theta\in\Theta$,

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

$$\langle\zeta_1(0)-\zeta_2(0),b(\zeta_1,\mu,\theta)-b(\zeta_2,\nu,\theta)\rangle\le\alpha_1\|\zeta_1-\zeta_2\|_\infty^2+\alpha_2\mathbb{W}_2(\mu,\nu)^2;$$

(**A2**) There exist $\beta_1,\beta_2>0$ such that

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

From [12, Theorem 3.1], (1.1) has a unique strong solution $(X^\varepsilon(t))_{t\ge-r_0}$ under the assumptions (**A1**) and (**A2**). For any $\zeta_1,\zeta_2\in\mathscr{C}$, $\mu,\nu\in\mathcal{P}_2(\mathscr{C})$, and $\theta\in\Theta$, if there exist $\alpha,\beta>0$ such that

$$\langle\zeta_1(0)-\zeta_2(0),b(\zeta_1,\mu,\theta)-b(\zeta_2,\mu,\theta)\rangle\le\alpha\|\zeta_1-\zeta_2\|_\infty^2$$

and

$$|b(\zeta_2,\mu,\theta)-b(\zeta_2,\nu,\theta)|\le\beta\mathbb{W}_2(\mu,\nu),$$

then (**A1**) holds.

Without loss of generality, we arbitrarily fix the time horizontal $T>0$ and assume that there exist positive integers n,M sufficiently large such that $\delta:=\frac{T}{n}=\frac{r_0}{M}$. Now we define the continuous-time tamed Euler-Maruyama (EM) scheme (see, e.g., [13]) associated with (1.1)

$$\mathrm{d}Y^\varepsilon(t)=b^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\theta)\mathrm{d}t+\varepsilon\,\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\mathrm{d}B(t),\qquad t>0 \tag{1.2}$$

with the initial value $Y^\varepsilon(t)=X^\varepsilon(t)=\xi(t)$ for any $t\in[-r_0,0]$, where

- $t_\delta:=\lfloor t/\delta\rfloor\delta$ for $t\ge0$;
- For any $\zeta\in\mathscr{C}$ and $\mu\in\mathcal{P}_2(\mathscr{C})$,

$$b^{(\delta)}(\zeta,\mu,\theta):=\frac{b(\zeta,\mu,\theta)}{1+\delta^\alpha|b(\zeta,\mu,\theta)|},\qquad\alpha\in(0,1/2]; \tag{1.3}$$

- For $k=0,1,\cdots,n$, $\overline{Y}^\varepsilon_{k\delta}=\{\overline{Y}^\varepsilon_{k\delta}(s):-r_0\le s\le0\}$, a $\mathscr{C}$-valued random variable, is defined by

$$\overline{Y}^\varepsilon_{k\delta}(s)=Y^\varepsilon((k-i)\delta)+\frac{s+i\delta}{\delta}\{Y^\varepsilon((k-i)\delta)-Y^\varepsilon((k-i-1)\delta)\} \tag{1.4}$$

  for any $s\in[-(i+1)\delta,-i\delta]$, $i=0,1,\cdots,M-1$, that is, $\overline{Y}^\varepsilon_{k\delta}$ is the linear interpolation of the points $(Y^\varepsilon(l\delta))_{l=k-M,\cdots,k}$.

We denote $(Y^\varepsilon_t)_{t\ge0}$ by the segment process generated by $(Y^\varepsilon(t))_{t\ge-r_0}$. It is worthy to point out that $\overline{Y}^\varepsilon_{t_\delta}\in\mathscr{C}$ is defined by (1.4) rather than by $\overline{Y}^\varepsilon_{t_\delta}(s)=\overline{Y}^\varepsilon(t_\delta+s)$ for any $s\in[-r_0,0]$.Based on the continuous-time tamed EM algorithm (1.2), we design the following contrast function

$$\Psi_{n,\varepsilon}(\theta)=\varepsilon^{-2}\delta^{-1}\sum_{k=1}^nP_k^*(\theta)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta), \tag{1.5}$$

in which, for $k=1,\cdots,n$,

$$P_k(\theta):=Y^\varepsilon(k\delta)-Y^\varepsilon((k-1)\delta)-b^{(\delta)}(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\delta,\ \widehat{\sigma}(\overline{Y}^\varepsilon_{k\delta}):=(\sigma\sigma^*)^{-1}(\overline{Y}^\varepsilon_{k\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{k\delta}}). \tag{1.6}$$

For more motivations on the construction of constrast function above, we refer to Ren-Wu [29]. To obtain the LSE of $\theta\in\Theta$, it is sufficient to choose an element $\widehat{\theta}_{n,\varepsilon}\in\Theta$ satisfying

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

Whence, for

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

one has

$$\Phi_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon})=\min_{\theta\in\Theta}\Phi_{n,\varepsilon}(\theta). \tag{1.7}$$

We shall rewrite $\widehat{\theta}_{n,\varepsilon}\in\Theta$ such that (1.7) holds true as

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

which is called the LSE of the unknown parameter $\theta\in\Theta$.

To discuss the consistency of LSE (see Theorem 1.1 below), we further suppose that, for any $\zeta,\zeta_1,\zeta_2\in\mathscr{C}$, $\mu,\nu\in\mathcal{P}_2(\mathscr{C})$, and $\theta\in\Theta$,

(**B1**) There exist $q_1,L_1>0$ such that

$$|b(\zeta_1,\mu,\theta)-b(\zeta_2,\nu,\theta)|\le L_1\Big\{(1+\|\zeta_1\|_\infty^{q_1}+\|\zeta_2\|_\infty^{q_1})\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mu,\nu)\Big\};$$

(**B2**) There exist $q_2,L_2>0$ such that

$$\sup_{\theta\in\overline{\Theta}}\|(\nabla_\theta b)(\zeta_1,\mu,\theta)-(\nabla_\theta b)(\zeta_2,\nu,\theta)\|\le L_2\Big\{(1+\|\zeta_1\|_\infty^{q_2}+\|\zeta_2\|_\infty^{q_2})\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mu,\nu)\Big\},$$

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

(**B3**) $(\sigma\sigma^*)(\zeta,\mu)$ is invertible, and there exist $q_3,L_3>0$ such that

$$\|(\sigma\sigma^*)^{-1}(\zeta_1,\mu)-(\sigma\sigma^*)^{-1}(\zeta_2,\nu)\|\le L_3\Big\{(1+\|\zeta_1\|_\infty^{q_3}+\|\zeta_2\|_\infty^{q_3})\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mu,\nu)\Big\};$$

(**B4**) There exists a constant $K>0$ such that

$$|\xi(t)-\xi(s)|\le K|t-s|,\quad t,s\in[-r_0,0],$$

where $\xi(\cdot)$ stands for the initial value of (1.1).

(**B5**) There exists a constant $K_0,p_0>0$ such that

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

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

In order to reveal the asymptotic distribution of LSE (see Theorem 1.2 below), we in addition assume that

(**C**) There exist $q_4,L_4>0$ such that

$$\begin{aligned}&\sup_{\theta\in\overline{\Theta}}\|(\nabla_\theta(\nabla_\theta b^*))(\zeta_1,\mu,\theta)-(\nabla_\theta(\nabla_\theta b^*))(\zeta_2,\nu,\theta)\|\\&\le L_4\Big\{(1+\|\zeta_1\|_\infty^{q_4}+\|\zeta_2\|_\infty^{q_4})\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mu,\nu)\Big\},\end{aligned}$$

where b^* means the transpose of b.

Next we consider the following deterministic path-dependent ordinary equation

$$\mathrm{d}X^0(t) = b(X_t^0, \mathscr{L}_{X_t^0}, \theta_0)\mathrm{d}t, \quad t > 0, \quad X_0^0 = \xi \in \mathscr{C}. \tag{1.8}$$

Under the assumption (**A1**), (1.8) is wellposed. In (1.8), $\mathscr{L}_{X_t^0}$ is indeed a Dirac's delta measure at the point X_t^0 as X_t^0 is deterministic. To unify the notation, we keep the notation $\mathscr{L}_{X_t^0}$ in lieu of $\delta_{X_t^0}$. We remark that (**B4**) is imposed to guarantee that the linear interpolation $\overline{Y}_{t_\delta}^{\varepsilon}$ tends to X_t^0 in the moment sense, see Lemma 2.2 below.

For any random variable $\zeta \in \mathscr{C}$ with $\mathscr{L}_\zeta \in \mathcal{P}_2(\mathscr{C})$, set

$$\Gamma(\zeta, \theta, \theta_0) := b(\zeta, \mathscr{L}_\zeta, \theta_0) - b(\zeta, \mathscr{L}_\zeta, \theta), \quad \Gamma^{(\delta)}(\zeta, \theta, \theta_0) := b^{(\delta)}(\zeta, \mathscr{L}_\zeta, \theta_0) - b^{(\delta)}(\zeta, \mathscr{L}_\zeta, \theta), \tag{1.9}$$

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

$$\Xi(\theta) = \int_0^T \Gamma^*(X_t^0, \theta, \theta_0)\widehat{\sigma}(X_t^0)\Gamma(X_t^0, \theta, \theta_0)\mathrm{d}t,$$

where $(X_t^0)_{t\geq 0}$ is the functional solution to (1.8).

The theorem below is concerned with the consistency of the LSE for the parameter $\theta \in \Theta$, which is the first contribution of our work.

Theorem 1.1. *Let* (**A1**) $-$ (**A2**) *and* (**B1**) $-$ (**B4**) *hold and assume further that* $\Xi(\theta) > 0$ *for* $\theta \neq \theta_0$. *Then*

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

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 $B \in \mathbb{R}^d$, define $A \circ B \in \mathbb{R}^p \otimes \mathbb{R}^p$ by

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

For any $\theta \in \Theta$, set

$$I(\theta) := \int_0^T (\nabla_\theta b)^*(X_t^0, \mathscr{L}_{X_t^0}, \theta)\widehat{\sigma}(X_t^0)(\nabla_\theta b)(X_t^0, \mathscr{L}_{X_t^0}, \theta)\mathrm{d}t, \tag{1.10}$$

$$K(\theta) : = -2\int_0^T (\nabla_\theta^{(2)} b^*)(X_t^0, \mathscr{L}_{X_t^0}, \theta) \circ \Big(\widehat{\sigma}(X_t^0)\Gamma(X_t^0, \theta, \theta_0)\Big)\mathrm{d}t, \tag{1.11}$$

where $(\nabla_\theta^{(2)} b^*) := (\nabla_\theta(\nabla_\theta b^*))$, and, for any random variable $\zeta \in \mathscr{C}$ with $\mathscr{L}_\zeta \in \mathcal{P}_2(\mathscr{C})$,

$$\Upsilon(\zeta, \theta_0) = (\nabla_\theta b)^*(\zeta, \mathscr{L}_\zeta, \theta_0)\widehat{\sigma}(\zeta)\sigma(\zeta, \mathscr{L}_\zeta). \tag{1.12}$$

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

Theorem 1.2. *Let the assumptions of Theorem 1.1 hold and suppose further that* (**C**) *holds and that* $I(\cdot)$ *and* $K(\cdot)$ *defined in* (1.10) *and* (1.11)*, respectively, are continuous. Then,*

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

as $\varepsilon \to 0$ *and* $n \to \infty$*, where* $\Upsilon(\cdot)$ *is given in* (1.12).

With contrast to the existing literature, the innovations of this paper lie in:

(i) The classical contrast function for LSE is based on EM algorithm. Whereas, under the monotone condition, the EM scheme no longer works. Hence in the present work we adopt a tamed EM method to establish the corresponding contrast function. The above is our first innovation.

