COMMUN. MATH. SCI.
Vol. 20, No. 1, pp. 265–296


# LEAST SQUARES ESTIMATION FOR DELAY MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS AND INTERACTING PARTICLE SYSTEMS*

MIN ZHU† AND YANYAN HU‡

**Abstract.** The aim of this paper is to solve the problem of parameter estimation for delay McKean-Vlasov stochastic differential equations (SDEs) with the coefficients exhibiting super-linear growth in the state component. Specifically, we propose a least squares estimator for an unknown parameter in the drift of a delay McKean-Vlasov SDEs with a small noise dispersion parameter by making use of time-discretized interacting particle systems and prove the weak convergence between the estimator and the true value, under suitable conditions. To achieve our main purposes on weak convergence, we give the approximation of the distribution of delay McKean-Vlasov SDEs at the discrete points and take advantage of calculating skills on the space of probability measures with finite order moments. Moreover, the asymptotic distribution of least squares estimator is derived via the properties of solutions for the corresponding interacting particle systems.

**Keywords.** McKean-Vlasov SDE; Interacting particle systems; Discrete observation; Least squares method; Consistency of LSE; Asymptotic distribution.

**AMS subject classifications.** 62F12; 62M05; 60G52; 60J75.

## 1. Introduction

The evolution of numerous stochastic systems depends not only on the microcosmic state of the particles, but also on the macrocosmic distribution of the particles. The McKean-Vlasov SDE is a kind of mathematical model, which can characterize the evolution of those stochastic systems. The pioneering work on McKean-Vlasov SDEs is initiated in [19], and McKean studied the propagation of chaos in physical systems of N-interacting particles related to Boltzmann's model for the statistical mechanics of rarefied gases in [20]. More concretely, McKean-Vlasov SDEs are a special class of SDEs, where the coefficients involved depend not only on the state process but also on their distribution. In response to the great needs, as a hot but difficult research topic, they have important application value in the fields of stochastic control, insurance, mathematical finance, to name a few; see, for instance, [5, 6]. McKean-Vlasov SDEs have been extensively investigated by many authors, and various results on wellposedness, Harnack inequalities, Bismut formula, ergodicity, and other quantitative and qualitative properties have been proposed (e.g. [7, 24, 27, 29]). In contrast to the general McKean-Vlasov SDEs, there has not been much research on path-dependent McKean-Vlasov SDEs, but these have begun to gain attention recently. For works on wellposedness and Harnack-type inequalities, we refer to [9, 10]. Huang and Yuan [11] showed the existence and uniqueness of strong solutions to distribution-dependent neutral SFDEs and gave the comparison theorem of these equations. Most of the previous works are concerned with path-dependent McKean-Vlasov SDEs which do not contain unknown parameters. However, in many practical applications, these models may contain unknown parame-

*Received: January 19, 2021; Accepted (in revised form): June 21, 2021. Communicated by Grigorios A. Pavliotis.
This paper is supported by the National Natural Science Foundation of China (No.11901188) and by the Scientific Research Funds of Hunan Provincial Education Department of China (No.19B156).
†College of Railway Transportation, Hunan University of Technology, Zhuzhou, 412007, Hunan, P.R. China (zhumin0107@hut.edu.cn).
‡Corresponding author. School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081, P.R. China (yanyanhu@bit.edu.cn).

ters. Hence, we want to estimate deterministic quantities of these unknown parameters for SDEs, especially, path-dependent McKean-Vlasov SDEs.

Based on discrete and continuous time observations, there have been a number of attempts in the literature to develop methods on the parametric estimation for SDEs, see, e.g., [3, 4, 16]. Beyond that, estimation for stochastic delay differential equations (SDDEs) has been studied from various points of view, we refer to Kuechler and Soerensen [14], who proposed an estimator of drift parameters for affine stochastic delay differential equations by discretization of the continuous-time likelihood function; Reiss [23] studied the problem of nonparametric estimation for affine SDDEs by continuous observation. Above all, the small diffusion asymptotic of SDEs has been discussed systematically and applied successively to real world problems; see, for instance, the monograph [15] for more details. In general, the parametric estimation relied on continuous-time observations, which is a mathematical idealization, and no measuring device can follow continuously the sample paths of the diffusion processes involved (cf. [26]). So, from a practical standpoint in parametric inference, it is more meaningful to explore asymptotic estimation for diffusion processes with small dispersions based on discrete observations. Whereas multiple methods have been proposed, the simplest and most natural solution seems to be the one based on the least squares estimation (LSE) in cases of large-scale scattered data, see, e.g., [13, 17, 18]. With regard to the LSE under various settings, we refer to, e.g., [17] for SDEs driven by small Lévy noises with Lipschitz condition with respect to the drift term, [12] for SDEs driven by small $\alpha$-stable noises with Lipschitz coefficients, and [21] for the $\alpha$-stable Ornstein-Uhlenbeck process with a constant drift.

Compared with the general SDEs, the corresponding issues for McKean-Vlasov SDEs are rare. Recent attempts towards parameter inference of McKean-Vlasov SDEs (cf. [25, 26, 30]) have led to renewed interests in the asymptotic theory of stochastic models. Inspired by their studies, we make a new attempt to study the problem of parameter estimation for delay McKean-Vlasov SDEs with a small dispersion. Moreover, there is no published LSE for delay McKean-Vlasov SDEs, to the best of our knowledge. What's more, for the problem of parameter estimation of delay McKean-Vlasov SDEs, the technique used for the general SDEs cannot directly be applied to obtain an asymptotically consistent estimation. This is because the McKean-Vlasov SDEs cannot be solved explicitly. A significant consequence of this fact is that we cannot obtain observations of distribution of the path at regular space time points directly in most of our arguments. So, whereas the mechanisms of LSE are often relatively simple, caution needs to be exercised when approximating the distribution at every step of the analysis. Indeed, in many situations, due to the complicated dependence structure among discrete points, results from the execution of LSEs may differ considerably from the standard SDEs, affecting both accuracy and precision of the LSE-based predictions.

References [25, 26] though, have succeeded in investigating parameter estimation for path-dependent McKean-Vlasov SDEs by an Euler-Maruyama type scheme. In particular, under the monotone condition, [26] studied LSE for path-dependent McKean-Vlasov SDEs by using continuous time tamed EM method. It is worth noting that they simulated the segment process by the linear interpolations between the points on the gridpoints and approximated its distribution directly using the law of the associated segment process. Even so, the distribution cannot be simulated by the computer. On the basis of macrocosmic property of the distribution of stochastic systems, we shall investigate parameter estimation for McKean-Vlasov SDEs by using an empirical distribution corresponding to stochastic interacting paricle systems to approximate the distributions at each step. This method, based on stochastic interacting paricle systems, has been successfully applied to the approximation of McKean-Vlasov SDEs in [1]. In the current work, by constructing an appropriate contrast function based on the associated interacting particle systems, we shall provide a new idea to derive the LSE consistency and asymptotic distribution for a class of McKean-Vlasov SDEs. Comparing with the existing results in the work, the innovations of our paper lie in two aspects:

(i) We introduce stochastic particle systems to simulate delay McKean-Vlasov SDEs, and establish the contrast function;

(ii) Our model is more applicable and practical as we are dealing with delay SDEs with super linear growth coefficients and which are distribution dependent.

## 2. Preliminaries and interacting particle system

Throughout this paper, the following notation and terminology will be used. For $m,d\in\mathbb{N}$, the set of all positive integers, let $(\mathbb{R}^d,\langle\cdot,\cdot\rangle,|\cdot|)$ be the $d$-dimensional Euclidean space with the inner product $\langle\cdot,\cdot\rangle$ inducing the norm $|\cdot|$ and $\mathbb{R}^d\times\mathbb{R}^m$ the collection of all $d\times m$ matrixes with real entries, which is endowed with the Hilbert-Schmidt norm $\|\cdot\|$. $\mathbf{0}\in\mathbb{R}^d$ denotes the zero vector. For a matrix $A$, $A^*$ denotes the transpose of $A$. Concerning a square matrix $A$, $A^{-1}$ means the inverse of $A$ provided that $\det(A)\neq 0$. For $p\in\mathbb{N}$, let $\Theta$ be an open bounded convex subset of $\mathbb{R}^p$, and $\bar{\Theta}$ the closure of $\Theta$. For $r>0$ and $x\in\mathbb{R}^p$, $B_r(x)$ represents the closed ball centered at $x$ with the radius $r$. $\lfloor a\rfloor$ stands for the integer part of the real number $a\geq 0$. For a random variable $\xi$, $\mathscr{L}_\xi$ denotes its law. For given $\tau>0$, $\mathscr{C}:=C([-\tau,0];\mathbb{R}^d)$ means the family of all continuous functions $\xi:[-\tau,0]\to\mathbb{R}^d$ with the uniform norm $\|\xi\|_\infty:=\sup_{-\tau\leq\theta\leq 0}|f(\theta)|$. For $p>0$, $\mathscr{P}_p(\mathbb{R}^d)$ stands for the space of all probability measures on $\mathbb{R}^d$ with the finite $p$-th moment, i.e., $\mu(|\cdot|^p):=\int_{\mathbb{R}^d}|x|^p\mu(\mathrm{d}x)<\infty$ for $\mu\in\mathscr{P}_p(\mathbb{R}^d)$. Define the $\mathbb{W}_p$-Wasserstein distance on $\mathscr{P}_p(\mathbb{R}^d)$ by

$$\mathbb{W}_p(\mu,\nu)=\inf_{\pi\in\mathcal{C}(\mu,\nu)}\Big(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^p\pi(\mathrm{d}x,\mathrm{d}y)\Big)^{\frac{1}{1\vee p}},\qquad \mu,\nu\in\mathscr{P}_p(\mathbb{R}^d),$$

where $\mathcal{C}(\mu,\nu)$ signifies the set of all couplings of $\mu$ and $\nu$. $L_p^0(\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued, $\mathscr{F}_0$-measurable random variables $X$ with $\mathbb{E}|X|^p<\infty$. Let $(W_t)_{t\geq 0}$ be an $m$-dimensional Brownian motion defined on the probability space $(\Omega,\mathscr{F},\mathbb{P})$ with the filtration $(\mathscr{F}_t)_{t\geq 0}$ satisfying the usual condition (i.e., $\mathscr{F}_0$ contains all $\mathbb{P}$-null sets and $\mathscr{F}_t=\mathscr{F}_{t+}:=\bigcap_{s>t}\mathscr{F}_s$).

For a fixed time horizon $T>0$ and scale parameter $\varepsilon\in(0,1)$, we consider a **delay McKean-Vlasov SDE** on $\mathbb{R}^d$

$$\begin{cases}\mathrm{d}X_t^\varepsilon=b(X_t^\varepsilon,X_{t-\tau}^\varepsilon,\mu_t^\varepsilon,\mu_{t-\tau}^\varepsilon,\theta)\mathrm{d}t+\varepsilon\sigma(X_t^\varepsilon,X_{t-\tau}^\varepsilon,\mu_t^\varepsilon,\mu_{t-\tau}^\varepsilon)\mathrm{d}W_t, & t\in(0,T],\\ X_s^\varepsilon=\xi(s),\ s\in[-\tau,0],\end{cases}\tag{2.1}$$

where $\mu_\cdot^\varepsilon:=\mathscr{L}_{X_\cdot^\varepsilon}$ denotes the law of $X_\cdot^\varepsilon$; $b:\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)\times\Theta\to\mathbb{R}^d$ and $\sigma:\mathbb{R}^d\times\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)\to\mathbb{R}^d\times\mathbb{R}^m$ are continuous. In (2.1), we assume that the drift term $b$ and the diffusion term $\sigma$ are known apart from the parameter $\theta\in\Theta$. We stipulate that $\theta_0\in\Theta$ is the true value of $\theta\in\Theta$.

For $i\in S_N:=\{1,\cdots,N\}$, $N\geq 1$, let $(X_0^i,W_t^i)$ be i.i.d copies of $(X_0,W_t)$. We introduce the stochastic interacting particle to approximate (2.1). First, consider the following

**stochastic non-interacting particle systems** associated with (2.1)

$$\begin{cases} \mathrm{d}X_t^{\varepsilon,i} = b(X_t^{\varepsilon,i}, X_{t-\tau}^{\varepsilon,i}, \mu_t^{\varepsilon,i}, \mu_{t-\tau}^{\varepsilon,i}, \theta)\mathrm{d}t + \varepsilon\sigma(X_t^{\varepsilon,i}, X_{t-\tau}^{\varepsilon,i}, \mu_t^{\varepsilon,i}, \mu_{t-\tau}^{\varepsilon,i})\mathrm{d}W_t^i, \quad t\in(0,T], \\ X_s^{\varepsilon,i} = \xi(s), \quad s\in[-\tau,0], \quad i\in\mathbf{S}_N, \end{cases} \tag{2.2}$$

where $\mu_\cdot^{\varepsilon,i} := \mathscr{L}_{X_\cdot^{\varepsilon,i}}$ denotes the law of $X_\cdot^{\varepsilon,i}$, $i\in\mathbf{S}_N$. By virtue of the weak uniqueness due to Theorem 3.1, it is easy to see that $\mu_\cdot^{\varepsilon} = \mu_\cdot^{\varepsilon,i}$, $i\in\mathbf{S}_N$. Let $\tilde{\mu}_\cdot^{\varepsilon,N}$ be the empirical distribution corresponding to $X_\cdot^{\varepsilon,1}, X_\cdot^{\varepsilon,2}, \cdots, X_\cdot^{\varepsilon,N}$, namely,

$$\tilde{\mu}_t^{\varepsilon,N}(dx) = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{\varepsilon,j}}(dx), \qquad t\geq -\tau.$$

Consider the following **deterministic ordinary differential equation**

$$\begin{cases} \mathrm{d}X_t^{0,i} = b(X_t^{0,i}, X_{t-\tau}^{0,i}, \mu_t^0, \mu_{t-\tau}^0, \theta_0)\mathrm{d}t, \qquad t>0, \\ \quad X_s^0 = \xi(s), \qquad s\in[-\tau,0], \qquad i\in\mathbf{S}_N, \end{cases} \tag{2.3}$$

where $\mu_\cdot^0 = \mu_\cdot^{0,i} := \mathscr{L}_{X_\cdot^{0,i}}$ denotes the law of $X_\cdot^{0,i}$.

Second, **stochastic interacting particle systems** can be described as

$$\begin{cases} \mathrm{d}X_t^{\varepsilon,i,N} = b(X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \mu_t^{\varepsilon,N}, \mu_{t-\tau}^{\varepsilon,N}, \theta)\mathrm{d}t + \varepsilon\sigma(X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \mu_t^{\varepsilon,N}, \mu_{t-\tau}^{\varepsilon,N})\mathrm{d}W_t^i, \\ \qquad\qquad t\in(0,T], \\ X_s^{\varepsilon,i,N} = \xi(s), \; s\in[-\tau,0], \quad i\in\mathbf{S}_N, \end{cases} \tag{2.4}$$

where $\mu_\cdot^{\varepsilon,N}$ stands for the empirical distribution corresponding to $X_\cdot^{\varepsilon,1,N}, \cdots, X_\cdot^{\varepsilon,N,N}$, namely,

$$\mu_t^{\varepsilon,N}(dx) = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{\varepsilon,j,N}}(dx), \qquad t\geq -\tau. \tag{2.5}$$

It is worth pointing out that (2.1), (2.2), (2.3) and (2.4) share the same initial data. Set

$$B(x,y,\theta_0,\theta) := b(x,y,\mu,\nu,\theta_0) - b(x,y,\mu,\nu,\theta)$$

and

$$\Lambda(x,y) := (\sigma\sigma^*)(x,y,\mu,\nu),$$

for any $x,\ y\in\mathbb{R}^d$ and $\mu,\ \nu\in\mathscr{P}(\mathbb{R}^d)$.

For a fixed time horizon $T>0$, we give a uniform time-discretization of $[-\tau,T]$ with mesh-size $\delta = \frac{T}{n} = \frac{\tau}{M}\in(0,1)$, where $n, M>1$. In order to approximate the measure $\mu_\cdot^\varepsilon$ and improve the simulation precision of (2.1), by virtue of the interacting particle system (2.4) we construct the following **contrast function**

$$\Psi_{n,\varepsilon}^{i,N}(\theta) = \varepsilon^{-2}\delta^{-1}\sum_{n=1}^n (P_k^{\varepsilon,i,N}(\theta))^*\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N}, X_{(k-1)\delta-\tau}^{\varepsilon,i,N})P_k^{\varepsilon,i,N}(\theta), \tag{2.6}$$

where

$$P_k^{\varepsilon,i,N}(\theta) = X_{k\delta}^{\varepsilon,i,N} - X_{(k-1)\delta}^{\varepsilon,i,N} - b(X_{(k-1)\delta}^{\varepsilon,i,N}, X_{(k-1)\delta-\tau}^{\varepsilon,i,N}, \mu_{(k-1)\delta}^{\varepsilon,N}, \mu_{(k-1)\delta-\tau}^{\varepsilon,N}, \theta)\delta, \tag{2.7}$$

for $k=1,2,\cdots,n$.

