

Local asymptotic normality for discretely observed McKean-Vlasov diffusions

Akram Heidari *

Mark Podolskij †

Abstract

We study the local asymptotic normality (LAN) property for the likelihood function associated with discretely observed d -dimensional McKean-Vlasov stochastic differential equations over a fixed time interval. The model involves a joint parameter in both the drift and diffusion coefficients, introducing challenges due to its dependence on the process distribution. We derive a stochastic expansion of the log-likelihood ratio using Malliavin calculus techniques and establish the LAN property under appropriate conditions. The main technical challenge arises from the implicit nature of the transition densities, which we address through integration by parts and Gaussian-type bounds. This work extends existing LAN results for interacting particle systems to the mean-field regime, contributing to statistical inference in non-linear stochastic models.

Keywords: LAN property, log-likelihood ratio, Malliavin calculus, McKean-Vlasov diffusions, parametric estimation.

AMS subject classifications: 62F12, 62E20, 62M05, 60G07, 60H10.

1 Introduction

The study of McKean-Vlasov stochastic differential equations (SDEs) have gained significant attention in recent years due to their wide-ranging applications in statistical physics, finance, and mean-field games among other fields [3, 7, 8, 14, 15, 18, 21, 28]. These equations are characterized by their dependence on the law of the solution, making them inherently nonlinear. In this work, we consider an i.i.d. array of d -dimensional processes, defined on a complete probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, governed by the McKean-Vlasov SDE:

$$\begin{cases} dX_t^{i,\theta} = b_{\theta_1}(X_t^{i,\theta}, \mu_t^\theta)dt + a_{\theta_2}(X_t^{i,\theta})dW_t^i & i = 1, \dots, N, \quad t \in [0, T] \\ \text{Law}(X_0^{1,\theta}, \dots, X_0^{N,\theta}) := \mu_0 \times \dots \times \mu_0 \end{cases} \quad (1.1)$$

where the unknown parameter $\theta := (\theta_1, \theta_2)$ is an element of the set $\Theta = \Theta_1 \times \Theta_2$, and $\Theta_j \subset \mathbb{R}, j = 1, 2$ are compact and convex sets with nonempty interior. The d -dimensional Brownian motions $(W^i)_{1 \leq i \leq N}$ are independent, μ_t^θ denotes the law of $X_t^{i,\theta}$, and

$$b : \Theta_1 \times \mathbb{R}^d \times \mathcal{P}_2 \mapsto \mathbb{R}^d, \quad a : \Theta_2 \times \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$$

are the drift and diffusion coefficient, respectively. Here, $\mathcal{P}_2$ denotes the set of probability measures on $\mathbb{R}^d$ with a finite second moment, endowed with the Wasserstein 2-metric

$$W_2(\mu, \lambda) := \left(\inf_{m \in \Gamma(\mu, \lambda)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 m(dx, dy) \right)^{\frac{1}{2}}$$

*Department of Mathematics, University of Luxembourg, Maison du Nombre, 6 Avenue de la Fonte, 4364 Esch-sur-Alzette, Luxembourg, Email: akram.heidari@uni.lu

†Department of Mathematics, University of Luxembourg, Maison du Nombre, 6 Avenue de la Fonte, 4364 Esch-sur-Alzette, Luxembourg, Email: mark.podolskij@uni.lu

where $\Gamma(\mu, \lambda)$ denotes the set of probability measures on the product space $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and λ .

In this paper, we aim to establish the local asymptotic normality (LAN) property for the likelihood function associated with the discrete observations

$$\left(X_{t_k}^{i, \theta} \right)_{k=1, \dots, n}^{i=1, \dots, N}, \quad (1.2)$$

where $t_k := Tk/n$ and $\Delta_n := T/n$ denotes the discretization step. We consider the asymptotic regime $\Delta_n \rightarrow 0$, $N \rightarrow \infty$, and the time horizon T being fixed. We recall that a sequence of statistical models $(\mathbb{P}_{m, \theta} : \theta \in \Theta \in \mathbb{R}^p)$ is said to be locally asymptotically normal if there exist matrices $r_m, \Sigma_\theta \in \mathbb{R}^{p \times p}$ such that for any $h \in \mathbb{R}^p$:

$$\log \frac{\mathbb{P}_{m, \theta + r_m^{-1} h}}{\mathbb{P}_{m, \theta}} = h^\top \mathcal{N} - \frac{1}{2} h^\top \Sigma_\theta h + o_{\mathbb{P}_{m, \theta}}(1), \quad \text{as } m \rightarrow \infty,$$

where $\mathcal{N} \sim \mathcal{N}(0, \Sigma_\theta)$. The LAN property is a crucial tool in asymptotic statistical inference, originally introduced by Le Cam. When the LAN property holds and the covariance matrix is invertible, minimax theorems can be applied to derive lower bounds on the asymptotic variance of estimators (see e.g. [22, 25, 26]).

Our work contributes to the growing field of statistical inference for McKean-Vlasov processes, providing new insights into the asymptotic properties of parameter estimators in a mean-field setting. Numerous parametric estimation methods for McKean-Vlasov diffusions have been studied in the literature under different sampling schemes in [1, 6, 9, 10, 16, 17, 24, 27, 31, 32]. Non-parametric approaches have also gained increasing attention, with notable contributions including [2, 4, 11, 12, 29].