(ii) For the classical setup, the discrete-time observations at the gridpoints are sufficient to construct the contrast function. Nevertheless, for our present model, the discrete-time observations are insufficient to establish the contrast function since the SDEs involved are path-dependent. In this paper, we overcome the difficulty mentioned by linear interpolation w.r.t. the discrete-time observations. The above is our second innovation.

(iii) Our model is much more applicable, which allow the coefficients to be distribution-dependent and weakly monotone. In particular, the drift terms are allowed to be singular (e.g., Hölder continuous). The above is our third innovation.

Now, we provide a concrete example to demonstrate Theorems 1.1 and 1.2.

Example 1.3. *For any $\varepsilon\in(0,1)$, consider the following scalar path-distribution dependent SDE*

(1.13)

$$\begin{aligned}\mathrm{d}X^\varepsilon(t)=&\Big\{\theta^{(1)}+\theta^{(2)}\int_{\mathscr{C}}\Big(-(X^\varepsilon(t))^3+X^\varepsilon(t)+\int_{-r_0}^0X^\varepsilon(t+\theta)\mathrm{d}\theta+\int_{\mathscr{C}}\zeta(\theta)\mathrm{d}\theta\Big)\mathscr{L}_{X_t^\varepsilon}(\mathrm{d}\zeta)\Big\}\mathrm{d}t\\&+\varepsilon\Big(1+\int_{-r_0}^0X^\varepsilon(t+\theta)|\mathrm{d}\theta\Big)\,\mathrm{d}B(t),\quad t\ge 0\end{aligned}$$

with the initial value $X_0^\varepsilon=\xi\in\mathscr{C}$ which is Lipschitz, where, for some $c_1<c_2$ and $c_3<c_4$, $\theta=(\theta^{(1)},\theta^{(2)})^*\in\Theta_0:=(c_1,c_2)\times(c_3,c_4)\subset\mathbb{R}_+^2$ is an unknown parameter with the true value $\theta_0=(\theta_0^{(1)},\theta_0^{(2)})^*\in\Theta_0$. Let $\widehat{\theta}_{n,\varepsilon}$ be the LSE for the unknown parameter $\theta=(\theta^{(1)},\theta^{(2)})^*\in\Theta_0$. Then,

$$\widehat{\theta}_{n,\varepsilon}\to\theta_0\qquad\text{in probability as }\varepsilon\to0\quad\text{and}\quad n\to\infty,$$

and

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

as $\varepsilon\to0$ and $n\to\infty$, where

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

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

$$\Upsilon(\zeta,\theta_0)=\frac{1}{1+|\zeta|}\begin{pmatrix}1\\ b_0(\zeta,\zeta)\end{pmatrix}.$$

The rest of the paper is organised as follows. In Section 2, we prove Theorem 1.1 on the basis of several auxiliary lemmas; Section 3 is devoted to show Theorem 1.2; Section 4, the final section, is for the proof of Example 1.3. Throughout this paper, we emphasise that $c>0$ is a generic constant whose value may change from line to line.

2 Proof of Theorem 1.1

To complete the proof of Theorem 1.1, we provide some technical lemmas. The lemma below expounds that the path associated with (1.2) is uniformly bounded in the p-th moment sense.

Lemma 2.1. *Let* (**A1**) *and* (**A2**) *hold. Then, for any* $p>0$ *there is a constant* $C_{p,T}>0$ *such that*

$$\sup_{0\le t\le T}\|X_t^0\|_\infty^p\le C_{p,T}(1+\|\xi\|_\infty^p), \tag{2.1}$$

and

$$\sup_{0\le t\le T}\mathbb{E}\Big(\sup_{-r_0\le s\le t}|Y^\varepsilon(s)|^p\Big)\le C_{p,T}(1+\|\xi\|_\infty^p). \tag{2.2}$$

Proof. With the assumption (**A1**) in hand, the proof of (2.1) can be achieved by the chain rule and the Gronwall inequality. We herein omit the details since it is standard. Now we turn to show the argument of (2.2). By Hölder's inequality, it suffices to verify that (2.2) holds for any $p>4$. By Itô's formula, we deduce that

$$\begin{aligned}|Y^\varepsilon(t)|^p&=|Y^\varepsilon(0)|^p+\int_0^t\Big\{p|Y^\varepsilon(s)|^{p-2}\langle Y^\varepsilon(s),b^{(\delta)}(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta)\rangle+\frac{p}{2}|Y^\varepsilon(s)|^{p-2}\|\sigma^*(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2\\&\quad+\frac{p(p-2)}{2}|Y^\varepsilon(s)|^{p-4}|\sigma^*(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})Y^\varepsilon(s)|^2\Big\}\mathrm{d}s\\&\quad+p\int_0^t|Y^\varepsilon(s)|^{p-2}\langle Y^\varepsilon(s),\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\mathrm{d}B(s)\rangle\\&\le p\int_0^t|Y^\varepsilon(s)|^{p-2}\langle Y^\varepsilon(s_\delta),b^{(\delta)}(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta)\rangle\mathrm{d}s\\&\quad+p\int_0^t|Y^\varepsilon(s)|^{p-2}\langle Y^\varepsilon(s)-Y^\varepsilon(s_\delta),b^{(\delta)}(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta)\rangle\mathrm{d}s\\&\quad+\frac{p(p-1)}{2}\int_0^t|Y^\varepsilon(s)|^{p-2}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2\mathrm{d}s\\&\quad+p\int_0^t|Y^\varepsilon(s)|^{p-2}\langle Y^\varepsilon(s),\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\mathrm{d}B(s)\rangle\\&=:\sum_{i=1}^4\Pi_i(t),\qquad t\in[0,T].\end{aligned}$$

Whence, for any $t\ge 0$ one has

$$\Upsilon(t):=\mathbb{E}\Big(\sup_{-r_0\le s\le t}|Y^\varepsilon(s)|^p\Big)\le\|\xi\|_\infty^p+\sum_{i=1}^4\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_i(s)\Big). \tag{2.3}$$

In the sequel, we are going to claim that

$$\Upsilon(t)\le 2\|\xi\|_\infty^p+c\,t+c\int_0^t\Upsilon(s)\mathrm{d}s. \tag{2.4}$$

If (2.4) was true, thus (2.2) follows directly from Gronwall's inequality. So, it remains to verify that (2.4) holds true.

Let $\zeta_0(s)\equiv \mathbf{0}\in\mathbb{R}^d$ for any $s\in[-r_0,0]$. For $\zeta\in\mathscr{C}$ and $\mu\in\mathcal{P}_2(\mathscr{C})$, we deduce from (**A1**) that

$$\begin{aligned}\langle\zeta(0),b(\zeta,\mu,\theta)\rangle&=\langle\zeta(0)-\zeta_0,b(\zeta,\mu,\theta)-b(\zeta_0,\delta_{\zeta_0},\theta)\rangle+\langle\zeta(0),b(\zeta_0,\delta_{\zeta_0},\theta)\rangle\\&\le\alpha_1\|\zeta\|_\infty^2+\alpha_2\mathbb{W}_2(\mu,\delta_{\zeta_0})^2+|\zeta(0)|^2+|b(\zeta_0,\delta_{\zeta_0},\theta)|^2\\&\le c\,(1+\|\zeta\|_\infty^2+\mathbb{W}_2(\mu,\delta_{\zeta_0})^2),\end{aligned}\tag{2.5}$$

and from (**A2**) that

$$\begin{aligned}\|\sigma(\zeta,\mu)\|^2&\le 2\beta_1\|\zeta\|_\infty^2+2\beta_2\mathbb{W}_2(\mu,\delta_{\zeta_0})^2+2\|\sigma(\zeta_0,\delta_{\zeta_0})\|^2\\&\le c\,(1+\|\zeta\|_\infty^2+\mathbb{W}_2(\mu,\delta_{\zeta_0})^2).\end{aligned}\tag{2.6}$$

According to (1.4), we obtain that

$$\begin{aligned}&\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty\\&=\max_{k=0,\cdots,M-1}\sup_{-(k+1)\delta\le s\le -k\delta}|\overline{Y}^\varepsilon_{t_\delta}(s)|\\&=\max_{k=0,\cdots,M-1}\sup_{-(k+1)\delta\le s\le -k\delta}\Big|\frac{s+(k+1)\delta}{\delta}Y^\varepsilon(t_\delta-k\delta)-\frac{s+k\delta}{\delta}Y^\varepsilon(t_\delta-(k+1)\delta)\Big|\\&\le 2\sup_{-r_0\le s\le t}|Y^\varepsilon(s)|.\end{aligned}\tag{2.7}$$

Furthermore, recall the Young inequality:

$$a^\alpha b^{1-\alpha}\le\alpha a+(1-\alpha)b,\qquad a,b\ge0,\quad \alpha\in[0,1],\tag{2.8}$$

and the fundamental fact that: for any $q>0$,

$$\mathbb{E}|B(t)|^q\le c\,t^{q/2}.\tag{2.9}$$

By virtue of (1.4), we notice that

$$\overline{Y}^\varepsilon_{t_\delta}(0)=Y^\varepsilon(t_\delta).\tag{2.10}$$

Then, by exploiting (2.5), (2.7) as well as (2.10), it follows from (2.8) and Hölder's inequality that

$$\begin{aligned}&\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_1(s)\Big)\\&=p\,\mathbb{E}\Big(\sup_{0\le s\le t}\int_0^s\frac{|Y^\varepsilon(u)|^{p-2}}{1+\delta^\alpha|b(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}},\theta)|}\langle Y^\varepsilon(u_\delta),b(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}},\theta)\rangle\mathrm{d}u\Big)\\&=p\,\mathbb{E}\Big(\sup_{0\le s\le t}\int_0^s\frac{|Y^\varepsilon(u)|^{p-2}}{1+\delta^\alpha|b(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}},\theta)|}\langle \overline{Y}^\varepsilon_{u_\delta}(0),b(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}},\theta)\rangle\mathrm{d}u\Big)\\&\le c\,\mathbb{E}\Big(\sup_{0\le s\le t}\int_0^s\frac{|Y^\varepsilon(u)|^{p-2}}{1+\delta^\alpha|b(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}},\theta)|}\Big\{1+\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\delta_{\zeta_0})^2\Big\}\mathrm{d}s\\&\le c\int_0^t\Big\{1+\mathbb{E}|Y^\varepsilon(s)|^p+\mathbb{E}\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^p\Big\}\mathrm{d}s\\&\le c\int_0^t\{1+\Upsilon(s)\}\mathrm{d}s.\end{aligned}\tag{2.11}$$

It is straightforward to see that, for any $\zeta \in \mathscr{C}$, $\mu \in \mathcal{P}_2(\mathscr{C})$, and $\theta \in \Theta$,

$$|b^{(\delta)}(\zeta,\mu,\theta)| = \frac{|b(\zeta,\mu,\theta)|}{1+\delta^\alpha|b(\zeta,\mu,\theta)|} \le \delta^{-\alpha}. \tag{2.12}$$

Taking (2.6) and (2.12) into consideration and making use of (2.9) and $\alpha \in (0,1/2]$, for any $q \ge 2$, we derive that

$$\begin{aligned}
\mathbb{E}|Y^\varepsilon(t)-Y^\varepsilon(t_\delta)|^q &\le c\Big\{\delta^{q(1-\alpha)} + \mathbb{E}\|\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\|^q\mathbb{E}|B(t)-B(t_\delta)|^q\Big\}\\
&\le c\Big\{\delta^{q(1-\alpha)} + \delta^{q/2}\mathbb{E}\|\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\|^q\Big\}\\
&\le c\,\delta^{q/2}\Big\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|^q + \mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^q\Big\}\\
&\le c\,\delta^{q/2}\Big\{1+\mathbb{E}\Big(\sup_{-r_0\le s\le t}|Y^\varepsilon(s)|^q\Big)\Big\},
\end{aligned} \tag{2.13}$$

where in the last procedure we have used Hölder's inequality and (2.7). Thus, taking advantage of (2.12) and (2.13) and employing Hölder's inequality yields that