According to the principle of least squares method, to achieve the least squares estimation of $\theta\in\Theta$, we need to seek an argument $\hat{\theta}^{i,N}_{n,\varepsilon}\in\Theta$ such that

$$\Psi^{i,N}_{n,\varepsilon}(\hat{\theta}^{i,N}_{n,\varepsilon})=\min_{\theta\in\Theta}\Psi^{i,N}_{n,\varepsilon}(\theta), \tag{2.8}$$

namely,

$$\hat{\theta}^{i,N}_{n,\varepsilon}=\arg\min_{\theta\in\Theta}\Psi^{i,N}_{n,\varepsilon}(\theta).$$

Let $\theta_0\in\Theta$ be the true value of $\theta$ and

$$\Phi^{i,N}_{n,\varepsilon}(\theta)=\varepsilon^2(\Psi^{i,N}_{n,\varepsilon}(\theta)-\Psi^{i,N}_{n,\varepsilon}(\theta_0)).$$

Then, from (2.8), one has

$$\hat{\theta}^{i,N}_{n,\varepsilon}=\arg\min_{\theta\in\Theta}\Phi^{i,N}_{n,\varepsilon}(\theta). \tag{2.9}$$

That is to say, $\hat{\theta}^{i,N}_{n,\varepsilon}$ satisfying (2.9) is called **LSE** of $\theta\in\Theta$.

To obtain the main results, we give the following assumptions. Let $K_i:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}_+$ such that

$$K_i(x,y)\le L_i(1+|x|^{r_i}+|y|^{r_i}),\qquad i=1,2,3,4,5, \tag{2.10}$$

for some constants $L_i>0$, $r_i\ge 1$ and any $x,y\in\mathbb{R}^d$. Furthermore, for any $x_i,y_i\in\mathbb{R}^d$ and $\mu_i,\nu_i\in\mathscr{P}_2(\mathbb{R}^d)$, $i=1,2$, we assume that

(**A1**) For any $\theta\in\bar{\Theta}$, there exists a $C_1>0$ such that

$$\langle x_1-x_2,b(x_1,y_1,\mu,\nu,\theta)-b(x_2,y_2,\mu,\nu,\theta)\rangle\le C_1\Big(|x_1-x_2|^2+|y_1-y_2|^2\Big);$$
$$|b(x,y,\mu_1,\nu_1,\theta)-b(x,y,\mu_2,\nu_2,\theta)|\le C_1\Big(\mathbb{W}_2(\mu_1,\mu_2)+\mathbb{W}_2(\nu_1,\nu_2)\Big);$$
$$|b(x_1,y_1,\mu,\nu,\theta)-b(x_2,y_2,\mu,\nu,\theta)|\le C_1|x_1-x_2|+K_1(y_1,y_2)|y_1-y_2|;$$

(**A2**) There exists a $C_2>0$ such that

$$\|\sigma(x_1,y_1,\mu,\nu)-\sigma(x_2,y_2,\mu,\nu)\|\le C_2|x_1-x_2|+K_2(y_1,y_2)|y_1-y_2|$$

and

$$\|\sigma(x,y,\mu_1,\nu_1)-\sigma(x,y,\mu_2,\nu_2)\|\le C_2\Big(\mathbb{W}_2(\mu_1,\mu_2)+\mathbb{W}_2(\nu_1,\nu_2)\Big);$$

(**A3**) $\sigma\sigma^*$ is invertible, and there exists a $C_3>0$ such that

$$\begin{aligned}&\|(\sigma\sigma^*)^{-1}(x_1,y_1,\mu_1,\nu_1)-(\sigma\sigma^*)^{-1}(x_2,y_2,\mu_2,\nu_2)\|\\ &\le C_3\Big(|x_1-x_2|+\mathbb{W}_2(\mu_1,\mu_2)+\mathbb{W}_2(\nu_1,\nu_2)\Big)+K_3(y_1,y_2)|y_1-y_2|;\end{aligned}$$

(**A4**) There exists a $C_4>0$ such that

$$\begin{aligned}&\sup_{\theta\in\bar{\Theta}}\|(\nabla_\theta b)(x_1,y_1,\mu_1,\nu_1,\theta)-(\nabla_\theta b)(x_2,y_2,\mu_2,\nu_2,\theta)\|\\ &\le C_4\Big(|x_1-x_2|+\mathbb{W}_2(\mu_1,\mu_2)+\mathbb{W}_2(\nu_1,\nu_2)\Big)+K_4(y_1,y_2)|y_1-y_2|,\end{aligned}$$

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

(**A5**) There exists a $C_5>0$ such that

$$\sup_{\theta\in\bar{\Theta}}\|(\nabla_\theta^{(2)}b^*)(x_1,y_1,\mu_1,\nu_1,\theta)-(\nabla_\theta^{(2)}b^*)(x_2,y_2,\mu_2,\nu_2,\theta)\|$$
$$\le C_5\Big(|x_1-x_2|+\mathbb{W}_2(\mu_1,\mu_2)+\mathbb{W}_2(\nu_1,\nu_2)\Big)+K_5(y_1,y_2)|y_1-y_2|,$$

where $\nabla_\theta^{(2)}b=:\nabla_\theta(\nabla_\theta b)$;

(**B1**) There exists a constant $q>p$ such that

$$\sup_{0\le t\le T}\mathbb{E}|X_t|^q<\infty.$$

The conditions (**A1**) and (**A2**) are used to guarantee the wellposedness of (2.1) and the corresponding stochastic interacting particle system (2.4) (see Theorem 3.1 and Lemma 5.1 below). Besides, conditions (**A3**) and (**B1**) also play an important role in the analysis of the consistency of the LSE for the unknown parameter $\theta$. Condition (**B1**) is set to ensure that strong convergence between stochastic interacting particle systems and non-interacting particle systems in the $p$-th moment holds, which improves the result in [1] on the convergence, in two aspects: First, the condition (**B1**) is more applicable than the conditions of [1, Theorem 1.4]; Second, in the current work we only need (**B1**) to hold for some $q>p$, wherein it is easier to seek a constant $q$ under conditions (**A1**) and (**A2**), and it is not confined to $p>4$ as [1]. Conditions (**A4**) and (**A5**) are used to establish the asymptotic distribution of the LSE.

## 3. Main results

Under the framework of non-Lipschitz condition, the tamed Euler scheme is adopted to establish contrast function for LSE in [26]. Here we approximate (2.1) by a particle system, and investigate the consistency and asymptotic distribution of the LSE under a super-linear condition. First, we show the following result on the strong wellposedness of (2.1), where the drift and diffusion terms have polynomial growth with respect to the delay variables.

THEOREM 3.1. *Assume that (**A1**) and (**A2**) hold, for any initial value $X_0^\varepsilon=\xi\in L^0_{p_1q_1}(\mathscr{C})$, where $p_1$ and $q_1$ will be shown in the proof, then (2.1) possesses a unique global strong solution $(X_t^\varepsilon)_{t\ge-\tau}$ with*

$$\mathbb{E}\Big(\sup_{-\tau\le t\le T}|X_t^\varepsilon|^p\Big)\le C<\infty,\qquad p\ge 2. \tag{3.1}$$

The strong wellposedness of McKean-Vlasov SDEs under various conditions has been studied largely, e.g., [22] with a drift of polynomial growth, [29] for continuity, monotonicity and growth of coefficients, [1] with Hölder continuous coefficients. In the meantime, strong well-posedness of path-dependent McKean-Vlasov SDEs has got more and more attention, e.g., [9] under the condition of continuity, monotonicity and growth of coefficients, [25] for one-side Lipschitz drifts and Lipschitz diffusions. Theorem 3.1 above shows that the delay McKean-Vlasov SDE has strong wellposedness when both the coefficients have super-linear growth.

The second result in the current work shows the consistency of the LSE with high frequency and small dispersion. In order to display this result, we analyse strong convergence between stochastic interacting particle systems and non-interacting particle systems corresponding to delay McKean-Vlasov SDEs (2.1) whenever the particle number goes to infinity and the stepsize closes to zero. For the sake of simplicity, we set

$$\Pi(\theta):=\int_0^T B^*(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0,\theta)\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})B(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0,\theta)\mathrm{d}t, \tag{3.2}$$

$$\Upsilon(x,y,\theta_0):=(\nabla_\theta b)^*(x,y,\mu,\nu,\theta_0)\Lambda^{-1}(x,y)\sigma(x,y,\mu,\nu),\quad x,y\in\mathbb{R}^d,\quad \mu,\nu\in\mathscr{P}_p(\mathbb{R}^d) \tag{3.3}$$

and

$$\begin{aligned} I(\theta):=\int_0^T &(\nabla_\theta b)^*(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta)\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})\\ &\times(\nabla_\theta b)(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta)\mathrm{d}t. \end{aligned} \tag{3.4}$$

THEOREM 3.2. *Under the conditions* (**A1**)-(**A3**) *and* (**B1**)*. If, for any* $\theta\in\Theta$, $\Pi(\theta)\geq 0$, *then*

$$\hat{\theta}_{n,\varepsilon}^{i,N}\to\theta_0\quad \text{in probability as}\quad N,n\to\infty\quad\text{and}\quad\varepsilon\to 0.$$

The last result focuses on the asymptotic distribution of the LSE $\hat{\theta}_{n,\varepsilon}^{i,N}$.

THEOREM 3.3. *Under the assumptions of Theorem 3.2, suppose that* (**A4**) *and* (**A5**) *hold. Then,*

$$\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon}^{i,N}-\theta_0)\to I^{-1}(\theta_0)\int_0^T\Upsilon(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0)\mathrm{d}W_t^i\qquad \mathbb{P}-a.s.$$

*as* $n,N\to\infty$ *and* $\varepsilon\to 0$*, where* $I(\cdot)$ *and* $\Upsilon(\cdot)$ *are continuous.*

## 4. An illustrative example

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

EXAMPLE 4.1. Let $\theta=(\theta^{(1)},\theta^{(2)})^*\in\Theta_0:=(c_1,c_2)\times(c_3,c_4)\subset\mathbb{R}^2$ for some $c_1<c_2$ and $c_3<c_4$. For any $\varepsilon\in(0,1)$, consider the following delay McKean-Vlasov SDE

$$\begin{aligned} \mathrm{d}X^\varepsilon(t)=&\Big\{\theta^{(1)}+\theta^{(2)}\Big(X_t^\varepsilon-(X_{t-\tau}^\varepsilon)^3+X_{t-\tau}^\varepsilon+\mathbb{E}X_{t-\tau}^\varepsilon\Big)\Big\}\mathrm{d}t\\ &+\varepsilon\Big\{1+|X_{t-\tau}^\varepsilon|^3+|X_{t-\tau}^\varepsilon|+\mathbb{E}|X_{t-\tau}^\varepsilon|\Big\}\mathrm{d}W(t),\quad t\in(0,T] \end{aligned} \tag{4.1}$$

with the initial value $X_0^\varepsilon=\xi$, where $\theta\in\Theta_0$ is an unknown parameter with the true value $\theta_0^*=(\theta_0^{(1)},\theta_0^{(2)})\in\Theta_0$. Let $\hat{\theta}_{n,\varepsilon}^{i,N}\in\Theta$ be the least squares estimation for the unknown parameter $\theta$. For any $x,y\in\mathbb{R}$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{R})$, set

$$\tilde{b}(x,y):=x-y^3+y+\mathbb{E}y, \tag{4.2}$$

$$b(x,y,\mu,\nu,\theta):=\theta^{(1)}+\theta^{(2)}\tilde{b}(x,y) \tag{4.3}$$

and

$$\sigma(x,y,\mu,\nu,\theta):=1+|y|^3+|y|+\mathbb{E}|y|. \tag{4.4}$$

Then, (4.1) can be reformulated as (2.1). Furthermore, according to Theorem 3.2 and Theorem 3.3, we get

$$\hat{\theta}_{n,\varepsilon}^{i,N}\to\theta_0$$

and

$$\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon}^{i,N}-\theta_0)\to I^{-1}(\theta_0)\int_0^T \Upsilon(X_s^{0,i},X_{s-\tau}^{0,i},\theta_0)\mathrm{d}W_s^i$$

in probability as $N,n\to\infty$ and $\varepsilon\to 0$. Here

$$I(\theta_0)=\int_0^T \frac{1}{(1+2|X_{s-\tau}^{0,i}|+|X_{s-\tau}^{0,i}|^3)^2}\begin{pmatrix} 1 & \tilde{b}(X_s^{0,i},X_{s-\tau}^{0,i}) \\ \tilde{b}(X_s^{0,i},X_{s-\tau}^{0,i}) & \tilde{b}(X_s^{0,i},X_{s-\tau}^{0,i})^2 \end{pmatrix}\mathrm{d}s$$

and

$$\Upsilon(X_s^{0,i},X_{s-\tau}^{0,i},\theta_0)=\frac{1}{1+2|X_{s-\tau}^{0,i}|+|X_{s-\tau}^{0,i}|^3}\Big(1,\tilde{b}(X_s^{0,i},X_{s-\tau}^{0,i})\Big)^*.$$

Next, concerning (4.1) we aim to examine that all the assumptions imposed on Theorem 3.2 and Theorem 3.3 apply very well. Indeed, by a direct calculation, for any $\mu,\nu,\mu_i,\nu_i\in\mathscr{P}_2(\mathbb{R})$ and $x,y,x_i,y_i\in\mathbb{R}$, $i=1,2$, it follows from (4.2), (4.3) and the Hölder inequality that there exists a constant $c>0$

$$\begin{aligned}|b(x,y,\mu_1,\nu_1,\theta)-b(x,y,\mu_2,\nu_2,\theta)|&=|\theta^{(2)}|\cdot|\mathbb{E}(y_1-y_2)|\le|\theta^{(2)}|\Big(\mathbb{E}|y_1-y_2|^2\Big)^{\frac{1}{2}}\\&\le c\mathbb{W}_2(\nu_1,\nu_2),\end{aligned}$$

$$\begin{aligned}\langle x_1-x_2,b(x_1,y_1,\mu,\nu,\theta)-b(x_2,y_2,\mu,\nu,\theta)\rangle&\le|\theta^{(2)}|\Big(|x_1-x_2|^2+|y_1-y_2|^2\Big)\\&\le c\Big(|x_1-x_2|^2+|y_1-y_2|^2\Big)\end{aligned}$$

and

$$\begin{aligned}|b(x_1,y_1,\mu,\nu,\theta)-b(x_2,y_2,\mu,\nu,\theta)|&\le|\theta^{(2)}|\cdot\Big(|x_1-x_2|+|y_1^3-y_2^3|+|y_1-y_2|\Big)\\&\le c\Big(|x_1-x_2|+|y_1-y_2|(1+|y_1|^2+|y_2|^2)\Big).\end{aligned}$$

On the other hand, it holds that, by (4.4),

$$|\sigma(x_1,y_1,\mu,\nu)-\sigma(x_2,y_2,\mu,\nu)|\le|y_1-y_2|(1+|y_1|^2+|y_2|^2)$$

and

$$|\sigma(x,y,\mu_1,\nu_1)-\sigma(x,y,\mu_2,\nu_2)|\le\mathbb{E}|y_1-y_2|\le c\mathbb{W}_2(\nu_1,\nu_2).$$

Hence, the assumptions (**A1**) and (**A2**) hold for (4.1). Next, note that

$$\begin{aligned}&|\sigma^{-2}(x_1,y_1,\mu_1,\nu_1)-\sigma^{-2}(x_2,y_2,\mu_2,\nu_2)|\\&=\Big|\frac{1}{(1+|y_1|^3+|y_1|+\mathbb{E}|y_1|)^2}-\frac{1}{(1+|y_2|^3+|y_2|+\mathbb{E}|y_2|)^2}\Big|\\&\le 4\Big||y_1|^3+|y_1|+\mathbb{E}|y_1|-|y_2|^3-|y_2|-\mathbb{E}|y_2|\Big|\\&\le c\Big(|y_1-y_2|(1+|y_1|^2+|y_2|^2)+\mathbb{W}_2(\nu_1,\nu_2)\Big).\end{aligned}$$

So, (**A3**) is fulfilled. Furthermore, it follows from (4.3) that

$$(\nabla_\theta b)(\zeta,\mu,\theta)=\Big(1,\tilde{b}(x,y)\Big)^* \quad \text{and} \quad (\nabla_\theta(\nabla_\theta b))(\zeta,\mu,\theta)=\mathbf{0}_{2\times 2}, \tag{4.5}$$

where $\mathbf{0}_{2\times 2}$ stands for the $2\times 2$-zero matrix. Thus, (4.2) yields that both (**A4**) and (**A5**) hold. As a consequence, concerning (4.1), the assumptions (**A1**)-(**A5**) hold, respectively. In terms of (2.6), the contrast function enjoys the form

$$\begin{aligned}\Psi_{n,\varepsilon}^{i,N}(\theta)=&\varepsilon^{-2}\delta^{-1}\sum_{k=1}^{n}\frac{1}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2}\\&\times\Big|X_{k\delta}^{\varepsilon,i,N}-X_{(k-1)\delta}^{\varepsilon,i,N}-\Big(\theta^{(1)}+\theta^{(2)}\tilde{b}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})\Big)\delta\Big|^2.\end{aligned}$$