A major challenge in establishing the local asymptotic normality property for McKean-Vlasov SDEs lies in the intractability of their transition densities, which complicates the analysis of the likelihood function's asymptotic behavior. To address this, we employ tools from Malliavin calculus—specifically, the integration by parts formula—to derive an explicit representation of the logarithmic derivative of the transition density. This methodology, originally introduced by Gobet in the context of classical diffusion models [19, 20], enables a stochastic expansion of the log-likelihood ratio and ultimately yields the desired LAN result. In contrast to Gobet's framework [20], which relies on the ergodicity of the underlying process, our approach does not require this assumption, as the asymptotic regime is instead driven by the growing number of particles.

We remark that the drift and diffusion coefficients are estimated at different asymptotic rates. Specifically, the drift parameter is estimated at rate $\sqrt{N}$, while the diffusion parameter is estimated at rate $\sqrt{N/\Delta_n}$. These findings are consistent with the recent results in [1], which develop a contrast-based estimation method for discretely observed particle systems and establish consistency and asymptotic normality of the resulting estimators. A closely related contribution is [13], which proves the LAN property for drift estimation in continuously observed d -dimensional McKean-Vlasov models. It is important to note, however, that in the latter setting the likelihood function admits a closed-form expression via the Girsanov theorem, which simplifies the analysis.

The paper is organized as follows. In Section 2 we introduce some notations, formulate the assumptions and state some technical lemmas we will use in order to show our main results. Section 3 is devoted to the preliminary results essential for proving the LAN property, such as an explicit expression for the derivative of the log-transition density using the Malliavin calculus, and the main theorem of the paper. Proofs are collected in Section 4.

2 Notation and assumptions

In this section, we introduce the notation and the main assumptions associated with the model (1.1).

2.1 Notation

Throughout the paper, we use a generic positive constant denoted by C , or C_q when it depends on an external parameter q . This constant is independent of n , N and θ , and may vary from line to line. The dimensions of the state space and the parameter space, as well as the time horizon $T > 0$, are fixed.

All vectors are understood as column vectors. We denote the Euclidean norm on $\mathbb{R}^d$ by $\|\cdot\|$, and for any vector $x \in \mathbb{R}^d$, we define $x^{\otimes 2} := xx^\top$; the r -th component of x is written as x_r or x^r . The trace of a matrix $A \in \mathbb{R}^{d \times d}$ is denoted by $\text{tr}(A)$. For a function $f : \mathbb{R}^d \times \Theta_j \rightarrow \mathbb{R}$, we denote by $\nabla_x f$ (resp. $\partial_{\theta_j} f$) the derivative with respect to x (resp. with respect to θ_j). A function $f : \mathbb{R}^d \times \mathcal{P}_l \rightarrow \mathbb{R}^d$ is said to have *polynomial growth* if there exist constants $k, l \geq 0$ such that

$$\|f(x, \mu)\| \leq C \left(1 + \|x\|^k + W_2^l(\mu, \delta_0)\right)$$

for all $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_l$, where $\mathcal{P}_l$ denotes the space of probability measures on $\mathbb{R}^d$ with finite l -th absolute moment. For $l \in [1, \infty)$, the Wasserstein- l distance between two probability measures μ and λ on $\mathbb{R}^d$ is defined as

$$W_l(\mu, \lambda) := \left(\inf_{m \in \Gamma(\mu, \lambda)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^l m(dx, dy) \right)^{1/l},$$

where $\Gamma(\mu, \lambda)$ denotes the set of all couplings of μ and λ .

We often suppress the dependence on θ and write X_t (or X_t^i) to denote the observed data. Moreover, we define conditional expectations

$$E_{t,x}^\theta[Z] := \mathbb{E}^\theta[Z | X_t^\theta = x] \quad \text{and} \quad \mathbb{E}_t^\theta[Z] := \mathbb{E}^\theta[Z | \mathcal{F}_t],$$

and introduce the conditional mean and covariance matrix of $X_{t_{k+1}}^{i,\theta}$ given the starting point $X_{t_k}^{i,\theta} = x \in \mathbb{R}^d$ by

$$\begin{aligned} m_{t_k, t_{k+1}}^\theta(x) &:= \mathbb{E}_{t_k, x}^\theta \left[X_{t_{k+1}}^{i,\theta} \right] \in \mathbb{R}^d, \\ V_{t_k, t_{k+1}}^\theta(x) &:= \mathbb{E}_{t_k, x}^\theta \left[\left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta(x) \right)^{\otimes 2} \right] \in \mathbb{R}^{d \times d}. \end{aligned} \tag{2.3}$$

For $\mathcal{F}_t$ -measurable random variables Y_t^i , $i = 1, \dots, N$, we will often use the notation

$$Y_t^i = R_t^i(\varepsilon), \quad \varepsilon = \varepsilon(n, N), \tag{2.4}$$

if $\varepsilon^{-1}Y_t^i$ is uniformly bounded (in (θ, i, t, n, N)) in $L^q(\Omega)$ for all $q \geq 1$, that is

$$\mathbb{E}^\theta[|\varepsilon^{-1}Y_t^i|^q]^{1/q} \leq C_q.$$

For a sequence of random variables $(Y_n)_{n \geq 1}$, we write $Y_n = O_{\mathbb{P}}(\varepsilon_n)$ (resp. $Y_n = o_{\mathbb{P}}(\varepsilon_n)$) when $\varepsilon_n^{-1}Y_n$ is stochastically bounded (resp. $\varepsilon_n^{-1}Y_n$ converges in probability to 0).