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}|\Pi_2(s)|\Big) &\le p\mathbb{E}\int_0^t |Y^\varepsilon(s)|^{p-2}|Y^\varepsilon(s)-Y^\varepsilon(s_\delta)|\cdot|b^{(\delta)}(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta)|\mathrm{d}s\\
&\le p\delta^{-\alpha}\int_0^t \mathbb{E}(|Y^\varepsilon(s)|^{p-2}|Y^\varepsilon(s)-Y^\varepsilon(s_\delta)|)\mathrm{d}s\\
&\le p\delta^{-\alpha}\int_0^t \Big(\mathbb{E}(|Y^\varepsilon(s)|^p)\Big)^{\frac{p-2}{p}}\Big(\mathbb{E}|Y^\varepsilon(s)-Y^\varepsilon(s_\delta)|^{\frac{p}{2}}\Big)^{\frac{2}{p}}\mathrm{d}s\\
&\le p\delta^{\frac12-\alpha}\int_0^t \Big(\mathbb{E}(|Y^\varepsilon(s)|^p)\Big)^{\frac{p-2}{p}}\Big\{1+\mathbb{E}\Big(\sup_{-r_0\le s\le t}|Y^\varepsilon(s)|^{\frac{p}{2}}\Big)\Big\}^{\frac{2}{p}}\mathrm{d}s\\
&\le c\int_0^t\{1+\Upsilon(s)\}\mathrm{d}s,
\end{aligned} \tag{2.14}$$

where in the last display we used $\alpha \in (0,1/2]$ and (2.8). Next, we observe that

$$\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_3(s)\Big) \le \frac{p(p-1)}{2}\int_0^t \mathbb{E}(|Y^\varepsilon(s)|^{p-2}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2)\mathrm{d}s. \tag{2.15}$$

Using Burkhold-Davis-Gundy's (BDG's for short) inequality and (2.8), we infer that

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_4(s)\Big) &\le p\,\mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^s |Y^\varepsilon(u)|^{p-2}\langle Y^\varepsilon(u),\sigma(\overline{Y}^\varepsilon_{u_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{u_\delta}})\mathrm{d}B(u)\rangle\Big|\Big)\\
&\le 4\sqrt{2}\,p\,\mathbb{E}\Big(\int_0^t |Y^\varepsilon(s)|^{2(p-2)}|\sigma^*(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})Y^\varepsilon(s)|^2\mathrm{d}s\Big)^{1/2}\\
&\le 4\sqrt{2}\,p\,\mathbb{E}\Big(\sup_{0\le s\le t}|Y^\varepsilon(s)|^p\int_0^t |Y^\varepsilon(s)|^{p-2}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2\mathrm{d}s\Big)^{1/2}\\
&\le \frac12\Upsilon(t) + 16p^2\int_0^t \mathbb{E}(|Y^\varepsilon(s)|^{p-2}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2)\mathrm{d}s.
\end{aligned} \tag{2.16}$$

Subsequently, one gets from (2.15) and (2.16) that

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_3(s)\Big)+\mathbb{E}\Big(\sup_{0\le s\le t}\Pi_4(s)\Big)\\
&\le \frac{1}{2}\Upsilon(t)+c\int_0^t \mathbb{E}(|Y^\varepsilon(s)|^{p-2}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2)\mathrm{d}s\\
&\le \frac{1}{2}\Upsilon(t)+c\int_0^t \{\mathbb{E}|Y^\varepsilon(s)|^{p}+\mathbb{E}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^p\}\mathrm{d}s\\
&\le \frac{1}{2}\Upsilon(t)+c\int_0^t \Big\{1+\mathbb{E}|Y^\varepsilon(s)|^{p}+\mathbb{E}\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^p+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\delta_{\zeta_0})^p\Big\}\mathrm{d}s\\
&\le \frac{1}{2}\Upsilon(t)+c\int_0^t \{1+\Upsilon(s)\}\mathrm{d}s,
\end{aligned}\tag{2.17}$$

where we have adopted (2.8) in the second inequality, used (2.6) in the last two step, and utilized Hölder's inequality, in addition to (2.7), in the last procedure. Substituting (2.11), (2.14), and (2.17) into (2.3) gives that

$$\Upsilon(t)\le\|\xi\|_\infty^p+\frac{1}{2}\Upsilon(t)+c\int_0^t\{1+\Upsilon(s)\}\mathrm{d}s.$$

As a consequence, (2.4) is now available. □

The following lemma shows that the linear interpolation $\overline{Y}^\varepsilon_{t_\delta}$ approaches X_t^0 in the mean square sense as ε and δ go to zero.

Lemma 2.2. *Assume* (**A1**), (**A2**), (**B1**) *and* (**B4**). *Then, for any $p>2$, there exists $c_p>0$*

$$\sup_{0\le t\le T}\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^p\le c_p(\delta^{p/2-1}+\varepsilon^2+\delta^{p\alpha}),\tag{2.18}$$

where $\alpha\in(0,1/2]$ is introduced in (1.3).

Proof. For any $p>2$ and $t\in[0,T]$, by Hölder's inequality and $Y_0^\varepsilon=X_0^0=\xi$, we find that

$$\begin{aligned}
&\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^p\\
&\le 3^{p-1}\,\mathbb{E}\|Y_t^\varepsilon-\overline{Y}^\varepsilon_{t_\delta}\|_\infty^p+3^{p-1}\,\mathbb{E}\|Y_t^\varepsilon-X_t^\varepsilon\|_\infty^p+3^{p-1}\,\mathbb{E}\|X_t^\varepsilon-X_t^0\|_\infty^p\\
&\le 3^{p-1}M\max_{k=0,\cdots,M-1}\mathbb{E}\Big(\sup_{-(k+1)\delta\le v\le -k\delta}|Y^\varepsilon(t+v)-\overline{Y}^\varepsilon_{t_\delta}(v)|^p\Big)\\
&\quad+3^{p-1}\,\mathbb{E}\Big(\sup_{0\le s\le t}|Y^\varepsilon(s)-X^\varepsilon(s)|^p\Big)+3^{p-1}\,\mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s)-X^0(s)|^p\Big)\\
&=:\Lambda_1(t,\varepsilon,\delta)+\Lambda_2(t,\varepsilon,\delta)+\Lambda_3(t,\varepsilon,\delta),
\end{aligned}\tag{2.19}$$

where $M>0$ such that $M\delta=r_0$. Hereinafter, we intend to estimate $\Lambda_i(t,\varepsilon,\delta)$, $i=1,2,3$, respectively. In the first place, we shall show that

$$\Lambda_1(t,\varepsilon,\delta)\le c\,\delta^{p/2-1},\qquad t\in[0,T].\tag{2.20}$$

For $t\in[0,T)$, there is an integer $k_0\ge 0$ such that $t\in[k_0\delta,(k_0+1)\delta)$. From (1.4), it follows that

$$\begin{aligned}
&\Lambda_1(t,\varepsilon,\delta)\\
&\le c\,M\max_{k=0,\cdots,M-1}\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0+1-k)\delta}|Y^\varepsilon(s)-Y^\varepsilon((k_0-k)\delta)|^p\Big)\\
&\quad+c\,M\max_{k=0,\cdots,M-1}\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0+1-k)\delta}|Y^\varepsilon(s)-Y^\varepsilon((k_0-k-1)\delta)|^p\Big)\\
&\le c\,M\max_{k=0,\cdots,M-1}\mathbb{E}\Big(\sup_{(k_0-k-1)\delta\le s\le(k_0+1-k)\delta}|Y^\varepsilon(s)-Y^\varepsilon((k_0-k-1)\delta)|^p\Big)\\
&\quad+c\,M\max_{k=0,\cdots,M-1}\mathbb{E}|Y^\varepsilon((k_0-k)\delta)-Y^\varepsilon((k_0-k-1)\delta)|^p.
\end{aligned}\tag{2.21}$$

In case of $k\ge k_0+1$, by virtue of (**B4**), one has

$$\Lambda_1(t,\varepsilon,\delta)\le c\,M\delta^p\le c\,r_0\delta^{p-1}.$$

In terms of (**B1**), for any $\zeta\in\mathscr{C}$ and $\mu\in\mathcal{P}_2(\mathscr{C})$,

$$|b(\zeta,\mu,\theta_0)|\le L_1\Big\{(1+\|\zeta\|_\infty^{q_1})\|\zeta\|_\infty+\mathbb{W}_2(\mu,\delta_{\zeta_0})\Big\}+|b(\zeta_0,\delta_{\zeta_0},\theta_0)|.\tag{2.22}$$

Let $k'\ge 0$ be an arbitrary integer. For any $t\in[k'\delta,(k'+2)\delta]$, note from BDG's inequality followed by Hölder's inequality that

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{k'\delta\le t\le(k'+2)\delta}|Y^\varepsilon(t)-Y^\varepsilon(k'\delta)|^p\Big)\\
&\le c\,\mathbb{E}\Big(\int_{k'\delta}^{(k'+2)\delta}|b^{(\delta)}(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta_0)|\mathrm{d}s\Big)^p+c\,\mathbb{E}\Big(\int_{k'\delta}^{(k'+2)\delta}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2\mathrm{d}s\Big)^{\frac{p}{2}}\\
&\le c\,\delta^{\frac{p}{2}-1}\int_{k'\delta}^{(k'+2)\delta}\Big\{\mathbb{E}|b(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta_0)|^p+\mathbb{E}\|\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^p\Big\}\mathrm{d}s,
\end{aligned}$$

where in the last display we have used the fact that

$$|b^{(\delta)}(\zeta,\mu,\theta_0)|\le|b(\zeta,\mu,\theta_0)|,\qquad\zeta\in\mathscr{C},\quad\mu\in\mathcal{P}_2(\mathscr{C}).\tag{2.23}$$

Subsequently, taking (2.2), (2.6) and (2.22) into account and making use of Hölder's inequality yields that

$$\mathbb{E}\Big(\sup_{k'\delta\le t\le(k'+2)\delta}|Y^\varepsilon(t)-Y^\varepsilon(k'\delta)|^p\Big)\le c\,\delta^{\frac{p}{2}-1}\int_{k'\delta}^{(k'+2)\delta}\Big\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{p(1+q_1)}\Big\}\mathrm{d}s\le c\,\delta^{\frac{p}{2}}.\tag{2.24}$$

Hence, it follows from (2.21) and (2.24) with $k'=k_0-k-1$ that

$$\Lambda_1(t,\varepsilon,\delta)\le c\,M\delta^{\frac{p}{2}}\le c\,\delta^{\frac{p}{2}-1}$$

provided that $k\le k_0-1$. Whenever $k=k_0$, we deduce from (2.21), (2.24) with $k'=0$ as well as (**B4**) that

$$\Lambda_1(t,\varepsilon,\delta)\le c\,M\mathbb{E}\Big(\sup_{0\le s\le\delta}|Y^\varepsilon(s)-Y^\varepsilon(0)|^p\Big)+c\,M\mathbb{E}\Big(\sup_{-\delta\le s\le 0}|Y^\varepsilon(s)-Y^\varepsilon(-\delta)|^p\Big)\le c\,\delta^{p/2-1}.$$

Next, we are going to claim that

$$\Lambda_3(t,\varepsilon,\delta) \le c\,\varepsilon^2, \qquad t\in[0,T]. \tag{2.25}$$