Note that

$$\begin{aligned}\frac{\partial}{\partial\theta^{(1)}}\Psi_{n,\varepsilon}^{i,N}(\theta)=&-2\varepsilon^{-2}\sum_{k=1}^{n}\frac{1}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2}\\&\times\Big\{X_{k\delta}^{\varepsilon,i,N}-X_{(k-1)\delta}^{\varepsilon,i,N}-\Big(\theta^{(1)}+\theta^{(2)}\tilde{b}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})\Big)\delta\Big\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta^{(2)}}\Psi_{n,\varepsilon}^{i,N}(\theta)=&-2\varepsilon^{-2}\sum_{k=1}^{n}\frac{1}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2}\\&\times\Big\{X_{k\delta}^{\varepsilon,i,N}-X_{(k-1)\delta}^{\varepsilon,i,N}-\Big(\theta^{(1)}+\theta^{(2)}\tilde{b}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})\Big)\delta\Big\}\\&\times\tilde{b}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N}).\end{aligned}$$

Setting

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

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

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

where

$$A_1:=\sum_{k=1}^{n}\frac{1}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2},$$

$$A_2:=\sum_{k=1}^{n}\frac{X_{k\delta}^{\varepsilon,i,N}-X_{(k-1)\delta}^{\varepsilon,i,N}}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2},$$

$$A_3:=\sum_{k=1}^{n}\frac{(X_{k\delta}^{\varepsilon,i,N}-X_{(k-1)\delta}^{\varepsilon,i,N})\tilde{b}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})}{(1+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|^3+|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|+\mathbb{E}|X_{(k-1)\delta-\tau}^{\varepsilon,i,N}|)^2},$$

$$A_4:=\sum_{k=1}^n \frac{\tilde{b}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})}{(1+|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|^3+|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|+\mathbb{E}|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|)^2}$$

and

$$A_5:=\sum_{k=1}^n \frac{\left(\tilde{b}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\right)^2}{(1+|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|^3+|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|+\mathbb{E}|X^{\varepsilon,i,N}_{(k-1)\delta-\tau}|)^2}.$$

In terms of Theorem 3.2, $\hat{\theta}^{i,N}_{n,\varepsilon}\to\theta$ in probability as $\varepsilon\to 0$ and $n\to\infty$. Next, from (4.5), it follows that

$$I(\theta_0)=\int_0^T \frac{1}{(1+2|X^{0,i}_{s-\tau}|+|X^{0,i}_{s-\tau}|^3)^2}\begin{pmatrix} 1 & \tilde{b}(X^{0,i}_s,X^{0,i}_{s-\tau}) \\ \tilde{b}(X^{0,i}_s,X^{0,i}_{s-\tau}) & \tilde{b}(X^{0,i}_s,X^{0,i}_{s-\tau})^2 \end{pmatrix}\mathrm{d}s,$$

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

$$\int_0^T \Upsilon(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0)\mathrm{d}W(s)=\int_0^T \frac{1}{1+2|X^{0,i}_{s-\tau}|+|X^{0,i}_{s-\tau}|^3}\begin{pmatrix} 1 \\ \tilde{b}(X^{0,i}_s,X^{0,i}_{s-\tau}) \end{pmatrix}\mathrm{d}W_s.$$

At last, according to Theorem 3.2, we conclude that

$$\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0)\to I^{-1}(\theta_0)\int_0^T \Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)\mathrm{d}W^i_t \quad \mathbb{P}-\text{a.s.}$$

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

## 5. Proof of main results

**5.1. Proof of Theorem 3.1.** The more popular methods to argue regarding the existence of solutions of McKean-Vlasov SDEs need to seek the convergence of the corresponding distribution-iterated SDEs; see, for instance [9,29]. However, it is hard to verify the convergence of the distribution-iterated SDEs for (2.1), due to the coefficients satisfying polynomial growth with respect to the delay variables. So, we will adopt an interval iteration method to overcome this difficulty in terms of the structure of (2.1).

*Proof.* **(Proof of Theorem 3.1.)** Under the conditions (**A1**) and (**A2**), firstly, we shall show the wellposedness of the delay McKean-Vlasov SDE (2.1). For each $t\in[0,\tau]$, (2.1) can be reformulated as

$$\begin{cases} \mathrm{d}X^\varepsilon_t=b(X^\varepsilon_t,\xi_{t-\tau},\mu^\varepsilon_t,\mu^0_{t-\tau},\theta)\mathrm{d}t+\varepsilon\sigma(X^\varepsilon_t,\xi_{t-\tau},\mu^\varepsilon_t,\mu^0_{t-\tau})\mathrm{d}W_t, \\ X^\varepsilon_s=\xi_s, \quad s\in[-\tau,0], \end{cases} \tag{5.1}$$

where $\mu^0_\cdot:=\mathscr{L}_{\xi_\cdot}$. Then (5.1) is a non-delay SDE. In terms of conditions (**A1**) and (**A2**), it holds that the coefficients of (5.1) are Lipschitz continuous, then this SDE has a unique strong solution on the interval $[0,\tau]$.
On the interval $t\in[\tau,2\tau]$, SDE (2.1) can be written as

$$\mathrm{d}X^\varepsilon_t=b(X^\varepsilon_t,X^{\varepsilon,(1)}_t,\mu^\varepsilon_t,\mu^1_t,\theta)\mathrm{d}t+\varepsilon\sigma(X^\varepsilon_t,X^{\varepsilon,(1)}_t,\mu^\varepsilon_t,\mu^1_t)\mathrm{d}W_t \tag{5.2}$$

with the initial value $X^{\varepsilon}_{\tau}$, where $X^{\varepsilon,(1)}_t = X^{\varepsilon}_{t-\tau}$ and $\mu^1_{\cdot} := \mathscr{L}_{X^{\varepsilon,(1)}_{\cdot}}$. Obviously, the delay McKean-Vlasov SDE (2.1) becomes a general McKean-Vlasov SDE with Lipschitiz condition. Then this Equation (5.2) has a unique strong solution on the interval $[\tau, 2\tau]$. Duplicating this procedure over the intervals $[n\tau, (n+1)\tau]$, where $2 < n \le \lfloor T/\tau \rfloor$.

In addition, for any $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$, by assumptions (**A1**) and (**A2**), it is easy to see that there is a constant $C > 0$ such that

$$|b(x,y,\mu,\nu,\theta)| \le C(1+|x|+|y|+|y|^{r_1+1}+\mathbb{W}_2(\mu,\delta_0)+\mathbb{W}_2(\nu,\delta_0)) \tag{5.3}$$

and

$$\|\sigma(x,y,\mu,\nu)\| \le C(1+|x|+|y|+|y|^{r_2+1}+\mathbb{W}_2(\mu,\delta_0)+\mathbb{W}_2(\nu,\delta_0)), \tag{5.4}$$

where $\delta_0$ is the Dirac measure at point $\mathbf{0} \in \mathbb{R}^n$.

Secondly, we shall show that the $p$-th moment of the solution is uniformly bounded in a finite time interval. In fact, set $(X^{\varepsilon}_t)_{t \ge -\tau}$ to be a solution to (2.1) with initial data $X^{\varepsilon}_0 = \xi \in L^0_{p_1 q_1}(\mathscr{C})$. Let $\tau_m = \inf\{t > 0 : |X^{\varepsilon}_t| \ge m\}$, for $m \ge 1$. Then, by the Burkhold-Davis-Gundy inequality and Hölder inequality, together with (5.3) and (5.4), one gets

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_m} |X^{\varepsilon}_s|^p\Big) \le{}& C\mathbb{E}\|\xi\|^p_{\infty} + C\mathbb{E}\Big(\int_0^{t\wedge\tau_m} |b(X^{\varepsilon}_s, X^{\varepsilon}_{s-\tau}, \mu^{\varepsilon}_s, \mu^{\varepsilon}_{s-\tau}, \theta)|^p \mathrm{d}s\Big) \\
&+ C\mathbb{E}\Big(\int_0^{t\wedge\tau_m} \|\sigma(X^{\varepsilon}_s, X^{\varepsilon}_{s-\tau}, \mu^{\varepsilon}_s, \mu^{\varepsilon}_{s-\tau})\|^p \mathrm{d}s\Big) \\
\le{}& C\Big\{1 + \mathbb{E}\int_0^{t\wedge\tau_m} (1+|X^{\varepsilon}_s|^p)\mathrm{d}s + \mathbb{E}\int_0^{t\wedge\tau_m} \Big(|X^{\varepsilon}_{s-\tau}|^{(r_1+1)p} \\
&+ |X^{\varepsilon}_{s-\tau}|^{(r_2+1)p}\Big)\mathrm{d}s + \mathbb{E}\int_0^{t\wedge\tau_m} (\mathbb{W}_2(\mu^{\varepsilon}_s,\delta_0)^p + \mathbb{W}_2(\mu^{\varepsilon}_{s-\tau},\delta_0)^p)\mathrm{d}s\Big\}.
\end{aligned}$$

Let $m \to \infty$, then we get

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t} |X^{\varepsilon}_s|^p\Big) \le{}& C\Big\{1 + \int_0^t \mathbb{E}|X^{\varepsilon}_s|^p \mathrm{d}s + \int_0^t (\mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_1 p} + \mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_2 p})\mathrm{d}s \\
&+ \int_0^t (\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon}_s,\delta_0)^p + \mathbb{E}\mathbb{W}_2(\mu^{\varepsilon}_{s-\tau},\delta_0)^p)\mathrm{d}s\Big\} \\
\le{}& C\Big\{1 + \int_0^t \mathbb{E}|X^{\varepsilon}_s|^p \mathrm{d}s + \int_0^t (\mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_1 p} + \mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_2 p})\mathrm{d}s\Big\},
\end{aligned}$$

where $\gamma_1 := r_1 + 1$ and $\gamma_2 := r_2 + 1$. Then Gronwall's inequality yields

$$\mathbb{E}\Big(\sup_{0\le s\le t} |X^{\varepsilon}_s|^p\Big) \le C\Big\{1 + \int_0^t (\mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_1 p} + \mathbb{E}|X^{\varepsilon}_{s-\tau}|^{\gamma_2 p})\mathrm{d}s\Big\}. \tag{5.5}$$

Set $q_1 := \gamma_1 \vee \gamma_2$ and

$$p_i = (\lfloor T/\tau \rfloor + 2 - i) p q_1^{\lfloor T/\tau \rfloor + 1 - i}, \quad i = 1, 2, \cdots, \lfloor T/\tau \rfloor + 1.$$

Then there exists a finite sequence $\{p_1, p_2, \cdots, p_{\lfloor T/\tau \rfloor + 1}\}$ such that

$$p_i \ge 2, \quad p_{i+1} q_1 < p_i \quad \text{and} \quad p_{\lfloor T/\tau \rfloor + 1} = p, \quad i = 1, 2, \cdots, \lfloor T/\tau \rfloor.$$

Further, for $X_0^\varepsilon=\xi\in L^0_{p_1q_1}(\mathscr{C})$, one has

$$\mathbb{E}\Big(\sup_{0\le s\le\tau}|X_s^\varepsilon|^{p_1}\Big)\le C(1+E\|\xi\|_\infty^{p_1\gamma_1}+E\|\xi\|_\infty^{p_1\gamma_2})\le C, \tag{5.6}$$

which leads to

$$\mathbb{E}\Big(\sup_{-\tau\le s\le\tau}|X_s^\varepsilon|^{p_1}\Big)\le\mathbb{E}\Big(\sup_{-\tau\le s\le 0}|X_s^\varepsilon|^{p_1}\Big)+\mathbb{E}\Big(\sup_{0\le s\le\tau}|X_s^\varepsilon|^{p_1}\Big)\le C.$$

It follows from (5.5), (5.6) and the Hölder inequality that

$$\begin{aligned}\mathbb{E}\Big(\sup_{-\tau\le s\le 2\tau}|X_s^\varepsilon|^{p_2}\Big)&\le C\Big\{1+\int_0^{2\tau}(\mathbb{E}|X_{s-\tau}^\varepsilon|^{p_2\gamma_1}+\mathbb{E}|X_{s-\tau}^\varepsilon|^{p_2\gamma_2})\mathrm{d}s\Big\}\\&\le C\Big\{1+\int_0^{\tau}\Big((\mathbb{E}|X_s^\varepsilon|^{p_1})^{p_2\gamma_1/p_1}+(\mathbb{E}|X_s^\varepsilon|^{p_1})^{p_2\gamma_2/p_1}\Big)\mathrm{d}s\Big\}\\&\le C<\infty.\end{aligned}$$

Carrying out the previous procedures gives (3.1). □

Remark 5.1. Obviously, $\mathbb{E}\Big(\sup_{-\tau\le t\le T}|X_t^0|^p\Big)\le C<\infty (p\ge 2)$ if the coefficients may be polynomial of any degree $r\ge 1$ with respect to the delay variables.

**5.2. Proof of Theorem 3.2.** Next, to derive the consistency of LSE, we display some auxiliary results in the form of lemmas.

Lemma 5.1. *Assume* (**A1**) *and* (**A2**). *Then stochastic interacting particle system* (2.4) *has a strong solution with*

$$\sup_{i\in\mathbf{S}_N}\mathbb{E}\Big(\sup_{-\tau\le t\le T}|X_t^{\varepsilon,i,N}|^p\Big)\le C<\infty,\qquad p\ge 2.$$

*Proof.* For $x:=(x_1,x_2,\cdots,x_N)\in\mathbb{R}^d\otimes\mathbb{R}^N$, $y:=(y_1,y_2,\cdots,y_N)\in\mathbb{R}^d\otimes\mathbb{R}^N$, $x_i,y_i\in\mathbb{R}^d$, $i=1,2,\cdots,N$, define

$$\mu_x^N=\frac{1}{N}\sum_{i=1}^N\delta_{x_i},\qquad \mu_y^N=\frac{1}{N}\sum_{i=1}^N\delta_{y_i},$$

$$\tilde b(x,y)=(b(x_1,y_1,\mu_x^N,\mu_y^N,\theta),\cdots,b(x_N,y_N,\mu_x^N,\mu_y^N,\theta))^*,$$

$$\tilde\sigma(x,y)=\mathrm{diag}\Big(\sigma(x_1,y_1,\mu_x^N,\mu_y^N),\cdots,\sigma(x_N,y_N,\mu_x^N,\mu_y^N)\Big)$$

and

$$\tilde W_t=\Big(W_t^1,\cdots,W_t^N\Big)^*.$$

Then, (2.4) can be redescribed as

$$\mathrm{d}X_t=\tilde b(X_t,X_{t-\tau})\mathrm{d}t+\varepsilon\tilde\sigma(X_t,X_{t-\tau})\mathrm{d}\tilde W_t,\quad t\ge 0. \tag{5.7}$$

Note that for any $x,y\in\mathbb{R}^d\otimes\mathbb{R}^N$

$$\mathbb{W}_2(\mu_x^N,\mu_y^N)\le\Big(\frac{1}{N}\sum_{j=1}^N|x_j-y_j|^2\Big)^{\frac12}.$$

This, together with the assumptions (**A1**) and (**A2**), for any $x,x',y,y' \in \mathbb{R}^d \otimes \mathbb{R}^N$, leads to

$$|\tilde{b}(x,y)-\tilde{b}(x',y')| \leq C|x-x'|+K_1(y_1,y_2)|y-y'|$$

and

$$|\tilde{\sigma}(x,y)-\tilde{\sigma}(x',y')| \leq C|x-x'|+K_2(y_1,y_2)|y-y'|.$$

Then by [2, Lemma 2.2], it can be readily seen that (5.7) admits a unique global strong solution with

$$\mathbb{E}\Big(\sup_{-\tau\leq t\leq T}|X_t|^p\Big)\leq C, \quad p\geq 2.$$

Consequently, we conclude that (2.4) has a unique strong solution with

$$\sup_{i\in\mathbf{S}^N}\mathbb{E}\Big(\sup_{-\tau\leq t\leq T}|X_t^{\varepsilon,i,N}|^p\Big)\leq C<\infty, \qquad p\geq 2.$$

□

REMARK 5.2. In [1, Lemma 3.1], Bao and Huang have investigated the question on the wellposedness of the stochastic $N$-interacting particle systems associated with McKean-Vlasov SDEs. We extend the idea used in [1] to the case of delay McKean-Vlasov SDEs.

