LEAST SQUARES ESTIMATION FOR
DISTRIBUTION-DEPENDENT STOCHASTIC DIFFERENTIAL
DELAY EQUATIONS

YANYAN HU AND FUBAO XI
School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081, China
MIN ZHU*
Hunan University of Technology, Zhuzhou 412007, Hunan, China
(Communicated by Feng-Yu Wang)

Abstract. The parametric estimation of drift parameter for distribution-dependent stochastic differential delay equations with a small diffusion is presented. The principle technique of our investigation is to construct an appropriate contrast function and carry out a limiting type of argument to show the consistency and convergence rate of the least squares estimator of the drift parameter via interacting particle systems. In addition, two examples are constructed to demonstrate the effectiveness of our work.

1. Introduction. The distribution-dependent stochastic differential equation (SDE) (also called McKean-Vlasov or mean-field SDE) is a kind of mathematical model, which can characterize the evolution of stochastic systems depending on the distribution and position of particles. This equation was first referenced by Kac [9, 10] as a stochastic toy model for the Vlasov kinetic equation of plasma, and then introduced by Mckean to model plasma dynamics. As a hot but difficult research topic, it has an important application value in the fields of stochastic control, insurance, mathematical finance, to name a few; see, for instance, [2, 3]. The distribution-dependent SDEs have been extensively investigated by many authors, and various results on well-posedness, Harnack inequalities, Bismut formula, ergodicity, and other quantitative and qualitative properties have been obtained (see e.g. [20, 18, 16, 4]).

In contrast to distribution-dependent SDEs without delay, path-distribution dependent SDEs have been much less studied, while these have begun to gain attention recently. For works on wellposedness and Harnack type inequalities, we refer to [6, 7]. Furthermore, Huang and Yuan [8] showed the existence and uniqueness of strong solutions to distribution dependent neutral SFDEs and gave the comparison theorem of these equations.

2020 Mathematics Subject Classification. Primary: 62F12, 62M05; Secondary: 60G52.
Key words and phrases. Distribution-dependent stochastic differential delay equations, least squares method, discrete observation, consistency, asymptotic distribution.

The second author is partially supported by the National Natural Science Foundation of China (No. 12071031). The third author is partially supported by the National Natural Science Foundation of China (No. 11901188) and the Scientific Research Funds of Hunan Provincial Education Department of China (19B156).

* Corresponding author.
Most of the previous works are concerned with path-distribution dependent SDEs which do not contain unknown parameters. However, in many practical applications, these models may contain unknown parameters. Hence, we want to estimate deterministic quantities of these unknown parameters for SDEs, especially, path-distribution dependent SDEs. Based on discrete and continuous-time observations, there is a lot of good literature on the parametric estimation for SDEs with small diffusions; see, e.g. [5, 12, 21]. Nevertheless, for path-distribution dependent SDEs, the related studies are rare. Recently, the paper [17] studied the least squares estimation for the drift parameter of path-distribution dependent SDEs involving a small dispersion parameter by constructing Euler-Maruyama discretion scheme. Inspired by their studies, we make a new attempt to study the problem of parameter estimation with a small dispersion for distribution-dependent stochastic differential delay equations (SDDEs), which is essentially point-delay McKean-Vlasov SDEs. At present, there has been no study on parametric estimation for distribution-dependent SDDEs.

As we know, in general, the parametric estimation relied on continuous-time observations is an ideal mathematical model, and no measuring device can follow continuously the sample paths of the diffusion processes involved. So, from a practical point of view, it is more meaningful to explore asymptotic estimation for diffusion processes with small dispersions based on discrete observations.

What’s more, for the general SDEs, the least squares estimation (LSE) is a popular parametric estimation method; see, e.g. [11, 13, 14, 16]. For distribution-dependent SDEs, this method cannot directly be applied. Because the fact is that we cannot directly obtain observations of distribution along the path at regularly space time points in most of our arguments. The key issue to simulate distribution-dependent SDEs lies in how to approximate the distribution at every step of the analysis. There is a need to seek executable numerical algorithms so that they can be simulated easily. Although Ren and Wu in [17] have succeeded in investigating parameter estimation for path-distribution dependent SDEs by linear interpolating at the discrete-time observations, which is different from our present model.

As one of the contributions in our work, stochastic non-interacting particle systems are introduced, and their solutions are identically distributed with respect to the target equations. So we only research the parameter estimation problem of stochastic non-interacting particle systems, which have the function of bridge connecting the target equations and stochastic interacting particle systems. For the SDDEs, the discrete-time observations at the gridpoints are sufficient to construct the contrast function. Nevertheless, for our present model, the discrete-time observations are insufficient to establish the contrast function because the SDDEs involved are distribution-dependent. In this article, we introduce the empirical distribution into stochastic interacting particle systems to discrete distribution, and it prevents loss of information and makes up the gap in [17]. Based on this preparation, we can continuous to apply the method of LSE to study the consistence and asymptotic property of unknown parameter for distribution-dependent SDDEs.

The remainder of the paper is organized as follows: Section 2 begins with the setup of the basic notation, reviews stochastic interacting particle systems and constructs the LSE $\hat{\theta}_{n,N}^{i}$. In section 3, several lemmas which lay the foundation for consistency analysis are given, and the consistency of the LSE is derived by the convergence of the contrast function. In Section 4, the asymptotic distribution
of the LSE converging to a normal distribution is shown, and two examples are established to demonstrate the applicability of our theory.

2. Preliminary. For integers $m, d > 0$, let $(\mathbb{R}^d, \langle \cdot, \cdot \rangle, |\cdot|)$ be the $d$-dimensional Euclidean space. $\mathbb{R}^d \otimes \mathbb{R}^m$ the collection of all $d \times m$ matrixes endowed with the Hilbert-Schmidt norm $\| \cdot \|$ (i.e. for a matrix $A$ and its transpose $A^*$, $\|A\| := \sqrt{\text{trace}(AA^*)}$). For a square matrix $A$, $A^{-1}$ means the inverse of $A$ provided that $\det(A) \neq 0$. $\mathbf{0} \in \mathbb{R}^d$ denotes the zero vector. Let $\Theta$ be an open bounded convex subset of $\mathbb{R}^d$, for some integer $d > 0$, and $\bar{\Theta}$ the closure of $\Theta$. $B_r(x)$ represents the closed ball centered at $x$ with the radius $r > 0$. $\delta_z$ denotes Dirac measure centered at the point $z$. $[a]$ means the integer part of the real number $a \geq 0$. For a random variable $\xi$, $\mathcal{L}_\xi$ denotes its law. For given $\tau > 0$, $\mathcal{C} := C([-\tau, 0]; \mathbb{R}^d)$ denotes the family of all continuous functions $\xi : [-\tau, 0] \to \mathbb{R}^d$ with the uniform norm $\|\xi\|_{\mathcal{C}} := \sup_{-\tau \leq \theta \leq 0} |\xi(\theta)|$. Here $\tau$ stands for the delay or time lag.

For $p > 0$, $\mathcal{P}_p(\mathbb{R}^d)$ stands for the space of all probability measures on $\mathbb{R}^d$ with finite $p$-th moment, i.e., $\mu(|\cdot|^p) := \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty$ for $\mu \in \mathcal{P}_p(\mathbb{R}^d)$. Define the $\mathcal{W}_p$-Wasserstein distance on $\mathcal{P}_p(\mathbb{R}^d)$ by

$$\mathcal{W}_p(\mu, \nu) = \inf_{\pi \in \mathcal{C}(\mu, \nu)} \left( \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^p \pi(dx, dy) \right)^{\frac{1}{p}}, \quad \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d),$$

where $\mathcal{C}(\mu, \nu)$ signifies the set of all couplings of $\mu$ and $\nu$. Let $\{\mathcal{W}_t\}_{t \geq 0}$ be an $m$-dimensional Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the filtration $(\mathcal{F}_t)_{t \geq 0}$ satisfying the usual condition (i.e., $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets and $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$).

For a differentiable function $V(x) = (V_1(x), \cdots, V_d(x))^* : \mathbb{R}^m \to \mathbb{R}^d$, define its gradient operator $(\nabla_x V)(x) \in \mathbb{R}^d \otimes \mathbb{R}^m$ w.r.t. $x = (x_1, \cdots, x_m)^* \in \mathbb{R}^m$ by

$$(\nabla_x V)(x) = \left( \begin{array}{ccc} \frac{\partial}{\partial x_1} V_1(x) & \frac{\partial}{\partial x_2} V_1(x) & \cdots & \frac{\partial}{\partial x_m} V_1(x) \\ \frac{\partial}{\partial x_1} V_2(x) & \frac{\partial}{\partial x_2} V_2(x) & \cdots & \frac{\partial}{\partial x_m} V_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} V_d(x) & \frac{\partial}{\partial x_2} V_d(x) & \cdots & \frac{\partial}{\partial x_m} V_d(x) \end{array} \right),$$

which enjoys the property $\nabla_x V^*(x) = (\nabla_x V)(x)^*$. For a matrix-valued function $V(x) = (V_{ij}(x))_{m \times d} : \mathbb{R} \to \mathbb{R}^m \otimes \mathbb{R}^d$ be differentiable, its derivative $\frac{\partial}{\partial x} V(x) \in \mathbb{R}^m \otimes \mathbb{R}^d$ w.r.t. $x \in \mathbb{R}$ admits the form

$$\frac{\partial}{\partial x} V(x) = \left( \begin{array}{ccc} \frac{\partial}{\partial x_1} V_{11}(x) & \frac{\partial}{\partial x_2} V_{11}(x) & \cdots & \frac{\partial}{\partial x_d} V_{11}(x) \\ \frac{\partial}{\partial x_1} V_{12}(x) & \frac{\partial}{\partial x_2} V_{12}(x) & \cdots & \frac{\partial}{\partial x_d} V_{12}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} V_{md}(x) & \frac{\partial}{\partial x_2} V_{md}(x) & \cdots & \frac{\partial}{\partial x_d} V_{md}(x) \end{array} \right).$$

If $V(x) = (V_{ij}(x))_{m \times d} : \mathbb{R}^m \to \mathbb{R}^m \otimes \mathbb{R}^d$ is differentiable, its gradient operator $(\nabla_x V)(x) \in \mathbb{R}^m \otimes \mathbb{R}^{md}$ w.r.t. the variable $x = (x_1, \cdots, x_m)^* \in \mathbb{R}^m$ is written as

$$(\nabla_x V)(x) = \left( \frac{\partial}{\partial x_1} V(x), \frac{\partial}{\partial x_2} V(x), \cdots, \frac{\partial}{\partial x_m} V(x) \right),$$

where $\frac{\partial}{\partial x_i} V(x)$ is defined as in (2.1). Moreover, for a differentiable function $V = (V_{ij})_{m \times d} : \mathbb{R}^m \to \mathbb{R}^d$, we have

$$(\nabla_x^2 V^*)(x) := (\nabla_x (\nabla_x V^*))(x) = (\nabla_x (\nabla_x V^*))(x).$$
For $A = (A_1, A_2, \cdots, A_m) \in \mathbb{R}^m \otimes \mathbb{R}^{md}$ with $A_k \in \mathbb{R}^m \otimes \mathbb{R}^d$, $k = 1, \cdots, m$ and $B \in \mathbb{R}^d$, let

$$A \circ B = (A_1 B, A_2 B, \cdots, A_m B) \in \mathbb{R}^m \otimes \mathbb{R}^m.$$ 

Next, we fix the time horizon $T > 0$ and emphasize that $C > 0$ is a generic constant which may change from line to line. For the scale parameter $\varepsilon \in (0, 1)$, we consider a distribution-dependent SDDE on $(\mathbb{R}^d, (\cdot, \cdot), |\cdot|)$

$$\begin{cases}
    dX^\varepsilon_t = b(X^\varepsilon_t, X^\varepsilon_{t-}, \mu^\varepsilon_t, \mu^\varepsilon_{t-}, \theta)dt + \varepsilon \sigma(X^\varepsilon_t, X^\varepsilon_{t-}, \mu^\varepsilon_t, \mu^\varepsilon_{t-})dW_t, & t \in (0, T], \\
    X^\varepsilon_s = \xi_s, & s \in [-\tau, 0],
\end{cases} \tag{2.3}$$

where $\mu^\varepsilon := \mathcal{L}_{X^\varepsilon}$ denotes the law of $X^\varepsilon$, $b : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \times \Theta \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m$. In (2.3), we assume that $b$ and $\sigma$ are both known functions apart from the parameter $\theta \in \Theta$ and denote the true value of $\theta$ by $\theta_0 \in \Theta$. We now give the definition of strong solution to (2.3).

**Definition 2.1.** A continuous adapted process $(X^\varepsilon_t)_{t \geq -\tau}$ on $\mathbb{R}^d$ is called a (strong) solution of (2.3), if

$$\int_0^t \mathbb{E}(|b(X^\varepsilon_s, X^\varepsilon_{s-}, \mu^\varepsilon, \mu^\varepsilon_{s-}, \theta)| + |\sigma(X^\varepsilon_s, X^\varepsilon_{s-}, \mu^\varepsilon, \mu^\varepsilon_{s-})|^2)|ds < \infty, \quad t \geq 0,$$

and $\mathbb{P}$-a.s.

$$X^\varepsilon_t = X^\varepsilon_0 + \int_0^t b(X^\varepsilon_s, X^\varepsilon_{s-}, \mu^\varepsilon, \mu^\varepsilon_{s-}, \theta)ds + \varepsilon \int_0^t \sigma(X^\varepsilon_s, X^\varepsilon_{s-}, \mu^\varepsilon, \mu^\varepsilon_{s-})dW_s, \quad t \geq 0.$$

**Remark 1.** If, for initial value $X^\varepsilon_0 = \xi$, $\mathbb{E}||\xi||^2_{\mathcal{L}^\varepsilon} < \infty$, by the Burkholder-Davis-Gundy inequality, Definition 2.1 implies $\mathbb{E}(\sup_{-\tau \leq t \leq T}|X^\varepsilon_t|^2) < \infty$. In Lemma 3.1 we will give the proof of this result in a different way.

Next, we introduce the stochastic interacting particle systems to approximate (2.3). Let $N > 0$ be an integer and $(X^\varepsilon_{0,i}, W^i_t)$ be i.i.d copies of $(X^\varepsilon_0, W_t)$, for $1 \leq i \leq N$. Firstly, for $S_N := \{1, \cdots, N\}$, consider the stochastic non-interacting particle systems

$$\begin{cases}
    dX^\varepsilon_{t,i} = b(X^\varepsilon_{t,i}, X^\varepsilon_{t-}, \mu^\varepsilon_t, \mu^\varepsilon_{t-}, \theta)dt \\
    \quad + \varepsilon \sigma(X^\varepsilon_{t,i}, X^\varepsilon_{t-}, \mu^\varepsilon_t, \mu^\varepsilon_{t-})dW^i_t, & t \in (0, T], \\
    X^\varepsilon_{0,i} = \xi_{s,i}, & s \in [-\tau, 0],
\end{cases} \tag{2.4}$$

where $\mu^\varepsilon_{t,i} := \mathcal{L}_{X^\varepsilon_{t,i}}$ denotes the law of $X^\varepsilon_{t,i}$, $i \in S_N$. By virtue of the weak uniqueness due to Theorem 3.1, it easy to see $\mu^\varepsilon_t = \mu^\varepsilon_{t,i}$, $i \in S_N$. Set $\tilde{\mu}^\varepsilon_N$ be the empirical distribution corresponding to $X^\varepsilon_{1,i}, X^\varepsilon_{2,i}, \cdots, X^\varepsilon_{N,i}$, namely,

$$\tilde{\mu}^\varepsilon_N = \frac{1}{N} \sum_{j=1}^{N} \delta_{X^\varepsilon_{t,j}}, \quad t \geq -\tau. \tag{2.5}$$

Secondly, the stochastic interacting particle systems can be described as

$$\begin{cases}
    dX^\varepsilon_{t,i,N} = b(X^\varepsilon_{t,i,N}, X^\varepsilon_{t-}, \mu^\varepsilon_{t-N}, \mu^\varepsilon_{t-}, \theta)dt \\
    \quad + \varepsilon \sigma(X^\varepsilon_{t,i,N}, X^\varepsilon_{t-}, \mu^\varepsilon_{t-N}, \mu^\varepsilon_{t-})dW^i_t, & t \in (0, T], \\
    X^\varepsilon_{0,i,N} = \xi_s, & s \in [-\tau, 0],
\end{cases} \tag{2.6}$$
where \( \mu_{\epsilon,N} \) stands for the empirical distribution corresponding to \( X_{\epsilon,1,N}, X_{\epsilon,2,N}, \ldots, X_{\epsilon,N,N} \), namely,

\[
\mu_{\epsilon,N} = \frac{1}{N} \sum_{j=1}^{N} \delta_{X_{\epsilon,j,N}^t}, \quad t \geq -\tau.
\]

(2.7)

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 \( \mu, \nu \in \mathcal{P}_p (\mathbb{R}^d) \).