2.2 Model assumptions

In order to get asymptotic properties of the likelihood ratio, it is necessary to put some additional conditions on the coefficients. We shall work under the following assumptions on (1.1):

A1. *The initial distribution μ_0 is sub-Gaussian, i.e. there exists a constant $\sigma > 0$ such that*

$$\mu_0(A) \leq C\Phi_\sigma(A), \quad \text{for any Borel set } A,$$

where Φ_σ denotes the distribution function of $\mathcal{N}_d(0, \sigma^2 \mathbb{I}_d)$.

A2. *The functions b_{θ_1} and a_{θ_2} are bounded uniformly in $\theta \in \Theta$. For all θ there exists $C > 0$ such that for all $(x, \mu), (y, \lambda) \in \mathbb{R}^d \times \mathcal{P}_2$,*

$$\|b_{\theta_1}(x, \mu) - b_{\theta_1}(y, \lambda)\| \leq C(\|x - y\| + W_2(\mu, \lambda)),$$

$$\|a_{\theta_2}(x) - a_{\theta_2}(y)\| \leq C\|x - y\|.$$

We remark that Assumption A2 guarantees existence and uniqueness of the McKean-Vlasov SDE (1.1). As for the regularity of the drift $b : \Theta_1 \times \mathbb{R}^d \times \mathcal{P}_2 \rightarrow \mathbb{R}^d$ the notion of linear differentiability, commonly used in the literature on McKean-Vlasov equations and mean-field games in order to quantify the smoothness of $\mu \mapsto b_{\theta_1}(x, \mu)$ as a mapping $\mathcal{P}_2 \rightarrow \mathbb{R}^d$, will be useful in our setting. We refer in particular to [12, Section 2] and the references therein.

Definition 1. *A mapping $f : \mathcal{P}_2 \rightarrow \mathbb{R}^d$ is said to have a linear functional derivative, if there exists $\partial_\mu f : \mathbb{R}^d \times \mathcal{P}_2 \rightarrow \mathbb{R}^d$ such that*

$$f(\mu) - f(\mu') = \int_0^1 \int_{\mathbb{R}^d} \partial_\mu f(y, \lambda\mu + (1-\lambda)\mu')(\mu - \mu')(dy)d\lambda$$

for every $(\mu, \mu') \in \mathcal{P}_2$ and $\partial_\mu f$ satisfies additional smoothness properties, which will be provided in the following assumption.

In the linear case $f(\mu) = \int_{\mathbb{R}^d} g(x)\mu(dx)$, where $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a μ -integrable function, we simply have $\partial_\mu f(y, \mu) = g(y)$.

A3. *Regularity of the derivatives:*

(I) *For all (x, μ) , the functions $b_{\theta_1}(x, \mu)$, $a_{\theta_2}(x)$ are in $C^2(\Theta_1; \mathbb{R}^d)$, $C^2(\Theta_2; \mathbb{R}^{d \times d})$ respectively. Furthermore, all their partial derivatives up to order three have polynomial growth, uniformly in θ .*

(II) *The first and second order derivatives in θ are locally Lipschitz in (x, μ) with polynomial weights, i.e. for all θ there exists $C > 0$, $k, l = 0, 1, \dots$ such that for all $r_1 + r_2 = 1, 2$ $(x, \mu), (x', \mu') \in \mathbb{R}^d \times \mathcal{P}_2$,*

$$\begin{aligned} & \|\partial_{\theta_1}^{r_1} \partial_{\theta_1}^{r_2} b_{\theta_1}(x, \mu) - \partial_{\theta_1}^{r_1} \partial_{\theta_1}^{r_2} b_{\theta_1}(x', \mu')\| + \|\partial_{\theta_2}^{r_1} \partial_{\theta_2}^{r_2} a_{\theta_2}(x) - \partial_{\theta_2}^{r_1} \partial_{\theta_2}^{r_2} a_{\theta_2}(x')\| \\ & \leq C(\|x - x'\| + W_1(\mu, \mu'))(1 + \|x\|^k + \|x'\|^k + W_1^l(\mu, \delta_0) + W_1^l(\mu', \delta_0)). \end{aligned}$$

(III) *The map $\mu \mapsto b_{\theta_1}(x, \mu)$ admits a functional derivative in the sense of Definition 1. There exists $k, k', l \geq 1$ such that*

$$\begin{aligned} & \|\partial_\mu b_{\theta_1}(x, y, \mu) - \partial_\mu b_{\theta_1}(x', y', \mu')\| \leq C(\|x - x'\| + \|y - y'\| + W_1(\mu, \mu')) \\ & \|\partial_\mu b_{\theta_1}(x, y, \mu)\| \leq C(1 + \|x\|^k + \|y\|^{k'} + W_l^l(\mu, \delta_0)). \end{aligned}$$

A4. *For every point θ in the interior of Θ , the coefficient functions are twice differentiable in x . Furthermore, the following estimates hold:*

$$(a) \quad \|\nabla_x b_{\theta_1}(x, \mu)\| + \|\nabla_x a_{\theta_2}(x)\| \leq C;$$

$$(b) \quad \|g(\cdot, x)\| \leq C(1 + \|x\|^q) \text{ for } g = \nabla_x^2 b, \nabla_x \partial_{\theta_1} b, \nabla_x^2 a, \nabla_x \partial_{\theta_2} a$$

for some positive constants C and q .