LEMMA 5.2. *Let (**A1**), (**A2**) and (**B1**) hold. Then, for initial value $X_0^\varepsilon=\xi\in L^0_{p_1q_1}(\mathscr{C})$, $p\geq 2$,*

$$\sup_{i\in\mathbf{S}_N}\mathbb{E}(\sup_{0\leq t\leq T}|X_t^{\varepsilon,i,N}-X_t^{\varepsilon,i}|^p)\leq C(C_N+C_N^{\frac{1}{2}}), \tag{5.8}$$

*where $C_N$ is a decreasing function with respect to $N$ and is defined as (5.16).*

*Proof.* For fixed $\lambda>1$ and arbitrary $\epsilon\in(0,1)$, there exists a continuous nonnegative function $\psi_{\lambda\epsilon}$, $x\geq 0$, with support $[\epsilon/\lambda,\epsilon]$ such that

$$\int_{\epsilon/\lambda}^{\epsilon}\psi_{\lambda\epsilon}(x)\mathrm{d}x=1 \qquad \text{and} \qquad 0\leq\psi_{\lambda\epsilon}(x)\leq\frac{2}{x\ln\lambda}, \quad x>0.$$

Let

$$\phi_{\lambda\epsilon}(x)=\int_0^x\int_0^y\psi_{\lambda\epsilon}(z)\mathrm{d}z\mathrm{d}y, \quad x>0.$$

Then $\phi_{\lambda\epsilon}$ is $C^2$ and satisfies

$$x-\epsilon\leq\phi_{\lambda\epsilon}(x)\leq x, \quad x>0 \tag{5.9}$$

and

$$0\leq\phi'_{\lambda\epsilon}(x)\leq 1, \quad \phi''_{\lambda\epsilon}(x)\leq\frac{2}{x\ln\lambda}\mathbf{1}_{[\epsilon/\lambda,\epsilon]}(x), \quad x>0. \tag{5.10}$$

Define

$$V_{\lambda\epsilon}(x)=\phi_{\lambda\epsilon}(|x|), \qquad x\in\mathbb{R}^d. \tag{5.11}$$

Then, by the definition of $\phi_{\lambda\epsilon}$, it holds that $V_{\lambda\epsilon}\in C^2(\mathbb{R}^d;\mathbb{R}^+)$. For $x\in\mathbb{R}^d$, a direct calculation leads to

$$\frac{\partial V_{\lambda\epsilon}(x)}{\partial x_i}=\phi'_{\lambda\epsilon}(|x|)\frac{x_i}{|x|}$$

and

$$\frac{\partial^2 V_{\lambda\epsilon}(x)}{\partial x_i\partial x_j}=\phi'_{\lambda\epsilon}(|x|)(\delta_{ij}|x|^2-x_ix_j)|x|^{-3}+\phi''_{\lambda\epsilon}(|x|)x_ix_j|x|^{-2},\quad i,j=1,2,\cdots,d,$$

where $\delta_{ij}=1$ if $i=j$ or otherwise 0. Set

$$(V_{\lambda\epsilon})_x(x):=\Big(\frac{\partial V_{\lambda\epsilon}(x)}{\partial x_1},\cdots,\frac{\partial V_{\lambda\epsilon}(x)}{\partial x_d}\Big)\quad\text{and}\quad(V_{\lambda\epsilon})_{xx}(x):=\Big(\frac{\partial^2 V_{\lambda\epsilon}(x)}{\partial x_i\partial x_j}\Big)_{d\times d},\qquad x\in\mathbb{R}^d.$$

According to (5.10) and (5.11), it holds

$$0\le|(V_{\lambda\epsilon})_x(x)|\le 1\quad\text{and}\quad\|(V_{\lambda\epsilon})_{xx}(x)\|\le 2d\Big(1+\frac{1}{\ln\lambda}\Big)\frac{1}{|x|}\mathbf{1}_{[\epsilon/\lambda,\epsilon]}(|x|),\quad x\in\mathbb{R}^d.\quad(5.12)$$

Set $Z^{i,N}_\cdot:=X^{\varepsilon,i,N}_\cdot-X^{\varepsilon,i}_\cdot$ and $\bar{Z}^{i,N}_\cdot:=(X^{\varepsilon,i,N}_\cdot,X^{\varepsilon,i}_\cdot)\in\mathbb{R}^{2d}$. For any $t\in[0,T]$, by Itô's formula, one gets

$$\begin{aligned}V_{\lambda\epsilon}(Z^{i,N}_t)=&\int_0^t\langle(V_{\lambda\epsilon})_x(Z^{i,N}_s),b(X^{\varepsilon,i,N}_s,X^{\varepsilon,i,N}_{s-\tau},\mu^{\varepsilon,N}_s,\mu^{\varepsilon,N}_{s-\tau},\theta)-b(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon}_s,\mu^{\varepsilon}_{s-\tau},\theta)\rangle\mathrm{d}s\\&+\frac{\varepsilon^2}{2}\int_0^t\mathrm{trace}\{(\sigma(X^{\varepsilon,i,N}_s,X^{\varepsilon,i,N}_{s-\tau},\mu^{\varepsilon,N}_s,\mu^{\varepsilon,N}_{s-\tau})-\sigma(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon}_s,\mu^{\varepsilon}_{s-\tau}))^*\\&\times(V_{\lambda\epsilon})_{xx}(Z^{i,N}_s)(\sigma(X^{\varepsilon,i,N}_s,X^{\varepsilon,i,N}_{s-\tau},\mu^{\varepsilon,N}_s,\mu^{\varepsilon,N}_{s-\tau})-\sigma(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon}_s,\mu^{\varepsilon}_{s-\tau}))\}\mathrm{d}s\\&+\varepsilon\int_0^t\langle(V_{\lambda\epsilon})_x(Z^{i,N}_s),\sigma(X^{\varepsilon,i,N}_s,X^{\varepsilon,i,N}_{s-\tau},\mu^{\varepsilon,N}_s,\mu^{\varepsilon,N}_{s-\tau})-\sigma(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon}_s,\mu^{\varepsilon}_{s-\tau})\mathrm{d}W^i_s\rangle\\=:&\sum_{i=1}^3 Q_i(t).\end{aligned}$$

By means of the assumption (**A1**), (5.12) and Hölder's inequality, we derive that, for any $t\in[0,T]$

$$\begin{aligned}&\mathbb{E}\Big(\sup_{0\le s\le t}|Q_1(s)|^p\Big)\\\le&\int_0^t\mathbb{E}|b(X^{\varepsilon,i,N}_s,X^{\varepsilon,i,N}_{s-\tau},\mu^{\varepsilon,N}_s,\mu^{\varepsilon,N}_{s-\tau},\theta)-b(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon}_s,\mu^{\varepsilon}_{s-\tau},\theta)|^p\mathrm{d}s\\\le&C\int_0^t\Big\{\mathbb{E}|Z^{i,N}_s|^p+\Big(\mathbb{E}K_1^{2p}(\bar{Z}^{i,N}_{s-\tau})\Big)^{\frac12}\Big(\mathbb{E}|Z^{i,N}_{s-\tau}|^{2p}\Big)^{\frac12}\\&+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_s,\mu^{\varepsilon}_s)^p+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{s-\tau},\mu^{\varepsilon}_{s-\tau})^p\Big\}\mathrm{d}s.\end{aligned}\qquad(5.13)$$

By means of the assumption (**A2**), (5.12) and Hölder's inequality, it holds that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|Q_2(s)|^p\Big)\le C\varepsilon^{2p}\mathbb{E}\int_0^t\frac{1}{|Z^{i,N}_s|^p}\Big\{&|Z^{i,N}_s|^{2p}+K_2^{2p}(\bar{Z}^{i,N}_{s-\tau})|Z^{i,N}_{s-\tau}|^{2p}+\mathbb{W}_2(\mu^{\varepsilon,N}_s,\mu^{\varepsilon}_s)^{2p}\\&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s-\tau},\mu^{\varepsilon}_{s-\tau})^{2p}\Big\}\mathbf{1}_{[\epsilon/\lambda,\epsilon]}(|Z^{i,N}_s|)\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&\le C\varepsilon^{2p}\int_0^t \Big\{\mathbb{E}|Z_s^{i,N}|^p+\frac{1}{\epsilon^p}\Big(\mathbb{E}K_2^{4p}(\bar{Z}_{s-\tau}^{i,N})\Big)^{\frac12}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{4p}\Big)^{\frac12}\\
&\quad+\frac{1}{\epsilon^p}\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^{2p}+\mathbb{E}|Z_s^{i,N}|^{-p}\mathbb{W}_2(\mu_s^{\varepsilon,N},\mu_s^{\varepsilon})^{2p}\mathbf{1}_{[\epsilon/\lambda,\epsilon]}(|Z_s^{i,N}|)\Big\}\mathrm{d}s\\
&\le C\varepsilon^{2p}\int_0^T \Big\{\mathbb{E}|Z_s^{i,N}|^p+\frac{1}{\epsilon^p}\Big(\mathbb{E}K_2^{4p}(\bar{Z}_{s-\tau}^{i,N})\Big)^{\frac12}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{4p}\Big)^{\frac12}\\
&\quad+\frac{1}{\epsilon^p}\Big(\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,N},\mu_s^{\varepsilon})^{2p}+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^{2p}\Big)\Big\}\mathrm{d}s. \qquad (5.14)
\end{aligned}
$$

By virtue of the assumption (**A2**), Burkhold-Davis-Gundy's inequality, Hölder's inequality and Young's inequality, we derive that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}|Q_3(s)|^p\Big)&\le C\varepsilon^p\int_0^t \Big\{\mathbb{E}|Z_s^{i,N}|^p+\Big(\mathbb{E}K_2^{2p}(\bar{Z}_{s-\tau}^{i,N})\Big)^{\frac12}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{2p}\Big)^{\frac12}\\
&\quad+\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,N},\mu_s^{\varepsilon})^p+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^p\Big\}\mathrm{d}s. \qquad (5.15)
\end{aligned}
$$

In addition, it follows from (2.10), (3.1) and Lemma 5.1 that

$$
\mathbb{E}K_1^{2p}(\bar{Z}_{s-\tau}^{i,N})+\mathbb{E}K_2^{4p}(\bar{Z}_{s-\tau}^{i,N})\le C.
$$

This, together with (5.13), (5.14) and (5.15), it holds from (5.9) that, for any $t\in[0,T]$ and $p\ge 2$

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}|Z_s^{i,N}|^p\Big)&\le 2^{p-1}\Big\{\epsilon^p+\mathbb{E}\Big(\sup_{0\le s\le t}V_{\lambda\epsilon}^p(Z_s^{i,N})\Big)\Big\}\\
&\le C\Big\{\epsilon^p+\int_0^t\Big\{\mathbb{E}|Z_s^{i,N}|^p+\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{2p}\Big)^{\frac12}+\frac{1}{\epsilon^p}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{4p}\Big)^{\frac12}\\
&\quad+\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,N},\mu_s^{\varepsilon})^p+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^p\\
&\quad+\frac{1}{\epsilon^p}\Big(\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,N},\mu_s^{\varepsilon})^{2p}+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^{2p}\Big)\Big\}\mathrm{d}s\\
&\le C\Big\{\epsilon^p+\int_0^t\Big\{\mathbb{E}|Z_s^{i,N}|^p+\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{2p}\Big)^{\frac12}+\frac{1}{\epsilon^p}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{4p}\Big)^{\frac12}\\
&\quad+\mathbb{E}\mathbb{W}_p(\tilde{\mu}_s^{\varepsilon,N},\mu_s^{\varepsilon})^p+\mathbb{E}\mathbb{W}_p(\tilde{\mu}_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^p\\
&\quad+\frac{1}{\epsilon^p}\Big(\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,N},\tilde{\mu}_{s-\tau}^{\varepsilon,N})^{2p}+\mathbb{E}\mathbb{W}_{2p}(\tilde{\mu}_{s-\tau}^{\varepsilon,N},\mu_{s-\tau}^{\varepsilon})^{2p}\\
&\quad+\mathbb{E}\mathbb{W}_{2p}(\tilde{\mu}_s^{\varepsilon,N},\mu_s^{\varepsilon})^{2p}\Big)\Big\}\mathrm{d}s,
\end{aligned}
$$

where, in the third inequality, we used the Hölder inequality and the fact that

$$
\mathbb{E}\mathbb{W}_2(\mu_\cdot^{\varepsilon,N},\mu_\cdot^{\varepsilon})^p\le\mathbb{E}\mathbb{W}_2(\mu_\cdot^{\varepsilon,N},\tilde{\mu}_\cdot^{\varepsilon,N})^p+\mathbb{E}\mathbb{W}_2(\tilde{\mu}_\cdot^{\varepsilon,N},\mu_\cdot^{\varepsilon})^p\le\mathbb{E}|Z_\cdot^{i,N}|^p+\mathbb{E}\mathbb{W}_2(\tilde{\mu}_\cdot^{\varepsilon,N},\mu_\cdot^{\varepsilon})^p
$$

since $(Z_\cdot^{i,N})_{1\le j\le N}$ are identically distributed. Moreover, according to [8, Theorem 1] and the assumption (**B1**), it holds that

$$
\mathbb{E}\mathbb{W}_p(\tilde{\mu}_\cdot^{\varepsilon,N},\mu_\cdot^{\varepsilon})^p\le\tilde{C}\begin{cases} N^{-1/2}+N^{\frac{p}{q}-1}, & \text{if } p>\dfrac{d}{2},\ q\ne 2p,\\ N^{-1/2}\log(1+N)+N^{\frac{p}{q}-1}, & \text{if } p=\dfrac{d}{2},\ q\ne 2p,\\ N^{-p/d}+N^{\frac{p}{q}-1}, & \text{if } p\in(0,\dfrac{d}{2}),\ q\ne\dfrac{d}{d-p}.\end{cases}
$$

$$=:C_N. \tag{5.16}$$

Thus, it follows from the Gronwall inequality that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|Z_s^{i,N}|^p\Big)\le C\Big\{\epsilon^p+\int_0^t\Big\{\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{2p}\Big)^{\frac{1}{2}}+\frac{1}{\epsilon^p}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{4p}\Big)^{\frac{1}{2}}+C_N+\frac{1}{\epsilon^p}C_N\Big\}\mathrm{d}s\Big\}.$$

Set, for any $p\ge 2$,

$$p_i=(\lfloor T/\tau\rfloor+2-i)p4^{\lfloor T/\tau\rfloor+1-i},\quad i=1,2,\cdots,\lfloor T/\tau\rfloor+1.$$

Then it is easy to see that

$$p_i\ge 2,\quad 4p_{i+1}<p_i\quad\text{and}\quad p_{\lfloor T/\tau\rfloor+1}=p,\quad i=1,2,\cdots,\lfloor T/\tau\rfloor. \tag{5.17}$$

For $s\in[0,\tau]$, $Z_{s-\tau}^{i,N}=0$, which, and taking $\epsilon=C_N^{\frac{1}{2p_1}}$ implies that

$$\mathbb{E}\Big(\sup_{0\le s\le\tau}|Z_s^{i,N}|^{p_1}\Big)\le C(C_N+C_N^{\frac{1}{2}}).$$

This fact, together with (5.17) and the Hölder inequality, implies

$$\begin{aligned}&\mathbb{E}\Big(\sup_{0\le s\le 2\tau}|Z_s^{i,N}|^{p_2}\Big)\\ &\le C\Big\{\epsilon^{p_2}+\int_0^{2\tau}\Big\{\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{p_1}\Big)^{\frac{p_2}{p_1}}+\frac{1}{\epsilon^{p_2}}\Big(\mathbb{E}|Z_{s-\tau}^{i,N}|^{p_1}\Big)^{\frac{2p_2}{p_1}}+C_N+\frac{1}{\epsilon^{p_2}}C_N\Big\}\mathrm{d}s\Big\}\\ &\le C(C_N+C_N^{\frac{1}{2}}),\end{aligned}$$

by setting $\epsilon=C_N^{\frac{1}{2p_2}}$. Repeating the previous procedures gives the desired assertion (5.8). □

REMARK 5.3. In terms of Lemma 5.2, it is desirable to measure the convergence between stochastic interacting particle systems and the corresponding non-interacting particle systems in the sense of the $p$-moment. This result plays an important role in the process of establishing the consistency of the LSE.

LEMMA 5.3. *Let* (**A1**), (**A2**) *and* (**B1**) *hold. Then, for initial value $X_0^\varepsilon=\xi\in L^0_{p_1q_1}(\mathscr{C})$, $p\ge 2$, there is $C>0$ such that*

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_{t_\delta}^{\varepsilon,i}-X_t^{0,i}|^p\Big)\le C\delta(\delta^{p-1}+\varepsilon^p)+C\varepsilon^p,\quad i\in\mathbf{S}_N, \tag{5.18}$$

*where* $t_\delta:=\lfloor t/\delta\rfloor\delta$.