Assume that there exist sufficiently large integers \( n, m \in \mathbb{N} \) such that \( \delta := \frac{T}{n} = \frac{\tau}{m} \) and \( \delta \in (0, 1) \). In this paper, our main work is to explore the LSE on the parameter \( \theta \in \Theta \) based on the corresponding stochastic interacting particle systems with a small dispersion \( \epsilon \) whenever the stepsize \( \delta \) approaches to zero and the particle number \( N \) goes to infinity. In order to improve the simulation precision of distribution-dependent SDDE (2.3), by virtue of the observations of stochastic interacting particle system at regularly spaced time points \( t_k = k\delta \) for \( k = 0, 1, \ldots, n \), we construct the following contrast function

\[
\Psi_{i,N}^{\epsilon,i,N}(\theta) = \epsilon^{-2} \delta^{-1} \sum_{k=1}^{n} (P_{k,i,N}^{\epsilon,i,N}(\theta))^* \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta-\tau}^{\epsilon,i,N}) P_{k,i,N}^{\epsilon,i,N}(\theta),
\]

(2.8)

where

\[
P_{k,i,N}^{\epsilon,i,N}(\theta) = X_{k\delta}^{\epsilon,i,N} - X_{(k-1)\delta}^{\epsilon,i,N} - b(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta-\tau}^{\epsilon,i,N}, \mu_{(k-1)\delta}, \mu_{(k-1)\delta-\tau}, \theta)\delta,
\]

(2.9)

for \( k = 1, 2, \ldots, n \).

Note that this contrast function is least squares quadratic. The method of least squares is about estimating parameters by minimizing the quadratic (2.8). Then, according to the principle of least squares method, to achieve the least squares estimation of \( \theta \in \Theta \), we need to seek an argument \( \hat{\theta}_{i,N}^{\epsilon,i,N} \in \Theta \) such that

\[
\Psi_{i,N}^{\epsilon,i,N}(\hat{\theta}_{i,N}^{\epsilon,i,N}) = \min_{\theta \in \Theta} \Psi_{i,N}^{\epsilon,i,N}(\theta),
\]

namely,

\[
\hat{\theta}_{i,N}^{\epsilon,i,N} = \arg \min_{\theta \in \Theta} \Psi_{i,N}^{\epsilon,i,N}(\theta).
\]

(2.10)

Let

\[
\Phi_{i,N}^{\epsilon,i,N}(\theta) = \epsilon^{-2} (\Psi_{i,N}^{\epsilon,i,N}(\theta) - \Psi_{i,N}^{\epsilon,i,N}(\theta_0)).
\]

Then, from (2.10), one has

\[
\Phi_{i,N}^{\epsilon,i,N}(\hat{\theta}_{i,N}^{\epsilon,i,N}) = \min_{\theta \in \Theta} \Phi_{i,N}^{\epsilon,i,N}(\theta),
\]

which implies

\[
\hat{\theta}_{i,N}^{\epsilon,i,N} = \arg \min_{\theta \in \Theta} \Phi_{i,N}^{\epsilon,i,N}(\theta).
\]

(2.11)

That is to say, \( \hat{\theta}_{i,N}^{\epsilon,i,N} \) satisfying (2.11) is called as the LSE of \( \theta \in \Theta \).

**Remark 2.** There have been extensive researches on parameter estimation of distribution-independent SDEs, for example, [5, 11, 12, 13, 14]. As is natural to expect, parameter estimation on distribution-dependent SDEs could be obtained by the least square method.
To obtain the main results, we give the following assumptions on the coefficients of (2.3). For any \( x_i, y_i \in \mathbb{R}^d \) and \( \mu_i, \nu_i \in \mathcal{P}_p(\mathbb{R}^d) \), \( i = 1, 2 \).

(A1) There exists \( K_1 > 0 \) such that
\[
\sup_{\theta \in \Theta} |b(x_1, y_1, \mu_1, \nu_1, \theta) - b(x_2, y_2, \mu_2, \nu_2, \theta)| \\
\leq K_1 \{|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2) + \mathbb{W}_2(\nu_1, \nu_2)\};
\]

(A2) There exists \( K_2 > 0 \) such that
\[
\|\sigma(x_1, y_1, \mu_1, \nu_1) - \sigma(x_2, y_2, \mu_2, \nu_2)\| \\
\leq K_2 \{|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2) + \mathbb{W}_2(\nu_1, \nu_2)\};
\]

(A3) \((\sigma\sigma^*) (x_i, y_i, \mu_i, \nu_i) (i = 1, 2)\) is invertible, and there exists \( K_3 > 0 \) such that
\[
\| (\sigma\sigma^*)^{-1} (x_1, y_1, \mu_1, \nu_1) - (\sigma\sigma^*)^{-1} (x_2, y_2, \mu_2, \nu_2)\| \\
\leq K_3 \{|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2) + \mathbb{W}_2(\nu_1, \nu_2)\};
\]

(A4) There exists \( K_4 > 0 \) such that
\[
\sup_{\theta \in \Theta} \|(\nabla_b \sigma)(x_1, y_1, \mu_1, \nu_1, \theta) - (\nabla_b \sigma)(x_2, y_2, \mu_2, \nu_2, \theta)\| \\
\leq K_4 \{|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2) + \mathbb{W}_2(\nu_1, \nu_2)\},
\]

where \((\nabla_b \sigma)\) means the gradient operator w.r.t. the fifth spatial variable;

(A5) There exists \( K_5 > 0 \) such that
\[
\sup_{\theta \in \Theta} \|(\nabla_b^{(2)} \sigma^*)(x_1, y_1, \mu_1, \nu_1, \theta) - (\nabla_b^{(2)} \sigma^*)(x_2, y_2, \mu_2, \nu_2, \theta)\| \\
\leq K_5 \{|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2) + \mathbb{W}_2(\nu_1, \nu_2)\},
\]

where \(\nabla_b^{(2)} \sigma^*\) is defined in (2.2).

3. The consistency of LSE. In the previous section, we constructed a contrast function via stochastic interacting particle systems. This section investigates the consistency of the least squares estimator by means of a series of lemmas. We start this section by introducing the following deterministic ordinary differential equation
\[
\begin{cases}
dX_{t}^{0, i} = b(X_{t}^{0, i}, X_{t-\tau}^{0, i}, \mu_{t-\tau}, \mu_0, \theta_0)dt, \; t > 0, \\
X_{s}^{0, i} = \xi(s), \; s \in [-\tau, 0], \; i \in \mathcal{S}_{N}.
\end{cases}
\]

Where \(\mu_{t} = \mu_{0, i} := \mathcal{L}_{X_{t}^{0, i}}\) denotes the law of \(X_{t}^{0, i}\). Under (A1), (3.1) admits a unique solution \((X_{t}^{0})_{t \geq -\tau}\). It is worth pointing out that (2.3), (2.4), (2.6) and (3.1) share the same initial data.

**Lemma 3.1.** Assume that (A1) and (A2) hold, for any initial value \(X_{0}^{*} = \xi \in C^{0}_{\mathcal{F}_{0}}([-\tau, 0]; \mathbb{R}^{d})\) with \(\xi \in \mathcal{P}_p(\mathbb{R}^d)\), \(p \geq 2\), (2.3) possesses a unique strong solution \((X_{t}^{*})_{t \geq -\tau}\) with
\[
\mathbb{E} \left( \sup_{-\tau \leq t \leq T} |X_{t}^{*}|^p \right) \leq C(1 + \mathbb{E}||\xi||_{p}^2) < \infty.
\]
Proof. We will prove this result by iterating in distribution which is similar to the argument of [1]. For each $k \geq 1$, consider the following distribution-iterated SDEs

\[
\begin{align*}
\frac{dX_t^{\varepsilon,(k)}}{dt} &= b(X_t^{\varepsilon,(k)}, X_{t-}^{\varepsilon,(k)}, \mu_t^{(k)}, \mu_{t-}^{(k)}, \theta)dt \\
& \quad + \varepsilon \sigma(X_t^{\varepsilon,(k)}, X_{t-}^{\varepsilon,(k)}, \mu_t^{(k)}, \mu_{t-}^{(k)})dW_t, \quad t \in (0, T], \\
X_s^{\varepsilon,(k)} &= X_s^{\varepsilon} = \xi(s), \quad s \in [-\tau, 0],
\end{align*}
\]  
(3.3)

where $\mu^{(k)} := X_{t-}^{\varepsilon,(k)}$. For $k = 0$, let $X_t^{\varepsilon,(0)} \equiv \xi$ and $\mu^{(0)} := X_{t-}^{\varepsilon,(0)}$.

First, we shall show (3.3) has a unique strong solution $(X_t^{\varepsilon,(k)})_{-\tau \leq t \leq T}$ with

\[
E \left( \sup_{-\tau \leq t \leq T} |X_t^{\varepsilon,(k)}|^p \right) \leq C < \infty, \quad k \geq 1.
\]  
(3.4)

For $k = 1$ and $t \geq 0$, set

\[
b(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}, \mu_t^{(0)}, \mu_{t-}^{(0)}, \theta) =: \tilde{b}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}, \theta)
\]

and

\[
\varepsilon \sigma(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}, \mu_t^{(0)}, \mu_{t-}^{(0)}) =: \tilde{\sigma}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}).
\]

Then (3.3) can be written as

\[
\frac{dX_t^{\varepsilon,(1)}}{dt} = \tilde{b}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}, \theta)dt + \tilde{\sigma}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)})dW_t, \quad X_0^{\varepsilon,(1)} = X_0^\varepsilon, \quad t \geq 0.
\]  
(3.5)

In view of (A1) and (A2), the coefficients $\tilde{b}$ and $\tilde{\sigma}$ are Lipschitzian, which, by the standard Banach fixed point theorem, implies that (3.5) has a unique and non-explosive strong solution $(X_t^{\varepsilon,(1)})_{t \geq -\tau}$ (see, e.g., [15, p.51]). In addition, for any $x, y \in \mathbb{R}^d$ and $\nu, \mu \in \mathcal{P}_p(\mathbb{R}^d)$, by assumptions (A1) - (A2), it is easy to see that there is a constant $L > 0$ such that

\[
|b(x, y, \mu, \nu, \theta)| \leq L \left(1 + |x| + |y| + W_2(\mu, \delta_0) + W_2(\nu, \delta_0)\right)
\]  
(3.6)

and

\[
\|\sigma(x, y, \mu, \nu)\| \leq L \left(1 + |x| + |y| + W_2(\mu, \delta_0) + W_2(\nu, \delta_0)\right).
\]  
(3.7)

For any $p \geq 2$, $T_M = \inf \{t > 0 : |X_t^{\varepsilon,(1)}| \geq M\}$, by virtue of (3.6), (3.7), the Burkholder-Davis-Gundy inequality and Hölder inequality, one has

\[
\begin{align*}
& E \left( \sup_{0 \leq t \leq T \wedge T_M} |X_t^{\varepsilon,(1)}|^p \right) \\
\leq & 3^{p-1} \left\{ \xi_0 \right\}^p + \mathbb{E} \left( \sup_{0 \leq t \leq T \wedge T_M} \left| \int_0^t \tilde{b}(X_s^{\varepsilon,(1)}, X_{s-}^{\varepsilon,(1)}, \theta)s \right|^p \right) \\
& + \mathbb{E} \left( \sup_{0 \leq t \leq T \wedge T_M} \left| \int_0^t \tilde{\sigma}(X_s^{\varepsilon,(1)}, X_{s-}^{\varepsilon,(1)})dW_s \right|^p \right) \\
\leq & C \left\{ 1 + \mathbb{E} \int_0^{T \wedge T_M} |\tilde{b}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}, \theta)|^p dt + \mathbb{E} \int_0^{T \wedge T_M} \left| \tilde{\sigma}(X_t^{\varepsilon,(1)}, X_{t-}^{\varepsilon,(1)}) \right|^p dt \right\} \\
\leq & C \left\{ 1 + \mathbb{E} \int_0^{T \wedge T_M} \left(1 + |X_t^{\varepsilon,(1)}|^p + |X_{t-}^{\varepsilon,(1)}|^p + W_2(\mu_t^{(0)}, \delta_0)^p + W_2(\mu_{t-}^{(0)}, \delta_0)^p \right) dt \right\} \\
\leq & C \left\{ 1 + \int_0^T \mathbb{E} \left( \sup_{0 \leq s \leq t \wedge T_M} |X_s^{\varepsilon,(1)}|^p \right) dt + \mathbb{E} \int_0^{T \wedge T_M} \left( W_2(\mu_t^{(0)}, \delta_0)^p + W_2(\mu_{t-}^{(0)}, \delta_0)^p \right) dt \right\}.
\end{align*}
\]
Next, applying the Gronwall inequality, one has

$$\mathbb{E}\left( \sup_{0 \leq t \leq T} |X_t^{\varepsilon,(1)}|^p \right) \leq C \left\{ 1 + \mathbb{E} \int_0^T \left( \mathbb{W}_2(\mu_t^{(0)}, \delta_0)^p + \mathbb{W}_2(\mu_{t-}^{(0)}, \delta_-)^p \right) dt \right\}.$$ 

Let $M \to \infty$, for any $p \geq 2$, we derive that

$$\mathbb{E}\left( \sup_{-T \leq t \leq T} |X_t^{\varepsilon,(1)}|^p \right) \leq C \left\{ 1 + \mathbb{E} \int_{-T}^T \left( \mathbb{W}_2(\mu_t^{(0)}, \delta_0)^p + \mathbb{W}_2(\mu_{t-}^{(0)}, \delta_-)^p \right) dt \right\} \leq (1 + \mathbb{E}\|\xi\|_p^p) \leq C < \infty.$$ 

So, for $k = 1$, (3.4) holds.

Next, we assume that the assertion (3.4) holds for $k = k_0$ for some $k_0 > 1$. Repeating the previous procedures with $(X_t^{\varepsilon,(k_0+1)}, X_t^{\varepsilon,(k_0+1)}, \mu_t^{(k_0)}, \mu_{t-}^{(k_0)})$ replacing $(X_t^{\varepsilon,(1)}, X_t^{\varepsilon,(1)}, \mu_t^{(0)}, \mu_{t-}^{(0)})$, we can show that (3.4) holds for $k = k_0 + 1$, and omit the details.