A5. (Regularity of the diffusion coefficient) The diffusion matrix a is symmetric, positive definite and satisfies an uniform ellipticity condition. That is, there exists a positive constant c such that

$$\forall (\theta_2, x) \in \Theta_2 \times \mathbb{R}^d, \quad \frac{1}{c} \mathbb{I}_d \leq a_{\theta_2}(x) \leq c \mathbb{I}_d.$$

We note that Assumptions A3-A5 are direct analogues of conditions imposed in [19, 20], which in particular imply uniform lower and upper bounds for the transition density established in Proposition 4.2.

2.3 Basic elements of Malliavin calculus

In this section, we present the fundamental concepts of Malliavin calculus that will be used throughout the remainder of the text.

Let $\mathbb{H} := L^2([0, T], \mathbb{R}^d)$ and consider a function $h \in \mathbb{H}$. We define the Itô integral of h with respect to a standard $\mathbb{R}^d$ -valued Brownian motion $B = (B_t)_{t \in [0, T]}$ by

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

Let $\mathcal{S}$ denote the class of *smooth random variables*, i.e., random variables of the form

$$F = f(B(h_1), \dots, B(h_k)),$$

where $h_i \in \mathbb{H}$ and $f : \mathbb{R}^k \rightarrow \mathbb{R}$ is a C_p^∞ -function (that is, a function which is infinitely differentiable and all its derivatives grow at most polynomially).

The *Malliavin derivative* of $F \in \mathcal{S}$ is defined as the $\mathbb{H}$ -valued random variable

$$\mathcal{D}F = \sum_{j=1}^k \partial_{x_j} f(B(h_1), \dots, B(h_k)) \cdot h_j.$$

The operator $\mathcal{D}$ is closable from $L^p(\Omega)$ to $L^p(\Omega; \mathbb{H})$ for any $p \geq 1$. Its closure has domain denoted by $\mathbb{D}^{1,p}$, which consists of all $F \in L^p(\Omega)$ such that $\mathcal{D}F \in L^p(\Omega; \mathbb{H})$, equipped with the norm

$$\|F\|_{1,p} := (\mathbb{E}[|F|^p] + \mathbb{E}[\|\mathcal{D}F\|_{\mathbb{H}}^p])^{1/p}.$$

We now introduce the *Skorohod integral* δ , which is defined as the adjoint of the Malliavin derivative. More precisely, δ is a linear operator from a subset of $L^2([0, T] \times \Omega; \mathbb{H})$ to $L^2(\Omega)$, characterized as follows:

(i) The domain of δ , denoted $\text{Dom}(\delta)$, consists of all processes $u \in L^2([0, T] \times \Omega; \mathbb{H})$ such that

$$\forall F \in \mathbb{D}^{1,2}, \quad |\mathbb{E}[\langle \mathcal{D}F, u \rangle_{\mathbb{H}}]| \leq c(u) \|F\|_{L^2(\Omega)}.$$

(ii) For $u \in \text{Dom}(\delta)$, the Skorohod integral $\delta(u) \in L^2(\Omega)$ is defined by the *integration by parts formula*

$$\forall F \in \mathbb{D}^{1,2}, \quad \mathbb{E}[F \delta(u)] = \mathbb{E}[\langle \mathcal{D}F, u \rangle_{\mathbb{H}}]. \quad (2.5)$$

It is well known that the Skorohod integral and the Itô integral coincide when both integrals exist. The following proposition summarizes key properties of the Skorohod integral.

Proposition 2.1. *Let $p > 1$. Then the following assertions hold:*

(i) *If $u \in \mathbb{D}^{1,p}(\mathbb{H})$, then $u \in \text{Dom}(\delta)$ and*

$$\|\delta(u)\|_p \leq c_p \left(\|u\|_{L^p(\Omega; \mathbb{H})} + \|\mathcal{D}u\|_{L^p(\Omega; \mathbb{H} \otimes \mathbb{H})} \right). \quad (2.6)$$

(ii) *Let $F \in \mathbb{D}^{1,2}$ and $u \in \text{Dom}(\delta)$ such that $\mathbb{E} \left[F^2 \int_0^T \|u_t\|^2 dt \right] < \infty$. Then, whenever the right-hand side is square-integrable, we have the product formula*

$$\delta(Fu) = F\delta(u) - \langle \mathcal{D}F, u \rangle_{\mathbb{H}}. \quad (2.7)$$