Following the argument to derive (2.2), we deduce that, for some constant $C_{p,T}>0$,

$$\sup_{0\le t\le T}\mathbb{E}\|X_t^\varepsilon\|_\infty^p \le C_{p,T}(1+\|\xi\|_\infty^p). \tag{2.26}$$

By the Itô formula and $X_0^\varepsilon = X_0^0=\xi$, we observe that

$$\begin{aligned}
&|X^\varepsilon(t)-X^0(t)|^p\\
&\le \int_0^t |X^\varepsilon(s)-X^0(s)|^{p-2}\Big\{p\langle X^\varepsilon(s)-X^0(s), b(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon},\theta_0)-b(X_s^0,\mathscr{L}_{X_s^0},\theta_0)\rangle\\
&\quad+\frac{\varepsilon^2p(p-1)}{2}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})\|^2\Big\}\mathrm{d}s + p\,\varepsilon\int_0^t |X^\varepsilon(s)-X^0(s)|^{p-2}\langle X^\varepsilon(s)-X^0(s),\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})\mathrm{d}B(s)\rangle.
\end{aligned}$$

Thus, by using BDG's inequality, Young's inequality and (2.8) and noting that $X_0^\varepsilon=X_0^0=\xi$, we infer from (**A1**) and (2.6) that

$$\begin{aligned}
\Lambda_3(t,\varepsilon,\delta) &\le c\int_0^t \Lambda_3(s,\varepsilon,\delta)\mathrm{d}s + c\,\varepsilon^2\int_0^t\{1+\mathbb{E}\|X_s^\varepsilon\|_\infty^p\}\mathrm{d}s\\
&\quad + c\,\varepsilon\mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s)-X^0(s)|^p\int_0^t |X^\varepsilon(s)-X^0(s)|^{p-2}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})\|^2\mathrm{d}s\Big)^{1/2}\\
&\le \frac{1}{2}\Lambda_3(t,\varepsilon,\delta) + c\int_0^t\Lambda_3(s,\varepsilon,\delta)\mathrm{d}s + c\,\varepsilon^2\int_0^t\{1+\mathbb{E}\|X_s^\varepsilon\|_\infty^p\}\mathrm{d}s.
\end{aligned}$$

So, one has

$$\Lambda_3(t,\varepsilon,\delta)\le c\int_0^t\Lambda_3(s,\varepsilon,\delta)\mathrm{d}s + c\,\varepsilon^2\int_0^t\{1+\mathbb{E}\|X_s^\varepsilon\|_\infty^p\}\mathrm{d}s.$$

Thus, (2.25) follows from (2.26) and Gronwall's inequality. Finally, we intend to verify that

$$\Lambda_2(t,\varepsilon,\delta)\le c\,(\delta^{p/2-1}+\delta^{2\alpha}), \qquad t\in[0,T]. \tag{2.27}$$

Also, by Itô's formula, we derive from $X_0^\varepsilon = Y_0^\varepsilon=\xi$ that

$$\begin{aligned}
&|X^\varepsilon(t)-Y^\varepsilon(t)|^p\\
&\le p\int_0^t |X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}\langle X^\varepsilon(s)-Y^\varepsilon(s), b(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon},\theta_0)-b(Y_s^\varepsilon,\mathscr{L}_{Y_s^\varepsilon},\theta_0)\rangle\mathrm{d}s\\
&\quad+p\int_0^t |X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}\langle X^\varepsilon(s)-Y^\varepsilon(s), b(Y_s^\varepsilon,\mathscr{L}_{Y_s^\varepsilon},\theta_0)-b(\overline{Y}_{s_\delta}^\varepsilon,\mathscr{L}_{\overline{Y}_{s_\delta}^\varepsilon},\theta_0)\rangle\mathrm{d}s\\
&\quad+p\int_0^t |X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}\langle X^\varepsilon(s)-Y^\varepsilon(s), b(\overline{Y}_{s_\delta}^\varepsilon,\mathscr{L}_{\overline{Y}_{s_\delta}^\varepsilon},\theta_0)-b^{(\delta)}(\overline{Y}_{s_\delta}^\varepsilon,\mathscr{L}_{\overline{Y}_{s_\delta}^\varepsilon},\theta_0)\rangle\mathrm{d}s\\
&\quad+\frac{\varepsilon^2p(p-1)}{2}\int_0^t |X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})-\sigma(\overline{Y}_{s_\delta}^\varepsilon,\mathscr{L}_{\overline{Y}_{s_\delta}^\varepsilon})\|^2\mathrm{d}s\\
&\quad+p\,\varepsilon\int_0^t |X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}\langle X^\varepsilon(s)-Y^\varepsilon(s),(\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})-\sigma(\overline{Y}_{s_\delta}^\varepsilon,\mathscr{L}_{\overline{Y}_{s_\delta}^\varepsilon}))\mathrm{d}B(s)\rangle\\
&=:\Xi_1(t)+\Xi_2(t)+\Xi_3(t)+\Xi_4(t)+\Xi_5(t).
\end{aligned}$$

In view of (**A1**), we deduce from Young's inequality that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}\Xi_1(s)\Big) &\le c\int_0^t \mathbb{E}\{|X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}(\|X_s^\varepsilon-Y_s^\varepsilon\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon},\mathscr{L}_{Y_s^\varepsilon})^2)\}\mathrm{d}s\\
&\le c\int_0^t \Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
\end{aligned}
\tag{2.28}
$$

Carrying out a similar argument to derive (2.20), for any $p>2$, we have

$$
\sup_{0\le t\le T}\mathbb{E}\|Y_t^\varepsilon-\overline{Y}^\varepsilon_{t_\delta}\|_\infty^p\le c\,\delta^{\frac{p}{2}-1}.
\tag{2.29}
$$

Taking (**A1**), (2.2) and (2.29) into consideration and applying Hölder's inequality that

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{0\le s\le t}|\Xi_2(s)|\Big)\\
&\le c\int_0^t\mathbb{E}|X^\varepsilon(s)-Y^\varepsilon(s)|^p\mathrm{d}s\\
&\quad + c\int_0^t\mathbb{E}\{(1+\|Y_s^\varepsilon\|_\infty^{pq_1}+\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{pq_1})\|Y_s^\varepsilon-\overline{Y}^\varepsilon_{s_\delta}\|_\infty^p+\mathbb{W}_2(\mathscr{L}_{Y_s^\varepsilon},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})^p\}\mathrm{d}s\\
&\le c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s+c\int_0^t\mathbb{E}\|Y_s^\varepsilon-\overline{Y}^\varepsilon_{s_\delta}\|_\infty^p\mathrm{d}s\\
&\quad + c\int_0^t\Big(1+\mathbb{E}\|Y_s^\varepsilon\|_\infty^{2pq_1}+\mathbb{E}\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{2pq_1}\Big)^{1/2}\Big(\mathbb{E}\|Y_s^\varepsilon-\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{2p}\Big)^{1/2}\mathrm{d}s\\
&\le c\,\delta^{\frac{p}{2}-1}+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
\end{aligned}
\tag{2.30}
$$

According to (1.3) and in view of (2.2) and (2.22), it follows from Hölder's inequality that

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{0\le s\le t}|\Xi_3(s)|\Big)\\
&\le c\int_0^t\mathbb{E}\Big\{|X^\varepsilon(s)-Y^\varepsilon(s)|^p+\frac{\delta^{p\alpha}|b(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta_0)|^{2p}}{(1+\delta^\alpha|b(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\theta_0)|)^p}\Big\}\mathrm{d}s\\
&\le c\int_0^t\Big\{\mathbb{E}|X^\varepsilon(s)-Y^\varepsilon(s)|^p+\delta^{2\alpha}\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{2p(1+q_1)}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}},\delta_{\zeta_0})^{2p}\}\Big\}\mathrm{d}s\\
&\le c\,\delta^{p\alpha}+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
\end{aligned}
\tag{2.31}
$$

Next, owing to $\varepsilon\in(0,1)$, (**A2**), and (2.20), one gets that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}\Xi_4(s)\Big) &\le c\int_0^t\mathbb{E}\{|X^\varepsilon(s)-Y^\varepsilon(s)|^{p-2}(\|X_s^\varepsilon-\overline{Y}^\varepsilon_{s_\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})^2)\}\mathrm{d}s\\
&\le c\int_0^t\{\mathbb{E}\|X_s^\varepsilon-Y_s^\varepsilon\|_\infty^p+\mathbb{E}\|Y_s^\varepsilon-\overline{Y}^\varepsilon_{s_\delta}\|_\infty^p\}\mathrm{d}s\\
&\le c\,\delta^{p/2-1}+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
\end{aligned}
\tag{2.32}
$$

Next, for $\varepsilon \in (0,1)$, BDG's inequality and Young's inequality (2.8), besides (2.32), give that

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{0\le s\le t}|\Xi_5(s)|\Big)\\
&\le c\,\mathbb{E}\Big(\sup_{0\le \le t}|X^\varepsilon(s)-Y^\varepsilon(s)|^p\int_0^t|X^\varepsilon(s)-Y^\varepsilon(s))|^{p-2}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})-\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^2\mathrm{d}s\Big)^{1/2}\\
&\le \frac{1}{2}\Lambda_2(t,\varepsilon,\delta)+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s+c\int_0^t\mathbb{E}\|\sigma(X_s^\varepsilon,\mathscr{L}_{X_s^\varepsilon})-\sigma(\overline{Y}^\varepsilon_{s_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{s_\delta}})\|^p\mathrm{d}s\\
&\le \frac{1}{2}\Lambda_2(t,\varepsilon,\delta)+c\,\delta^{p/2-1}+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
\end{aligned}\tag{2.33}
$$

Thus, (2.28), (2.30)-(2.33) yield that

$$
\Lambda_2(t,\varepsilon,\delta)\le \frac{1}{2}\Lambda_2(t,\varepsilon,\delta)+c\,(\delta^{p/2-1}+\delta^{p\alpha})+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
$$

Namely,

$$
\Lambda_2(t,\varepsilon,\delta)\le c\,(\delta^{p/2-1}+\delta^{p\alpha})+c\int_0^t\Lambda_2(s,\varepsilon,\delta)\mathrm{d}s.
$$

As a result, we obtain from Gronwall's inequality that

$$
\Lambda_2(t,\varepsilon,\delta)\le c\,(\delta^{p/2-1}+\delta^{p\alpha}).\tag{2.34}
$$

Inserting (2.20), (2.25), and (2.34) back into (2.19) leads to the desired assertion (2.18). □

The lemma below plays a crucial role in revealing the asymptotic behavior of the LSE of the unknown parameter $\theta\in\Theta$.

Lemma 2.3. *Let* (**A1**) – (**A2**) *and* (**B1**) – (**B4**) *hold. Then,*

$$
\begin{aligned}
&\delta\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\\
&\to \Xi(\theta):=\int_0^T\Gamma(X_t^0,\theta,\theta_0)^*\widehat{\sigma}(X_t^0)\Gamma(X_s^0,\theta,\theta_0)\mathrm{d}t
\end{aligned}\tag{2.35}
$$

in L^1 *uniformly w.r.t.* $\theta\in\overline{\Theta}$ *as* $\varepsilon\to0$ *and* $\delta\to0$ *(i.e.,* $n\to\infty$*). Moreover,*

$$
\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0)\to0\tag{2.36}
$$

in probability uniformly w.r.t. $\theta\in\overline{\Theta}$ *as* $\varepsilon\to0$.