Set

$$Z_t^{\varepsilon,(k)} := X_t^{\varepsilon,(k)} - X_t^{\varepsilon,(k-1)},$$

$$B_t^{\varepsilon,(k)} := b(X_t^{\varepsilon,(k)}, X_t^{\varepsilon,(k)}, \mu_t^{(k-1)}, \mu_{t-}^{(k-1)}, \theta) - b(X_t^{\varepsilon,(k-1)}, X_t^{\varepsilon,(k-1)}, \mu_t^{(k-2)}, \mu_{t-}^{(k-2)}, \theta)$$

and

$$\Sigma_t^{\varepsilon,(k)} := \sigma(X_t^{\varepsilon,(k)}, X_t^{\varepsilon,(k)}, \mu_t^{(k-1)}, \mu_{t-}^{(k-1)}) - \sigma(X_t^{\varepsilon,(k-1)}, X_t^{\varepsilon,(k-1)}, \mu_t^{(k-2)}, \mu_{t-}^{(k-2)}).$$

By the Itô formula, one has

$$|Z_t^{\varepsilon,(k+1)}|^2 = 2 \int_0^t \langle Z_s^{\varepsilon,(k+1)}, B_s^{\varepsilon,(k+1)} \rangle ds + \varepsilon^2 \int_0^t \text{trace}[(\Sigma_s^{\varepsilon,(k+1)})^* (\Sigma_s^{\varepsilon,(k+1)})] ds$$

$$+ 2\varepsilon \int_0^t \langle Z_s^{\varepsilon,(k+1)}, \Sigma_s^{\varepsilon,(k+1)} \rangle dW_s \leq 2 \int_0^t |Z_s^{\varepsilon,(k+1)}||B_s^{\varepsilon,(k+1)}| ds + \varepsilon^2 \int_0^t \|\Sigma_s^{\varepsilon,(k+1)}\|^2 ds$$

$$+ 2\varepsilon \int_0^t \langle Z_s^{\varepsilon,(k+1)}, \Sigma_s^{\varepsilon,(k+1)} \rangle dW_s.$$
For $I_3$, by the Burkholder-Davis-Gundy inequality, the Young inequality and the assumption (A2), one has
\[
I_3 \leq 8 \sqrt{2} \varepsilon E \left( \int_0^t \left| Z_s^{(k+1)} \right|^2 \left| \Sigma_s \right| ds \right)^{1/2} \\
\leq \frac{1}{2} E \left( \sup_{0 \leq s \leq t} \left| Z_s^{(k+1)} \right|^2 \right) + C \varepsilon^2 E \left( \int_0^t \left| \Sigma_s \right| ds \right) \\
+ \mathbb{W}_2(\mu_s^{(k)}, \mu_s^{(k-1)})^2 + \mathbb{W}_2(\mu_s, \mu_s^{(k)})^2 ds.
\]
Plugging these inequalities into (3.8), we get
\[
E \left( \sup_{0 \leq s \leq t} \left| Z_s^{(k+1)} \right|^2 \right) \leq C(1 + \varepsilon^2) \int_0^t \left[ E \left( \sup_{0 \leq r \leq s} \left| Z_r^{(k+1)} \right|^2 \right) + E \left| Z_s^{(k)} \right|^2 \right] ds.
\]
Then, the Gronwall inequality leads to
\[
E \left( \sup_{0 \leq s \leq t} \left| Z_s^{(k+1)} \right|^2 \right) \leq C(1 + \varepsilon^2) e^{(1 + \varepsilon^2)t} \int_0^t E \left| Z_s^{(k)} \right|^2 ds \\
\leq Ct(1 + \varepsilon^2) e^{(1 + \varepsilon^2)t} E \left( \sup_{0 \leq s \leq t} \left| Z_s^{(k)} \right|^2 \right).
\]
Taking $t_0 > 0$ such that $Ct_0(1 + \varepsilon^2) e^{(1 + \varepsilon^2)t_0} \leq e^{-1}$, one has, for $k \geq 1$,
\[
E \left( \sup_{0 \leq s \leq t_0} \left| Z_s^{(k+1)} \right|^2 \right) \leq e^{-1} E \left( \sup_{0 \leq s \leq t_0} \left| Z_s^{(k)} \right|^2 \right). \tag{3.9}
\]
Setting $k = 0$ leads to
\[
E \left( \sup_{0 \leq s \leq t_0} \left| Z_s^{(1)} \right|^2 \right) \leq 2 E \left( \sup_{0 \leq s \leq t_0} \left| X_s^{(1)} \right|^2 \right) + 2 E \left| X_s^{(0)} \right|^2 \\
\leq 4 E \left( \sup_{0 \leq s \leq t_0} \left| X_s^{(1)} \right|^2 \right) \leq C < \infty.
\]
This, in addition to (3.9), yields that
\[
E \left( \sup_{0 \leq s \leq t_0} \left| Z_s^{(k+1)} \right|^2 \right) \leq e^{-k} E \left( \sup_{0 \leq s \leq t_0} \left| Z_s^{(1)} \right|^2 \right) \leq 4 e^{-k} E \left( \sup_{0 \leq s \leq t_0} \left| X_s^{(1)} \right|^2 \right) =: C e^{-k}.
\]
So, there exists an $\mathcal{F}_t$-adapted continuous stochastic process $(X_t^\varepsilon)_{t \in [0,t_0]}$ with $X_0^\varepsilon = \xi$ and $\mu_t^\varepsilon = \mathcal{L}X_t^\varepsilon$ such that
\[
\lim_{k \to \infty} \sup_{0 \leq s \leq t_0} W_2(\mu_t^\varepsilon, \mu_t^{(k)}) \leq \lim_{k \to \infty} \left( E \left( \sup_{0 \leq s \leq t_0} \left| X_t^{(k)} - X_t^\varepsilon \right|^2 \right) \right)^{1/2} = 0. \tag{3.10}
\]
By virtue of (3.3), for $k \geq 1$,
\[
X_t^{(k)} = \xi + \int_0^t b(X_s^{(k)}, X_s^{(k)}, \mu_s^{(k-1)}, \mu_s^{(k-1)}, \theta) ds \\
+ \varepsilon \int_0^t \sigma(X_s^{(k)}, X_s^{(k)}, \mu_s^{(k-1)}, \mu_s^{(k-1)}) dW_s.
\]
Thus, by (3.10), together with the assumptions (A1) and (A2), using the dominated convergence theorem leads that $\mathbb{P}$-a.s.
\[
X_t^\varepsilon = \xi + \int_0^t b(X_s, X_s^{\varepsilon-}, \mu_s^\varepsilon, \mu_s^{\varepsilon-}, \theta) ds + \varepsilon \int_0^t \sigma(X_s, X_s^{\varepsilon-}, \mu_s^\varepsilon, \mu_s^{\varepsilon-}) dW_s, \quad t \in [0, t_0].
\]
Henceforth, the existence of solution to (2.3) up to time $t_0$ is available, and (3.10) leads to
$$
\mathbb{E}\left( \sup_{-\tau \leq t \leq t_0} |X_t|^2 \right) \leq C < \infty.
$$
Further, this result for $(X_t^\tau)_{t \in [-\tau, t_0 \wedge T]}$ is true. So, in view of the arbitrariness of $T > 0$, by solving the equation piecewise in time, we deduce that (2.3) admits a strong solution $(X_t^\tau)_{t \geq -\tau}$ with
$$
\mathbb{E}\left( \sup_{-\tau \leq t \leq T} |X_t|^2 \right) \leq C < \infty. \quad (3.11)
$$
Finally, we shall prove the uniqueness of (2.3). For the same initial data, we assume that $(X_t^\tau)_{t \geq 0}$ and $(Y_t^\tau)_{t \geq 0}$ are solutions to (2.3). Set
$$
\Delta_t := X_t^\tau - Y_t^\tau \text{ and } \tilde{\mu}_t := \mathcal{L}_t, \quad t \geq 0.
$$
By the assumptions (A1) and (A2), the Itô formula and Hölder inequality, we get
$$
\mathbb{E}|\Delta_t|^2 \leq 2\mathbb{E}\int_0^t |\Delta_s| |b(X_s^\tau, X_s^{\tau - \cdot}, \mu_s^\tau, \mu_s^{\tau - \cdot}, \theta) - b(Y_s^\tau, Y_s^{\tau - \cdot}, \tilde{\mu}_s^\tau, \tilde{\mu}_s^{\tau - \cdot}, \theta)| ds
+ 4K_2\varepsilon^2 \mathbb{E}\int_0^t \left( |\Delta_s|^2 + |\Delta_s^{\tau - \cdot}|^2 + \mathbb{W}_2(\mu_s^\tau, \tilde{\mu}_s^\tau)^2 + \mathbb{W}_2(\mu_s^{\tau - \cdot}, \tilde{\mu}_s^{\tau - \cdot})^2 \right) ds
$$
$$
\leq \mathbb{E}\int_0^t |\Delta_s|^2 ds + \mathbb{E}\int_0^t |b(X_s^\tau, X_s^{\tau - \cdot}, \mu_s^\tau, \mu_s^{\tau - \cdot}, \theta) - b(Y_s^\tau, Y_s^{\tau - \cdot}, \tilde{\mu}_s^\tau, \tilde{\mu}_s^{\tau - \cdot}, \theta)|^2 ds
+ 16K_2\varepsilon^2 \mathbb{E}\int_0^t |\Delta_s|^2 ds
$$
$$
\leq \mathbb{E}\int_0^t |\Delta_s|^2 ds + 16(K_1^2 + K_2^2\varepsilon^2) \int_0^t \mathbb{E}|\Delta_s|^2 ds
\leq C(1 + K_1^2 + K_2^2\varepsilon^2) \int_0^t \mathbb{E}|\Delta_s|^2 ds. \quad (3.12)
$$
This, together with the Gronwall inequality and (3.11), leads to the uniqueness of (2.3). By iterating piecewise in time, we deduce that (2.3) has a unique strong solution $(X_t^\tau)_{t \geq -\tau}$.

In addition, in view of the assumptions (A1) and (A2), using a standard calculation and the fact $\xi \in C^{\beta}_b([-\tau, 0]; \mathbb{R}^d)$ with $\mathcal{L}_\xi \in \mathcal{P}_p(\mathbb{R}^d)$, we conclude that (3.2) holds.

**Lemma 3.2.** Assume (A1) and (A2). Then, for initial value $X^\xi_{-\tau} \in C^{\beta}_b([-\tau, 0]; \mathbb{R}^d)$ with $\mathcal{L}_\xi \in \mathcal{P}_p(\mathbb{R}^d)$, (2.6) has a strong solution with
$$
\sup_{t \in \mathbb{N}} \mathbb{E}\left( \sup_{-\tau \leq t \leq T} |X_t^\xi|_p \right) < \infty, \quad p \geq 2.
$$

**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
$$
\mu^{X,N} = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}, \quad \mu^{Y,N} = \frac{1}{N} \sum_{i=1}^N \delta_{Y_i},
$$
$$
\hat{b}(X, Y) = (b(X_1, Y_1, \mu^{X,N}, \mu^{Y,N}, \theta), \cdots, b(X_N, Y_N, \mu^{X,N}, \mu^{Y,N}, \theta))^*,
$$
$$
\hat{\sigma}(X, Y) = \text{diag}(\sigma(X_1, Y_1, \mu^{X,N}, \mu^{Y,N}), \cdots, \sigma(X_N, Y_N, \mu^{X,N}, \mu^{Y,N}))
$$
and
\[ \tilde{W}_t = \left( W^{1}_t, \ldots, W^{N}_t \right)^{\ast}. \]

Then, (2.6) can be redescribed as
\[ dX_t = \tilde{b}(X_t, X_{t-})dt + \varepsilon \tilde{\sigma}(X_t, X_{t-})dW_t, \quad t \geq 0. \]

By the assumptions (A1) and (A2), it can be readily seen that, for any \( X, Y \in \mathbb{R}^d \otimes \mathbb{R}^N \), the coefficients \( \tilde{b} \) and \( \tilde{\sigma} \) satisfy Lipschitz condition, which implies that (3.13) has a unique strong solution. The proof is complete. \( \square \)

**Lemma 3.3.** Let (A1) and (A2) hold. For initial value \( X^\tau_0 = \xi \in C^0_{\mathbb{P}, \mathbb{F}}([-\tau, 0]; \mathbb{R}^d) \) with \( \mathcal{L}_\xi \in \mathcal{P}(\mathbb{R}^d) \), \( p > 4 \), it holds that
\[ \sup_{-\tau \leq t \leq T} \mathbb{E}|X^{\tau, i, N}_t - X^{\tau, i}_t|^2 \leq C_N(1 + \varepsilon^2), \quad i \in S_N, \]
where \( C_N \) is a decreasing function with respect to \( N \) and is defined as (3.17).

**Proof.** By (2.4), (2.6), the Itô isometry and Hölder inequality, we derive that, for any \( t \in [0, T] \),
\[ \mathbb{E}|Z^{\tau, i, N}_t|^2 := \mathbb{E}|X^{\tau, i, N}_t - X^{\tau, i}_t|^2 \]
\[ \leq C\mathbb{E} \int_0^t |b(X^{\tau, i, N}_s, X^{\tau, i, N}_s, \mu^{\tau, N}_s, \mu^{\tau, N}_s, \theta) - b(X^{\tau, i}_s, X^{\tau, i}_s, \mu^{\tau, N}_s, \mu^{\tau, N}_s, \theta)|^2 ds \]
\[ + C\varepsilon^2 \mathbb{E} \int_0^t \| \sigma(X^{\tau, i, N}_s, X^{\tau, i, N}_s, \mu^{\tau, N}_s, \mu^{\tau, N}_s) - \sigma(X^{\tau, i}_s, X^{\tau, i}_s, \mu^{\tau, N}_s, \mu^{\tau, N}_s) \|^2 ds \]
\[ \leq C(t + \varepsilon^2) \mathbb{E} \int_0^t \{ |X^{\tau, i, N}_s - X^{\tau, i}_s|^2 + |X^{\tau, i, N}_s - X^{\tau, i}_s|^2 + \mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)^2 \]
\[ + \mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)^2 \} ds \]
\[ \leq C(t + \varepsilon^2) \int_0^t \{ \mathbb{E}|Z^{\tau, i, N}_s|^2 + \mathbb{E}|Z^{\tau, i, N}_s|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \} ds, \]
where in the fourth step we have used the assumptions (A1) and (A2). Moreover, by means of [1, Lemma 3.2] and (2.5), one gets
\[ \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \leq \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \]
\[ \leq \mathbb{E}|Z^{\tau, i, N}_s|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \]
\[ \leq \mathbb{E}|Z^{\tau, i, N}_s|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2. \]

Thus, we infer that
\[ \mathbb{E}|Z^{\tau, i, N}_t|^2 \leq C(t + \varepsilon^2) \int_0^t \{ \mathbb{E}|Z^{\tau, i, N}_s|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \} ds, \]
which, by the Gronwall inequality, implies
\[ \mathbb{E}|Z^{\tau, i, N}_t|^2 \leq C \exp\{ C(t + \varepsilon^2) t \} (t + \varepsilon^2) \int_0^t \{ \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 + \mathbb{E}|\mathbb{W}_2(\mu^{\tau, N}_s, \mu^{\tau, N}_s)|^2 \} ds \]
\[ \leq C_N(t + \varepsilon^2). \]
Here in the second step we have used [3, Theorem 5.8] and
\[ C_N := C \begin{cases} 
N^{-1/2}, & \text{if } d < 4, \\
N^{-1/2} \log N, & \text{if } d = 4, \\
N^{-2/d}, & \text{if } d > 4. 
\end{cases} \tag{3.17} \]
This leads to
\[ \sup_{0 \leq t \leq T} \mathbb{E}|Z^i_t|^2 \leq C_N(1 + \varepsilon^2)T \leq C_N(1 + \varepsilon^2). \]

**Remark 3.** This result reveals that stochastic interacting particle system convergence strongly to the non-interacting particle system when the step $\delta$ goes to zero and the particle number goes to infinity.

**Remark 4.** Here the assumption on the $p$-th moment of the initial data is assigned to guarantee that Glivenko-Cantelli convergence under the Wasserstein distance is available, which is important to the asymptotic behavior of LSE based on stochastic interacting particle systems.

**Lemma 3.4.** Let (A1) and (A2) hold. Then, for any initial value $X_0 = \xi \in C_b^0([-\tau, 0]; \mathbb{R}^d)$ with $\mathcal{L}_t \in \mathcal{P}_p(\mathbb{R}^d)$, $p > 4$, there is a constant $C > 0$ such that
\[ \sup_{0 \leq t \leq T} \mathbb{E}|X_t^{\tau,i} - X_t^{0,i}|^2 \leq C\delta(\delta + \varepsilon^2)(1 + C_N(1 + \varepsilon^2)) + C\varepsilon^2 + C_N\varepsilon^2(1 + \varepsilon^2), \quad i \in \mathcal{S}^N, \tag{3.18} \]
where $t_\delta := \lceil t / \delta \rceil$.