3 Main results

In this section, we present the main results of the paper. We begin by noting that for any $s < t$, the conditional law of $X_t^{\theta,i}$ given $X_s^{\theta,i} = x$ admits a strictly positive transition density (see Proposition 4.2 and [20, Proposition 1.2]), denoted by $p^\theta(s, t, x, y)$, which is differentiable with respect to the parameter θ . We define the perturbed parameter as

$$(\theta_1^+, \theta_2^+) := \left(\theta_1^0 + \frac{u}{\sqrt{N}}, \theta_2^0 + \frac{v}{\sqrt{N/\Delta_n}} \right), \quad \theta^+ := (\theta_1^+, \theta_2^+), \quad (3.8)$$

and exploit the Markov property of the processes $(X_t^{\theta,i})_{t \geq 0}$, $i = 1, \dots, N$, to express the log-likelihood ratio between the measures $\mathbb{P}^{\theta^+}$ and $\mathbb{P}^{\theta^0}$ as

$$z(\theta^0, \theta^+) := \log \frac{d\mathbb{P}^{\theta^+}}{d\mathbb{P}^{\theta^0}}(X_{t_k})_{k=1, \dots, n} = \sum_{k=1}^n \sum_{i=1}^N \log \left(\frac{p^{\theta^+}}{p^{\theta^0}} \right) (t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i). \quad (3.9)$$

From the expansion (3.9), we obtain the identity

$$z(\theta^0, \theta^+) = \sum_{k=1}^n \sum_{i=1}^N (\zeta_k^{i, \theta_1} + \zeta_k^{i, \theta_2}), \quad (3.10)$$

where the individual components are given by

$$\begin{aligned} \zeta_k^{i, \theta_1} &:= \frac{u}{\sqrt{N}} \int_0^1 \frac{\partial_{\theta_1} p^{\theta_1(l), \theta_2^+}}{p^{\theta_1(l), \theta_2^+}} (t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) dl, \\ \zeta_k^{i, \theta_2} &:= \frac{v}{\sqrt{N/\Delta_n}} \int_0^1 \frac{\partial_{\theta_2} p^{\theta_1^0, \theta_2(l)}}{p^{\theta_1^0, \theta_2(l)}} (t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) dl. \end{aligned} \quad (3.11)$$

Here, we use the notation

$$\theta_1(l) := \theta_1^0 + \frac{lu}{\sqrt{N}}, \quad \theta_2(l) := \theta_2^0 + \frac{lv}{\sqrt{N/\Delta_n}}. \quad (3.12)$$

In the next step, we present an expansion formula for the transition density expressions appearing in (3.11). To this end, we introduce the notation ($t \in [0, \Delta_n]$)

$$X_{t_k+t}^{i, \theta} = X_{t_k}^{i, \theta} + \int_0^t b_{\theta_1}(X_{t_k+s}^{i, \theta}, \mu_{t_k+s}^\theta) ds + \int_0^t a_{\theta_2}(X_{t_k+s}^{i, \theta}) dB_s^i,$$

where the process B^i represents the Brownian motion W^i shifted in time by t_k . Furthermore, for $t \in [0, \Delta_n]$ we introduce derivative processes which satisfy the following stochastic differential equations:

$$\begin{aligned}\partial_{\theta_1} X_t^{i,\theta} &= \int_0^t \left(\partial_{\theta_1} b_{\theta_1}(X_{t_k+s}^{i,\theta}, \mu_{t_k+s}^\theta) + \nabla_x b_{\theta_1}(X_{t_k+s}^{i,\theta}, \mu_{t_k+s}^\theta) \partial_{\theta_1} X_s^{i,\theta} \right. \\ &\quad \left. + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}(X_{t_k+s}^{i,\theta}, y, \mu_{t_k+s}^\theta) \partial_{\theta_1} \mu_{t_k+s}^\theta(dy) \right) ds + \sum_{r=1}^d \int_0^t \nabla_x a_{\theta_2}^r(X_{t_k+s}^{i,\theta}) \partial_{\theta_1} X_s^{i,\theta} dB_{r,s}^i, \\ \partial_{\theta_2} X_t^{i,\theta} &= \int_0^t \left(\nabla_x b_{\theta_1}(X_{t_k+s}^{i,\theta}, \mu_{t_k+s}^\theta) \partial_{\theta_2} X_s^{i,\theta} + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}(X_{t_k+s}^{i,\theta}, y, \mu_{t_k+s}^\theta) \partial_{\theta_2} \mu_{t_k+s}^\theta(dy) \right) ds \\ &\quad + \sum_{r=1}^d \int_0^t \left(\partial_{\theta_2} a_{\theta_2}^r(X_{t_k+s}^{i,\theta}) + \nabla_x a_{\theta_2}^r(X_{t_k+s}^{i,\theta}) \partial_{\theta_2} X_s^{i,\theta} \right) dB_{r,s}^i.\end{aligned}\tag{3.13}$$

We remark that the third term in the representation of $\partial_{\theta_1} X_t^{i,\theta}$ is indeed finite due to Assumption (A3)(III) and the fact that the density of $\partial_{\theta_1} \mu_t^\theta$ has a sub-Gaussian density due to Proposition 4.2. The same argument applies to the corresponding term in the representation of $\partial_{\theta_2} X_t^{i,\theta}$.