Proof. Observe that

$$
\begin{aligned}
&\delta\sum_{k=1}^{n}(\Gamma^{(\delta)})^*(\overline{Y}^{\varepsilon}_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^{\varepsilon}_{(k-1)\delta})\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{(k-1)\delta},\theta,\theta_0)-\int_0^T\Gamma^*(X_t^0,\theta,\theta_0)\widehat{\sigma}(X_t^0)\Gamma(X_t^0,\theta,\theta_0)\mathrm{d}t\\
&=\int_0^T\Big\{(\Gamma^{(\delta)})^*(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)-\Gamma^*(X_t^0,\theta,\theta_0)\widehat{\sigma}(X_t^0)\Gamma(X_t^0,\theta,\theta_0)\Big\}\mathrm{d}t\\
&=\int_0^T\Big(\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)-\Gamma(X_t^0,\theta,\theta_0)\Big)^*\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)\mathrm{d}t\\
&\quad+\int_0^T\Gamma(X_t^0,\theta,\theta_0)^*\Big(\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})-\widehat{\sigma}(X_t^0)\Big)\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)\mathrm{d}t\\
&\quad+\int_0^T\Gamma(X_t^0,\theta,\theta_0)^*\widehat{\sigma}(X_t^0)\Big(\Gamma^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta,\theta_0)-\Gamma(X_t^0,\theta,\theta_0)\Big)\mathrm{d}t\\
&=:J_1(\varepsilon,\delta,\theta)+J_2(\varepsilon,\delta,\theta)+J_3(\varepsilon,\delta,\theta).
\end{aligned}
$$

From (**B1**) and (2.22), a direct calculation shows that, for any random variables $\zeta_1,\zeta\in\mathscr{C}$ with $\mathscr{L}_{\zeta_1},\mathscr{L}_{\zeta_2}\in\mathcal{P}_2(\mathscr{C})$,

$$
\begin{aligned}
&|\Gamma^{(\delta)}(\zeta_1,\theta,\theta_0)-\Gamma(\zeta_2,\theta,\theta_0)|\\
&=|b^{(\delta)}(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta_0)+b(\zeta_2,\mathscr{L}_{\zeta_2},\theta)-b^{(\delta)}(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\\
&\le|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta_0)|+|b(\zeta_2,\mathscr{L}_{\zeta_2},\theta)-b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\\
&\quad+|b^{(\delta)}(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)|+|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)-b^{(\delta)}(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\\
&=|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta_0)|+|b(\zeta_2,\mathscr{L}_{\zeta_2},\theta)-b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\\
&\quad+\delta^\alpha\Big|\frac{|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)|}{1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)|}b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)\Big|+\delta^\alpha\Big|\frac{|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}{1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\Big|\\
&\le|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)-b(\zeta_2,\mathscr{L}_{\zeta_2},\theta_0)|+|b(\zeta_2,\mathscr{L}_{\zeta_2},\theta)-b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\\
&\quad+\delta^\alpha\{|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta_0)|^2+|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|^2\}\\
&\le c\Big\{(1+\|\zeta_1\|_\infty^{q_1}+\|\zeta_2\|_\infty^{q_1})\|\zeta_1-\zeta_2\|_\infty+\mathbb{W}_2(\mathscr{L}_{\zeta_1},\mathscr{L}_{\zeta_2})\Big\}\\
&\quad+c\,\delta^\alpha\Big\{1+\|\zeta_1\|_\infty^{2(1+q_1)}+\mathbb{W}_2(\mathscr{L}_{\zeta_1},\delta_{\zeta_0})^2\Big\}.
\end{aligned}\tag{2.37}
$$

Next, for a random variable $\zeta\in\mathscr{C}$ with $\mathscr{L}_\zeta\in\mathcal{P}_2(\mathscr{C})$, by (2.22) and (2.23), it follows that

$$
|\Gamma^{(\delta)}(\zeta,\theta,\theta_0)|+|\Gamma(\zeta,\theta,\theta_0)|\le c\Big\{1+\|\zeta\|_\infty^{1+q_1}+\mathbb{W}_2(\mathscr{L}_\zeta,\delta_{\zeta_0})\Big\}\tag{2.38}
$$

and, due to (**B3**), that

$$
\|\widehat{\sigma}(\zeta)\|\le\|\widehat{\sigma}(\zeta)-\widehat{\sigma}(0)\|+\|\widehat{\sigma}(0)\|\le c\Big\{1+\|\zeta\|_\infty^{1+q_3}+\mathbb{W}_2(\mathscr{L}_\zeta,\delta_{\zeta_0})\Big\}.\tag{2.39}
$$

Consequently, combining (2.37) with (2.38) and(2.39), for $q := q_1 \vee q_3$, we deduce from (2.1) that

$$
\begin{aligned}
&|J_1(\varepsilon,\delta,\theta)| + |J_3(\varepsilon,\delta,\theta)| \\
&\le c\int_0^T \Big\{(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{q_1}+\|X^0_s\|_\infty^{q_1})\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty^p+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\mathscr{L}_{X^0_t}) \\
&\quad +\delta^{p\alpha}\Big(1+\|\overline{Y}^\varepsilon_{s_\delta}\|_\infty^{2(1+q_1)}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^2\Big)\Big\} \\
&\quad \times\Big\{1+\|X^0_t\|_\infty^{1+q_1}+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_1}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big\} \\
&\quad \times\Big\{1+\|X^0_t\|_\infty^{1+q_3}+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_3}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big\}\mathrm{d}t \\
&\le c\int_0^T \Big\{(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{q})\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty+\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty^2}\Big\}\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q)}\Big\}\mathrm{d}s \\
&\quad + c\,\delta^{p\alpha}\int_0^T\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)}\Big\}\mathrm{d}t.
\end{aligned}
$$

This, by exploiting (2.2) and (2.19) and using Hölder's inequality, gives that

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{\theta\in\overline{\Theta}}|J_1(\varepsilon,\delta,\theta)|\Big)+\mathbb{E}\Big(\sup_{\theta\in\overline{\Theta}}|J_3(\varepsilon,\delta,\theta)|\Big) \\
&\le c\int_0^T\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty^2}\Big\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{8(1+q)}\Big\}\mathrm{d}t \\
&\quad + c\,\delta^{\alpha}\int_0^T\Big\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)}\Big\}\mathrm{d}t \\
&\to 0
\end{aligned}
\tag{2.40}
$$

as $\varepsilon\to0$ and $\delta\to0$. Next, making use of (**B3**) and (2.38), we derive that

$$
\begin{aligned}
|J_2(\varepsilon,\delta,\theta)| \le c\int_0^T&(1+\|X^0_t\|_\infty^{1+q_1})\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_1}+\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2}\Big) \\
&\times\Big((1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{q_3}+\|X^0_t\|_\infty^{q_3})\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty+\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty^2}\Big)\mathrm{d}t.
\end{aligned}
$$

Again, using (2.1), (2.2) and (2.18) and utilizing Hölder's inequality gives that

$$
\begin{aligned}
\mathbb{E}|J_2(\varepsilon,\delta)| &\le c\int_0^T\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X^0_t\|_\infty^2}\Big\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)}\Big\}\mathrm{d}t \\
&\to 0
\end{aligned}
\tag{2.41}
$$

as $\varepsilon\to0$ and $\delta\to0$. Hence, (2.35) follows immediately from (2.40) and (2.41).

In the sequel, we are going to show that (2.36) holds. In terms of (1.2), we obtain that

$$
\begin{aligned}
&\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0) \\
&=\varepsilon\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\sigma(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}})(B(k\delta)-B((k-1)\delta)) \\
&=\varepsilon\int_0^T(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{t_\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{t_\delta})\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\mathrm{d}B(t) \\
&=:\Pi(T,\varepsilon,\theta).
\end{aligned}
\tag{2.42}
$$

By the BDG inequality and the Hölder inequality, we derive from (2.6), (2.38), and (2.39) that for $p > 2$

$$\begin{aligned}\mathbb{E}|\Pi(T,\varepsilon,\theta)|^p &\le c\,\varepsilon^p\,\mathbb{E}\Big(\int_0^T |(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{t_\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{t_\delta})\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})|^2\mathrm{d}t\Big)^{p/2}\\
&\le c\,\varepsilon^p\int_0^T \mathbb{E}\{|\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\theta,\theta_0)|^2\cdot\|\widehat{\sigma}(\overline{Y}^\varepsilon_{t_\delta})\|^2\cdot\|\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\|^2\}^{p/2}\mathrm{d}t\\
&\le c\,\varepsilon^p\int_0^T \mathbb{E}\Big\{\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^2\Big)^{p/2}\\
&\quad\times\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q_3)}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^2\Big)^{p/2}\\
&\quad\times\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q_1)}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^2\Big)^{p/2}\Big\}\mathrm{d}t\\
&\le c\,\varepsilon^p\int_0^T\{1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4p(1+q)}\}\mathrm{d}t\\
&\le c\,\varepsilon^p.\end{aligned}$$

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 (**B5**), (2.6), and (2.39) that for $p > 2$

$$\begin{aligned}&\mathbb{E}|\Pi(T,\varepsilon,\theta_1)-\Pi(T,\varepsilon,\theta_2)|^p\\
&\le c\,\varepsilon^p\int_0^T\mathbb{E}\{|\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\theta_1,\theta_0)-\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\theta_2,\theta_0)|^2\cdot\|\widehat{\sigma}(\overline{Y}^\varepsilon_{t_\delta})\|^2\cdot\|\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\|^2\}^{p/2}\mathrm{d}t\\
&\le c\,\varepsilon^p\int_0^T\mathbb{E}\{|b(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\theta_1)-b(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\theta_2)|^2\cdot\|\widehat{\sigma}(\overline{Y}^\varepsilon_{t_\delta})\|^2\cdot\|\sigma(\overline{Y}^\varepsilon_{t_\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}})\|^2\}^{p/2}\mathrm{d}t\\
&\le c\,\varepsilon^p|\theta_1-\theta_2|^p.\end{aligned}$$

As a consequence, we obtain (2.36) from [22, Theorem 20, p378]. □

So far, with Lemma 2.3 in hand, we are in the position to complete the

Proof of Theorem 1.1. A direction calculation shows that

$$\begin{aligned}&\Phi_{n,\varepsilon}(\theta)\\
&=\delta^{-1}\sum_{k=1}^n\Big\{P_k^*(\theta)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta)-P_k^*(\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0)\Big\}\\
&=\delta^{-1}\sum_{k=1}^n\Big\{\Big(P_k(\theta_0)+(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\delta\Big)^*\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\Big(P_k(\theta_0)+\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{t_{k-1}},\theta,\theta_0)\delta\Big)\\
&\quad-P_k^*(\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0)\Big\}\\
&=2\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0)\\
&\quad+\delta\sum_{k=1}^n(\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0).\end{aligned}\tag{2.43}$$

By virtue of Lemma 2.3, we therefore infer that

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

Next, for any $\kappa>0$, due to $\Xi(\cdot)>0$,

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

Furthermore, one has $-\Phi_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon}) \geq -\Phi_{n,\varepsilon}(\theta_0)=0$. Consequently, we deduce from [35, Theorem 5.9] with $M_n(\cdot)=-\Phi_{n,\varepsilon}(\cdot)$ and $M(\cdot)=-\Xi(\cdot)$ therein that $\widehat{\theta}_{n,\varepsilon}\to\theta_0$ in probability as $\varepsilon\to0$ and $n\to\infty$. We henceforth complete the proof. □

3 Proof of Theorem 1.2

Before we start to finish the argument of Theorem 1.2, we also need to prepare some auxiliary lemmas below. For any random variable $\zeta\in\mathscr{C}$ with $\mathscr{L}_\zeta\in\mathcal{P}_2(\mathscr{C})$, set

$$\Upsilon^{(\delta)}(\zeta,\theta) := (\nabla_\theta b^{(\delta)})^*(\zeta,\mathscr{L}_\zeta,\theta)\widehat{\sigma}(\zeta)\sigma(\zeta,\mathscr{L}_\zeta).$$

Lemma 3.1. *Let* (**A1**) – (**A2**) *and* (**B1**) – (**B4**) *hold. Then,*

$$\varepsilon^{-1}(\nabla_\theta\Phi_{n,\varepsilon})(\theta) \to -2\int_0^T \Upsilon(X_t^0,\theta)\mathrm{d}B(t) \qquad \textit{in probability} \tag{3.1}$$

whenever $\varepsilon\to0$ *and* $n\to\infty$, *where* $\Upsilon(\cdot,\cdot)$ *is introduced in* (1.12).