**Proof.** For any $t \in [0, T]$, it holds that
\[ \mathbb{E}|X_t^{\tau,i} - X_t^{0,i}|^2 \leq 2\mathbb{E}|X_t^{\tau,i} - X_t^{\tau,i}|^2 + 2\mathbb{E}|X_t^{\tau,i} - X_t^{0,i}|^2 \tag{3.19} \]
and
\[ \mathbb{E}|X_t^{\tau,i}|^2 = \mathbb{E}|X_t^{\tau,i} - X_t^{\tau,i,N} + X_t^{\tau,i,N}|^2 \leq 2\mathbb{E}|X_t^{\tau,i} - X_t^{\tau,i,N}|^2 + 2\mathbb{E}|X_t^{\tau,i}|^2. \tag{3.20} \]
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, $k_0 = \lceil t / \delta \rceil$. Next, the Hölder inequality and Itô isometry, together with (3.6), (3.7) and (3.20), yield that
\[ \mathbb{E}|X_{t_\delta}^{\tau,i} - X_t^{\tau,i}|^2 \\
= \mathbb{E}|X_{t_\delta}^{\tau,i} - X_t^{\tau,i}|^2 \\
\leq 2\delta \mathbb{E}\int_{k_0\delta}^t |b(X_s^{\tau,i}, X_s^{\tau,i}, \mu_s^{\tau,i}, \mu_s^{\tau,i}, \theta)|^2 ds + 2\varepsilon^2 \mathbb{E}\int_{k_0\delta}^t \|\sigma(X_s^{\tau,i}, X_s^{\tau,i}, \mu_s^{\tau,i}, \mu_s^{\tau,i}, \mu_s^{\tau,i})\|^2 ds \\
\leq C(\delta + \varepsilon^2) \delta \int_{k_0\delta}^t \left\{ 1 + \mathbb{E}|X_s^{\tau,i}|^2 + \mathbb{E}|X_s^{\tau,i}|^2 + \mathbb{E}W_2(\mu_s^{\tau,i}, \delta_0)^2 + \mathbb{E}W_2(\mu_s^{\tau,i}, \delta_0)^2 \right\} ds \\
\leq C(\delta + \varepsilon^2) \int_{k_0\delta}^t \left\{ 1 + \mathbb{E}|X_s^{\tau,i}|^2 + \mathbb{E}|X_s^{\tau,i}|^2 \right\} ds \\
\leq C\delta(\delta + \varepsilon^2)(1 + C_N(1 + \varepsilon^2)), \tag{3.21} \]
in which the last step we have used Lemma 3.2 and 3.3. Moreover, using the same technique as (3.21), we derive that
\[ \mathbb{E}|X_t^{\tau,i} - X_t^{0,i}|^2 \]
\begin{align*}
\leq & Ct \int_0^t \left( |X_s^{\varepsilon,i} - X_0^{\varepsilon,i}|^2 + |X_s^{\varepsilon,i} - X_0^{\varepsilon,i}|^2 + \mathbb{W}_2(\mu_s^{\varepsilon,i}, \mu_s^0)^2 + \mathbb{W}_2(\mu_s^{\varepsilon,i}, \mu_s^{\varepsilon,i})^2 \right) ds \\
& + C\varepsilon^2 \mathbb{E} \int_0^t \|\sigma(X_s^{\varepsilon,i}, X_s^{\varepsilon,i}, \mu_s^{\varepsilon,i}, \mu_s^{\varepsilon,i})\|^2 ds \\
\leq & Ct \int_0^t \left( \mathbb{E}|X_s^{\varepsilon,i} - X_0^{\varepsilon,i}|^2 + \mathbb{E}|X_s^{\varepsilon,i} - X_0^{\varepsilon,i}|^2 \right) ds \\
& + C\varepsilon^2 \int_0^t \left( 1 + \mathbb{E}|X_s^{\varepsilon,i}|^2 + \mathbb{E}|X_s^{\varepsilon,i}|^2 \right) ds \\
\leq & Ct \int_0^t \mathbb{E}|X_s^{\varepsilon,i} - X_0^{\varepsilon,i}|^2 ds + C\varepsilon^2 (t + 1) + C\varepsilon^2 \int_0^t \mathbb{E}|X_s^{\varepsilon,i}|^2 ds,
\end{align*}

which, by the Gronwall inequality, implies to

\[ \mathbb{E}|X_t^{\varepsilon,i} - X_t^{0,\varepsilon,i}|^2 \leq \exp\{Ct^2\} \{ C\varepsilon^2 (t + 1) + C\varepsilon^2 \int_0^t \mathbb{E}|X_s^{\varepsilon,i}|^2 ds \} \]

This, for \( t \in [0, T] \), together with Lemmas 3.2, 3.3 and (3.20), implies that
\[
\sup_{0 \leq t \leq T} \mathbb{E}|X_t^{\varepsilon,i} - X_t^{0,\varepsilon,i}|^2 \\
\leq \exp\{CT^2\} \{ C\varepsilon^2 (T + 1) + C\varepsilon^2 \int_0^T \mathbb{E}|X_s^{\varepsilon,i}|^2 ds \} \\
\leq \exp\{CT^2\} \{ C\varepsilon^2 + C\varepsilon^2 \int_0^T (\mathbb{E}|X_s^{\varepsilon,i} - X_s^{0,i,N}|^2 + \mathbb{E}|X_s^{\varepsilon,i,N}|^2) ds \} \\
\leq & C\varepsilon^2 + C\varepsilon^2 \int_0^T (\mathbb{E}|X_s^{\varepsilon,i} - X_s^{0,i,N}|^2 + \mathbb{E}|X_s^{\varepsilon,i,N}|^2) ds \\
\leq & C N \varepsilon^2 (1 + \varepsilon^2) + C\varepsilon^2.
\]

(3.22)

Plugging (3.21) and (3.22) into (3.19) yields (3.18).

\[ \square \]

**Lemma 3.5.** Let (A1)-(A3) hold. Then, for \( X_0^{\varepsilon} = \xi \in \mathcal{C}_b\left([-\tau, 0]; \mathbb{R}^d\right) \) with \( \mathcal{L}\xi \in \mathcal{P}_p(\mathbb{R}^d) \), \( p > 4 \),

\[ \Phi_{n,\varepsilon}^{\varepsilon,i,N}(1) := \sum_{k=1}^{n} B^*(X^{\varepsilon,i,N}_{(k-1)\delta}, X^{\varepsilon,i,N}_{(k-1)\delta}, \theta_0, \theta) \Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta}, X^{\varepsilon,i,N}_{(k-1)\delta}, \theta_0, \theta) P_k^{\varepsilon,i,N}(\theta_0) \to 0, \]

in \( L^1 \) as \( \varepsilon \to 0 \) and \( n, N \to \infty \).

**Proof.** According to (2.7), we get
\[
\mathbb{W}_2(\mu_s^{\varepsilon,i,N}, \delta_0)^2 \leq \frac{1}{N} \sum_{i=1}^{N} |X_s^{\varepsilon,i,N}|^2, \quad s \geq -\tau.
\]

(3.23)

In view of (2.6) and (2.9), one has
\[
\Phi_{n,\varepsilon}^{\varepsilon,i,N}(1) := \sum_{k=1}^{n} B^*(X^{\varepsilon,i,N}_{(k-1)\delta}, X^{\varepsilon,i,N}_{(k-1)\delta}, \theta_0, \theta) \Lambda^{-1}(X^{\varepsilon,i,N}_{(k-1)\delta}, X^{\varepsilon,i,N}_{(k-1)\delta}, \theta_0, \theta) \\
\times \sigma(X^{\varepsilon,i,N}_{(k-1)\delta}, X^{\varepsilon,i,N}_{(k-1)\delta}, \mu_{(k-1)\delta}, \mu_{(k-1)\delta}) (W^{i}_{k\delta} - W^{i}_{(k-1)\delta}) \\
= \varepsilon \int_0^T B^*(X^{\varepsilon,i,N}_{s\delta}, X^{\varepsilon,i,N}_{s\delta}, \theta_0, \theta) \Lambda^{-1}(X^{\varepsilon,i,N}_{s\delta}, X^{\varepsilon,i,N}_{s\delta}) \\
\times \sigma(X^{\varepsilon,i,N}_{s\delta}, X^{\varepsilon,i,N}_{s\delta}, \mu_{s\delta}, \mu_{s\delta}) dW^{i}_{s\delta},
\]
where \( s_k := [s/\delta] \delta \). This, together with the Hölder inequality and Itô isometry, further implies that

\[
\mathbb{E} \left| \Phi_{n,\varepsilon}^{i,N(1)} \right| \leq \varepsilon \left( \mathbb{E} \left| \int_0^T B^*(X_{s_k}, X_{s_k-\tau}, \theta_0, \theta) \Lambda^{-1}(X_{s_k}, X_{s_k-\tau}) \right|^2 \right)^{\frac{1}{2}} \times \mathbb{E} \left( \left| \int_0^T \sigma(X_{s_k}, X_{s_k-\tau}, \mu^{\varepsilon,N}, \mu^{\varepsilon,N}_s) dW_s \right|^2 \right)^{\frac{1}{2}}
\leq \varepsilon \left( \mathbb{E} \left| \int_0^T |B^*(X_{s_k}, X_{s_k-\tau}, \theta_0, \theta)|^2 \right|^2 \right)^{\frac{1}{2}} \times \mathbb{E} \left( \left| \int_0^T \sigma(X_{s_k}, X_{s_k-\tau}, \mu^{\varepsilon,N}, \mu^{\varepsilon,N}_s) dW_s \right|^2 \right)^{\frac{1}{2}}.
\]

On the other hand, for any \( x, y \in \mathbb{R}^d \) and \( \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d) \), by the assumption (A3), it is easy to see that there is a constant \( L > 0 \) such that

\[
\left\| (\sigma \sigma^*)^{-1}(x, y, \mu, \nu) \right\| \leq L \left\{ 1 + |x| + |y| + \mathbb{W}_2(\mu, \delta_0) + \mathbb{W}_2(\nu, \delta_0) \right\}.
\]

Now, it follows from (3.6) that

\[
|B^*(X_{s_k}, X_{s_k-\tau}, \theta_0, \theta)|^2 \leq C \left\{ 1 + |X_{s_k}|^2 + |X_{s_k-\tau}|^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 \right\}.
\]

Due to (3.7), we obtain

\[
\left\| \sigma(X_{s_k}, X_{s_k-\tau}, \mu^{\varepsilon,N}, \mu^{\varepsilon,N}_s) \right\|^2 \leq C \left\{ 1 + |X_{s_k}|^2 + |X_{s_k-\tau}|^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 \right\}.
\]

From (3.25), one has

\[
\left\| \Lambda^{-1}(X_{s_k}, X_{s_k-\tau}) \right\|^2 \leq C \left\{ 1 + |X_{s_k}|^2 + |X_{s_k-\tau}|^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 \right\}.
\]

Substituting these inequalities into (3.24), by the Hölder inequality and using (3.23) lead to

\[
\mathbb{E} \left| \Phi_{n,\varepsilon}^{i,N(1)} \right| \leq C \varepsilon \left\{ \mathbb{E} \left( \int_0^T \left\{ 1 + |X_{s_k}|^2 + |X_{s_k-\tau}|^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \delta_0)^2 \right\}^3 \right) \right\}^{\frac{1}{2}} \leq C \varepsilon \left\{ \int_0^T \left( 1 + \mathbb{E}|X_{s_k}|^6 + \mathbb{E}|X_{s_k-\tau}|^6 + \left( \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_{s_k-j}|^6 \right) \right) \right\}^{\frac{1}{2}} \leq C \varepsilon,
\]

where the last step is due to Lemma 3.2. Hence, the desired result holds by taking \( \varepsilon \) sufficiently small and \( n, N \) sufficiently large.

\[\square\]

**Lemma 3.6.** Let (A1)-(A3) hold. Then, for \( X_0 = \xi \in \mathcal{C}_b([-\tau, 0]; \mathbb{R}^d) \) with \( \mathcal{L}_\xi \in \mathcal{P}_p(\mathbb{R}^d) \), \( p > 4 \),

\[
\delta \sum_{k=1}^n B^*(X_{(k-1)\delta}, X_{(k-1)\delta-\tau}, \theta_0, \theta) \Lambda^{-1}(X_{(k-1)\delta}, X_{(k-1)\delta-\tau})
\]

\[\square\]
in $L^1$ as $\varepsilon \to 0$, $N \to \infty$ and $\delta \to 0$, where

$$\Pi(\theta) := \int_0^T B^*(X_{t+i}^0, X_{t-i}^0, \theta_0, \theta) \Lambda^{-1}(X_{t+i}^0, X_{t-i}^0) B(X_{t+i}^0, X_{t-i}^0, \theta_0, \theta) dt.$$

**Proof.** Obviously,

$$\Phi_{n,\varepsilon}(2)(\theta)$$

$$= \delta \sum_{k=1}^n B^*(X_{k-1}^\varepsilon, X_{k+1}^\varepsilon, \theta_0, \theta) \Lambda^{-1}(X_{k-1}^\varepsilon, X_{k+1}^\varepsilon) B(X_{k-1}^\varepsilon, X_{k+1}^\varepsilon, \theta_0, \theta)$$

Thus, by calculating directly, one has

$$\Phi_{n,\varepsilon}(2)(\theta) - \Pi(\theta)$$

In addition, for any $x_i, y_i \in \mathbb{R}^d$ and $\mu_i, \nu_i \in \mathcal{P}_p(\mathbb{R}^d)$, $i = 1, 2$, it follows from the assumption $(A1)$ that

$$|B(x_1, y_1, \theta_0, \theta) - B(x_2, y_2, \theta_0, \theta)| \leq 2K_1 \{|x_1 - x_2| + |y_1 - y_2| + \mathcal{W}_2(\mu_1, \mu_2) + \mathcal{W}_2(\nu_1, \nu_2)\}.$$  \hspace{1cm} (3.27)

This, together with (3.15), (3.16), (3.23) and (3.25), implies that

$$\mathbb{E}\left|B(X_{s_i}^\varepsilon, X_{s_{i-1}}^\varepsilon, \theta_0, \theta) - B(X_{s_i}^0, X_{s_{i-1}}^0, \theta_0, \theta)\right|^2$$

$$\leq C\mathbb{E}\left|B(X_{s_i}^\varepsilon, X_{s_{i-1}}^\varepsilon, \theta_0, \theta) - B(X_{s_i}^0, X_{s_{i-1}}^0, \theta_0, \theta)\right|^2$$

$$+ C\mathbb{E}\left|B(X_{s_i}^i, X_{s_{i-1}}^i, \theta_0, \theta) - B(X_{s_i}^0, X_{s_{i-1}}^0, \theta_0, \theta)\right|^2$$