We now present an explicit formula for the Malliavin derivative $\mathcal{D}_s X_t^{i,\theta} \in \mathbb{R}^{d \times d}$. To this end, we introduce the $\mathbb{R}^{d \times d}$ -valued stochastic process $(Y_t^{i,\theta})_{t \in [0, \Delta_n]}$, defined by

$$Y_t^{i,\theta} := \mathbb{I}_d + \int_0^t \nabla_x b_{\theta_1}(X_{t_k+s}^{i,\theta}, \mu_{t_k+s}^\theta) Y_s^{i,\theta} ds + \sum_{r=1}^d \int_0^t \nabla_x a_{\theta_2}^r(X_{t_k+s}^{i,\theta}) Y_s^{i,\theta} dB_{r,s}^{r,i},\tag{3.14}$$

where $a_{\theta_2}^r$ denotes the r th column of the matrix-valued function a_{θ_2} . It can be shown that $Y_t^{i,\theta}$ is invertible for all $t \geq 0$, and the Malliavin derivative satisfies the following identity (see [30, Eq. (2.59)]):

$$\mathcal{D}_s X_{t_k+t}^{i,\theta} = Y_t^{i,\theta} (Y_s^{i,\theta})^{-1} a_{\theta_2}(X_{t_k+s}^{i,\theta}) \mathbb{1}_{[0,t]}(s).\tag{3.15}$$

Our first result provides an identity for the terms ζ_k^{i,θ_1} and ζ_k^{i,θ_2} introduced in (3.11).

Proposition 3.1. *Assume that Assumptions A1–A5 hold. Then, for $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$, the transition density $p^\theta(t_k, t_{k+1}, x, y)$ is absolutely continuous with respect to θ , and we have the representation*

$$\begin{aligned}\frac{\partial_{\theta_1} p^\theta}{p^\theta}(t_k, t_{k+1}, x, y) &= \frac{1}{\Delta_n} \mathbb{E}_{t_k, x}^\theta \left[\sum_{r=1}^d \delta \left(\partial_{\theta_1} X_{r, \Delta_n}^{i,\theta} U_r^i \right) \middle| X_{t_{k+1}}^{i,\theta} = y \right], \\ \frac{\partial_{\theta_2} p^\theta}{p^\theta}(t_k, t_{k+1}, x, y) &= \frac{1}{\Delta_n} \mathbb{E}_{t_k, x}^\theta \left[\sum_{r=1}^d \delta \left(\partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} U_r^i \right) \middle| X_{t_{k+1}}^{i,\theta} = y \right],\end{aligned}\tag{3.16}$$

where U_r^i denotes the r -th column of the $\mathbb{R}^{d \times d}$ -valued process

$$U_s^i := \left(\mathcal{D}_s X_{t_{k+1}}^{i,\theta} \right)^{-1} = a_{\theta_2}^{-1}(X_{t_k+s}^{i,\theta}) Y_s^{i,\theta} \left(Y_{\Delta_n}^{i,\theta} \right)^{-1}, \quad s \in [0, \Delta_n].$$

The application of Malliavin calculus to derive the transformation in (3.16) is essential for establishing the LAN property of the statistical model. This formula was originally shown by Gobet (see, e.g., [19, Proposition 4.1]) in the setting of classical stochastic differential equations, and the same reasoning applies in our framework, so we omit the proof. However, in contrast to classical SDEs, the derivative process $\partial_{\theta_1} X^{i,\theta}$ includes an additional functional derivative term $\partial_\mu b_{\theta_1}$, which captures the dependence of the laws μ_t^θ on the parameter θ .

The next result applies Proposition 3.1 to derive an asymptotic expansion of the log-likelihood function $z(\theta^0, \theta^+)$.

Proposition 3.2. *Assume that Assumptions A1–A5 hold. Define the quantities*

$$z_t^\theta(x) := \partial_{\theta_1} b_{\theta_1}(x, \mu_t^\theta) + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}(x, y, \mu_t^\theta) \partial_{\theta_1} \mu_t^\theta(dy) \quad (3.17)$$

and

$$\begin{aligned} \widehat{\zeta}_k^{i, \theta_1} &:= \frac{u}{\sqrt{N}} \int_0^1 z_{t_k}^{\theta_1(l), \theta_2^+}(X_{t_k}^i)^\top \left[a_{\theta_2^+}^{-2}(X_{t_k}^i) \left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_1(l), \theta_2^+}(X_{t_k}^i) \right) \right] dl, \\ \widehat{\zeta}_k^{i, \theta_2} &:= \frac{v}{\sqrt{N\Delta_n}} \int_0^1 \text{tr} \left[\partial_{\theta_2} a_{\theta_2(l)}(X_{t_k}^i) a_{\theta_2(l)}^{-1}(X_{t_k}^i) \right. \\ &\quad \times \left. \left(\left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) \right)^{\otimes 2} - V_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) \right) a_{\theta_2(l)}^{-2}(X_{t_k}^i) \right] dl. \end{aligned}$$

Then, as $N, n \rightarrow \infty$, we have for $j = 1, 2$:

$$\sum_{k=1}^n \sum_{i=1}^N \left(\zeta_k^{i, \theta_j} - \widehat{\zeta}_k^{i, \theta_j} \right) \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Remark 3.3 (Interacting particle systems). A model closely related to (1.1) is the *interacting particle system* described by

$$\begin{cases} dX_t^{i, N, \theta} = b_{\theta_1}(X_t^{i, N, \theta}, \mu_t^N) dt + a_{\theta_2}(X_t^{i, N, \theta}) dW_t^i, & i = 1, \dots, N, \quad t \in [0, T], \\ \text{Law}(X_0^{1, N, \theta}, \dots, X_0^{N, N, \theta}) := \mu_0^{\otimes N}, \end{cases} \quad (3.18)$$

where $\mu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i, N, \theta}}$ denotes the *empirical measure* of the particle system $(X_t^{1, N, \theta}, \dots, X_t^{N, N, \theta})$. Parametric estimation techniques for such interacting particle systems have been studied, for instance, in [1, 13, 31]. In many cases, statistical inference methods developed for the particle system also extend to its mean-field limit (1.1).