*Proof.* For any $t\in[0,T]$,

$$|X_{t_\delta}^{\varepsilon,i}-X_t^{0,i}|^p\le 2^{p-1}|X_{t_\delta}^{\varepsilon,i}-X_t^{\varepsilon,i}|^p+2^{p-1}|X_t^{\varepsilon,i}-X_t^{0,i}|^p. \tag{5.19}$$

Now, for any $t\in[0,T]$, there exists an integer $k_0\in[0,n-1]$ such that $t\in[k_0\delta,(k_0+1)\delta]$. Obviously, $k_0=[t/\delta]$. Next, the Hölder inequality and Burkhold-Davis-Gundy inequality, together with (**A1**) and (**A2**), yield that

$$\mathbb{E}\Big(\sup_{0\le t\le T}|X_{t_\delta}^{\varepsilon,i}-X_t^{\varepsilon,i}|^p\Big)=\mathbb{E}\Big(\sup_{0\le t\le T}|X_{k_0\delta}^{\varepsilon,i}-X_t^{\varepsilon,i}|^p\Big)$$

$$
\begin{aligned}
&\le 2^{p-1}\delta^{p-1}\mathbb{E}\int_{k_0\delta}^{T}|b(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i},\theta)|^p\mathrm{d}s+2^{p-1}\varepsilon^pC_{p,T}\\
&\quad\times\mathbb{E}\int_{k_0\delta}^{T}\|\sigma(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i})\|^p\mathrm{d}s\\
&\le C(2\delta)^{p-1}\int_{k_0\delta}^{T}\Big\{1+\mathbb{E}|X_s^{\varepsilon,i}|^p+\mathbb{E}|X_{s-\tau}^{\varepsilon,i}|^p+\mathbb{E}|X_{s-\tau}^{\varepsilon,i}|^{p(r_1+1)}+\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,i},\delta_0)^p\\
&\quad+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,i},\delta_0)^p\Big\}\mathrm{d}s+C\varepsilon^p\int_{k_0\delta}^{T}\Big\{1+\mathbb{E}|X_s^{\varepsilon,i}|^p+\mathbb{E}|X_{s-\tau}^{\varepsilon,i}|^p+\mathbb{E}|X_{s-\tau}^{\varepsilon,i}|^{p(r_2+1)}\\
&\quad+\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,i},\delta_0)^p+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,i},\delta_0)^p\Big\}\mathrm{d}s\\
&\le C(\delta^{p-1}+\varepsilon^p)\int_{k_0\delta}^{T}\Big\{1+\mathbb{E}|X_s^{\varepsilon,i}|^p+\mathbb{E}|X_{s-\tau}^{\varepsilon,i}|^{pq_1}\Big\}\mathrm{d}s\\
&\le C\delta(\delta^{p-1}+\varepsilon^p). \qquad (5.20)
\end{aligned}
$$

Moreover, set $Z_t^i := X_t^{\varepsilon,i}-X_t^{0,i}$ and $\bar{Z}_\cdot^i := (X_\cdot^{\varepsilon,i},X_\cdot^{0,i})\in\mathbb{R}^{2d}$. Applying the Itô formula implies that

$$
\begin{aligned}
&V_{\lambda\epsilon}(Z_t^i)\\
&=\int_0^t\langle (V_{\lambda\epsilon})_x(Z_s^i), b(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i},\theta)-b(X_s^{0,i},X_{s-\tau}^{0,i},\mu_s^0,\mu_{s-\tau}^0,\theta)\rangle\mathrm{d}s\\
&\quad+\frac{\varepsilon^2}{2}\int_0^t\mathrm{trace}\{(\sigma(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i}))^*(V_{\lambda\epsilon})_{xx}(Z_s^i)\sigma(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i})\}\mathrm{d}s\\
&\quad+\varepsilon\int_0^t\langle (V_{\lambda\epsilon})_x(Z_s^i),\sigma(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon},\mu_{s-\tau}^{\varepsilon})\mathrm{d}W_s^i\rangle\\
&=:\sum_{i=1}^3\bar{Q}_i(t).
\end{aligned}
$$

By means of the assumption (**A1**), (5.12) and Hölder's inequality, we derive that, for any $t\in[0,T]$

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Q}_1(s)|^p\Big)&\le\int_0^t\mathbb{E}|b(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i},\theta)-b(X_s^{0,i},X_{s-\tau}^{0,i},\mu_s^0,\mu_{s-\tau}^0,\theta)|^p\mathrm{d}s\\
&\le C\int_0^t\Big\{\mathbb{E}|Z_s^i|^p+\Big(\mathbb{E}K_1^{2p}(\bar{Z}_{s-\tau}^i)\Big)^{\frac12}\Big(\mathbb{E}|Z_{s-\tau}^i|^{2p}\Big)^{\frac12}+\mathbb{E}\mathbb{W}_2(\mu_s^{\varepsilon,i},\mu_s^0)^p\\
&\quad+\mathbb{E}\mathbb{W}_2(\mu_{s-\tau}^{\varepsilon,i},\mu_{s-\tau}^0)^p\Big\}\mathrm{d}s\\
&\le C\int_0^t\Big\{\mathbb{E}|Z_s^i|^p+\mathbb{E}|Z_{s-\tau}^i|^p+\Big(\mathbb{E}|Z_{s-\tau}^i|^{2p}\Big)^{\frac12}\Big\}\mathrm{d}s. \qquad (5.21)
\end{aligned}
$$

By means of the assumption (**A2**), the elementary inequality, the Hölder inequality and (5.12), it holds that

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{0\le s\le t}|\bar{Q}_2(s)|^p\Big)\le\frac{\varepsilon^{2p}}{2}\mathbb{E}\int_0^t\{\|(V_{\lambda\epsilon})_{xx}(Z_s^i)\|\|\sigma(X_s^{\varepsilon,i},X_{s-\tau}^{\varepsilon,i},\mu_s^{\varepsilon,i},\mu_{s-\tau}^{\varepsilon,i})\|^2\}^p\mathrm{d}s\\
&\le C\varepsilon^{2p}\mathbb{E}\int_0^t\frac{1}{|Z_s^i|^p}\Big\{1+|Z_s^i|^{2p}+|Z_{s-\tau}^i|^{2p}+|Z_{s-\tau}^i|^{2p(r_2+1)}+\mathbb{W}_2(\mu_s^{\varepsilon,i},\mu_s^0)^{2p}
\end{aligned}
$$

$$+\mathbb{W}_2(\mu^{\varepsilon,i}_{s-\tau},\mu^0_{s-\tau})^{2p}\Big\}\mathbf{1}_{[\epsilon/\lambda,\epsilon]}(|Z^i_s|)\mathrm{d}s$$
$$\leq C\varepsilon^{2p}\int_0^t\Big\{\epsilon^p+\mathbb{E}|Z^i_s|^p+\frac{1}{\epsilon^p}\Big(1+\mathbb{E}|Z^i_{s-\tau}|^{2p}+\mathbb{E}|Z^i_{s-\tau}|^{2p(r_2+1)}\Big)\Big\}\mathrm{d}s. \tag{5.22}$$

By virtue of the assumption (**A2**), Burkhold-Davis-Gundy's inequality, Hölder's inequality and the elementary inequality, we derive that

$$\mathbb{E}\Big(\sup_{0\leq s\leq t}|\bar{Q}_3(s)|^p\Big)\leq C\varepsilon^p\mathbb{E}\int_0^t\|\sigma(X^{\varepsilon,i}_s,X^{\varepsilon,i}_{s-\tau},\mu^{\varepsilon,i}_s,\mu^{\varepsilon,i}_{s-\tau})\|^p\mathrm{d}s$$
$$\leq C\varepsilon^p\mathbb{E}\int_0^t\Big\{1+|X^{\varepsilon,i}_s|^p+|X^{\varepsilon,i}_{s-\tau}|^p+|X^{\varepsilon,i}_{s-\tau}|^{p(r_2+1)}+\mathbb{W}_2(\mu^{\varepsilon,i}_s,\mu^0_s)^p+\mathbb{W}_2(\mu^{\varepsilon,i}_{s-\tau},\mu^0_{s-\tau})^p\Big\}\mathrm{d}s$$
$$\leq C\varepsilon^p\int_0^t\Big\{1+\mathbb{E}|Z^i_s|^p+\mathbb{E}|Z^i_{s-\tau}|^p+\mathbb{E}|Z^i_{s-\tau}|^{p(r_2+1)}\Big\}\mathrm{d}s. \tag{5.23}$$

Furthermore, in view of (5.21), (5.22) and (5.23), we derive that, for any $t\in[0,T]$ and $p\geq 2$,

$$\mathbb{E}\Big(\sup_{0\leq s\leq t}|Z^i_s|^p\Big)\leq 2^{p-1}\Big\{\epsilon^p+\mathbb{E}\Big(\sup_{0\leq s\leq t}V^p_{\lambda\epsilon}(Z^i_s)\Big)\Big\}$$
$$\leq C\Big\{\epsilon^p+\int_0^t\Big\{\mathbb{E}|Z^i_s|^p+\mathbb{E}|Z^i_{s-\tau}|^p+\Big(\mathbb{E}|Z^i_{s-\tau}|^{2p}\Big)^{\frac{1}{2}}+\epsilon^p\varepsilon^{2p}$$
$$+\frac{\varepsilon^{2p}}{\epsilon^p}\Big(1+\mathbb{E}|Z^i_{s-\tau}|^{2p}+\mathbb{E}|Z^i_{s-\tau}|^{2p(r_2+1)}\Big)+\varepsilon^p+\varepsilon^p\mathbb{E}|Z^i_{s-\tau}|^{p(r_2+1)}\Big\}\mathrm{d}s.$$

Then, the Gronwall inequality implies that

$$\mathbb{E}\Big(\sup_{0\leq s\leq t}|Z^i_s|^p\Big)\leq C\Big\{\epsilon^p+\int_0^t\Big\{\epsilon^p\varepsilon^{2p}+\varepsilon^p+\mathbb{E}|Z^i_{s-\tau}|^p+\Big(\mathbb{E}|Z^i_{s-\tau}|^{2p}\Big)^{\frac{1}{2}}+\frac{\varepsilon^{2p}}{\epsilon^p}\Big(1+\mathbb{E}|Z^i_{s-\tau}|^{2p}$$
$$+\mathbb{E}|Z^i_{s-\tau}|^{2p(r_2+1)}\Big)+\varepsilon^p\mathbb{E}|Z^i_{s-\tau}|^{p(r_2+1)}\Big\}\mathrm{d}s. \tag{5.24}$$

Set, for any $p\geq 2$,

$$p_i=(\lfloor T/\tau\rfloor+2-i)p(2r_2+2)^{\lfloor T/\tau\rfloor+1-i},\quad i=1,2,\cdots,\lfloor T/\tau\rfloor+1.$$

Then it is easy to see that

$$p_i\geq 2,\quad 2(r_2+1)p_{i+1}<p_i\quad\text{and}\quad p_{\lfloor T/\tau\rfloor+1}=p,\quad i=1,2,\cdots,\lfloor T/\tau\rfloor. \tag{5.25}$$

For $s\in[0,\tau]$, $Z^i_{s-\tau}=0$, which, and taking $\epsilon=\varepsilon$ implies that

$$\mathbb{E}\Big(\sup_{0\leq s\leq\tau}|Z^i_s|^{p_1}\Big)\leq C\varepsilon^{p_1}.$$

This fact together with (5.24), (5.25) and the Hölder inequality implies, by setting $\epsilon=\varepsilon$,

$$\mathbb{E}\Big(\sup_{0\leq s\leq 2\tau}|Z^i_s|^{p_2}\Big)\leq C\Big\{\epsilon^{p_2}+\int_0^{2\tau}\Big\{\epsilon^{p_2}\varepsilon^{2p_2}+\varepsilon^{p_2}+\mathbb{E}|Z^i_{s-\tau}|^{p_2}+\Big(\mathbb{E}|Z^i_{s-\tau}|^{2p_2}\Big)^{\frac{1}{2}}$$
$$+\frac{\varepsilon^{2p_2}}{\epsilon^{p_2}}\Big(1+\mathbb{E}|Z^i_{s-\tau}|^{2p_2}+\mathbb{E}|Z^i_{s-\tau}|^{2p_2(r_2+1)}\Big)+\varepsilon^{p_2}\mathbb{E}|Z^i_{s-\tau}|^{p_2(r_2+1)}\Big\}\mathrm{d}s$$

$$
\begin{aligned}
&\leq C\Big\{\epsilon^{p_2}+\int_0^{2\tau}\Big\{\epsilon^{p_2}\varepsilon^{2p_2}+\varepsilon^{p_2}+\Big(\mathbb{E}|Z^i_{s-\tau}|^{p_1}\Big)^{\frac{p_2}{p_1}}+\Big(\mathbb{E}|Z^i_{s-\tau}|^{p_1}\Big)^{\frac{p_2}{p_1}}\\
&\quad+\frac{\varepsilon^{2p_2}}{\epsilon^{p_2}}\Big(1+\Big(\mathbb{E}|Z^i_{s-\tau}|^{p_1}\Big)^{\frac{2p_2}{p_1}}+\Big(\mathbb{E}|Z^i_{s-\tau}|^{p_1}\Big)^{\frac{2p_2(r_2+1)}{p_1}}\Big)\\
&\quad+\varepsilon^{p_2}\Big(\mathbb{E}|Z^i_{s-\tau}|^{p_1}\Big)^{\frac{p_2(r_2+1)}{p_1}}\Big\}\mathrm{d}s\\
&\leq C\varepsilon^{p_2}.
\end{aligned}
$$

Following the previous procedures implies that

$$
\mathbb{E}\Big(\sup_{0\leq t\leq T}|Z^i_t|^p\Big)\leq C\varepsilon^p. \tag{5.26}
$$

Plugging (5.20) and (5.26) into (5.19) yields (5.18). ∎

LEMMA 5.4. *Let* (**A1**), (**A2**) *and* (**A3**) *hold. Then, for any initial value* $X^\varepsilon_0=\xi\in L^0_{p_1q_1}(\mathscr{C})$,

$$
\Phi^{i,N,(1)}_{n,\varepsilon}:=\sum_{k=1}^n B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})P^{\varepsilon,i,N}_k(\theta_0)\to 0,
$$

*in* $L^1$ *as* $\varepsilon\to 0$ *and* $n,N\to\infty$.

*Proof.* According to (2.5), we get

$$
\mathbb{W}_2(\mu^{\varepsilon,N}_s,\delta_0)^2\leq\frac{1}{N}\sum_{i=1}^N|X^{\varepsilon,i,N}_s|^2,\qquad s\geq-\tau. \tag{5.27}
$$

In view of (2.4) and (2.7), one has

$$
\begin{aligned}
\Phi^{i,N,(1)}_{n,\varepsilon}&=\varepsilon\sum_{k=1}^n B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\\
&\quad\times\sigma(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau})(W^i_{k\delta}-W^i_{(k-1)\delta})\\
&=\varepsilon\int_0^T B^*(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\\
&\quad\times\sigma(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\mu^{\varepsilon,N}_{s_\delta},\mu^{\varepsilon,N}_{s_\delta-\tau})\mathrm{d}W^i_s,
\end{aligned}
$$

where $s_\delta:=\lfloor s/\delta\rfloor\delta$. This, together with the Hölder inequality and [6, Theorem 7.1], further implies that

$$
\begin{aligned}
\mathbb{E}|\Phi^{i,N,(1)}_{n,\varepsilon}|&=\varepsilon\mathbb{E}\Big|\int_0^T B^*(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\\
&\quad\times\sigma(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\mu^{\varepsilon,N}_{s_\delta},\mu^{\varepsilon,N}_{s_\delta-\tau})\mathrm{d}W^i_s\Big|\\
&\leq C\varepsilon\Big(\mathbb{E}\int_0^T|B^*(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)|^2\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\|^2\\
&\quad\times\|\sigma(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\mu^{\varepsilon,N}_{s_\delta},\mu^{\varepsilon,N}_{s_\delta-\tau})\|^2\mathrm{d}s\Big)^{\frac{1}{2}}.
\end{aligned}\tag{5.28}
$$

One the other hand, for any $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}(\mathbb{R}^d)$, $i=1,2$, by the assumption (**A3**), it is easy to see that there is a constant $\bar{L}>0$ such that

$$\|(\sigma\sigma^*)^{-1}(x,y,\mu,\nu)\|\le\bar{L}\Big\{1+|x|+|y|+|y|^{r_3+1}+\mathbb{W}_2(\mu,\delta_0)+\mathbb{W}_2(\nu,\delta_0)\Big\}. \tag{5.29}$$

Now, it follows from (5.3) that

$$|B^*(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)|^2\le C\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{2(r_1+1)}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)^2+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)^2\Big\}. \tag{5.30}$$

Due to (5.4), we obtain

$$\|\sigma(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\mu^{\varepsilon,N}_{s_\delta},\mu^{\varepsilon,N}_{s_\delta-\tau})\|^2\le C\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{2(r_2+1)}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)^2+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)^2\Big\}. \tag{5.31}$$

From (5.29), one has

$$\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\|^2\le C\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{2(r_3+1)}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)^2+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)^2\Big\}. \tag{5.32}$$

Substituting these inequalities into (5.28), by the Hölder inequality and using (5.27) lead to

$$\begin{aligned}
&\mathbb{E}|\Phi^{i,N,(1)}_{n,\varepsilon}|\\
&\le C\varepsilon\Big\{\mathbb{E}\int_0^T\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^8+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^8+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{8q_1}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)^8+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)^8\\
&\quad+|X^{\varepsilon,i,N}_{s_\delta}|^4+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^4+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{4(r_3+1)}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)^4+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)^4\Big\}\mathrm{d}s\Big\}^{\frac12}\\
&\le C\varepsilon\Big\{\mathbb{E}\int_0^T\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^8+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^8+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{8q_1}+\Big(\frac1N\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{s_\delta}|^8\Big)\\
&\quad+\Big(\frac1N\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{s_\delta-\tau}|^8\Big)+|X^{\varepsilon,i,N}_{s_\delta}|^4+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^4+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{4(r_3+1)}+\Big(\frac1N\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{s_\delta}|^4\Big)\\
&\quad\quad+\Big(\frac1N\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{s_\delta-\tau}|^4\Big)\Big\}\mathrm{d}s\Big\}^{\frac12}\\
&\le C\varepsilon,
\end{aligned}$$

where the last step is due to Lemma 5.1. Hence, the desired result holds by taking $\varepsilon$ sufficiently small and $n,N$ sufficiently large. □

LEMMA 5.5. *Let* (**A1**)-(**A3**) *and* (**B1**) *hold. Then, for any initial value* $X^\varepsilon_0=\xi\in L^0_{p_1q_1}(\mathscr{C})$,

$$\delta\sum_{k=1}^nB^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})$$

$$\times B(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\to\Pi(\theta)$$

*in* $L^1$ *as* $\varepsilon\to0$, $N\to\infty$ *and* $\delta\to0$, *where* $\Pi(\theta)$ *is defined in* (3.2).