$$\leq C\left\{\mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^\varepsilon - X_{s_i}^\varepsilon|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2 + \mathbb{E}|X_{s_i}^i - X_{s_i}^i|^2ight\}.$$
\[
\leq C_N(1 + \varepsilon^2) + C_N + C\delta(\delta + \varepsilon^2)(1 + C_N(1 + \varepsilon^2)) + C\varepsilon^2 + C_N\varepsilon^2(1 + \varepsilon^2), \quad (3.28)
\]
where in the last step we have used Lemmas 3.3 and 3.4. Moreover, by (3.25) and (3.26), one has
\[
\mathbb{E}\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau})\|_p^2 : |B(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}, \theta_0, \theta_0)|^2
\leq C\mathbb{E}\left\{1 + |X^{\varepsilon,i,N}_{s}|^2 + |X^{\varepsilon,i,N}_{s-\tau}|^2 + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s}, \delta_0)^2 + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s-\tau}, \delta_0)^2\right\}^2
\leq C\left\{1 + \mathbb{E}|X^{\varepsilon,i,N}_{s}|^4 + \mathbb{E}|X^{\varepsilon,i,N}_{s-\tau}|^4 + \frac{1}{N}\sum_{j=1}^{N} \mathbb{E}|X^{\varepsilon,j,N}_{s}|^4 + \frac{1}{N}\sum_{j=1}^{N} \mathbb{E}|X^{\varepsilon,j,N}_{s-\tau}|^4\right\}
\leq C,
\]  
(3.29)
where in the second step used (3.23). Then, the Hölder inequality implies that
\[
\mathbb{E}|J_1|
\leq \int_0^T \left(\mathbb{E}\left[|B(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}, \theta_0, \theta_0) - B(X^{0,i}_{s}, X^{0,i}_{s-\tau}, \theta_0, \theta_0)|^2\right]\right) \frac{1}{2} ds
\leq \int_0^T \left(\mathbb{E}\left[\|\Lambda^{-1}(X^{\varepsilon,i,N}_{[s/\delta]}, X^{\varepsilon,i,N}_{[s/\delta]-\delta})\|_p^2 : |B(X^{\varepsilon,i,N}_{[s/\delta]}, X^{\varepsilon,i,N}_{[s/\delta]-\delta}, \theta_0, \theta_0)|^2\right]\right) \frac{1}{2} ds
\leq \int_0^T \left(\mathbb{E}\left[\Lambda^{-1}(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau})\right] : |B(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}, \theta_0, \theta_0)|^2\right) \frac{1}{2} ds
\leq C\left(\mathbb{E}\left[\Lambda^{-1}(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau})\right] : |B(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}, \theta_0, \theta_0)|^2\right) \frac{1}{2}.
\]  
(3.30)
Here in the second step used (3.28) and (3.29), and in the last step utilized lemma 3.2. For the estimate of \( J_2 \), we first seek some inequalities of the integrands. By (3.6), we find out
\[
|B^*(X^{0,i}_{s}, X^{0,i}_{s-\tau}, \theta_0, \theta_0)|
\leq L\left\{1 + |X^{0,i}_{s}| + |X^{0,i}_{s-\tau}| + \mathbb{W}_2(\mu^{0,i}_{s}, \delta_0) + \mathbb{W}_2(\mu^{0,i}_{s-\tau}, \delta_0)\right\}
\]  
(3.31)
and
\[
|B^*(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}, \theta_0, \theta_0)|
\leq L\left\{1 + |X^{\varepsilon,i,N}_{s}| + |X^{\varepsilon,i,N}_{s-\tau}| + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s}, \delta_0) + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s-\tau}, \delta_0)\right\}.
\]
By means of the assumption (A3), one gets
\[
\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}) - \Lambda^{-1}(X^{0,i}_{s}, X^{0,i}_{s-\tau})\|
\leq C\left\{|X^{\varepsilon,i,N}_{s} - X^{0,i}_{s}| + |X^{\varepsilon,i,N}_{s-\tau} - X^{0,i}_{s-\tau}| + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s}, \mu^{0,i}_{s}) + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s-\tau}, \mu^{0,i}_{s-\tau}) + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s}, \mu^{\varepsilon,i,N}_{s-\tau}) + \mathbb{W}_2(\mu^{\varepsilon,i,N}_{s-\tau}, \mu^{\varepsilon,i,N}_{s})\right\}
\]
and
\[
\|\Lambda^{-1}(X^{\varepsilon,i,N}_{s}, X^{\varepsilon,i,N}_{s-\tau}) - \Lambda^{-1}(X^{0,i}_{s}, X^{0,i}_{s-\tau})\|
\leq C\left\{|X^{\varepsilon,i,N}_{s} - X^{0,i}_{s}| + |X^{\varepsilon,i,N}_{s-\tau} - X^{0,i}_{s-\tau}| + \mathbb{W}_2(\mu^{\varepsilon,i}_{s}, \mu^{0,i}_{s}) + \mathbb{W}_2(\mu^{\varepsilon,i}_{s-\tau}, \mu^{0,i}_{s-\tau}) + \mathbb{W}_2(\mu^{\varepsilon,i}_{s}, \mu^{\varepsilon,i}_{s-\tau}) + \mathbb{W}_2(\mu^{\varepsilon,i}_{s-\tau}, \mu^{\varepsilon,i}_{s})\right\}.
\]
In view of the above results obtained, we find out
\[
|J_2| \leq \int_0^T \left\{ |B^\ast(X^{0,i}_s, X^{N,i}_s, \theta_0, \theta)| \right\} \left\{ \|\Lambda^{-1}(X^{\varepsilon,i,N}_s, X^{\varepsilon,i,N}_s) - \Lambda^{-1}(X^{\varepsilon,i}_s, X^{\varepsilon,i}_s)\| s
\right.
\]
\[
+ \|\Lambda^{-1}(X^{\varepsilon,i}_s, X^{\varepsilon,i}_s) - \Lambda^{-1}(X^{0,i}_s, X^{0,i}_s)\| \right\} \left\{ |B(X^{\varepsilon,i,N}_s, X^{\varepsilon,i,N}_s, \theta_0, \theta)| ds
\right.
\]
\[
\leq C\int_0^T \left\{ 1 + |X^{0,i}_s| + |X^{0,i}_s| + (\mathbb{E}|X^{0,i}_s|^2)^{1/2} + (\mathbb{E}|X^{0,i}_s|^2)^{1/2} \right\} \left\{ |X^{\varepsilon,i,N}_s - X^{\varepsilon,i}_s| \right.
\]
\[
+ |X^{\varepsilon,i}_s - X^{0,i}_s| + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \nu^{\varepsilon,N}_s) + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \mu^{\varepsilon,i}_s) \right\} \right\} \left\{ |X^{\varepsilon,i,N}_s - X^{\varepsilon,i}_s| \right.
\]
\[
+ |X^{\varepsilon,i}_s - X^{0,i}_s| + (\mathbb{E}|X^{\varepsilon,i}_s - X^{0,i}_s|^2)^{1/2} + (\mathbb{E}|X^{\varepsilon,i}_s - X^{0,i}_s|^2)^{1/2} \right\}
\]
\[
\times \left\{ 1 + |X^{\varepsilon,i,N}_s| + |X^{\varepsilon,i,N}_s| + \left( \frac{N}{N} \sum_{j=1}^N |X^{\varepsilon,i,N}_j|^2 \right)^{1/2} + \left( \frac{1}{N} \sum_{j=1}^N |X^{\varepsilon,i,N}_j|^2 \right)^{1/2} \right\} ds.
\]
This, by Hölder inequality, we obtain that
\[
\mathbb{E}|J_2| \leq C\int_0^T \left\{ \mathbb{E}|X^{\varepsilon,i,N}_s - X^{\varepsilon,i}_s|^2 + \mathbb{E} \mathbb{W}_2(\mu^{\varepsilon,N}_s, \mu^{\varepsilon,i}_s)^2 + \mathbb{E} \mathbb{W}_2(\mu^{\varepsilon,N}_s, \mu^{\varepsilon,i}_s)^2 \right\}
\]
\[
+ \mathbb{E}|X^{\varepsilon,i}_s - X^{0,i}_s|^{1/2} \left( 1 + \mathbb{E}|X^{\varepsilon,i,N}_s|^2 + \mathbb{E}|X^{\varepsilon,i,N}_s|^2 \right) + \left( \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X^{\varepsilon,i,N}_j|^2 \right)
\]
\[
\leq C\int_0^T (C_N(1+\varepsilon^2) + \delta(\delta+\varepsilon^2)(1+C_N(1+\varepsilon^2)) + \varepsilon^2 + C_N\varepsilon^2(1+\varepsilon^2)) ds, \quad (3.32)
\]
where the last step we have used the inequalities (3.14) and (3.18). Moreover, the inequality (3.25) leads to
\[
\|\Lambda^{-1}(X^{0,i}_s, X^{0,i}_s)\| \leq \hat{L}\left\{ 1 + |X^{0,i}_s| + |X^{0,i}_s| + \mathbb{W}_2(\mu^{0,i}_s, \delta_0) + \mathbb{W}_2(\mu^{0,i}_s, \delta_0) \right\}. \quad (3.33)
\]
Thus, it follows from (3.28), (3.31) and (3.33) that
\[
\mathbb{E}|J_3| \leq C\mathbb{E} \int_0^T \left\{ 1 + |X^{0,i}_s| + |X^{0,i}_s| + (\mathbb{E}|X^{0,i}_s|^2)^{1/2} + (\mathbb{E}|X^{0,i}_s|^2)^{1/2} \right\}
\]
\[
\times \left\{ |X^{\varepsilon,i,N}_s - X^{\varepsilon,i}_s| + |X^{\varepsilon,i,N}_s - X^{\varepsilon,i}_s| + |X^{\varepsilon,i}_s - X^{0,i}_s| \right\}
\]
\[
+ |X^{\varepsilon,i}_s - X^{0,i}_s| + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \mu^{\varepsilon,i}_s) + \mathbb{W}_2(\mu^{\varepsilon,N}_s, \mu^{\varepsilon,i}_s) \right\} ds
\]
\[
\leq C T(C_N(1+\varepsilon^2) + \delta(\delta+\varepsilon^2)(1+C_N(1+\varepsilon^2)) + \varepsilon^2 + C_N\varepsilon^2(1+\varepsilon^2))^{1/2}. \quad (3.34)
\]
Therefore, from (3.30), (3.32) and (3.34), we conclude that the desired result holds.

\[\square\]

**Theorem 3.7.** Under the conditions of Lemma 3.6. If, for any \( \theta \in \Theta, \Pi(\theta) \geq 0 \), then
\[\hat{\theta}_{n,\varepsilon} \rightarrow \theta_0 \text{ in probability as } \varepsilon \rightarrow 0 \text{ and } N, n \rightarrow \infty.\]
Proof. According to (2.8) and (2.9), we arrive at that

\[
\Phi_{n,\varepsilon}^{i,N}(\theta) = \varepsilon^2 (\Psi_{n,\varepsilon}^{i,N}(\theta) - \Psi_{n,\varepsilon}^{i,N}(\theta_0)) \\
= \delta^{-1} \sum_{k=1}^{n} \left\{ (P_k^{\varepsilon,i,N}(\theta_0))^* \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) P_k^{\varepsilon,i,N}(\theta) \\
- (P_k^{\varepsilon,i,N}(\theta_0))^* \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) P_k^{\varepsilon,i,N}(\theta_0) \right\} \\
= \delta^{-1} \sum_{k=1}^{n} \left\{ (P_k^{\varepsilon,i,N}(\theta_0) + b(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \mu_{(k-1)\delta,\tau}, \theta) \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) \\
\times \left( P_k^{\varepsilon,i,N}(\theta) + \delta (b(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \mu_{(k-1)\delta,\tau}, \theta) \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) \\
- b(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \mu_{(k-1)\delta,\tau}, \theta) \right) - (P_k^{\varepsilon,i,N}(\theta_0))^* \\
\times \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) P_k^{\varepsilon,i,N}(\theta_0) \right\} \\
= 2 \sum_{k=1}^{n} B^*(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \theta_0, \theta) \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) P_k^{\varepsilon,i,N}(\theta) \\
\times \left( P_k^{\varepsilon,i,N}(\theta) + \delta (b(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \mu_{(k-1)\delta,\tau}, \theta) \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) \\
- b(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta, \mu_{(k-1)\delta,\tau}, \theta) \right) - (P_k^{\varepsilon,i,N}(\theta_0))^* \\
\times \Lambda^{-1}(X_{(k-1)\delta,\tau}, X_{(k-1)\delta,\tau}^\delta) P_k^{\varepsilon,i,N}(\theta_0) \right\} \\
= 2\Phi_{n,\varepsilon}^{i,N}(1)(\theta) + \Phi_{n,\varepsilon}^{i,N}(2)(\theta).
\]

In view of Lemmas 3.5 and 3.6, together with the Chebyshev’s inequality, we deduce that

\[
\sup_{\theta \in \Theta} | - \Phi_{n,\varepsilon}^{i,N}(\theta) - (-\Pi(\theta)) | \to 0 \text{ in probability.} \tag{3.35}
\]

Here $\Pi(\cdot)$ is defined in Lemma 3.6. According to (2.11), we find out $0 = \Phi_{n,\varepsilon}^{i,N}(\theta_0) \geq \Phi_{n,\varepsilon}^{i,N}(\theta_{n,\varepsilon})$, i.e., $0 = -\Phi_{n,\varepsilon}^{i,N}(\theta_0) \leq -\Phi_{n,\varepsilon}^{i,N}(\theta_{n,\varepsilon})$. In addition, due to $\Pi(\cdot) \geq 0$, we get

\[
\sup_{|\theta - \theta_0| \geq \iota} (-\Pi(\theta)) < -\Pi(\theta_0) = 0, \text{ for any } \iota > 0. \tag{3.36}
\]

In terms of [19, Theorem 5.9], and combining with (3.35) and (3.36), we deduce that $\theta_{n,\varepsilon} \to \theta_0$ in probability as $N,n \to \infty$ and $\varepsilon \to 0$. \hfill $\square$

4. The asymptotic distribution of LSE. In this section, we are concerned with the asymptotic distribution of the LSE $\hat{\theta}_{n,\varepsilon}^{i,N}$. Set

\[
\Upsilon(x, y, \theta_0) := (\nabla_{\theta} b)^*(x, y, \mu, \nu, \theta_0) \Lambda^{-1}(x, y) \sigma(x, y, \mu, \nu),
\]

\[
x, y \in \mathbb{R}^d, \mu, \nu \in \mathcal{P}_d(\mathbb{R}^d)
\]

and

\[
I(\theta) := \int_0^T (\nabla_{\theta} b)^*(X_t^{0,i}, X_t^{0,i}, \mu_t^{0,i}, \mu_t^{0,i}, \theta) \Lambda^{-1}(X_t^{0,i}, X_t^{0,i}, \mu_t^{0,i}, \mu_t^{0,i}, \theta) dt.
\]
Theorem 4.1. Under the assumptions of Theorem 3.7, suppose that (A4) and (A5) hold. Then,
\[ \varepsilon^{-1}(\hat{g}_{n,\varepsilon} - \theta_0) \to I^{-1}(\theta_0) \int_0^T \Upsilon(X_t^{0,i}, X_t^{0,i}, \theta_0) dW_t \quad \mathbb{P} \text{-a.s.} \]
as \( n, N \to \infty \) and \( \varepsilon \to 0 \), where \( I(\cdot) \) and \( \Upsilon(\cdot) \) are continuous.

For the proof of Theorem 4.1, we need to prepare for several lemmas below.

Lemma 4.2. Assume that (A1)-(A5) hold. Then, for \( X_0^\varepsilon = \xi \in \mathcal{C}_{p>0}^d([-\tau, 0]; \mathbb{R}^d) \) with \( \mathcal{L}_\xi \in \mathcal{P}_p(\mathbb{R}^d) \), \( p > 4 \),
\[ \int_0^T \Upsilon(X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) dW_t^i \to \int_0^T \Upsilon(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) dW_t^i \quad \mathbb{P} \text{-a.s.} \]as \( \varepsilon \to 0 \), \( \delta \to 0 \) and \( N \to \infty \). Moreover,
\[ \varepsilon^{-1}(\nabla \Phi_{n,\varepsilon}^i(\theta_0)) \to -2 \int_0^T \Upsilon(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) dW_t^i \quad \mathbb{P} \text{-a.s.} \]
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 \( \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d) \), there exists a constant \( L_1 > 0 \) such that
\[ \sup_{\theta \in \mathcal{I}} \| (\nabla \Phi)(x, y, \mu, \nu, \theta) \| \leq L_1 \{ 1 + |x| + |y| + \mathbb{W}_2(\mu, \delta_0) + \mathbb{W}_2(\nu, \delta_0) \}. \]

We first claim that
\[ \int_0^T \| \Upsilon(X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) - \Upsilon(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) \|^2 dt \to 0 \quad \mathbb{P} \text{-a.s.} \]
as \( \varepsilon \to 0 \), \( \delta \to 0 \) and \( N \to \infty \). According to (4.1), one gets
\[ \| \Upsilon(X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) - \Upsilon(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) \|^2 \]
\[ \leq 3 \left\{ |(\nabla \Phi)^* (X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \mu_{t,-

\begin{align*}
\varepsilon^{-1}(\hat{g}_{n,\varepsilon} - \theta_0) & \to I^{-1}(\theta_0) \int_0^T \Upsilon(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) dW_t^i \quad \mathbb{P} \text{-a.s.} \\
& \text{as } n, N \to \infty \text{ and } \varepsilon \to 0, \text{ where } I(\cdot) \text{ and } \Upsilon(\cdot) \text{ are continuous.}
\end{align*}

For the first term \( G_1 \), from the assumption (A4) we first give the below result
\[ \| (\nabla \Phi)^* (X_t^{\varepsilon,i,N}, X_{t-\tau}^{\varepsilon,i,N}, \mu_{t,-

\begin{align*}
& \leq C \left\{ |X_t^{\varepsilon,i,N} - X_t^{0,i}|^2 + |X_{t-\tau}^{\varepsilon,i,N} - X_{t-\tau}^{0,i}|^2 + \frac{1}{N} \sum_{j=1}^N |X_t^{j,i,N} - X_t^{0,i}|^2 \\
& \quad + \frac{1}{N} \sum_{j=1}^N |X_{t-\tau}^{j,i,N} - X_{t-\tau}^{0,i}|^2 \right\}.
\end{align*}
This, combining (3.7) with (3.25), leads to
\[
G_1 \leq C \left\{ 1 + |X_{t_0}^{\epsilon,i,N}|^2 + |X_{t_1}^{\epsilon,i,N}|^2 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N}|^2 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{j-\tau}}^{\epsilon,j,N}|^2 \right\} \bigg( \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg)^2
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg) . \tag{4.8}
\]
For the second term $G_2$, by the assumption (A3), (3.7) and (4.5), we get
\[
G_2 \leq C \left\{ 1 + |X_{t_0}^{\epsilon,i,N}|^2 + |X_{t_1}^{\epsilon,i,N}|^2 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N}|^2 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{j-\tau}}^{\epsilon,j,N}|^2 \right\} \bigg( \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg)^2
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg) . \tag{4.9}
\]
For the third term $G_3$, by the assumption (A2), (3.25) and (4.5), one has
\[
G_3 \leq C \left\{ |X_{t_0}^{\epsilon,i,N} - X_{t_1}^{0,i}|^2 + |X_{t_1}^{\epsilon,i,N} - X_{t_{1-\tau}}^{0,i}|^2 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \right\}
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg) . \tag{4.10}
\]
It follows from (4.8), (4.9) and the H"older inequality that
\[
\int_0^T (G_1 + G_2) dt
\]
\[
\leq C \int_0^T \left\{ 1 + |X_{t_0}^{\epsilon,i,N}|^4 + |X_{t_1}^{\epsilon,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{j-\tau}}^{\epsilon,j,N}|^4 \right\} \bigg( \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg)^2
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg) dt
\]
\[
\leq C \int_0^T \left\{ |X_{t_1}^{\epsilon,i,N} - X_{t_1}^{0,i}|^2 \left\{ 1 + |X_{t_0}^{\epsilon,i,N}|^4 + |X_{t_1}^{\epsilon,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N}|^4 \right\} \right.
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N}|^4 \bigg) \bigg( \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg)^2
\]
\[
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_j}^{\epsilon,j,N} - X_{t_{j-\tau}}^{0,i}|^2 \bigg) . \tag{4.11}
\]
\[
\frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 + \left( \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}}^{0,i}|^2 \right) \left( 1 + |X_{t_{s_j}}^{ε,i,N}|^4 \right) \\
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 \left( 1 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 \right) + \left( \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}}^{0,i}|^2 \right) \\
\times \left\{ 1 + |X_{t_{s_j}}^{ε,i,N}|^4 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 \right\} dt \\
= 4 \sum_{i=1}^{4} P_i.
\]