However, establishing the LAN property for the interacting particle system under *discrete-time observations* proves to be significantly more challenging, and no general methods are currently available in the literature. A central difficulty lies in the derivation of lower and upper bounds for the transition densities, which are instrumental in proving Proposition 3.2—a key step in the LAN analysis.

In the case of model (1.1), where the particles are i.i.d., it suffices to obtain bounds for the d -dimensional transition density $p^\theta(t, s, x, y)$ (cf. Proposition 4.2). In contrast, for interacting particle systems, one cannot reduce the analysis to marginal densities, and must instead handle the full dN -dimensional joint transition density $\mathbf{p}^\theta(t, s, x, y)$. Unfortunately, such bounds in high dimensions are not yet available in the literature, posing a major obstacle in the LAN analysis for the particle system.

We finally note that these mathematical challenges are absent in the setting of *continuous observations* $(X_t^{i, N, \theta})_{t \in [0, T]}$, where the LAN property for the drift component has been successfully established in [13].

Using the expansion in Proposition 3.2, we now derive the LAN property for the statistical model (1.1).

Theorem 3.4. *Assume that Assumptions A1–A5 hold. Then, as $N, n \rightarrow \infty$,*

$$z(\theta^0, \theta^+) \xrightarrow{\mathbb{P}^{\theta^0} - \text{law}} \begin{pmatrix} u \\ v \end{pmatrix}^\top \mathcal{N}^{\theta^0} - \frac{1}{2} \begin{pmatrix} u \\ v \end{pmatrix}^\top \Sigma^{\theta^0} \begin{pmatrix} u \\ v \end{pmatrix}, \quad (3.19)$$

where the matrix $\Sigma^{\theta^0} \in \mathbb{R}^{2 \times 2}$ is given by

$$\Sigma^{\theta^0} = \begin{pmatrix} \Sigma_b^{\theta^0} & 0 \\ 0 & \Sigma_a^{\theta^0} \end{pmatrix},$$

with

$$\begin{aligned} \Sigma_b^{\theta^0} &= \int_0^T \int_{\mathbb{R}^d} z_s^{\theta^0}(x)^\top a_{\theta_2^0}^{-2}(x) z_s^{\theta^0}(x) \mu_s^{\theta^0}(dx) ds, \\ \Sigma_a^{\theta^0} &= 2 \int_0^T \int_{\mathbb{R}^d} \text{tr} \left(\partial_{\theta_2} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \partial_{\theta_2} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \right) \mu_s^{\theta^0}(dx) ds. \end{aligned}$$

Let us now comment on the implications of Theorem 3.4. The article [1] investigates parametric estimation of the drift and diffusion coefficients in one-dimensional particle systems of the form (3.18), observed at discrete time points. The proposed estimator, based on minimising a contrast function, is shown to be asymptotically normal. Under the additional condition $N\Delta_n \rightarrow 0$, the asymptotic covariance matrix takes the same form as Σ^{θ^0} .

However, a key distinction arises when comparing the particle system model (3.18) to the i.i.d. observation scheme of the McKean–Vlasov system (1.1). In the particle system setting, the linear functional derivative $\partial_\mu b_{\theta_1}$ does not appear in the asymptotic covariance. Consequently, the quantity $z_t^\theta(x)$ is replaced by the simpler expression $\partial_{\theta_1} b_{\theta_1}(x, \mu_t^\theta)$. This structural difference has already been highlighted in [13] in the context of continuously observed particle systems.

For ease of exposition, we have considered the case of univariate parameters. Nevertheless, as shown in [20] for ergodic SDEs, the LAN property extends naturally to the multi-parameter setting. While the notation becomes much more cumbersome, the mathematical complexity does not increase significantly. Assume, for instance, that $\theta_1 \in \Theta_1 \subset \mathbb{R}^p$ and $\theta_2 \in \Theta_2 \subset \mathbb{R}^q$. Then the LAN property holds with asymptotic covariance matrix

$$\Sigma^{\theta^0} = \begin{pmatrix} \Sigma_b^{\theta^0} & 0 \\ 0 & \Sigma_a^{\theta^0} \end{pmatrix} \in \mathbb{R}^{(p+q) \times (p+q)},$$

where $\Sigma_b^{\theta^0} \in \mathbb{R}^{p \times p}$ and $\Sigma_a^{\theta^0} \in \mathbb{R}^{q \times q}$ are defined by

$$\begin{aligned} (\Sigma_b^{\theta^0})_{kl} &= \int_0^T \int_{\mathbb{R}^d} z_{k,s}^{\theta^0}(x)^\top a_{\theta_2^0}^{-2}(x) z_{l,s}^{\theta^0}(x) \mu_s^{\theta^0}(dx) ds, \\ (\Sigma_a^{\theta^0})_{kl} &= 2 \int_0^T \int_{\mathbb{R}^d} \text{tr} \left(\partial_{\theta_{k,2}} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \partial_{\theta_{l,2}} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \right) \mu_s^{\theta^0}(dx) ds, \end{aligned}$$

and

$$z_{k,t}^\theta(x) := \partial_{\theta_{k,1}} b_{\theta_1}(x, \mu_t^\theta) + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}(x, y, \mu_t^\theta) \partial_{\theta_{k,1}} \mu_t^\theta(dy).$$