*Proof.* Obviously,

$$\begin{aligned}\Phi^{i,N,(2)}_{n,\varepsilon}(\theta):=&\delta\sum_{k=1}^{n}B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\\&\times B(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\\=&\int_0^T B^*(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\mathrm{d}s.\end{aligned}$$

Thus, by calculating directly, one has

$$\begin{aligned}&\Phi^{i,N,(2)}_{n,\varepsilon}(\theta)-\Pi(\theta)\\&=\int_0^T\Big\{B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)-B(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0,\theta)\Big\}^*\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\\&\quad\times B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\mathrm{d}s+\int_0^T B^*(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0,\theta)\Big\{\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\\&\quad-\Lambda^{-1}(X^{0,i}_s,X^{0,i}_{s-\tau})\Big\}B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)\mathrm{d}s+\int_0^T B^*(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0,\theta)\\&\quad\times\Lambda^{-1}(X^{0,i}_s,X^{0,i}_{s-\tau})\Big\{B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)-B(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0,\theta)\Big\}\mathrm{d}s\\&=:\sum_{i=1}^{3}J_i.\end{aligned}$$

In addition, for any $x_i,y_i\in\mathbb{R}^d$ and $\mu_{x_i},\mu_{y_i}\in\mathscr{P}_2(\mathbb{R}^d)$, $i=1,2$, it follows from (**A1**) that

$$\begin{aligned}|B(x_1,y_1,\theta_0,\theta)-B(x_2,y_2,\theta_0,\theta)|\le& C\{|x_1-x_2|+(1+|y_1|^{r_1}+|y_2|^{r_1})|y_1-y_2|\\&+\mathbb{W}_2(\mu_{x_1},\mu_{x_2})+\mathbb{W}_2(\mu_{y_1},\mu_{y_2})\}.\end{aligned}\tag{5.33}$$

This leads to

$$\begin{aligned}&|B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)-B(X^{0,i}_s,X^{0,i}_{s-\tau},\theta_0,\theta)|\\&\le C\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{r_1}+|X^{0,i}_{s-\tau}|^{r_1})|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s-\tau}|+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)\\&\quad+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\}.\end{aligned}\tag{5.34}$$

Set $q_2=(r_1+1)\vee(r_3+1)$. By (5.30) and (5.32), one has

$$\begin{aligned}&\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau})\|\cdot|B(X^{\varepsilon,i,N}_{s_\delta},X^{\varepsilon,i,N}_{s_\delta-\tau},\theta_0,\theta)|\\&\le C\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^2+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{2q_2}+\frac{1}{N}\sum_{j=1}^{N}|X^{\varepsilon,j,N}_{s_\delta}|^2+\frac{1}{N}\sum_{j=1}^{N}|X^{\varepsilon,j,N}_{s_\delta-\tau}|^2\Big\},\end{aligned}\tag{5.35}$$

where we have used (5.27). Then, the Hölder inequality implies that

$$\mathbb{E}|J_1|\le C\mathbb{E}\Big|\int_0^T\Big\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{r_1}+|X^{0,i}_{s-\tau}|^{r_1})|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s-\tau}|$$

$$
\begin{aligned}
&+\mathbb{W}_2(\mu_{s_\delta}^{\varepsilon,N},\mu_s^{0,i})+\mathbb{W}_2(\mu_{s_\delta-\tau}^{\varepsilon,N},\mu_{s-\tau}^{0,i})\Big\}\\
&\times\Big\{1+|X_{s_\delta}^{\varepsilon,i,N}|^2+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^2+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^{2q_2}+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta}^{\varepsilon,j,N}|^2+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta-\tau}^{\varepsilon,j,N}|^2\Big\}\mathrm{d}s\Big|\\
\le& C\mathbb{E}\Big|\int_0^T\Big\{|X_{s_\delta}^{\varepsilon,i,N}-X_s^{0,i}|\Big(1+|X_{s_\delta}^{\varepsilon,i,N}|^2+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^{2q_2}+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta}^{\varepsilon,j,N}|^2\\
&+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta-\tau}^{\varepsilon,j,N}|^2\Big)+|X_{s_\delta-\tau}^{\varepsilon,i,N}-X_{s-\tau}^{0,i}|\Big(1+|X_{s_\delta}^{\varepsilon,i,N}|^2+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^{2q_2}+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta}^{\varepsilon,j,N}|^2\\
&+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta-\tau}^{\varepsilon,j,N}|^2\Big)^2+\Big(\mathbb{W}_2(\mu_{s_\delta}^{\varepsilon,N},\mu_s^{0,i})+\mathbb{W}_2(\mu_{s_\delta-\tau}^{\varepsilon,N},\mu_{s-\tau}^{0,i})\Big)\\
&\times\Big(1+|X_{s_\delta}^{\varepsilon,i,N}|^2+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^{2q_2}+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta}^{\varepsilon,j,N}|^2+\frac{1}{N}\sum_{j=1}^{N}|X_{s_\delta-\tau}^{\varepsilon,j,N}|^2\Big)\Big\}\mathrm{d}s\Big|\\
\le& C\int_0^T\Big\{\sqrt{\mathbb{E}|X_{s_\delta}^{\varepsilon,i,N}-X_s^{0,i}|^2}+\Big(\mathbb{E}\mathbb{W}_2(\mu_{s_\delta}^{\varepsilon,N},\mu_s^{0,i})^2+\mathbb{E}\mathbb{W}_2(\mu_{s_\delta-\tau}^{\varepsilon,N},\mu_{s-\tau}^{0,i})^2\Big)^{\frac{1}{2}}\Big\}\mathrm{d}s\\
\le& C\int_0^T\Big\{(C_N+C_N^{\frac{1}{2}})+C\delta(\delta+\varepsilon^2)\Big\}^{\frac{1}{2}}\mathrm{d}s, \qquad (5.36)
\end{aligned}
$$

where, in the first step we used (5.34) and (5.35), and in the last step we utilized Lemmas 5.2 and 5.3. To obtain the estimate of $J_2$, we firstly seek some inequalities of the integrands. By (5.3), we find out

$$
\begin{aligned}
|B^*(X_s^{0,i},X_{s-\tau}^{0,i},\theta_0,\theta)|\le L\Big\{&1+|X_s^{0,i}|+|X_{s-\tau}^{0,i}|+|X_{s-\tau}^{0,i}|^{(r_1+1)}+\mathbb{W}_2(\mu_s^{0,i},\delta_0)\\
&+\mathbb{W}_2(\mu_{s-\tau}^{0,i},\delta_0)\Big\}. \qquad (5.37)
\end{aligned}
$$

By means of the assumption (**A3**), one gets

$$
\begin{aligned}
&\|\Lambda^{-1}(X_{s_\delta}^{\varepsilon,i,N},X_{s_\delta-\tau}^{\varepsilon,i,N})-\Lambda^{-1}(X_s^{0,i},X_{s-\tau}^{0,i})\|\\
\le& C_3\Big\{|X_{s_\delta}^{\varepsilon,i,N}-X_s^{0,i}|+(1+|X_{s_\delta-\tau}^{\varepsilon,i,N}|^{r_3}+|X_{s-\tau}^{0,i}|^{r_3})|X_{s_\delta-\tau}^{\varepsilon,i,N}-X_{s_\delta-\tau}^{0,i}|+\mathbb{W}_2(\mu_{s_\delta}^{\varepsilon,N},\mu_s^{0,i})\\
&+\mathbb{W}_2(\mu_{s_\delta-\tau}^{\varepsilon,N},\mu_{s-\tau}^{0,i})\Big\},
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\mathbb{W}_2(\mu_{s_\delta}^{\varepsilon,N},\mu_s^{0,i})^2+\mathbb{E}\mathbb{W}_2(\mu_{s_\delta-\tau}^{\varepsilon,N},\mu_{s-\tau}^{0,i})^2\\
\le&\mathbb{E}|X_{s_\delta}^{\varepsilon,i,N}-X_{s_\delta}^{\varepsilon,i}|^2+\mathbb{E}\mathbb{W}_2(\tilde{\mu}_{s_\delta}^{\varepsilon,N},\mu_{s_\delta}^{\varepsilon})^2+\mathbb{E}|X_{s_\delta-\tau}^{\varepsilon,i,N}-X_{s_\delta-\tau}^{\varepsilon,i}|^2+\mathbb{E}\mathbb{W}_2(\tilde{\mu}_{s_\delta-\tau}^{\varepsilon,N},\mu_{s_\delta-\tau}^{\varepsilon})^2\\
&+\mathbb{E}|X_{s_\delta}^{\varepsilon,i}-X_s^{0,i}|^2+\mathbb{E}|X_{s_\delta-\tau}^{\varepsilon,i}-X_{s-\tau}^{0,i}|^2
\end{aligned}
$$

and

$$
\mathbb{E}|X_{s_\delta}^{\varepsilon,i,N}-X_s^{0,i}|^2\le 2\mathbb{E}|X_{s_\delta}^{\varepsilon,i,N}-X_s^{\varepsilon,i}|^2+2\mathbb{E}|X_{s_\delta}^{\varepsilon,i}-X_s^{0,i}|^2. \qquad (5.38)
$$

In view of the results obtained above, we find out

$$
\begin{aligned}
&\mathbb{E}|J_2|\\
\le& C\mathbb{E}\Big|\int_0^T\Big\{1+|X_s^{0,i}|+|X_{s-\tau}^{0,i}|+|X_{s-\tau}^{0,i}|^{(r_1+1)}+\mathbb{W}_2(\mu_s^{0,i},\delta_0)+\mathbb{W}_2(\mu_{s-\tau}^{0,i},\delta_0)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&\times\Big\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{r_3}+|X^{0,i}_{s-\tau}|^{r_3})|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s_\delta-\tau}|+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\Big\}\times\Big\{1+|X^{\varepsilon,i,N}_{s_\delta}|+|X^{\varepsilon,i,N}_{s_\delta-\tau}|+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{(r_1+1)}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)\Big\}\mathrm{d}s\Big|\\
\le& C\mathbb{E}\int_0^T\Big|\Big\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|\Big(1+|X^{\varepsilon,i,N}_{s_\delta}|+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{q_2}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)\Big)\\
&+\Big(1+|X^{\varepsilon,i,N}_{s_\delta}|+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{q_1}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)\Big)^2|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s_\delta-\tau}|\\
&+\Big(\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\Big)\Big(1+|X^{\varepsilon,i,N}_{s_\delta}|+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{q_2}+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\delta_0)\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\delta_0)\Big)\Big\}\Big|\mathrm{d}s\\
\le& C\int_0^T\Big\{\sqrt{\mathbb{E}|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|^2}+\sqrt{\mathbb{E}|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s_\delta-\tau}|^2}+\Big(\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)^2\\
&+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})^2\Big)^{\frac12}\Big\}\mathrm{d}s\\
\le& C\int_0^T\Big\{(C_N+C_N^{\frac12})+C\delta(\delta+\varepsilon^2)\Big\}^{\frac12}\mathrm{d}s, 
\end{aligned}\tag{5.39}
$$

where, in the last step we have used the inequalities (5.8) and (5.18). Moreover, the inequality (5.29) leads to

$$
\|\Lambda^{-1}(X^{0,i}_s,X^{0,i}_{s-\tau})\|\le\bar L\Big\{1+|X^{0,i}_s|+|X^{0,i}_{s-\tau}|+|X^{0,i}_{s-\tau}|^{(r_3+1)}+\mathbb{W}_2(\mu^{0,i}_s,\delta_0)+\mathbb{W}_2(\mu^{0,i}_{s-\tau},\delta_0)\Big\},\tag{5.40}
$$

which, together with (5.34) and (5.37), further leads to

$$
\begin{aligned}
\mathbb{E}|J_3|\le& C\mathbb{E}\Big|\int_0^T\Big\{1+|X^{0,i}_s|+|X^{0,i}_{s-\tau}|+|X^{0,i}_{s-\tau}|^{q_2}+\mathbb{W}_2(\mu^{0,i}_s,\delta_0)+\mathbb{W}_2(\mu^{0,i}_{s-\tau},\delta_0)\Big\}^2\\
&\Big\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{r_1}+|X^{0,i}_{s-\tau}|^{r_1})|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s-\tau}|\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\Big\}\mathrm{d}s\Big|\\
\le& C\mathbb{E}\int_0^T\Big\{|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{r_1}+|X^{0,i}_{s-\tau}|^{r_1})|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s-\tau}|\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)+\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\Big\}\mathrm{d}s\\
\le& C\int_0^T\Big\{\mathbb{E}|X^{\varepsilon,i,N}_{s_\delta}-X^{0,i}_s|+(1+\mathbb{E}|X^{\varepsilon,i,N}_{s_\delta-\tau}|^{2r_1})^{\frac12}\sqrt{\mathbb{E}|X^{\varepsilon,i,N}_{s_\delta-\tau}-X^{0,i}_{s-\tau}|^2}\\
&+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta},\mu^{0,i}_s)+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{s_\delta-\tau},\mu^{0,i}_{s-\tau})\Big\}\mathrm{d}s\\
\le& C\int_0^T\Big\{(C_N+C_N^{\frac12}+\delta(\delta+\varepsilon^2))^{\frac12}+C_N+C_N^{\frac12}+\delta(\delta+\varepsilon^2)\Big\}\mathrm{d}s.
\end{aligned}\tag{5.41}
$$

Therefore, from (5.36), (5.39) and (5.41), we conclude that the desired result holds. □

*Proof.* **(Proof of Theorem 3.2.)**

$$
\Phi^{i,N}_{n,\varepsilon}(\theta)=\varepsilon^2(\Psi^{i,N}_{n,\varepsilon}(\theta)-\Psi^{i,N}_{n,\varepsilon}(\theta_0))
$$

$$
\begin{aligned}
=&\delta^{-1}\sum_{k=1}^{n}\Big\{\Big(P_k^{\varepsilon,i,N}(\theta_0)+b(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau},\theta_0)\delta\\
&-b(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau},\theta)\delta\Big)^*\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\\
&\times\Big(P_k^{\varepsilon,i,N}(\theta_0)+\delta(b(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau},\theta_0)\\
&-b(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau},\theta)\Big)-(P_k^{\varepsilon,i,N}(\theta_0))^*\\
&\times\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})P_k^{\varepsilon,i,N}(\theta_0)\Big\}\\
=&2\sum_{k=1}^{n}B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})P_k^{\varepsilon,i,N}(\theta_0)\\
&+\delta\sum_{k=1}^{n}B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\\
&\times B^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\theta_0,\theta)\\
=:&2\Phi^{i,N,(1)}_{n,\varepsilon}(\theta)+\Phi^{i,N,(2)}_{n,\varepsilon}(\theta).
\end{aligned}
$$

In view of Lemmas 5.4 and 5.5, together with the Chebyshev inequality, we deduce that

$$\sup_{\theta\in\Theta}|-\Phi^{i,N}_{n,\varepsilon}(\theta)-(-\Pi(\theta))|\to 0 \quad \text{in probability.} \tag{5.42}$$

According to (2.9), we find out $0=\Phi^{i,N}_{n,\varepsilon}(\theta_0)\geq\Phi^{i,N}_{n,\varepsilon}(\hat\theta^{i,N}_{n,\varepsilon})$, i.e., $0=-\Phi^{i,N}_{n,\varepsilon}(\theta_0)\leq -\Phi^{i,N}_{n,\varepsilon}(\hat\theta^{i,N}_{n,\varepsilon})$. In addition, due to $\Pi(\cdot)\geq 0$, we get

$$\sup_{|\theta-\theta_0|\geq\iota}(-\Pi(\theta))<-\Pi(\theta_0)=0, \qquad \text{for any } \ \iota>0. \tag{5.43}$$

In terms of [28, Theorem 5.9], and combining with (5.42) and (5.43), we deduce that $\hat\theta^{i,N}_{n,\varepsilon}\to\theta_0$ in probability as $N,n\to\infty$ and $\varepsilon\to 0$. ∎

**5.3. Proof of Theorem 3.3.** To make the deduction of the asymptotic distribution of the LSE $\hat\theta^{i,N}_{n,\varepsilon}$ clearer, we divide the proof of Theorem 3.3 into several auxiliary lemmas.