Proof. By the chain rule, one infers from (1.2) and (2.43) that

$$\begin{aligned}
&\varepsilon^{-1}(\nabla_\theta\Phi_{n,\varepsilon})(\theta)\\
&= 2\,\varepsilon^{-1}\sum_{k=1}^n(\nabla_\theta\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\Big\{P_k(\theta_0)+\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\delta\Big\}\\
&= 2\,\varepsilon^{-1}\sum_{k=1}^n(\nabla_\theta\Gamma^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta)\\
&= -2\sum_{k=1}^n(\nabla_\theta b^{(\delta)})^*(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\sigma(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}})\\
&\quad\times(B(k\delta)-B((k-1)\delta))\\
&= -2\int_0^T\Upsilon^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\theta)\mathrm{d}B(t),
\end{aligned} \tag{3.2}$$

where in the last two display we used the fact that

$$(\nabla_\theta\Gamma^{(\delta)})(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0) = -(\nabla_\theta b^{(\delta)})(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta). \tag{3.3}$$

To achieve (3.1), in terms of [6, Theorem 2.6, P.63], it is sufficient to claim that

$$\int_0^T \|\Upsilon^{(\delta)}(\overline{Y}^\varepsilon_{t_\delta},\theta)-\Upsilon(X_t^0,\theta)\|^2\mathrm{d}t\to0 \qquad \text{in probability} \tag{3.4}$$

as $\varepsilon\to0$ and $\delta\to0$. Observe that

$$
\begin{aligned}
&\Upsilon^{(\delta)}(\overline{Y}^{\varepsilon}_{t_\delta},\theta)-\Upsilon(X^0_t,\theta)\\
&=(\nabla_\theta b^{(\delta)})^*(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}},\theta)\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})\sigma(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}})-(\nabla_\theta b)^*(X^0_t,\mathscr{L}_{X^0_t},\theta)\widehat{\sigma}(X^0_t)\sigma(X^0_t,\mathscr{L}_{X^0_t})\\
&=\{(\nabla_\theta b^{(\delta)})^*(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}},\theta)-(\nabla_\theta b)^*(X^0_t,\mathscr{L}_{X^0_t},\theta)\}\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})\sigma(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}})\\
&\quad+(\nabla_\theta b)^*(X^0_t,\mathscr{L}_{X^0_t},\theta)\{\widehat{\sigma}(\overline{Y}^{\varepsilon}_{t_\delta})-\widehat{\sigma}(X^0_t)\}\sigma(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}})\\
&\quad+(\nabla_\theta b)^*(X^0_t,\mathscr{L}_{X^0_t},\theta)\widehat{\sigma}(X^0_t)\{\sigma(\overline{Y}^{\varepsilon}_{t_\delta},\mathscr{L}_{\overline{Y}^{\varepsilon}_{t_\delta}})-\sigma(X^0_t,\mathscr{L}_{X^0_t})\}\\
&=:\Sigma_1(t,\varepsilon,\delta)+\Sigma_2(t,\varepsilon,\delta)+\Sigma_3(t,\varepsilon,\delta).
\end{aligned}
$$

By a straightforward calculation, for any random variable $\zeta\in\mathscr{C}$ with $\mathscr{L}_\zeta\in\mathcal{P}_2(\mathscr{C})$, one has

$$
\begin{aligned}
(\nabla_\theta b^{(\delta)})(\zeta,\mathscr{L}_\zeta,\theta)&=\nabla_\theta\Big(\frac{b(\zeta,\mu,\theta)}{1+\delta^\alpha|b(\zeta,\mu,\theta)|}\Big)\\
&=\frac{(\nabla_\theta b)(\zeta,\mu,\theta)}{1+\delta^\alpha|b(\zeta,\mu,\theta)|}-\frac{\delta^\alpha(bb^*)(\zeta,\mu,\theta)(\nabla_\theta b)(\zeta,\mu,\theta)}{|b(\zeta,\mu,\theta)|(1+\delta^\alpha|b(\zeta,\mu,\theta)|)^2}.
\end{aligned}\tag{3.5}
$$

Next, for any random variables $\zeta_1,\zeta_2\in\mathscr{C}$ with $\mathscr{L}_{\zeta_1},\mathscr{L}_{\zeta_2}\in\mathcal{P}_2(\mathscr{C})$, it follows from (3.5) that

$$
\begin{aligned}
&\|(\nabla_\theta b^{(\delta)})^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)-(\nabla_\theta b)^*(\zeta_2,\mathscr{L}_{\zeta_2},\theta)\|\\
&=\Big\|\frac{(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)}{1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}-(\nabla_\theta b)^*(\zeta_2,\mathscr{L}_{\zeta_2},\theta)-\frac{\delta^\alpha(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)}{(1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|)^2|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}\Big\|\\
&=\Big\|\frac{(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)-(\nabla_\theta b)^*(\zeta_2,\mathscr{L}_{\zeta_2},\theta)}{1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}-\frac{\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|(\nabla_\theta b)^*(\zeta_2,\mathscr{L}_{\zeta_2},\theta)}{1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}\\
&\quad-\frac{\delta^\alpha(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)}{(1+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|)^2|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|}\Big\|\\
&\le\|(\nabla_\theta b)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)-(\nabla_\theta b)(\zeta_2,\mathscr{L}_{\zeta_2},\theta)\|\\
&\quad+\delta^\alpha|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|\cdot\{\|(\nabla_\theta b)(\zeta_2,\mathscr{L}_{\zeta_2},\theta)\|+\|(\nabla_\theta b)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\|\},
\end{aligned}\tag{3.6}
$$

where in the last step we utilized the facts that $\|A\|=\|A^*\|$ for a matrix A and that

$$
\begin{aligned}
&\|(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\|^2\\
&=\mathrm{trace}\Big(((\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta))^*(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\Big)\\
&=\mathrm{trace}\Big((bb^*)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta))((\nabla_\theta b)(\nabla_\theta b)^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\Big)\\
&=\mathrm{trace}\Big(((\nabla_\theta b)(\nabla_\theta b)^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)(\zeta_1,\mathscr{L}_{\zeta_1},\theta)(bb^*)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta))\Big)\\
&=|b(\zeta_1,\mathscr{L}_{\zeta_1},\theta)|^4\|(\nabla_\theta b)^*(\zeta_1,\mathscr{L}_{\zeta_1},\theta)\|^2.
\end{aligned}
$$

Moreover, from (**B2**), one has

$$
\|(\nabla_\theta b)(\zeta_2,\mathscr{L}_{\zeta_2},\theta)\|\le c\Big\{1+\|\zeta_2\|_\infty^{1+q_2}+\mathbb{W}_2(\mathscr{L}_{\zeta_2},\delta_{\zeta_0})\Big\}.\tag{3.7}
$$

Now, taking (**B2**), (3.6), and (3.7), in addition to (2.6) and (2.39), into account yields that

$$\begin{aligned}\|\Sigma_1(t,\varepsilon,\delta)\| \le c\Big\{&(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{q_2}+\|X_t^0\|_\infty^{q_2})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\mathscr{L}_{X_t^0})\\&+\delta^\alpha\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_1}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big)\Big(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_2}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big)\Big\}\\&\times\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{1+q_3}+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big\}\times\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})\Big\}.\end{aligned}$$

For $q := q_1 \vee q_2 \vee q_3$, simple calculations and (2.2) give that

$$\begin{aligned}\|\Sigma_1(t,\varepsilon,\delta)\| &\le c\Big\{(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{q})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty+\sqrt{\mathbb{E}\|\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}}-X_t^0\|_\infty^2}\Big\}\\&\quad\times\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q)}+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2\Big\}\\&\quad+c\delta^\alpha\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q)}+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2\Big\}^2\\&\le c(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty\\&\quad+c(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q)})\sqrt{\mathbb{E}\|\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}}-X_t^0\|_\infty^2}+c\delta^\alpha(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)})\\&=:\widetilde{\Lambda}_1(t,\varepsilon,\delta)+\widetilde{\Lambda}_2(t,\varepsilon,\delta)+\widetilde{\Lambda}_3(t,\varepsilon,\delta).\end{aligned}$$

For any $\rho > 0$, by virtue of Hölder's inequality, together with (2.2) and (2.18), it follows that

$$\begin{aligned}&\mathbb{P}\Big(\int_0^T\|\widetilde{\Lambda}_1(t,\varepsilon,\delta)\|^2\mathrm{d}t\ge\rho\Big)\\&\le\mathbb{P}\Big(c\int_0^T(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{8(1+q)})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2\mathrm{d}t\ge\rho\Big)\\&\le\mathbb{P}\Big(c\int_0^T(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{9(1+q)})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty\mathrm{d}t\ge\rho\Big)\\&\le\frac{c}{\rho}\int_0^T(1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{18(1+q)})\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2}\mathrm{d}t\\&\to0\end{aligned}\tag{3.8}$$

whenever $\varepsilon \to 0$ and $\delta \to 0$. On the other hand, by means of (2.2), and (2.18), it follows that

$$\begin{aligned}\mathbb{E}\widetilde{\Lambda}_2^2(t,\varepsilon,\delta)+\mathbb{E}\widetilde{\Lambda}_3^2(t,\varepsilon,\delta)&\le c(1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(1+q)})\mathbb{E}\|\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}}-X_t^0\|_\infty^2+c\delta^\alpha(1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{8(1+q)})\\&\le c(\delta^\beta+\varepsilon^2+\delta^\alpha)\\&\to0\end{aligned}\tag{3.9}$$

as $\varepsilon \to 0$ and $\delta \to 0$. As a consequence, we infer from (3.8) and (3.9) that

$$\int_0^T\|\Sigma_1(t,\varepsilon,\delta)\|^2\mathrm{d}t\to0\qquad\text{in probability}\tag{3.10}$$

when $\varepsilon\to0$ and $\delta\to0$. Next, taking advantage of (**A2**), (**B3**), (2.6), and (3.7) leads to

$$
\begin{aligned}
&\|\Sigma_2(t,\varepsilon,\delta)\|^2+\|\Sigma_3(t,\varepsilon,\delta)\|^2\\
&\le c\Big\{1+\|X_t^0\|_\infty^{2(1+q)}\Big\}\\
&\quad\times\Big\{(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2q}+\|X_t^0\|_\infty^{2q})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\mathscr{L}_{X_t^0})^2\Big\}\\
&\quad\times(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2+\mathbb{W}_2(\mathscr{L}_{\overline{Y}^\varepsilon_{t_\delta}},\delta_{\zeta_0})^2)\\
&\le c\Big\{(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(1+q)})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2\Big\}\\
&\quad\times\Big\{1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2\Big\}\\
&\le c(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{2(2+q)})\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty+c(1+\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2)\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2\\
&=:\Xi_1(t,\varepsilon,\delta)+\Xi_2(t,\varepsilon,\delta),
\end{aligned}
$$

in which we adopted (2.2) in the last procedure. Via Hölder's inequality, we obtain from (2.2) and (2.18) that

$$
\mathbb{E}\Xi_1(t,\varepsilon,\delta)\le c(1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^{4(2+q)})\sqrt{\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2}\to0 \tag{3.11}
$$

as $\varepsilon\to0$ and $\delta\to0$. Also, by (2.2) and (2.18), one has

$$
\mathbb{E}\Xi_2(t,\varepsilon,\delta)\le c(1+\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}\|_\infty^2)\mathbb{E}\|\overline{Y}^\varepsilon_{t_\delta}-X_t^0\|_\infty^2\to0 \tag{3.12}
$$

provided that $\varepsilon\to0$ and $\delta\to0$. Therefore, (3.11) and (3.12) lead to

$$
\mathbb{E}\|\Sigma_2(t,\varepsilon,\delta)\|^2+\mathbb{E}\|\Sigma_3(t,\varepsilon,\delta)\|^2\to0 \tag{3.13}
$$

if $\varepsilon\to0$ and $\delta\to0$. At last, the desired assertion (3.1) holds from (3.10) and (3.13). □