For any \( \epsilon > 0 \), from the above inequality, one has

\[
\mathbb{P}\left( \int_0^T (G_1 + G_2) dt \geq \epsilon \right) \leq \mathbb{P}(P_1 \geq \epsilon/4) + \mathbb{P}(P_2 \geq \epsilon/4) + \mathbb{P}(P_3 \geq \epsilon/4) + \mathbb{P}(P_4 \geq \epsilon/4).
\]

For any \( K > 0, \ i \in S_N \), in view of the Chebyshev inequality, we arrive at

\[
\mathbb{P}(P_1 \geq \epsilon/4) \\
= \mathbb{P}\left( C \int_0^T |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{ε,i,N}|^2 \left\{ 1 + |X_{t_{s_j}}^{ε,i,N}|^4 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N}|^4 \\
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 \right\} dt \geq \epsilon/4 \right) \\
\leq \mathbb{P}\left( C \int_0^T |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N} - X_{t_{s_{j-1}}+\tau}^{0,i}|^4 \\
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N} - X_{t_{s_{j-1}}+\tau}^{0,i}|^4 \right) \left\{ 1 + |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 \right\} \left\{ 1 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N} - X_{t_{s_{j-1}}+\tau}^{0,i}|^4 \right\} dt \geq \epsilon/8 \right) \\
+ \mathbb{P}\left( C(1 + K^4) \int_0^T |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^2 dt \geq \epsilon/8 \right) \\
\leq \mathbb{P}\left( C \int_0^T \left\{ 1 + |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N} - X_{t_{s_{j-1}}+\tau}^{0,i}|^4 \\
+ \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N} - X_{t_{s_{j-1}}+\tau}^{0,i}|^4 \right\}^2 dt \geq \epsilon/8 \right) \\
+ \mathbb{P}\left( C(1 + K^4) \int_0^T |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^2 dt \geq \epsilon/8 \right) \\
\leq \frac{C}{\epsilon} \int_0^T \mathbb{E} \left\{ 1 + |X_{t_{s_j}}^{ε,i,N}|^8 + |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^8 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_j}}^{ε,i,N}|^8 + \frac{1}{N} \sum_{j=1}^{N} |X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^8 \right\} \\
\times \left\{ 1 + |X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^4 \right\} dt + \frac{C(1 + K^4)}{\epsilon} \int_0^T \mathbb{E}|X_{t_{s_j}}^{ε,i,N} - X_{t_{s_j}+\tau}^{0,i}|^2 dt \\
\leq \frac{C}{\epsilon} \int_0^T \left\{ 1 + \mathbb{E}|X_{t_{s_j}}^{ε,i,N}|^4 + \mathbb{E}|X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} \mathbb{E}|X_{t_{s_j}}^{ε,i,N}|^4 + \frac{1}{N} \sum_{j=1}^{N} \mathbb{E}|X_{t_{s_{j-1}}+\tau}^{ε,i,N}|^4 \right\}^{1/2}.
\]
In addition, by means of (3.14) and (3.18), we have

\[
\mathbb{E}|X_{t}^{\epsilon,i,N} - X_{t}^{0,i}|^2 \\
\leq 2\mathbb{E}|X_{t}^{\epsilon,i,N} - X_{t}^{\epsilon,i}|^2 + 2\mathbb{E}|X_{\epsilon,i} - X_{t}^{0,i}|^2 \\
\leq C_\delta(1 + \epsilon^2) + C_\delta(\delta + \epsilon^2)(1 + C_N(1 + \epsilon^2)) + C_\epsilon^2 + C_\epsilon^2(1 + \epsilon^2) \\
\to 0,
\]

(4.11)
as \epsilon \to 0, \delta \to 0 \text{ and } N \to \infty. \text{ So, it is easy to see that}

\[
\mathbb{P}(P_1 \geq \epsilon/4) \to 0,
\]
as \epsilon \to 0, \delta \to 0 \text{ and } N \to \infty. \text{ Carrying out a similar argument as the above analysis, we get}

\[
\mathbb{P}(P_2 \geq \epsilon/4) \to 0,
\]
as \epsilon \to 0, \delta \to 0 \text{ and } N \to \infty. \text{ Moreover, for any } K > 0, i \in S_N, \text{ in view of the Chebyshev inequality, one has}

\[
\mathbb{P}(P_3 \geq \epsilon/4) \\
\leq \mathbb{P}\left( \int_0^T \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^2 \left\{ 1 + |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^4 + |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^4 \right\} dt \geq \epsilon/8 \right) \\
+ \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^4 + \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^4 \right\} 1_{\{|X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^4 \geq \epsilon/8 \}} dt \geq \epsilon/8 \right) \\
\leq \mathbb{P}\left( C \int_0^T \left\{ 1 + |X_{t}^{\epsilon,i,N}|^8 + |X_{t}^{\epsilon,i,N}|^8 \right\} \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N}|^8 + \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N}|^8 \right) \\
\times 1_{\{|X_{t}^{\epsilon,i,N} - X_{t}^{0,i}|^4 \geq \epsilon/8 \}} + \mathbb{P}\left( (1 + K^4) \int_0^T \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^2 dt \geq \epsilon/8 \right) \\
\leq \frac{C}{\epsilon} \int_0^T \left( \mathbb{P}\left( 1 + |X_{t}^{\epsilon,i,N}|^8 + |X_{t}^{\epsilon,i,N}|^8 \right) \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N}|^8 + \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N}|^8 \right)^{1/2} \\
\times \left( \mathbb{P}(\{|X_{t}^{\epsilon,i,N} - X_{t}^{0,i}|^4 \geq \epsilon/8 \}) \right)^{1/2} dt + \mathbb{P}(1 + K^4) \int_0^T \frac{1}{N} \sum_{j=1}^N |X_{t}^{\epsilon,j,N} - X_{t}^{0,j}|^2 dt \geq \epsilon/8 \right)
\[ \frac{C}{\epsilon K} \int_0^T \left( \mathbb{E}[X_{\varepsilon}^{i,N} - X_t^{0,i}]^2 \right)^{1/2} dt + \frac{C(1 + K^4)}{\epsilon} \int_0^T \frac{1}{N} \sum_{j=1}^N \mathbb{E}[X_{\varepsilon}^{j,N} - X_t^{0,j}]^2 dt. \]

This, together with (4.11), implies that
\[ \mathbb{P}(P_3 \geq \epsilon/4) \to 0, \]
as \( \epsilon \to 0, \delta \to 0 \) and \( N \to \infty \). Similarly, for any \( K > 0 \), \( i \in S_N \), one has
\[ \mathbb{P}(P_4 \geq \epsilon/4) \to 0, \]
as \( \epsilon \to 0, \delta \to 0 \) and \( N \to \infty \). Therefore,
\[ \int_0^T (G_1 + G_2)dt \to 0, \quad \mathbb{P} \text{- a.s.,} \quad (4.12) \]
as \( \epsilon \to 0, \delta \to 0 \) and \( N \to \infty \). On the other hand, thanks to (4.10) and (4.11), it follows that
\[ \mathbb{P} \left( \int_0^T G_3 dt \geq \epsilon \right) \leq \frac{C}{\epsilon} \int_0^T \left( \mathbb{E}[X_{\varepsilon}^{i,N} - X_t^{0,i}]^2 + \mathbb{E}[X_{\varepsilon}^{j,N} - X_t^{0,j}]^2 + \frac{1}{N} \sum_{j=1}^N \mathbb{E}[X_{\varepsilon}^{j,N} - X_t^{0,j}]^2 \right) dt \to 0, \]
as \( \epsilon \to 0, \delta \to 0 \) and \( N \to \infty \). Hence,
\[ \int_0^T G_3 dt \to 0, \quad \mathbb{P} \text{- a.s.,} \quad (4.13) \]
as \( \epsilon \to 0, \delta \to 0 \) and \( N \to \infty \). As a consequence, (4.6) follows from (4.12) and (4.13). What’s more, for any \( \rho > 0 \) and \( \epsilon > 0 \), owing to (4.6), one gets
\[ \mathbb{P} \left( \left| \int_0^T \mathbb{Y}(X_{\varepsilon}^{i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) - \mathbb{Y}(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) \right| dt \geq \rho \right) \leq \mathbb{P} \left( \int_0^T \| \mathbb{Y}(X_{\varepsilon}^{i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) - \mathbb{Y}(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) \|^2 dt \geq \rho^2 \epsilon \right) + \epsilon, \]
which, together with the arbitrariness of \( \epsilon \) and (4.6), implies that (4.3) holds. And by a simple calculation, one gets
\[ \varepsilon^{-1}(\nabla \phi \Phi_{n,\varepsilon}^{i,N})(\theta_0) \]
\[ = -\frac{2}{n} \sum_{k=1}^n (\nabla \phi)^* (X_{(k-1)\delta}^{\varepsilon,i,N}, X_{(k-1)\delta-\tau}^{\varepsilon,i,N}, \mu_{(k-1)\delta}^{\varepsilon,N}, \mu_{(k-1)\delta-\tau}^{\varepsilon,N}, \theta_0) \]
\times \Lambda^{-1}(X_{(k-1)\delta}^{\varepsilon,i,N}, X_{(k-1)\delta-\tau}^{\varepsilon,i,N}) \sigma(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}) \delta W_k \]
\[ = -2 \int_0^T \mathbb{Y}(X_{\varepsilon}^{i,N}, X_{t-\tau}^{\varepsilon,i,N}, \theta_0) W_t \]
\[ \to -2 \int_0^T \mathbb{Y}(X_t^{0,i}, X_{t-\tau}^{0,i}, \theta_0) W_t, \quad \mathbb{P} \text{- a.s.,} \]
whenever \( \varepsilon \to 0, n, N \to \infty \).
Lemma 4.3. Under the assumptions of Theorem 4.1,
\[
(\nabla_\theta^{(2)} \Phi_{n,\epsilon}^i(\theta)) \to K(\theta) := K(\theta) + 2I(\theta), \text{ } \mathbb{P} - \text{ a.s.,}
\]
(4.14)
as \epsilon \to 0, n \to \infty and N \to \infty, where \((\nabla_\theta^{(2)} \Phi_{n,\epsilon}^i)\) and \(I(\theta)\) are defined as in (2.2) and (4.2), respectively,
\[
K(\theta) := -2 \int_0^T (\nabla_\theta^{(2)} b^s)(X_{t_1}^{0,i}, X_{t_1}^{0,i}, \mu_{t_1}^{0,i}, \mu_{t_1}^{0,i}) \circ \left\{ \Lambda^{-1}(X_{t_1}^{0,i}, X_{t_1}^{0,i}) B(X_{t_1}^{0,i}, X_{t_1}^{0,i}, \theta_0, \theta) \right\} dt.
\]
(4.15)

Proof. Clearly,
\[
(\nabla_\theta^{(2)} \Phi_{n,\epsilon}^i(\theta)) = (\nabla_\theta(\nabla_\theta \Phi_{n,\epsilon}^i(\theta)))(\theta)
\]
(4.14)
\[
= -2 \sum_{k=1}^{n} (\nabla_\theta b^s)(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}, \mu_{(k-1)\delta}^{\epsilon,N}, \mu_{(k-1)\delta}^{\epsilon,N}) \circ \left\{ \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}) P_{\kappa}^{\epsilon,i,N}(\theta) \right\}
\]
\[
- 2 \sum_{k=1}^{n} (\nabla_\theta b^s)(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}, \mu_{(k-1)\delta}^{\epsilon,N}, \mu_{(k-1)\delta}^{\epsilon,N}) \circ \left\{ \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}) (\nabla_\theta P_{\kappa}^{\epsilon,i,N}(\theta)) \right\}
\]
\[
= -2 \sum_{k=1}^{n} (\nabla_\theta b^s)(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}, \mu_{(k-1)\delta}^{\epsilon,N}, \mu_{(k-1)\delta}^{\epsilon,N}) \circ \left\{ \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}) \right\}
\]
\[
- 2 \sum_{k=1}^{n} (\nabla_\theta b^s)(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}, \mu_{(k-1)\delta}^{\epsilon,N}, \mu_{(k-1)\delta}^{\epsilon,N}) \circ \left\{ \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}) \right\}
\]
\[
\circ \left\{ \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta}^{\epsilon,i,N}) (\nabla_\theta P_{\kappa}^{\epsilon,i,N}(\theta)) \right\}
\]
\[
:= P_{i1} + P_{i2}.
\]

For any \(x, y \in \mathbb{R}^d\) and \(\mu, \nu \in \mathcal{P}_d(\mathbb{R}^d)\), notice from (A5) that
\[
\sup_{\theta \in \Theta} \| (\nabla_\theta^{(2)} b^s)(x, y, \mu, \nu, \theta) \| \leq C \{ 1 + |x| + |y| + \mathbb{W}_2(\mu, \delta_0) + \mathbb{W}_2(\nu, \delta_0) \}.
\]
(4.16)

For the first term \(P_{i1}\), by (4.16), one arrives at
\[
\mathbb{E}[P_{i1}] \leq 2\epsilon \left( \mathbb{E} \left| \int_0^T (\nabla_\theta^{(2)} b^s)(X_{t_1}^{\epsilon,i,N}, X_{t_1}^{\epsilon,i,N}, \mu_{t_1}^{\epsilon,N}, \mu_{t_1}^{\epsilon,N}) \circ \left\{ \Lambda^{-1}(X_{t_1}^{\epsilon,i,N}, X_{t_1}^{\epsilon,i,N}) \right\} \circ dW_{t_1}^2 \right|^2 \right)^{1/2}
\]
\[
\leq 2\epsilon \left( \mathbb{E} \left| \int_0^T \| (\nabla_\theta^{(2)} b^s)(X_{t_1}^{\epsilon,i,N}, X_{t_1}^{\epsilon,i,N}, \mu_{t_1}^{\epsilon,N}, \mu_{t_1}^{\epsilon,N}) \|^2 \| \Lambda^{-1}(X_{t_1}^{\epsilon,i,N}, X_{t_1}^{\epsilon,i,N}) \| \circ dW_{t_1}^2 \right|^2 \right)^{1/2}
\]
\[
\times \sigma(X_{t_1}^{\epsilon,i,N}, X_{t_1}^{\epsilon,i,N}, \mu_{t_1}^{\epsilon,N}, \mu_{t_1}^{\epsilon,N}) \circ dW_{t_1}^2 \right)^{1/2}
\]
\[ \leq C \varepsilon \left( \int_0^T \left( 1 + \mathbb{E}|X_{t,s}^{\varepsilon,i,N}|^6 + \mathbb{E}|X_{t,s}^{\varepsilon,i,N}|^6 + \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_{t,s}^{\varepsilon,j,N}|^6 \right) dt \right)^{1/2} \]

\[ \leq C \varepsilon \to 0, \text{ as } \varepsilon \to 0, \ \delta \to 0 \text{ and } N \to \infty. \]