LEMMA 5.6. Assume that (**A1**)-(**A5**) and (**B1**) hold. Then, for $X_0^\varepsilon=\xi\in L^0_{p_1q_1}(\mathscr{C})$,

$$\int_0^T\Upsilon(X^{\varepsilon,i,N}_{t_\delta},X^{\varepsilon,i,N}_{t_\delta-\tau},\theta_0)\mathrm{d}W^i_t\longrightarrow\int_0^T\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)\mathrm{d}W^i_t \quad \mathbb{P}-\text{a.s.} \tag{5.44}$$

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

$$\varepsilon^{-1}(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\theta_0)\to -2\int_0^T\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)W^i_t \quad \mathbb{P}-\text{a.s.}$$

whenever $\varepsilon\to 0$ and $n,N\to\infty$.

*Proof.* In view of (**A4**), we see that, for any $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$, there exists a constant $C>0$ such that

$$\sup_{\theta\in\bar\Theta}\|(\nabla_\theta b)(x,y,\mu,\nu,\theta)\|\leq C\{1+|x|+|y|+|y|^{r_4+1}+\mathbb{W}_2(\mu,\delta_0)+\mathbb{W}_2(\nu,\delta_0)\}. \tag{5.45}$$

We first claim that

$$\int_0^T \|\Upsilon(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0)-\Upsilon(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0)\|^2\mathrm{d}t\to 0 \qquad \mathbb{P}-\text{a.s.} \tag{5.46}$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. According to (3.3), one gets

$$\begin{aligned}
&\|\Upsilon(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0)-\Upsilon(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0)\|^2\\
&\le 3\|\{(\nabla_\theta b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta_0)-(\nabla_\theta b)^*(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta_0)\}\\
&\quad\times\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\sigma(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N})\|^2\\
&\quad+3\|(\nabla_\theta b)^*(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta_0)\times\{\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\\
&\quad-\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})\}\sigma(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N})\|^2\\
&\quad+3\|(\nabla_\theta b)^*(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^0,\mu_{t-\tau}^0,\theta_0)\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})\\
&\quad\times\{\sigma(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N})-\sigma(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i})\}\|^2=:\sum_{k=1}^3 G_k.
\end{aligned}$$

For the first term $G_1$, from the assumption (**A4**) we first give the below result

$$\begin{aligned}
&\|(\nabla_\theta b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta_0)-(\nabla_\theta b)^*(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta_0)\|^2\\
&\le C\Big\{|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2+(1+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2r_4}+|X_{t-\tau}^{0,i}|^{2r_4})|X_{t_\delta-\tau}^{\varepsilon,i,N}-X_{t-\tau}^{0,i}|^2+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\mu_t^{0,i})^2\\
&\quad+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\mu_{t-\tau}^{0,i})^2\Big\}.
\end{aligned}$$

This, combining (5.31) with (5.32), leads to

$$\begin{aligned}
G_1&\le C\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2(r_2+1)}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^2+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^2\Big\}\\
&\quad\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2(r_3+1)}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^2+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^2\Big\}\\
&\quad\times\Big\{|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}-X_{t-\tau}^{0,i}|^2(1+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2r_4}+|X_{t-\tau}^{0,i}|^{2r_4})\\
&\quad+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\mu_t^{0,i})^2+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\mu_{t-\tau}^{0,i})^2\Big\}\\
&\le C|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^4+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{q_3}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^4+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^4\Big\}\\
&\quad+C|X_{t_\delta-\tau}^{\varepsilon,i,N}-X_{t-\tau}^{0,i}|^2\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^8+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2q_3}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^8+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^8\Big\}\\
&\quad+C\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\mu_t^{0,i})^2\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^4+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{q_3}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^4+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^4\Big\}\\
&\quad+C\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\mu_{t-\tau}^{0,i})^2\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^4+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{q_3}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^4+\mathbb{W}_2(\mu_{t_\delta-\tau}^{\varepsilon,N},\delta_0)^4\Big\}\\
&=:\sum_{k=1}^4\Sigma_k,
\end{aligned}\tag{5.47}$$

where $q_3=4((r_2+1)\vee(r_3+1))\vee(2r_4)$.

For any $\rho>0$ and $i\in\mathcal{S}_N$, in view of the Chebyshev inequality and (5.38), we arrive at

$$\mathbb{P}\Big(\int_0^T\|\Sigma_1\|\mathrm{d}t\ge\rho\Big)\le\mathbb{P}\Big(C\int_0^T|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2\Big\{1+|X_{t_\delta}^{\varepsilon,i,N}|^4+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{q_3}+\mathbb{W}_2(\mu_{t_\delta}^{\varepsilon,N},\delta_0)^4$$

$$
\begin{aligned}
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta-\tau},\delta_0)^4\Big\}\mathrm{d}t\geq\rho\Big)\\
\leq&\frac{C}{\rho}\int_0^T\Big(\mathbb{E}\Big\{1+|X^{\varepsilon,i,N}_{t_\delta}|^8+|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2q_3}+\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta},\delta_0)^8\\
&+\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta-\tau},\delta_0)^8\Big\}\Big)^{\frac{1}{2}}\Big(\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}-X^{0,i}_t|^4\Big)^{\frac{1}{2}}\mathrm{d}t\\
\leq&\frac{C}{\rho}\int_0^T\Big(1+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}|^8+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2q_3}+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta}|^8\\
&+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta-\tau}|^8\Big)^{\frac{1}{2}}\Big(\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}-X^{0,i}_t|^4\Big)^{\frac{1}{2}}\mathrm{d}t\longrightarrow 0
\end{aligned}\tag{5.48}
$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. Using the same idea like in the above, we get

$$
\mathbb{P}\Big(\int_0^T\|\Sigma_2\|\mathrm{d}t\geq\rho\Big)\longrightarrow 0,\tag{5.49}
$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. At the same time, it follows from the Hölder inequality that

$$
\begin{aligned}
\mathbb{E}\|\Sigma_3\|\leq&C\Big(\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta},\mu^{0,i}_t)^4\Big)^{\frac{1}{2}}\Big(1+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}|^8+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2q_3}+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta},\delta_0)^8\\
&+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta-\tau},\delta_0)^8\Big)^{\frac{1}{2}}\\
\leq&C\Big(\mathbb{E}\mathbb{W}_4(\mu^{\varepsilon,N}_{t_\delta},\tilde{\mu}^{\varepsilon,N}_{t_\delta})^4+\mathbb{E}\mathbb{W}_4(\tilde{\mu}^{\varepsilon,N}_{t_\delta},\mu^{\varepsilon}_{t_\delta})^4+\mathbb{E}\mathbb{W}_2(\mu^{\varepsilon}_{t_\delta},\mu^{0,i}_t)^4\Big)^{\frac{1}{2}}\Big(1+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}|^8\\
&+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2q_3}+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta}|^8+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta-\tau}|^8\Big)^{\frac{1}{2}}\\
\leq&C\Big(\mathbb{E}|Z^{i,N}_{t_\delta}|^4+C_N+\mathbb{E}|X^{\varepsilon,i}_{t_\delta}-X^{0,i}_t|^4\Big)^{\frac{1}{2}}\Big(1+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}|^8+\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2q_3}\\
&+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta}|^8+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta-\tau}|^8\Big)^{\frac{1}{2}}\\
\leq&C\Big(C_N+C_N^{\frac{1}{2}}+\delta(\delta^{p-1}+\varepsilon^p)+\varepsilon^p\Big)^{\frac{1}{2}}\longrightarrow 0,
\end{aligned}\tag{5.50}
$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. Similarly, one has

$$
\mathbb{E}\|\Sigma_3\|\longrightarrow 0,\tag{5.51}
$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. Consequently, from (5.47)-(5.50), we get

$$
\int_0^T|G_1|\mathrm{d}t\longrightarrow 0,\quad\text{in probability},\tag{5.52}
$$

when $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$.

For the second term $G_2$, following a similar line of argument as (5.52), we get

$$
\int_0^T|G_2|\mathrm{d}t\longrightarrow 0,\quad\text{in probability},\tag{5.53}
$$

when $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$.

For the third term $G_3$, by the assumption (**A2**), (5.40) and (5.45), one has

$$G_3\le C\Big\{|X^{\varepsilon,i,N}_{t_\delta}-X^{0,i}_t|^2+(1+|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2r_2}+|X^{0,i}_{t-\tau}|^{2r_2})|X^{\varepsilon,i,N}_{t_\delta-\tau}-X^{0,i}_{t-\tau}|^2 \\ +\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta},\mu^{0,i}_t)^2+\mathbb{W}_2(\mu^{\varepsilon,N}_{t_\delta-\tau},\mu^{0,i}_{t-\tau})^2\Big\}. \tag{5.54}$$

On the other hand, thanks to (5.38) and (5.54), it follows that

$$\mathbb{P}\Big(\int_0^T G_3\mathrm{d}t\ge\epsilon\Big) \\ \le\frac{C}{\epsilon}\int_0^T\Big\{\mathbb{E}|X^{\varepsilon,i,N}_{t_\delta}-X^{0,i}_t|^2+\mathbb{E}(1+|X^{\varepsilon,i,N}_{t_\delta-\tau}|^{2r_2}+|X^{0,i}_{t-\tau}|^{2r_2})|X^{\varepsilon,i,N}_{t_\delta-\tau}-X^{0,i}_{t-\tau}|^2 \\ +\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta}-X^{0,j}_t|^2+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X^{\varepsilon,j,N}_{t_\delta-\tau}-X^{0,j}_{t-\tau}|^2\Big\}\mathrm{d}t\longrightarrow 0,$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. Hence,

$$\int_0^T G_3\mathrm{d}t\longrightarrow 0,\quad\text{in probability}, \tag{5.55}$$

as $\varepsilon\to 0$, $\delta\to 0$ and $N\to\infty$. As a consequence, (5.46) follows from (5.52), (5.53) and (5.55). What's more, for any $\rho>0$ and $\epsilon>0$, owing to (5.46), one gets

$$\mathbb{P}\Big(\Big|\int_0^T(\Upsilon(X^{\varepsilon,i,N}_{t_\delta},X^{\varepsilon,i,N}_{t_\delta-\tau},\theta_0)-\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0))\mathrm{d}W^i_t\Big|\ge\rho\Big) \\ \le\mathbb{P}\Big(\int_0^T\|\Upsilon(X^{\varepsilon,i,N}_{t_\delta},X^{\varepsilon,i,N}_{t_\delta-\tau},\theta_0)-\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)\|^2\mathrm{d}t\ge\rho^2\epsilon\Big)+\epsilon,$$

which, together with the arbitrariness of $\epsilon$ and (5.46), implies that (5.44) holds. And by a simple calculation, one gets

$$\varepsilon^{-1}(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\theta_0) \\ =-2\sum_{k=1}^n(\nabla_\theta b)^*(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau},\theta_0) \\ \times\Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau})\sigma(X^{\varepsilon,i,N}_{(k-1)\delta},X^{\varepsilon,i,N}_{(k-1)\delta-\tau},\mu^{\varepsilon,N}_{(k-1)\delta},\mu^{\varepsilon,N}_{(k-1)\delta-\tau})\delta W^i_k \\ =-2\int_0^T\Upsilon(X^{\varepsilon,i,N}_{t_\delta},X^{\varepsilon,i,N}_{t_\delta-\tau},\theta_0)\dot{\mathrm{W}}^i_t\to-2\int_0^T\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)\dot{\mathrm{W}}^i_t,\qquad\mathbb{P}-\text{a.s.},$$

whenever $\varepsilon\to 0$ and $n,N\to\infty$. □

LEMMA 5.7. *Under the assumptions of Theorem 3.3,*

$$(\nabla^{(2)}_\theta\Phi^{i,N}_{n,\varepsilon})(\theta)\longrightarrow\bar{K}(\theta):=K(\theta)+2I(\theta),\qquad\mathbb{P}-a.s., \tag{5.56}$$

*as* $\varepsilon\to 0$, $\delta\to 0$ *and* $N\to\infty$, *where* $I(\theta)$ *is defined in* (3.4), *and*

$$K(\theta):=-2\int_0^T(\nabla^{(2)}_\theta b^*)(X^{0,i}_t,X^{0,i}_{t-\tau},\mu^{0,i}_t,\mu^{0,i}_{t-\tau}) \\ \circ\Big\{\Lambda^{-1}(X^{0,i}_t,X^{0,i}_{t-\tau})B(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0,\theta)\Big\}\mathrm{d}t.$$

*Proof.*

$$
\begin{aligned}
&(\nabla_\theta^{(2)}\Phi_{n,\varepsilon}^{i,N})(\theta)=(\nabla_\theta(\nabla_\theta\Phi_{n,\varepsilon}^{i,N}))(\theta)\\
&=-2\sum_{k=1}^n(\nabla_\theta^{(2)}b)^*(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\\
&\quad\circ\Big\{\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})P_k^{\varepsilon,i,N}(\theta)\Big\}\\
&\quad-2\sum_{k=1}^n(\nabla_\theta b)^*(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\\
&\qquad\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})(\nabla_\theta P_k^{\varepsilon,i,N})(\theta)\\
&=-2\sum_{k=1}^n(\nabla_\theta^{(2)}b)^*(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\\
&\qquad\circ\Big\{\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})P_k^{\varepsilon,i,N}(\theta_0)\Big\}\\
&\quad-2\delta\sum_{k=1}^n\Big\{(\nabla_\theta^{(2)}b)^*(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\\
&\qquad\circ\Big\{\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})\\
&\quad\times B(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta)\Big\}-(\nabla_\theta b)^*(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\\
&\quad\times\Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N})(\nabla_\theta b)(X_{(k-1)\delta}^{\varepsilon,i,N},X_{(k-1)\delta-\tau}^{\varepsilon,i,N},\mu_{(k-1)\delta}^{\varepsilon,N},\mu_{(k-1)\delta-\tau}^{\varepsilon,N},\theta)\Big\}\\
&=:\Pi_1+\Pi_2.
\end{aligned}
$$

For any $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_p(\mathbb{R}^d)$, notice from (**A5**) that

$$
\sup_{\theta\in\bar\Theta}\|(\nabla_\theta^{(2)}b^*)(x,y,\mu,\nu,\theta)\|\le C\{1+|x|+|y|+|y|^{r_5+1}+\mathbb{W}_2(\mu,\delta_0)+\mathbb{W}_2(\nu,\delta_0)\}. \tag{5.57}
$$

Set $q_4=(r_3+1)\vee(r_5+1)$. For the first term $\Pi_1$, by (5.32) and (5.57), one arrives at

$$
\begin{aligned}
\mathbb{E}|\Pi_1|&\le 2\varepsilon\Big(\mathbb{E}\int_0^T\|(\nabla_\theta^{(2)}b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)\|^2\|\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\\
&\qquad\times\sigma(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N})\|^2\mathrm{d}t\Big)^{1/2}\\
&\le C\varepsilon\Big(\mathbb{E}\int_0^T\Big(1+|X_{t_\delta}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^2+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2q_4}+\frac{1}{N}\sum_{j=1}^N|X_{t_\delta}^{\varepsilon,j,N}|^2\\
&\qquad+\frac{1}{N}\sum_{j=1}^N|X_{t_\delta-\tau}^{\varepsilon,j,N}|^2\Big)^3\mathrm{d}t\Big)^{1/2}\\
&\le C\varepsilon\Big(\int_0^T\Big(1+\mathbb{E}|X_{t_\delta}^{\varepsilon,i,N}|^6+\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{6q_4}+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta}^{\varepsilon,j,N}|^6+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,j,N}|^6\Big)\mathrm{d}t\Big)^{1/2}\\
&\le C\varepsilon\to0,\quad\text{as }\ \varepsilon\to0,\quad\delta\to0\text{ and }N\to\infty.
\end{aligned}
$$

For the second term $\Pi_2$, we infer that

$$\begin{aligned}\Pi_2=&-2\int_0^T(\nabla_\theta^{(2)}b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)\circ(\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\\&\times B(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta))\mathrm{d}t\\&+2\int_0^T(\nabla_\theta b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\\&\times(\nabla_\theta b)(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)\mathrm{d}t\\=:&\,H_1+H_2.\end{aligned}$$