Lemma 3.2. *Let* (**A1**) – (**A3**), (**B1**) – (**B4**), *and* (**C**) *hold. Then*

$$
(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta)\to K_0(\theta):=K(\theta)+I(\theta)\qquad\text{in probability} \tag{3.14}
$$

as $n\to\infty$ and $\varepsilon\to0$, where $I(\cdot)$ and $K(\cdot)$ are introduced in (1.10) *and* (1.11), *respectively.*

Proof. From (3.2) and (3.3), we deduce that

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

For any random variable $\zeta \in \mathscr{C}$ with $\mathscr{L}_\zeta \in \mathcal{P}_2(\mathscr{C})$, by the chain rule, we infer from (3.5) that

$$
\begin{aligned}
\Big(\nabla_\theta^{(2)}(b^{(\delta)})^*\Big)(\zeta,\mathscr{L}_\zeta,\theta) &= \bigg(\nabla_\theta\bigg(\frac{(\nabla_\theta b^*)}{1+\delta^\alpha|b|}\bigg)\bigg)(\zeta,\mathscr{L}_\zeta,\theta)\\
&\quad - \delta^\alpha\bigg(\nabla_\theta\bigg(\frac{(\nabla_\theta b^*)(bb^*)}{|b|(1+\delta^\alpha|b|)^2}\bigg)\bigg)(\zeta,\mathscr{L}_\zeta,\theta)\\
&= (\nabla_\theta^{(2)}b^*)(\zeta,\mathscr{L}_\zeta,\theta) - \delta^\alpha\Theta_1(\zeta,\mathscr{L}_\zeta,\theta).
\end{aligned}
\tag{3.15}
$$

Next, the chain rule shows that

$$
\begin{aligned}
\Theta_1(\zeta,\mathscr{L}_\zeta,\theta) := \Bigg(&\frac{|b|(\nabla_\theta^{(2)}b^*)}{1+\delta^\alpha|b|} + \frac{\Big(b^*(\frac{\partial}{\partial\theta_1}b)(\nabla_\theta b)^*,\cdots,b^*(\frac{\partial}{\partial\theta_p}b)(\nabla_\theta b)^*\Big)_{p\times pd}}{|b|(1+\delta^\alpha|b|)^2}\\
&+ \frac{\Big((\frac{\partial}{\partial\theta_1}(\nabla_\theta b^*))(bb^*),\cdots,(\frac{\partial}{\partial\theta_p}(\nabla_\theta b)^*)(bb^*)\Big)_{p\times pd}}{|b|(1+\delta^\alpha|b|)^2}\\
&+ \frac{\Big((\nabla_\theta b)^*((\frac{\partial}{\partial\theta_1}b)b^*+b\frac{\partial}{\partial\theta_1}b^*),\cdots,(\nabla_\theta b)^*((\frac{\partial}{\partial\theta_p}b)b^*+b\frac{\partial}{\partial\theta_p}b^*)\Big)_{p\times pd}}{|b|(1+\delta^\alpha|b|)^2}\\
-\frac{1+3\delta^\alpha|b|}{|b|^3(1+\delta^\alpha|b|)^3}&\bigg(\bigg(b^*\Big(\frac{\partial}{\partial\theta_1}b\Big)\bigg)(\nabla_\theta b)^*(bb^*),\cdots,\bigg(b^*\Big(\frac{\partial}{\partial\theta_p}b\Big)\bigg)(\nabla_\theta b)^*(bb^*)\bigg)\Bigg)_{p\times pd}(\zeta,\mathscr{L}_\zeta,\theta).
\end{aligned}
$$

Thanks to (3.5), it follows that

$$
\Big((\nabla_\theta b^{(\delta)})^*\widehat{\sigma}(\zeta)(\nabla_\theta b^{(\delta)})\Big)(\zeta,\mathscr{L}_\zeta,\theta) = \Big((\nabla_\theta b^*)\widehat{\sigma}(\zeta)(\nabla_\theta b)\Big)(\zeta,\mathscr{L}_\zeta,\theta) - \delta^\alpha\Theta_2(\zeta,\mathscr{L}_\zeta,\theta),
\tag{3.16}
$$

where

$$
\begin{aligned}
\Theta_2(\zeta,\mathscr{L}_\zeta,\theta) : = \bigg(&\frac{(2|b|+\delta^\alpha|b|^2)(\nabla_\theta b^*)\widehat{\sigma}(\zeta)(\nabla_\theta b)}{(1+\delta^\alpha|b|)^2} + \frac{(\nabla_\theta b^*)\widehat{\sigma}(\zeta)(b\,b^*)(\nabla_\theta b)}{|b|(1+\delta^\alpha|b|)^3}\\
&+ \frac{(\nabla_\theta b^*)(b\,b^*)\widehat{\sigma}(\zeta)(\nabla_\theta b)}{|b|(1+\delta^\alpha|b|)^3} - \delta^\alpha\frac{(\nabla_\theta b^*)(b\,b^*)\widehat{\sigma}(\zeta)(b\,b^*)(\nabla_\theta b)}{|b|^2(1+\delta^\alpha|b|)^4}\bigg)(\zeta,\mathscr{L}_\zeta,\theta).
\end{aligned}
$$

Thus, taking (3.16) and (3.15) into consideration yields that

$$
\begin{aligned}
(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta) &= -2\delta\sum_{k=1}^n(\nabla_\theta^{(2)}b^*)(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\circ\Big(\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})\Gamma^{(\delta)}(\overline{Y}^\varepsilon_{(k-1)\delta},\theta,\theta_0)\Big)\\
&\quad + 2\delta\sum_{k=1}^n\Big((\nabla_\theta b^*)\widehat{\sigma}(\zeta)(\nabla_\theta b)\Big)(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\\
&\quad - 2\sum_{k=1}^n(\nabla_\theta^{(2)}b^*)(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\circ\Big(\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta_0)\Big)\\
&\quad - 2\delta^\alpha\sum_{k=1}^n\Theta_1(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\circ\Big(\widehat{\sigma}(\overline{Y}^\varepsilon_{(k-1)\delta})P_k(\theta)\Big)\\
&\quad - \delta^{1+\alpha}\sum_{k=1}^n\Theta_2(\overline{Y}^\varepsilon_{(k-1)\delta},\mathscr{L}_{\overline{Y}^\varepsilon_{(k-1)\delta}},\theta)\\
&=: \sum_{i=1}^5 I_i(n,\varepsilon).
\end{aligned}
$$

By following the argument to derive (2.35), we deduce from (**A3**) that

$$I_1(n,\varepsilon)\to K(\theta)\quad\text{and}\quad I_2(n,\varepsilon)\to I(\theta),\qquad\text{in probability}\tag{3.17}$$

as $\varepsilon\to0$ and $\delta\to0$. Notice from (**A3**) and (2.22) that

$$\begin{aligned}\|\Theta_1\|(\zeta,\mathscr{L}_\zeta,\theta)&\le c\Big((|b|+\|\nabla_\theta^{(2)}b\|+(1+3|b|)\|\nabla_\theta^{(2)}b\|)\|\nabla_\theta^{(2)}b\|\Big)(\zeta,\mathscr{L}_\zeta,\theta)\\&\le c(1+\|\zeta\|_\infty^{4(1+q)}+\mathcal{W}_2(\mathscr{L}_\zeta,\delta_{\zeta_0})^4).\end{aligned}\tag{3.18}$$

On the other hand, owing to (3.7), (2.39), and (2.22), one has

$$\begin{aligned}\|\Theta_2\|(\zeta,\mathscr{L}_\zeta,\theta)&\le 2\Big(|b|\;\|\nabla_\theta b\|^2\|\widehat{\sigma}(\zeta)\|(1+2|b|)\Big)(\zeta,\mathscr{L}_\zeta,\theta)\\&\le c(1+\|\zeta\|_\infty^{4(1+q)}+\mathcal{W}_2(\mathscr{L}_\zeta,\delta_{\zeta_0})^4).\end{aligned}\tag{3.19}$$

Thus, by mimicking the argument of (2.36), we obtain from (3.18) that

$$I_3(n,\varepsilon)\to0\quad\text{and}\quad I_4(n,\varepsilon)\to0\quad\text{in probability}\tag{3.20}$$

as $\varepsilon\to0$ and $\delta\to0$. Furthermore, (2.2) and (3.19) enable us to get that

$$I_5(n,\varepsilon)\to0\quad\text{in probability}\tag{3.21}$$

whenever $\varepsilon\to0$ and $\delta\to0$. Thus, the desired assertion (3.14) follows from (3.17), (3.20), as well as (3.21). □

Now, we move forward to complete the

Proof of Theorem 1.2. With Lemmas 3.1 and 3.2 at hand, the proof of Theorem 1.2 is parallel to that of [29, Theorem 4.1]. Whereas, to make the content self-contained, we give an outline of the proof. In terms of Theorem 1.1, there exists a sequence $\eta_{n,\varepsilon}\to0$ as $\varepsilon\to0$ 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

$$(\nabla_\theta\Phi_{n,\varepsilon})(\widehat{\theta}_{n,\varepsilon})=(\nabla_\theta\Phi_{n,\varepsilon})(\theta_0)+D_{n,\varepsilon}(\widehat{\theta}_{n,\varepsilon}-\theta_0),\quad \widehat{\theta}_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0)\tag{3.22}$$

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 $\widehat{\theta}_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0)$,

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

This, together with Lemma 3.2 and continuity of $K_0(\cdot)$, gives that

$$D_{n,\varepsilon} \to K_0(\theta_0) \quad \text{in probability} \tag{3.23}$$

as $\varepsilon \to 0$ and $n \to \infty$. By following the exact line of [21, 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 3.2, one has

$$\lim_{\varepsilon \to 0, n \to \infty} \mathbb{P}\Big(\sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta) - K_0(\theta_0)\| \le \frac{\alpha}{2} \Big) = 1. \tag{3.24}$$

On the other hand, recall that

$$\lim_{\varepsilon \to 0, n \to \infty} \mathbb{P}\Big(\widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0) \Big) = 1. \tag{3.25}$$

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

$$\begin{aligned} 1 \ge \mathbb{P}(\Gamma_{n,\varepsilon}) \ge \mathbb{P}\Big(&\sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)}\Phi_{n,\varepsilon})(\theta) - K_0(\theta_0)\| \le \frac{\alpha}{2} \Big) \\ &+ \mathbb{P}\Big(\widehat{\theta}_{n,\varepsilon} \in B_{\eta_{n,\varepsilon}}(\theta_0) \Big) - 1. \end{aligned} \tag{3.26}$$

Thus, taking advantage of (3.24), (3.25) as well as (3.26), we deduce from Sandwich theorem that

$$\mathbb{P}(\mathscr{D}_{n,\varepsilon}) \ge \mathbb{P}(\Gamma_{n,\varepsilon}) \to 1 \tag{3.27}$$

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 (3.22) 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 3.1, (3.23) as well as (3.27) and by noting $K_0(\theta_0) = I(\theta_0)$. We therefore complete the proof. □

4 Proof of Example 1.3

Proof of Example 1.3. It is sufficient to check all of assumptions in Theorems 1.1 and Theorem 1.2 are fulfilled.