For the second term \(P_{l_2}\),

\[ P_{l_2} = -2 \int_0^T (\nabla^2 \theta) \ast (X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \mu_{t,s}^{\varepsilon,N}, \mu_{t,s}^{\varepsilon,N}, \theta) \circ (\Lambda^{-1}(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N})) \times B(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \theta_0, \theta) dt \]

\[ + 2 \int_0^T (\nabla \theta) \ast (X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \mu_{t,s}^{\varepsilon,N}, \mu_{t,s}^{\varepsilon,N}, \theta) \Lambda^{-1}(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}) \times (\nabla \theta)(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \mu_{t,s}^{\varepsilon,N}, \mu_{t,s}^{\varepsilon,N}, \theta) dt \]

\[ =: H_1 + H_2. \]

Taking into consideration Lemma 3.5 and (A5) yields that

\[ H_1 - K(\theta) \]

\[ = -2 \int_0^T (\nabla^2 \theta) \ast (X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \mu_{t,s}^{\varepsilon,N}, \mu_{t,s}^{\varepsilon,N}, \theta) \circ (\Lambda^{-1}(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N})) \times B(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \theta_0, \theta) \]

\[ - (\nabla \theta) \ast (X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N}, \mu_{t,s}^{0,i,N}, \mu_{t,s}^{0,i,N}, \theta) \circ (\Lambda^{-1}(X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N})) B(X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N}, \theta_0, \theta) \] \]

\[ = -2 \int_0^T \left( (\nabla^2 \theta) \ast (X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \mu_{t,s}^{\varepsilon,N}, \mu_{t,s}^{\varepsilon,N}, \theta) \right) \times B(X_{t,s}^{\varepsilon,i,N}, X_{t,s}^{\varepsilon,i,N}, \theta_0, \theta) \]

\[ + \left( \nabla \theta \ast (X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N}, \mu_{t,s}^{0,i,N}, \mu_{t,s}^{0,i,N}, \theta) \circ (\Lambda^{-1}(X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N})) B(X_{t,s}^{0,i,N}, X_{t,s}^{0,i,N}, \theta_0, \theta) \right) \]

\[ =: \sum_{i=1}^3 M_3. \]

For the term \(M_1\), thanks to (A5), (3.26) and (3.25), it follows from the Hölder inequality that

\[ \mathbb{E}|M_1| \leq C \int_0^T \left( \mathbb{E}|X_{t,s}^{\varepsilon,i,N} - X_{t,s}^{0,i}|^2 + \mathbb{E}|X_{t,s}^{\varepsilon,i,N} - X_{t,s}^{0,i}|^2 \right) \]

\[ + \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_{t,s}^{\varepsilon,j,N} - X_{t,s}^{0,j}|^2 + \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_{t,s}^{\varepsilon,j,N} - X_{t,s}^{0,j}|^2 \] \[ \left( \frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_{t,s}^{\varepsilon,j,N} - X_{t,s}^{0,j}|^2 \right)^{1/2} dt, \]
in which we have used the result in lemma 3.2. Then, according to (4.11), one has
\[ \mathbb{E}|M_1| \to 0 \text{ as } \varepsilon \to 0, \delta \to 0, N \to \infty. \]
Carrying out the similar arguments as above, we have
\[ \mathbb{E}|M_2| \to 0 \text{ as } \varepsilon \to 0, \delta \to 0, N \to \infty \]
and
\[ \mathbb{E}|M_3| \to 0 \text{ as } \varepsilon \to 0, \delta \to 0, N \to \infty. \]
As a result, we conclude that
\[ H_1 \to K(\theta) \mathbb{P} - \text{a.s. as } \varepsilon \to 0, \delta \to 0, N \to \infty. \tag{4.17} \]
Again, carrying out analogous arguments to derive (4.17), we obtain
\[ H_2 \to 2I(\theta) \mathbb{P} - \text{a.s. as } \varepsilon \to 0, \delta \to 0, N \to \infty. \tag{4.18} \]
Therefore, the desired assertion is complete by (4.17) and (4.18) immediately. \( \Box \)

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

**Proof of Theorem 4.1.** According to the result of Theorem 3.7, there exists a sequence \( \eta_{n,\varepsilon}^i \to 0 \) as \( N, n \to \infty \) and \( \varepsilon \to 0 \) such that \( \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0) \subset \Theta, \mathbb{P}\text{-a.s.} \), that is to say,
\[ \mathbb{P}\left( \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0) \right) \to 1, \text{ as } n, N \to \infty, \varepsilon \to 0, \tag{4.19} \]
where \( B_{\eta_{n,\varepsilon}^i}(\theta_0) \) represents the closed ball centered at \( \theta_0 \) with the radius \( \eta_{n,\varepsilon}^i \). Then, it is easy to see that
\[ (\nabla_{\theta} \hat{\Phi}_{n,\varepsilon}^i)(\hat{\theta}_{n,\varepsilon}^i) = (\nabla_{\theta} \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) + F_{n,\varepsilon}^i(\hat{\theta}_{n,\varepsilon}^i - \theta_0), \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0) \tag{4.20} \]
with
\[ F_{n,\varepsilon}^i := \int_0^1 (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0 + v(\hat{\theta}_{n,\varepsilon}^i - \theta_0)) dv, \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0), \]
by the Taylor expansion. In what follows we intend to deduce that
\[ F_{n,\varepsilon}^i \to \bar{K}(\theta_0) \mathbb{P} - \text{a.s.} \tag{4.21} \]
as \( n, N \to \infty \) and \( \varepsilon \to 0 \). Note that, for \( \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0) \),
\[ \| F_{n,\varepsilon}^i - \bar{K}(\theta_0) \|
\leq \| F_{n,\varepsilon}^i - (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) \| + \| (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) - \bar{K}(\theta_0) \|
\leq \int_0^1 \| (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0 + v(\hat{\theta}_{n,\varepsilon}^i - \theta_0)) - (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) \| dv + \| (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) - \bar{K}(\theta_0) \|
\leq \sup_{\theta \in B_{\eta_{n,\varepsilon}^i}(\theta_0)} \| (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta) - \bar{K}(\theta) \| + \sup_{\theta \in B_{\eta_{n,\varepsilon}^i}(\theta_0)} \| \bar{K}(\theta) - \bar{K}(\theta_0) \|
\quad + 2\| (\nabla_{\theta}^2 \hat{\Phi}_{n,\varepsilon}^i)(\theta_0) - \bar{K}(\theta_0) \|, \]
where \( \bar{K}(\cdot) \) is defined in (4.14). This, together with Lemma 3.4 and the continuity of \( \bar{K}(\cdot) \), yields that (4.21) holds. Next we show the asymptotic distribution of \( \hat{\theta}_{n,\varepsilon}^i \).
Let
\[ \mathcal{F}_{n,\varepsilon}^i = \{ F_{n,\varepsilon}^i \text{ is invertible, } \hat{\theta}_{n,\varepsilon}^i \in B_{\eta_{n,\varepsilon}^i}(\theta_0) \}. \]
By Lemma 3.4, one gets, for some positive constant $\alpha$,
\[ P\left( \sup_{\theta \in B_{n,N}(\theta_0)} \left\| (\nabla_\theta (2) \Phi_{n,e}^{i,N}) (\theta) - \bar{K}(\theta_0) \right\| \leq \frac{\alpha}{2} \right) \rightarrow 1 \]  
(4.22)
as $n, N \rightarrow \infty$ and $\varepsilon \rightarrow 0$. What’s more, by following the line of [13, Theorem 2.2], we can deduce that $F_{n,e}^{i,N}$ is invertible on the set
\[ \Gamma_{n,e}^{i,N} := \left\{ \sup_{\theta \in B_{n,e}(\theta_0)} \left\| (\nabla_\theta (2) \Phi_{n,e}^{i,N}) (\theta) - \bar{K}(\theta_0) \right\| \leq \frac{\alpha}{2}, \bar{\theta}_{n,e}^{i} \in B_{n,e}(\theta_0) \right\}, \]
Clearly,
\[ 1 \geq P(\Gamma_{n,e}^{i,N}) \geq P\left( \sup_{\theta \in B_{n,e}(\theta_0)} \left\| (\nabla_\theta (2) \Phi_{n,e}^{i,N}) (\theta) - K_0(\theta_0) \right\| \leq \frac{\alpha}{2} \right) \]
\[ + P\left( \bar{\theta}_{n,e}^{i} \in B_{n,e}(\theta_0) \right) - 1. \]  
(4.23)
Thus, taking advantage of (4.22), (4.19) as well as (4.23), we deduce
\[ P(F_{n,e}^{i,N}) \geq P(\Gamma_{n,e}^{i,N}) \rightarrow 1 \text{ as } n, N \rightarrow \infty \text{ as } \varepsilon \rightarrow 0. \]  
(4.24)
Let
\[ U_{n,e}^{i,N} = F_{n,e}^{i,N} \mathbf{1}_{F_{n,e}^{i,N}} + I_p \mathbf{1}_{(\mathcal{F}_{n,e}^{i,N})^c}, \]
where $I_p$ is a $p \times p$ identity matrix. It follows from (4.20) that
\[ \varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0) \]
\[ = (\varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0)) \mathbf{1}_{F_{n,e}^{i,N}} + (\varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0)) \mathbf{1}_{(\mathcal{F}_{n,e}^{i,N})^c} \]
\[ = (U_{n,e}^{i,N})^{-1} F_{n,e}^{i,N} (\varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0)) \mathbf{1}_{F_{n,e}^{i,N}} + (\varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0)) \mathbf{1}_{(\mathcal{F}_{n,e}^{i,N})^c} \]
\[ = - \varepsilon^{-1}(U_{n,e}^{i,N})^{-1} ((\nabla_\theta \Phi_{n,e}^{i,N})(\bar{\theta}_{n,e}^{i,N}) - (\nabla_\theta \Phi_{n,e}^{i,N})(\theta_0)) \mathbf{1}_{F_{n,e}^{i,N}} + (\varepsilon^{-1}(\bar{\theta}_{n,e}^{i,N} - \theta_0)) \mathbf{1}_{(\mathcal{F}_{n,e}^{i,N})^c} \]
\[ \rightarrow I^{-1}(\theta_0) \int_0^T \mathbf{Y}(X_t^{0,i}, X_t^{0,i}, \bar{\theta}_0) dW_t \text{ as } n, N \rightarrow \infty, \varepsilon \rightarrow 0, \]
where in the forth step we have used the Fermat’s theorem and dropped the term $(\nabla_\theta \Phi_{n,e}^{i,N})(\bar{\theta}_{n,e}^{i,N})$, and in the last step utilized Lemma 3.2, (4.14), (4.21), and (4.24). The desired conclusion is obtained.

Furthermore, in order to demonstrate the applicability of our results, two examples are constructed.

**Example 4.4.** Let $\theta \in \Theta_0 \subset \mathbb{R}$. For any $\varepsilon \in (0, 1)$, consider a scalar distribution-dependent SDDE
\[ dX^\varepsilon_t = \left( X^\varepsilon_t + \int_{\mathbb{R}} x \mu^\varepsilon_t(dx) + \int_{\mathbb{R}} y \mu^\varepsilon_{t-}(dy) + \theta \right) dt \]
\[ + \varepsilon \left( 1 + |X^\varepsilon_t| + |X^\varepsilon_{t-}| + \int_{\mathbb{R}} |x| \mu^\varepsilon_t(dx) + \int_{\mathbb{R}} |y| \mu^\varepsilon_{t-}(dy) \right) dW_t, \quad t \in (0, T], \]
\[ X^\varepsilon_s = \xi(s), \quad s \in [-\tau, 0]. \]
Set
\[ b(X^\varepsilon_t, X^\varepsilon_{t-}, \mu^\varepsilon_t, \mu^\varepsilon_{t-}, \theta) := X^\varepsilon_t + X^\varepsilon_{t-} + \int_{\mathbb{R}} x \mu^\varepsilon_t(dx) + \int_{\mathbb{R}} y \mu^\varepsilon_{t-}(dy) + \theta \]
and
\[ \sigma(X_{1}^{\epsilon}, X_{1-\tau}^{\epsilon}, \mu_{1}^{\epsilon}, \mu_{2}^{\epsilon-\tau}) := 1 + |X_{1}^{\epsilon}| + |X_{1-\tau}^{\epsilon}| + \int_{\mathbb{R}} |x| \mu_{1}^{\epsilon}(dx) + \int_{\mathbb{R}} |y| \mu_{2}^{\epsilon-\tau}(dy). \]

Then, (4.25) can be reformulated as (2.3) with \( X_{0}^{\epsilon} = \xi \in C_{\mathcal{F}_{0}}([-\tau, 0], \mathbb{R}^{d}) \) and \( \xi \in \mathcal{P}_{p}(\mathbb{R}^{d}), p > 4 \). Obviously, for any \( x_{i}, y_{i} \in \mathbb{R} \) and \( \mu_{i}, \nu_{i} \in \mathcal{P}_{p}(\mathbb{R}), i = 1, 2 \), (4.25) satisfies the assumptions (A1)-(A5). Indeed,
\[ |b(x_{1}, y_{1}, \mu_{1}, \nu_{1}, \theta) - b(x_{2}, y_{2}, \mu_{2}, \nu_{2}, \theta)| \]
\[ \leq |x_{1} - x_{2}| + |y_{1} - y_{2}| + \left| \int_{\mathbb{R}} x_{1} \mu_{1}(dx) - \int_{\mathbb{R}} y_{2} \mu_{2}(dy) \right| + \left| \int_{\mathbb{R}} x_{1} \nu_{1}(dx) - \int_{\mathbb{R}} y_{2} \nu_{2}(dy) \right| \]
\[ \leq |x_{1} - x_{2}| + |y_{1} - y_{2}| + W_{2}(\mu_{1}, \mu_{2}) + W_{2}(\nu_{1}, \nu_{2}). \]

Similarly,
\[ |\sigma(x_{1}, y_{1}, \mu_{1}, \nu_{1}) - \sigma(x_{2}, y_{2}, \mu_{2}, \nu_{2})| \leq |x_{1} - x_{2}| + |y_{1} - y_{2}| + W_{2}(\mu_{1}, \mu_{2}) + W_{2}(\nu_{1}, \nu_{2}) \]
and
\[ |\sigma^{-2}(x_{1}, y_{1}, \mu_{1}, \nu_{1}) - \sigma^{-2}(x_{2}, y_{2}, \mu_{2}, \nu_{2})| \]
\[ = \left| \frac{1}{(1 + |x_{1}| + |y_{1}| + \int_{\mathbb{R}} |x| \mu_{1}(dx) + \int_{\mathbb{R}} |y| \nu_{1}(dy))^{2}} - \frac{1}{(1 + |x_{2}| + |y_{2}| + \int_{\mathbb{R}} |x| \mu_{2}(dx) + \int_{\mathbb{R}} |y| \nu_{2}(dy))^{2}} \right| \]
\[ \leq 2 \left| \frac{1}{1 + |x_{1}| + |y_{1}| + \int_{\mathbb{R}} |x| \mu_{1}(dx) + \int_{\mathbb{R}} |y| \nu_{1}(dy)} - \frac{1}{1 + |x_{2}| + |y_{2}| + \int_{\mathbb{R}} |x| \mu_{2}(dx) + \int_{\mathbb{R}} |y| \nu_{2}(dy)} \right| \]
\[ \leq 2 \left| x_{1} + \int_{\mathbb{R}} |x| \mu_{1}(dx) + \int_{\mathbb{R}} |y| \nu_{1}(dy) - |x_{2}| - |y_{2}| - \int_{\mathbb{R}} |y| \nu_{2}(dy) \right| \]
\[ \leq 2 \{ |x_{1} - x_{2}| + |y_{1} - y_{2}| + W_{2}(\mu_{1}, \mu_{2}) + W_{2}(\nu_{1}, \nu_{2}) \}. \]

Furthermore,
\[ \nabla_{\theta} b(x, y, \mu, \nu, \theta) = x + y + \int_{\mathbb{R}} x \mu(dx) + \int_{\mathbb{R}} y \nu(dy) \]
and \( \nabla_{\theta}^{(2)} b(x, y, \mu, \nu, \theta) = 0 \), (4.26) for any \( x, y \in \mathbb{R} \) and \( \mu, \nu \in \mathcal{P}_{p}(\mathbb{R}) \). The contrast function admits the form below
\[ \Psi_{n, \epsilon}(\theta) \]
\[ = \epsilon^{-2} \delta^{-1} \sum_{k=1}^{n} \Lambda^{-1}(X_{(k-1)\delta}^{\epsilon,i,N}, X_{(k-1)\delta-\tau}^{\epsilon,i,N})(P_{k}^{\epsilon,i,N}(\theta))^{2} \]
\[ = \epsilon^{-2} \delta^{-1} \sum_{k=1}^{n} \left( 1 + |X_{(k-1)\delta}^{\epsilon,i,N}| + |X_{(k-1)\delta-\tau}^{\epsilon,i,N}| + \int_{\mathbb{R}} |x| \mu_{(k-1)\delta}^{\epsilon,N}(dx) + \int_{\mathbb{R}} |x| \mu_{(k-1)\delta-\tau}^{\epsilon,N}(dx) \right) \]
\[ \times |X_{k\delta}^{\epsilon,i,N} - X_{(k-1)\delta}^{\epsilon,i,N}|^{2} \]
\[ \int_{\mathbb{R}} x \mu_{(k-1)\delta}^{\epsilon,N}(dx) + \int_{\mathbb{R}} x \mu_{(k-1)\delta-\tau}^{\epsilon,N}(dx) + \theta)^{2}. \]