Taking into consideration Lemma 5.4 and (**A5**) yields that

$$\begin{aligned}&H_1-K(\theta)\\=&-2\int_0^T\Big((\nabla_\theta^{(2)}b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)\circ(\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})\\&\times B(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta))\\&-(\nabla_\theta^{(2)}b^*)(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta)\circ(\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})B(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0,\theta))\Big)\mathrm{d}t\\=&-2\int_0^T\Big(((\nabla_\theta^{(2)}b)^*(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\mu_{t_\delta}^{\varepsilon,N},\mu_{t_\delta-\tau}^{\varepsilon,N},\theta)-(\nabla_\theta^{(2)}b^*)(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta))\\&\circ(\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})B(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta))+(\nabla_\theta^{(2)}b^*)(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta)\\&\circ(\Lambda^{-1}(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N})-\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i}))B(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta))\\&+(\nabla_\theta^{(2)}b^*)(X_t^{0,i},X_{t-\tau}^{0,i},\mu_t^{0,i},\mu_{t-\tau}^{0,i},\theta)\circ(\Lambda^{-1}(X_t^{0,i},X_{t-\tau}^{0,i})(B(X_{t_\delta}^{\varepsilon,i,N},X_{t_\delta-\tau}^{\varepsilon,i,N},\theta_0,\theta)\\&-B(X_t^{0,i},X_{t-\tau}^{0,i},\theta_0,\theta))\Big)\mathrm{d}t\\=:&\sum_{i=1}^3 M_3.\end{aligned}$$

For the term $M_1$, thanks to (**A5**), (5.3) and (5.29), it follows from the Hölder inequality that

$$\begin{aligned}\mathbb{E}|M_1|\le&\, C\int_0^T\Big(\mathbb{E}|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2+\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}-X_{t-\tau}^{0,i}|^2(1+|X_{t-\tau}^{0,i}|^{2r_5}+|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{2r_5})\\&+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta}^{\varepsilon,j,N}-X_t^{0,j}|^2+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,j,N}-X_{t-\tau}^{0,j}|^2\Big)^{1/2}\\&\times\Big(1+\mathbb{E}|X_{t_\delta}^{\varepsilon,i,N}|^4+\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}|^4+\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{4q_2}+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta}^{\varepsilon,j,N}|^4\\&+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,j,N}|^4\Big)^{1/2}\mathrm{d}t\\\le&\, C\int_0^T\Big(\mathbb{E}|X_{t_\delta}^{\varepsilon,i,N}-X_t^{0,i}|^2+(\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}-X_{t-\tau}^{0,i}|^4)^{\frac12}(1+\mathbb{E}|X_{t-\tau}^{0,i}|^{4r_5}\\&+\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,i,N}|^{4r_5})^{\frac12}+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta}^{\varepsilon,j,N}-X_t^{0,j}|^2+\frac{1}{N}\sum_{j=1}^N\mathbb{E}|X_{t_\delta-\tau}^{\varepsilon,j,N}-X_{t-\tau}^{0,j}|^2\Big)^{1/2}\mathrm{d}t,\end{aligned}$$

where, in the second step we have used the result in Lemma 5.1. Then, according to (5.38), one has

$$\mathbb{E}|M_1| \to 0 \quad \text{as} \;\; \varepsilon \to 0, \;\; \delta \to 0, \;\; N \to \infty.$$

By (**A3**), (5.29), (5.33) and (5.57) and carrying out similar arguments, one has

$$\mathbb{E}|M_2| \to 0 \quad \text{as} \;\; \varepsilon \to 0, \;\; \delta \to 0, \;\; N \to \infty$$

and

$$\mathbb{E}|M_3| \to 0 \quad \text{as} \;\; \varepsilon \to 0, \;\; \delta \to 0, \;\; N \to \infty.$$

As a result, we conclude that

$$H_1 \to K(\theta) \;\; \mathbb{P}-\text{a.s.} \;\; \text{as} \;\; \varepsilon \to 0, \;\; \delta \to 0, \;\; N \to \infty. \tag{5.58}$$

Again, carrying out analogous arguments to derive (5.58), we obtain

$$H_2 \to 2I(\theta) \quad \mathbb{P}-\text{a.s.} \;\; \text{as} \;\; \varepsilon \to 0, \;\; \delta \to 0, \;\; N \to \infty. \tag{5.59}$$

Therefore, the desired assertion is complete by (5.58) and (5.59) immediately. □

Now we start to finish the argument of Theorem 3.3 on the basis of the previous lemmas.

*Proof.* **(Proof of Theorem 3.3.)** According to the result of Theorem 3.2, there exists a sequence $\eta_{n,\varepsilon}^{i,N} \to 0$ as $N, n \to \infty$ and $\varepsilon \to 0$ such that $\hat{\theta}_{n,\varepsilon}^{i,N} \in B_{\eta_{n,\varepsilon}^{i,N}}(\theta_0) \subset \Theta$, $\mathbb{P}$-a.s., that is to say,

$$\mathbb{P}\Big(\hat{\theta}_{n,\varepsilon}^{i,N} \in B_{\eta_{n,\varepsilon}^{i,N}}(\theta_0)\Big) \to 1, \quad \text{as} \;\; n, N \to \infty, \;\; \varepsilon \to 0. \tag{5.60}$$

Then, it is easy to see that

$$(\nabla_\theta \Phi_{n,\varepsilon}^{i,N})(\hat{\theta}_{n,\varepsilon}^{i,N}) = (\nabla_\theta \Phi_{n,\varepsilon}^{i,N})(\theta_0) + F_{n,\varepsilon}^{i,N}(\hat{\theta}_{n,\varepsilon}^{i,N} - \theta_0), \quad \hat{\theta}_{n,\varepsilon}^{i,N} \in B_{\eta_{n,\varepsilon}^{i,N}}(\theta_0) \tag{5.61}$$

with

$$F_{n,\varepsilon}^{i,N} := \int_0^1 (\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0 + v(\hat{\theta}_{n,\varepsilon}^{i,N} - \theta_0))\mathrm{d}v, \qquad \hat{\theta}_{n,\varepsilon}^{i,N} \in B_{\eta_{n,\varepsilon}^{i,N}}(\theta_0),$$

owing to the Taylor expansion. In what follows we intend to deduce that

$$F_{n,\varepsilon}^{i,N} \to \bar{K}(\theta_0) \qquad \mathbb{P}-\text{a.s.} \tag{5.62}$$

as $n, N \to \infty$ and $\varepsilon \to 0$. Indeed, for $\hat{\theta}_{n,\varepsilon}^{i,N} \in B_{\eta_{n,\varepsilon}^{i,N}}(\theta_0)$,

$$\begin{aligned}
&\|F_{n,\varepsilon}^{i,N} - \bar{K}(\theta_0)\| \le \|F_{n,\varepsilon}^{i,N} - (\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0)\| + \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0) - \bar{K}(\theta_0)\| \\
&\le \int_0^1 \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0 + v(\hat{\theta}_{n,\varepsilon}^{i,N} - \theta_0)) - (\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0)\|\mathrm{d}v + \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0) - \bar{K}(\theta_0)\| \\
&\le \sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta) - \bar{K}(\theta)\| + \sup_{\theta \in B_{\eta_{n,\varepsilon}}(\theta_0)} \|\bar{K}(\theta) - \bar{K}(\theta_0)\| \\
&\quad + 2\|(\nabla_\theta^{(2)} \Phi_{n,\varepsilon}^{i,N})(\theta_0) - \bar{K}(\theta_0)\|,
\end{aligned}$$

where $\bar{K}(\cdot)$ is shown in (5.56). This, together with Lemma 5.7 and the continuity of $\bar{K}(\cdot)$, yields that (5.62) holds. Next we show the asymptotic distribution of $\hat{\theta}^{i,N}_{n,\varepsilon}$. Let

$$\mathcal{F}^{i,N}_{n,\varepsilon}=\{F^{i,N}_{n,\varepsilon} \text{ is invertible },\hat{\theta}^{i,N}_{n,\varepsilon}\in B_{\eta^{i,N}_{n,\varepsilon}}(\theta_0)\}.$$

By Lemma 5.7, one gets, for some positive constant $\alpha$,

$$\mathbb{P}\Big(\sup_{\theta\in B_{\eta^{i,N}_{n,\varepsilon}}(\theta_0)}\|(\nabla^{(2)}_\theta\Phi^{i,N}_{n,\varepsilon})(\theta)-\bar{K}(\theta_0)\|\leq\frac{\alpha}{2}\Big)\to 1 \tag{5.63}$$

as $n,N\to\infty$ and $\varepsilon\to 0$. What's more, by following the line of [17, Theorem 2.2], we can deduce that $F^{i,N}_{n,\varepsilon}$ is invertible on the set

$$\Gamma^{i,N}_{n,\varepsilon}:=\Big\{\sup_{\theta\in B_{\eta^{i,N}_{n,\varepsilon}}(\theta_0)}\|(\nabla^{(2)}_\theta\Phi^{i,N}_{n,\varepsilon})(\theta)-\bar{K}(\theta_0)\|\leq\frac{\alpha}{2},\ \ \hat{\theta}^{i,N}_{n,\varepsilon}\in B_{\eta^{i,N}_{n,\varepsilon}}(\theta_0)\Big\}.$$

Clearly,

$$1\geq\mathbb{P}(\Gamma^{i,N}_{n,\varepsilon})\geq\mathbb{P}\Big(\sup_{\theta\in B_{\eta_{n,\varepsilon}}(\theta_0)}\|(\nabla^{(2)}_\theta\Phi^{i,N}_{n,\varepsilon})(\theta)-K_0(\theta_0)\|\leq\frac{\alpha}{2}\Big)+\mathbb{P}\Big(\hat{\theta}^{i,N}_{n,\varepsilon}\in B_{\eta_{n,\varepsilon}}(\theta_0)\Big)-1. \tag{5.64}$$

Thus, taking advantage of (5.63), (5.60) as well as (5.64), we deduce

$$\mathbb{P}(\mathcal{F}^{i,N}_{n,\varepsilon})\geq\mathbb{P}(\Gamma^{i,N}_{n,\varepsilon})\to 1\quad\text{as}\ \ n,N\to\infty,\ \ \varepsilon\to 0. \tag{5.65}$$

Let

$$U^{i,N}_{n,\varepsilon}=F^{i,N}_{n,\varepsilon}\mathbf{1}_{\mathcal{F}^{i,N}_{n,\varepsilon}}+I_p\mathbf{1}_{(\mathcal{F}^{i,N}_{n,\varepsilon})^c}\ ,$$

where $I_p$ is a $p\times p$ identity matrix. It follows from (5.61) that

$$\begin{aligned}\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0)=&(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{\mathcal{F}^{i,N}_{n,\varepsilon}}+(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{(\mathcal{F}^{i,N}_{n,\varepsilon})^c}\\ =&(U^{i,N}_{n,\varepsilon})^{-1}F^{i,N}_{n,\varepsilon}(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{\mathcal{F}^{i,N}_{n,\varepsilon}}+(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{(\mathcal{F}^{i,N}_{n,\varepsilon})^c}\\ =&\varepsilon^{-1}(U^{i,N}_{n,\varepsilon})^{-1}\{(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\hat{\theta}^{i,N}_{n,\varepsilon})-(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\theta_0)\}\mathbf{1}_{\mathcal{F}^{i,N}_{n,\varepsilon}}\\ &+(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{(\mathcal{F}^{i,N}_{n,\varepsilon})^c}\\ =&-\varepsilon^{-1}(U^{i,N}_{n,\varepsilon})^{-1}(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\theta_0)\mathbf{1}_{\mathcal{F}^{i,N}_{n,\varepsilon}}+(\varepsilon^{-1}(\hat{\theta}^{i,N}_{n,\varepsilon}-\theta_0))\mathbf{1}_{(\mathcal{F}^{i,N}_{n,\varepsilon})^c}\\ \to& I^{-1}(\theta_0)\int_0^T\Upsilon(X^{0,i}_t,X^{0,i}_{t-\tau},\theta_0)\mathrm{d}W^i_t\qquad\text{as}\ \ n,N\to\infty,\ \ \varepsilon\to 0,\end{aligned}$$

where, in the fourth step we have used Fermat's lemma and dropped the term $(\nabla_\theta\Phi^{i,N}_{n,\varepsilon})(\hat{\theta}^{i,N}_{n,\varepsilon})$, and in the last step we have utilized Lemma 5.1, (5.56), (5.62), and (5.65). The desired conclusion is obtained. □

REFERENCES

[1] J. Bao and X. Huang, *Approximations of McKean-Vlasov stochastic differential equations with irregular coefficients*, J. Theor. Probab., 2021. 1, 2, 3, 5.2

[2] J. Bao and C. Yuan, *Convergence rate of EM scheme for SDDEs*, Proc. Amer. Math. Soc., 141:3231–3243, 2013. 5.2

[3] M. Barczy, M.B. Alaya, A. Kebaier, and G. Pap, *Asymptotic properties of maximum likelihood estimator for the growth rate for a jump-type CIR process based on continuous time observations*, Stoch. Process. Their Appl., 128:1135–1164, 2017. 1

[4] M. Barczy and G. Pap, *Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations*, Statistics, 50:389–417, 2015. 1

[5] R. Buckdahn, J. Li, and J. Ma, *A mean-field stochastic control problem with partial observations*, Ann. Appl. Probab., 27:3201–3245, 2017. 1

[6] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I. Mean Field FBSDEs, Control, and Games*, Probability Theory and Stochastic Modelling, Springer, Cham., 2018. 1, 5.2

[7] A. Eberle, A. Guillin, and R. Zimmer, *Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes*, Trans. Amer. Math. Soc., 371:7135–7173, 2019. 1

[8] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Relat. Fields, 3:707–738, 2015. 5.2

[9] X. Huang, M. Röckner, and F.-Y. Wang, *Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs*, Discrete Contin. Dyn. Syst., 39(6):3017–3035, 2019. 1, 3, 5.1

[10] X. Huang and F.-Y. Wang, *Distribution dependent SDEs with singular coefficients*, Stoch. Process. Their Appl., 129:4747–4770, 2019. 1

[11] X. Huang and C. Yuan, *Comparison theorem for distribution dependent neutral SFDEs*, J. Evol. Eqs., 21:653–670, 2021. 1

[12] H. Long, *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(3):645–663, 2010. 1

[13] R.A. Kasonga, *The consistency of a non-linear least squares estimator from diffusion processes*, Stoch. Process. Their Appl., 2:263–275, 1988. 1

[14] U. Kuechler and M. Soerensen, *A simple estimator for discrete-time samples from affine stochastic delay differential equations*, Stat. Inf. Stoch. Process., 13:125–132, 2010. 1

[15] Y.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer, London, 2004. 1

[16] H. Long, C. Ma, and Y. Shimizu, *Least squares estimators for stochastic differential equations driven by small Lévy noises*, Stoch. Process. Their Appl., 127:1475–1495, 2017. 1

[17] H. Long, Y. Shimizu, and W. Sun, *Least squares estimators for discretely observed stochastic processes driven by small Lévy noises*, J. Multivar. Anal., 116:422–439, 2013. 1, 5.3

[18] C. Ma, *A note on least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises*, Stat. Probab. Lett., 80:1528–1531, 2010. 1

[19] H.P. McKean Jr., *A class of Markov processes associated with nonlinear parabolic equations*, Proc. Natl. Acad. Sci. U.S.A., 56:1907–1911, 1966. 1

[20] H.P. McKean, *Propagation of chaos for a class of nonlinear parabolic equations*, in Stochastic Differential Equations, Lecture Series in Diff. Eqs., 7:41–57, 1967. 1

[21] Y. Pan and L. Yan, *The least squares estimation for the $\alpha$-Stable Ornstein-Uhlenbeck Process with constant drift*, Methodol. Comput. Appl. Probab., 21:1165–1182, 2019. 1

[22] G.D. Reis, W. Salkeld, and J. Tugaut, *Freidlin-Wentzell LDPs in path space for McKean-Vlasov equations and the functional iterated logarithm law*, Ann. Appl. Probab., 29(3):1487–1540, 2019. 3

[23] M. Reiss, *Adaptive estimation for affine stochastic delay differential equations*, Bernoulli, 11(1):67–102, 2005. 1

[24] P. Ren and F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications*, J. Differ. Equ., 8:4745–4777, 2019. 1

[25] P. Ren and J. Wu, *Least squares estimator for path-dependent McKean-Vlasov SDEs via discrete-time observations*, Acta Math. Sci., 39:691–716, 2019. 1, 3

[26] P. Ren and J. Wu, *Least squares estimation for path-distribution dependent stochastic differential equations*, arXiv preprint, arXiv:1802.00820, 2018. 1, 3

[27] M. Röckner and X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts*, Bernoulli, 27(2):1131–1158, 2021. 1

[28] A.W. Van der Vaart, *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 3, 1998. 5.2

[29] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations*, Stoch. Process. Their Appl., 128(2):595–621, 2018. 1, 3, 5.1

[30] J. Wen, X. Wang, S. Mao, and X. Xiao, *Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application*, Appl. Math. Comput., 274:237–246, 2016. 1