For any $\zeta, \zeta' \in \mathscr{C}$, $\mu \in \mathcal{P}_2(\mathscr{C})$ and $\theta = (\theta^{(1)}, \theta^{(2)})^* \in \Theta_0$, set

$$b_0(\zeta, \zeta') := -\zeta^3(0) + \zeta(0) + \int_{-r_0}^0 \zeta(v)\mathrm{d}v + \int_{-r_0}^0 \zeta'(v)\mathrm{d}v, \tag{4.1}$$

$$b(\zeta, \mu, \theta) := \theta^{(1)} + \theta^{(2)} \int_{\mathscr{C}} b_0(\zeta, \zeta')\mu(\mathrm{d}\zeta') \quad \text{and} \quad \sigma(\zeta) := \sigma(\zeta, \mu) := 1 + \int_{-r_0}^0 |\zeta(v)|\mathrm{d}v.$$

Then, (2.24) can be reformulated as (1.1). By (4.1) and Hölder's inequality, we find out some constants $c_1, c_2 > 0$ such that

$$\begin{aligned}
&\langle \zeta_1(0) - \zeta_2(0), b(\zeta_1, \mu, \theta) - b(\zeta_2, \mu, \theta)\rangle \\
&= \theta^{(2)} \int_{\mathscr{C}} \langle \zeta_1(0) - \zeta_2(0), b_0(\zeta_1, \zeta) - b_0(\zeta_2, \zeta)\rangle \mu_t(\mathrm{d}\zeta) \\
&\le c_1 \Big\{ |\zeta_1(0) - \zeta_2(0)|^2 + \int_{-r_0}^0 |\zeta_1(v) - \zeta_2(v)|^2 \mathrm{d}v \Big\} \\
&\le c_2 \|\zeta_1 - \zeta_2\|_\infty^2, \qquad \mu \in \mathcal{P}_2(\mathscr{C}), \quad \zeta_1, \zeta_2 \in \mathscr{C}
\end{aligned} \tag{4.2}$$

Next, we deduce from (4.1) that for some constant $c_3 > 0$,

$$\begin{aligned}
|b(\zeta, \mu, \theta) - b(\zeta, \nu, \theta)| &\le \theta^{(2)} \Big| \int_{\mathscr{C}} b_0(\zeta, \zeta_1)\mu(\mathrm{d}\zeta_1) - \int_{\mathscr{C}} b_0(\zeta, \zeta_2)\nu(\mathrm{d}\zeta_2) \Big| \\
&\le \theta^{(2)} \int_{\mathscr{C}} \int_{\mathscr{C}} |b_0(\zeta, \zeta_1) - b_0(\zeta, \zeta_2)| \pi(\mathrm{d}\zeta_1, \mathrm{d}\zeta_2) \\
&\le c_3 \, \mathbb{W}_2(\mu, \nu), \qquad \zeta \in \mathscr{C}, \quad \mu, \nu \in \mathcal{P}_2(\mathscr{C}),
\end{aligned}$$

in which $\pi \in \mathcal{C}(\mu, \nu)$. Therefore, (**A1**) holds true. Next, for any $\zeta_1, \zeta_2 \in \mathscr{C}$ and $\mu, \nu \in \mathcal{P}_2(\mathscr{C})$, we obtain that

$$|\sigma(\zeta_1, \mu) - \sigma(\zeta_2, \nu)| \le \int_{-r_0}^0 |\zeta_1(\theta) - \zeta_2(\theta)|\mathrm{d}\theta \le r_0 \|\zeta_1 - \zeta_2\|_\infty.$$

So (**A2**) is satisfied. For any $\zeta_1, \zeta_2, \zeta^{(1)}, \zeta^{(2)} \in \mathscr{C}$, note that

$$\begin{aligned}
&|b_0(\zeta_1, \zeta^{(1)}) - b_0(\zeta_2, \zeta^{(2)})| \\
&\le |\zeta_1^3(0) - \zeta_2^3(0)| + |\zeta_1(0) - \zeta_2(0)| + \int_{-r_0}^0 |\zeta_1(v) - \zeta_2(v)|\mathrm{d}v + \int_{-r_0}^0 |\zeta^{(1)}(v) - \zeta^{(2)}(v)|\mathrm{d}v \\
&\le c_4 (1 + \zeta_1^2(0) + \zeta_2^2(0)) |\zeta_1(0) - \zeta_2(0)| + r_0 \|\zeta_1 - \zeta_2\|_\infty + r_0 \|\zeta^{(1)} - \zeta^{(2)}\|_\infty \\
&\le c_5 (1 + \|\zeta_1\|_\infty^2 + \|\zeta_2\|_\infty^2) \|\zeta_1 - \zeta_2\|_\infty + r_0 \|\zeta^{(1)} - \zeta^{(2)}\|_\infty
\end{aligned} \tag{4.3}$$

for some constants $c_4, c_5 > 0$. Next, we have

$$(\nabla_\theta b)(\zeta, \mu, \theta) = \Big(1, \int_{\mathscr{C}} b_0(\zeta, \zeta')\mu(\mathrm{d}\zeta')\Big)^* \quad \text{and} \quad (\nabla_\theta(\nabla_\theta b))(\zeta, \mu, \theta) = \mathbf{0}_{2\times 2}, \tag{4.4}$$

where $\mathbf{0}_{2\times 2}$ stands for the 2×2-zero matrix. Thus, (4.3) and (4.4) enable us to deduce that (**B2**) and (**C**) hold, respectively. Furthermore, due to (4.3), we find that

$$\begin{aligned}|b(\zeta_1,\mu,\theta)-b(\zeta_2,\nu,\theta)| &\le \theta^{(2)}\Big|\int_{\mathscr{C}} b_0(\zeta_1,\zeta^{(1)})\mu(\mathrm{d}\zeta^{(1)})-\int_{\mathscr{C}} b_0(\zeta_2,\zeta^{(2)})\nu(\mathrm{d}\zeta^{(2)})\Big|\\ &\le \theta^{(2)}\int_{\mathscr{C}}\int_{\mathscr{C}}|b_0(\zeta_1,\zeta^{(1)})-\int_{\mathscr{C}} b_0(\zeta_2,\zeta^{(2)})|\pi(\mathrm{d}\zeta^{(1)},\mathrm{d}\zeta^{(2)})\\ &\le c_6(1+\|\zeta_1\|_\infty^2+\|\zeta_2\|_\infty^2)\|\zeta_1-\zeta_2\|_\infty+c_6\mathbb{W}_2(\mu,\nu).\end{aligned}$$

Therefore, we infer that (**B1**) holds. Next, observe that

$$|\sigma^{-2}(\zeta_1,\mu)-\sigma^{-2}(\zeta_2,\nu)|\le c_7\|\zeta_1-\zeta_2\|_\infty$$

for some $c_7>0$. Consequently, (**B3**) is true.

The discrete-time EM scheme associated with (2.24) is given by

$$Y^\varepsilon(t_k)=Y^\varepsilon(t_{k-1})+\Big(\theta^{(1)}+\theta^{(2)}\int_{\mathscr{C}} b_0(\widehat{Y}^\varepsilon_{t_{k-1}},\zeta)\mathscr{L}_{\widehat{Y}^\varepsilon_{t_{k-1}}}(\mathrm{d}\zeta)\Big)\delta+\varepsilon\,\sigma(\widehat{Y}^\varepsilon_{t_{k-1}})\triangle B_k,\quad k\ge 1, \tag{4.5}$$

with $Y^\varepsilon(t)=X^\varepsilon(t)=\xi(t),t\in[-r_0,0]$, where $(\widehat{Y}^\varepsilon_{t_k})$ is defined as in (1.4). According to (1.5), the contrast function admits the form below

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

Observe that

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

and

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

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

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

where

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

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

and

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

In terms of Theorem 1.1, $\widehat{\theta}_{n,\varepsilon} \to \theta$ in probability as $\varepsilon \to 0$ and $n \to \infty$. Next, from (4.4), it follows that

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

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

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

At last, according to Theorem 1.2, we conclude that

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

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

□

Acknowledgment The authors are grateful to the editor and the referee for their valuable suggestions and comments on the original manuscript which improved the presentation of our paper enoumously.

References

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

[2] Buckdahn, R., Li, J., Peng, S., Rainer, C., Mean-field stochastic differential equations and associated PDEs, *Ann. Probab.*, **45** (2017), 824–878.

[3] Butkovsky, O., On ergodic properties of nonlinear Markov chains and stochastic McKean-Vlasov equations, arXiv:1311.6367v3.

[4] Eberle, A., Guillin, A., Zimmer, R., Quantitative Harris type theorems for diffusions and McKean-Vlasov processes, arXiv:1606.06012v2.

[5] Dos Reis, G., Salkeld, W., Tugaut, J., Freidlin-Wentzell LDPs in path space for McKean-Vlasov equations and the functional iterated logarithm law, arXiv:1708.04961.

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

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

[8] Hammersley, W., Šiška, D., Szpruch, L., McKean-Vlasov SDE under measure dependent Lyapunov conditions, arXiv:1802.03974v1.

[9] Hairer, M., Mattingly, J. C., Scheutzow, M., Asymptotic coupling and a general form of Harris theorem with applications to stochastic delay equations, *Probab. Theory Related Fields*, **149** (2011), 223–259.

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

[11] Huang, X., Liu, C., Wang, F.-Y., Order preservation for path-distribution dependent SDEs, arXiv:1710.08569, to appear in *Communications on Pure and Applied Analysis.*

[12] Huang, X., Röckner, M., Wang, F.-Y., Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, arXiv:1709.00556.

[13] Hutzenthaler, M., Jentzen, A., Kloeden, P. E., Strong convergence of an explicit numerical method for SDEs with non-globally Lipschitz continuous coefficients,*Ann. Appl. Probab.*, **22** (2012), 1611–1641.

[14] Itô, K., Nisio, M., On stationary solutions of a stochastic differential equation, *J. Math. Kyoto Univ.*, **4** (1964), 1–75.

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

[16] 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.

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

[18] Long, H., Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises, *Statist. Probab. Lett.*, **79** (2009), 2076–2085.

[19] Long, H., Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations, *Acta Math. Sci. Ser. B Engl. Ed.*, **30** (2010), 645–663.

[20] 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.

[21] 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.

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

[23] Ma, C., A note on "Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises", *Statist. Probab. Lett.*, **80** (2010), 1528–1531.

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

[25] McKean, H. P., A class of Markov processes associated with nonlinear parabolic equations, *Proceedings of the National Academy of Sciences*,**56** (1966), 1907–1911.

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

[27] Mohammed, S.-E. A., *Stochastic Functional Differential Equations*, Pitman, Boston, 1984.

[28] Prakasa Rao, B. L. S., *Statistical Inference for Diffusion Type Processes*, Arnold, London, Oxford University Press, New York, 1999.

[29] Ren, P., Wu, J.-L., Least squares estimation for path-distribution dependent stochastic differential equations, arXiv:1802.00820.

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

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

[32] Sznitman, A.-S., Topics in propagation of chaos, Springer, 1991.

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

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

[35] van der Vaart, A. W., Asymptotic Statistics, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, 1998.

[36] 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.

[37] Wang, F.-Y., Estimates for invariant probability measures of degenerate SPDEs with singular and path-dependent drifts, arXiv:1701.01248v4, to appear in *Probab. Theory Related Fields.*

[38] Wang, F.-Y., Distribution dependent SDEs for Landau type equations, *Stochastic Process. Appl.*,**128** (2018), 595–621.