In order to obtain the LSE \( \hat{\theta}_{n, \epsilon} \) of the unknown parameter \( \theta \in \Theta_{0} \), we need to find the stable point of the contrast function. Let
\[ \nabla_{\theta} \Psi_{n, \epsilon}(\theta) \]
Consider the following distribution-dependent SDDE, for \( \theta \) the true value

\[
\theta_{n,\varepsilon} = \arg\min_{\theta} \left( \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\left( 1 + |X_{(k-1)\delta}^{\varepsilon,N}| + |X_{(k-1)\delta-\tau}^{\varepsilon,N}| + \int_{\mathbb{R}} |x| \mu_{(k-1)\delta}^{\varepsilon,N}(dx) + \int_{\mathbb{R}} |x| \mu_{(k-1)\delta-\tau}^{\varepsilon,N}(dx) \right)^2 \right)
\]

This yields that

\[
\tilde{\theta}_{n,\varepsilon} = \frac{X_{k\delta}^{\varepsilon,N} - X_{(k-1)\delta}^{\varepsilon,N}}{\delta} \left( X_{(k-1)\delta}^{\varepsilon,N} + X_{(k-1)\delta-\tau}^{\varepsilon,N} + \int_{\mathbb{R}} x \mu_{(k-1)\delta}^{\varepsilon,N}(dx) + \int_{\mathbb{R}} x \mu_{(k-1)\delta-\tau}^{\varepsilon,N}(dx) \right).
\]

In terms of Theorem 3.7, \( \tilde{\theta}_{n,\varepsilon} \to \theta \) in probability as \( \varepsilon \to 0 \) and \( n, N \to \infty \). Next, from (4.1), (4.2) and (4.26), one gets

\[
I(\theta_0) = \int_0^T \Delta^{-1}(X_t^{0,i}, X_t^{0,i-\tau})(\nabla_x b)^2(X_t^{0,i}, X_t^{0,i-\tau}, \mu_t^{0,i}, \mu_t^{0,i}, \theta_0)dt
\]

and

\[
\int_0^T \Upsilon(X_t^{0,i}, X_t^{0,i-\tau}, \theta_0) dW_t^i
\]

At last, according to Theorem 4.1, we conclude that

\[
\varepsilon^{-1}(\tilde{\theta}_{n,\varepsilon} - \theta_0) \to I^{-1}(\theta_0) \int_0^T \Upsilon(X_t^{0,i}, X_t^{0,i-\tau}, \theta_0) dW_t^i, \quad \mathbb{P} \text{-a.s.}
\]

as \( n, N \to \infty, \varepsilon \to 0 \).

**Example 4.5.** Consider the following distribution-dependent SDDE, for \( \varepsilon \in (0,1) \),

\[
\begin{align*}
\left\{ 
\begin{array}{l}
\text{d}X_t^{\varepsilon} = \left( \theta_1 + \theta_2 \left( \int_{\mathbb{R}} b_1(X_t^{\varepsilon}, x) \mu_t^{\varepsilon}(dx) + \int_{\mathbb{R}} b_2(X_t^{\varepsilon-\tau}, x) \mu_{t-\tau}^{\varepsilon}(dx) \right) \right) dt \\
\quad \quad + \varepsilon(1 + |X_t^{\varepsilon}| + |X_t^{\varepsilon-\tau}|) dW_t, \quad t \in (0, T],
\end{array}
\right.
\end{align*}
\]

(4.27)

where \( \theta = (\theta_1, \theta_2) \in \Theta_0 := (a_1, a_2) \times (a_3, a_4) \subset \mathbb{R}^2 \) is an unknown parameter with the true value \( \theta_0 \in \Theta_0 \), \( X_0^\varepsilon = \xi \in C_0^\varepsilon([-\tau, 0]; \mathbb{R}^d) \) with \( \mathcal{X}_\varepsilon \in \mathcal{P}_p(\mathbb{R}^d) \), \( p > 4 \), and \( b_1, b_2 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \) satisfy the global Lipschitz condition, that is, for any \( x_i, y_i \in \mathbb{R}, i = 1, 2 \)

\[
|b_1(x_1, x_2) - b_1(y_1, y_2)| \leq K(|x_1 - y_1| + |x_2 - y_2|), \quad K > 0.
\]

(4.28)
In this example, let
\[ b(X^+_{i,x}, X^-_{i,x}, \mu_i, \mu_i^-) = \theta_1 + \theta_2 \left( \int_R b_1(X^+_{i,x}, x) \mu_i^+ (dx) + \int_R b_2(X^-_{i,x}, x) \mu_i^- (dx) \right) \]
(4.29)
and
\[ \sigma(X^+_{i,x}, X^-_{i,x}, \mu_i, \mu_i^-) = 1 + |X^+_{i,x}| + |X^-_{i,x}|. \]
(4.30)
Then (4.27) can be rewritten as (2.3). For any \( x_i, y_i \in \mathbb{R} \) and \( \mu_i, \nu_i \in P_p(\mathbb{R}), i = 1, 2 \), by (4.28), one has
\[
|b(x_1, y_1, \mu_1, \nu_1, \theta) - b(x_2, y_2, \mu_2, \nu_2, \theta)|
\leq |\theta_2| \left( \int_R \int_R |b_1(x_1, x) - b_1(x_2, x)| \mu_1 (dx) \nu_2 (dx) \right.
\quad+
\left. \int_R \int_R |b_2(y_1, y) - b_1(y_2, y)| \nu_1 (dy) \nu_2 (dy) \right)
\leq K([a_3] \vee [a_4]) \{ |x_1 - x_2| + |y_1 - y_2| + W_1(\mu_1, \mu_2) + W_2(\nu_1, \nu_2) \}
\leq K([a_3] \vee [a_4]) \{ |x_1 - x_2| + |y_1 - y_2| + W_1(\mu_1, \mu_2) + W_2(\nu_1, \nu_2) \},
\]
so that the assumptions (A1)-(A3) hold for (4.27). Moreover, according to (2.1) and (4.29), a direct calculation implies that
\[
(\nabla_\theta)(X^+_{i,x}, X^-_{i,x}, \mu_i, \mu_i^-) = \left( 1, \int_R b_1(X^+_{i,x}, x) \mu_i^+ (dx) + \int_R b_2(X^-_{i,x}, x) \mu_i^- (dx) \right)
\]
(4.31)
and
\[
(\nabla_\theta)(X^+_{i,x}, X^-_{i,x}, \mu_i, \mu_i^-) = 0_{2 \times 2},
\]
(4.32)
where \( 0_{2 \times 2} \) stands for the \( 2 \times 2 \) zero matrix. Hence, (4.31) and (4.32) yield that both (A4) and (A5) hold. In view of (2.8), the contrast function admits the form below
\[
\Psi_{n, \epsilon}^\tau(\theta)
\]
\[ = \varepsilon^{-2} \delta^{-3} \sum_{k=1}^n \frac{1}{(1 + |X^\epsilon_{k,i,N}^{+,\delta} + |X^\epsilon_{k,i,N}^{-,\delta} - \tau|)^2} \left| X^\epsilon_{k,i,N}^{+,\delta} - X^\epsilon_{k,i,N}^{-,\delta} \right|^2
\quad- \delta \left( \theta_1 + \theta_2 \left( \int_R b_1(X^\epsilon_{k,i,N}^{+,\delta}, x) \mu_i^{+,\delta} (dx) + \int_R b_2(X^\epsilon_{k,i,N}^{-,\delta}, x) \mu_i^{-,\delta} (dx) \right) \right)^2.\]
Similarly,
\[
\frac{\partial}{\partial \theta_1} \Psi_{n,\varepsilon}^{\alpha,i,N}(\theta) = -2 \varepsilon^{-2} \sum_{k=1}^{n} \frac{1}{(1 + |X_{(k-1)}^{\varepsilon,i,N}| + |X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau)|)^2} \left\{ X_{\delta}^{\varepsilon,i,N} - X_{(k-1)}^{\varepsilon,i,N} \right\} \right.
- \delta \left( \theta_1 + \theta_2 \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right) \right),
\]
and
\[
\frac{\partial}{\partial \theta_2} \Psi_{n,\varepsilon}^{\alpha,i,N}(\theta) = -2 \varepsilon^{-2} \sum_{k=1}^{n} \frac{1}{(1 + |X_{(k-1)}^{\varepsilon,i,N}| + |X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau)|)^2} \left\{ X_{\delta}^{\varepsilon,i,N} - X_{(k-1)}^{\varepsilon,i,N} \right\} \right.
- \delta \left( \theta_1 + \theta_2 \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right) \right) \times \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right).
\]

From the following equation
\[
\frac{\partial}{\partial \theta_1} \Psi_{n,\varepsilon}^{\alpha,i,N}(\theta) = \frac{\partial}{\partial \theta_2} \Psi_{n,\varepsilon}^{\alpha,i,N}(\theta) = 0,
\]
it is easy to obtain the LSE \( \hat{\theta}_{n,\varepsilon}^{\alpha,i,N} = (\hat{\theta}_{n,\varepsilon,1}^{\alpha,i,N}, \hat{\theta}_{n,\varepsilon,2}^{\alpha,i,N})^* \) of the unknown parameter \( \theta = (\theta_1, \theta_2)^* \in \Theta_0 \) possesses the formula
\[
\hat{\theta}_{n,\varepsilon,1}^{\alpha,i,N} = \frac{\alpha_2 \alpha_5 - \alpha_3 \alpha_4}{\delta (\alpha_1 \alpha_5 - \alpha_4^2)}\quad \text{and} \quad \hat{\theta}_{n,\varepsilon,2}^{\alpha,i,N} = \frac{\alpha_1 \alpha_3 - \alpha_2 \alpha_4}{\delta (\alpha_1 \alpha_5 - \alpha_4^2)},
\]
where
\[
\alpha_1 := \sum_{k=1}^{n} \frac{1}{(1 + |X_{(k-1)}^{\varepsilon,i,N}| + |X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau)|)^2} X_{\delta}^{\varepsilon,i,N} - X_{(k-1)}^{\varepsilon,i,N}
\]
\[
\alpha_2 := \sum_{k=1}^{n} \frac{1}{(1 + |X_{(k-1)}^{\varepsilon,i,N}| + |X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau)|)^2} |X_{\delta}^{\varepsilon,i,N} - X_{(k-1)}^{\varepsilon,i,N}|
\]
\[
\alpha_3 := \sum_{k=1}^{n} \left( X_{\delta}^{\varepsilon,i,N} - X_{(k-1)}^{\varepsilon,i,N} \right) \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right)
\]
\[
\alpha_4 := \sum_{k=1}^{n} \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right)
\]
\[
\alpha_5 := \sum_{k=1}^{n} \left( \int_{\mathbb{R}} b_1(X_{(k-1)}^{\varepsilon,i,N} (\delta_1), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) + \int_{\mathbb{R}} b_2(X_{(k-1)}^{\varepsilon,i,N} (\delta_1 - \tau), x) \mu^{\varepsilon,i,N}_{(k-1)}(dx) \right)^2.
\]

In view of Theorem 3.7, \( \hat{\theta}_{n,\varepsilon}^{\alpha,i,N} \rightarrow \theta \) in probability as \( \varepsilon \rightarrow 0 \) and \( n, N \rightarrow \infty \). Set \( \beta := b_1(X_{t}^{0,i}, X_{t-\tau}^{0,i}) + b_2(X_{t-\tau}^{0,i}, X_{t}^{0,i}) \), then it follows from (4.2) and (4.31) that
\[
I(\theta_0) = \int_0^T \frac{1}{1 + |X_{t-\tau}^{0,i}| + |X_{t}^{0,i}|} \left( \begin{array}{c} \beta \\ \beta^2 \end{array} \right)
\]
and
\[
\int_0^T \Psi(X_{t}^{0,i}, X_{t-\tau}^{0,i}, \theta_0) dW_t = \int_0^T \frac{1}{1 + |X_{t-\tau}^{0,i}| + |X_{t}^{0,i}|} \left( \begin{array}{c} 1 \\ \beta \end{array} \right) dW_t.
\]
In conclusion, by Theorem 4.1, we immediately get

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

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

Acknowledgments. We would like to thank the anonymous referees and editors for their constructive comments and corrections.

REFERENCES

[1] J. Bao and X. Huang, Approximations of McKean-Vlasov stochastic differential equations with irregular coefficients, J. Theor. Probab., (2021), 29 pp.
[2] R. Buckdahn, J. Li and J. Ma, A mean-field stochastic control problem with partial observations, Ann. Appl. Probab., 27 (2017), 3201–3245.
[3] R. Carmona and F. Delarue, Probabilistic Theory of Mean Field Games with Applications I. Mean Field FBSDEs, Control, and Games, 1st edition, Springer, Cham, 2018.
[4] R. Buckdahn, J. Li and J. Ma, Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes, Trans. Amer. Math. Soc., 371 (2019), 7135–7173.
[5] Y. Hu and H. Long, Least squares estimator for Ornstein-Uhlenbeck processes driven by $\alpha$-stable motions, Stochastic Process. Appl., 119 (2009), 2465–2480.
[6] X. Huang, Path-Distribution Dependent SDEs with Singular Coefficients, Electron. J. Probab., 26 (2021), 1–21.
[7] X. Huang, M. Röckner and F. Wang, Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, Discrete Contin. Dyn. Syst., 39 (2019), 3017–3035.
[8] X. Huang, M. Röckner and F. Wang, Comparison theorem for distribution-dependent neutral SFDEs, J. Evol. Equ., 21 (2021), 653–670.
[9] M. Kac, Foundations of kinetic theory, Proc. Third Berkeley Symp. on Math. Stat. and Prob., 3 (1956), 171–197.
[10] M. Kac, Probability and Related Topics in the Physical Sciences, 1st edition, Interscience Publishers, New York, 1960.
[11] R. A. Kasonga, The consistency of a non-linear least squares estimator from diffusion processes, Stochastic Process. Appl., 2 (1988), 263–275.
[12] H. Long, C. Ma and Y. Shimizu, Least squares estimators for stochastic differential equations driven by small Lévy noises, Stochastic Process. Appl., 127 (2017), 1475–1495.
[13] H. Long, Y. Shimizu and W. Sun, Least squares estimators for discretely observed stochastic processes driven by small Lévy noises, J. Multivariate Anal., 116 (2013), 422–439.
[14] C. Ma, A note on ‘Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises’, Statist. Probab. Lett., 80 (2010), 1528–1531.
[15] M. Kac, Stochastic differential equations and applications, 2nd edition, Horwood Publishing Limited, Chichester, 2008.
[16] P. Ren and F. Wang Bismut formula for Lions derivative of distribution dependent SDEs and applications, J. Differ. Equ., 267 (2019), 4745–4777.
[17] P. Ren and J. Wu, Least squares estimation for path-distribution dependent stochastic differential equations, Appl. Math. Comput., 410 (2021), 126457.
[18] M. Röckner and X. Zhang, Well-posedness of distribution dependent SDEs with singular drifts, Appl. Math. Comput., 2 (2021), 1131–1158.
[19] A. W. van der Vaart, Asymptotic statistics, Cambridge Series in Statistical and Probabilistic Mathematics, 3rd edition, Cambridge University Press, Cambridge, 1998.
[20] F. Wang, Distribution-dependent SDEs for Landau type equations, Stochastic Process. Appl., 128 (2018), 595–621.
[21] N. Yoshida, Asymptotic expansion for statistics related to small diffusions, J. Japan Statist. Soc., 22 (1992), 139–159.

Received July 2021; revised December 2021; early access February 2022.

E-mail address: yanyanhu@bit.edu.cn
E-mail address: xife@bit.edu.cn
E-mail address: zhumin0107@hut.edu.cn