We omit the full details of the multi-parameter extension, referring the interested reader to the cited literature for exposition of technicalities.

4 Proofs

4.1 Preliminary results

In this section, we collect several auxiliary results that will be instrumental in establishing the main theorems of this paper.

To begin with, under Assumptions A3 and A4, the random variables $X_t^{i,\theta}$, $Y_t^{i,\theta}$, $(Y_t^{i,\theta})^{-1}$, $\partial_{\theta_1} X_t^{i,\theta}$, and $\partial_{\theta_2} X_t^{i,\theta}$ belong to the Malliavin space $\mathbb{D}^{1,p}$ for any $t \in [0, T]$ and $p \geq 1$ (see [30, Section 2.2]). Moreover, the following uniform estimates hold:

$$\mathbb{E}^\theta \left[\sup_{0 \leq t \leq T} \|Z_t\|^p \right] + \sup_{r \in [0, T]} \mathbb{E}^\theta \left[\sup_{r \leq t \leq T} \|\mathcal{D}_r Z_t\|^p \right] \leq C, \quad (4.20)$$

for $Z_t \in \{X_t^{i,\theta}, Y_t^{i,\theta}, (Y_t^{i,\theta})^{-1}\}$ and $t \in [0, T]$.

Next, we recall a collection of moment bounds adapted from [1].

Lemma 4.1. *Assume that Assumptions A1–A2 hold. Then, for all $p \geq 1$, $0 \leq s < t \leq T$ with $t - s \leq 1$, $i \in \{1, \dots, N\}$, and $N \in \mathbb{N}$, the following estimates are satisfied:*

- (i) $\sup_{t \in [0, T]} \mathbb{E}[|X_t^{i,\theta,N}|^p] < C$, and moreover, $\sup_{t \in [0, T]} \mathbb{E}[W_p^q(\mu_t^\theta, \delta_0)] < C$ for $p \leq q$.
- (ii) $\mathbb{E}[|X_t^{i,N} - X_s^{i,N}|^p] \leq C(t - s)^{p/2}$.
- (iii) $\mathbb{E}[W_2^p(\mu_t^\theta, \mu_s^\theta)] \leq C(t - s)^{p/2}$.

In the next step, we present a technique which reduces the McKean-Vlasov system introduced in (1.1) to standard SDEs with time-varying coefficients. Consider the SDE

$$d\tilde{X}_t^\theta = b_{\theta_1}(\tilde{X}_t^\theta, \mu_t^\theta)dt + a_{\theta_2}(\tilde{X}_t^\theta)dW_t, \quad \tilde{X}_0 \sim \mu_0,$$

obtained by freezing the laws $(\mu_t)_{t \in [0, T]}$ in the original model (1.1). Under Assumption A2, the stochastic process $\tilde{X}^\theta$ follows an SDE with bounded coefficients. We also note that the processes $\tilde{X}^\theta$ and $X^{\theta,i}$ have the same law due to uniqueness of the solution of (1.1). Consequently, we can transfer certain results from standard SDE setting to McKean-Vlasov diffusion. In particular, we obtain the following upper and lower bounds of Aronson type for the transition density $p^\theta(t_k, t_{k+1}, x, y)$ (see [20, Proposition 1.2]).

Proposition 4.2. *Assume that Assumptions A2–A5 are satisfied. Then there exist constants $c > 1$ and $L > 1$ such that for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ and $t_k, t_{k+1} \in [0, T]$, the transition density $p^\theta(t_k, t_{k+1}, x, y)$ satisfies the following Gaussian-type bounds:*

$$p^\theta(t_k, t_{k+1}, x, y) \leq \frac{L}{\Delta_n^{d/2}} \exp\left(-\frac{\|x - y\|^2}{c\Delta_n}\right) \exp(c\Delta_n\|x\|^2), \quad (4.21)$$

$$p^\theta(t_k, t_{k+1}, x, y) \geq \frac{1}{L\Delta_n^{d/2}} \exp\left(-c\frac{\|x - y\|^2}{\Delta_n}\right) \exp(-c\Delta_n\|x\|^2). \quad (4.22)$$

Moreover, for any $m > 1$, there exist constants $c > 1$, $L > 1$, and $q > 0$ such that

$$\mathbb{E}_{t_k, x}^{\bar{\theta}} \left[\left| \frac{\partial_{\theta_1} p^\theta}{p^\theta}(t_k, t_{k+1}, x, X_{t_{k+1}}^i) \right|^m \right] \leq L\Delta_n^{m/2} \exp(c\Delta_n\|x\|^2)(1 + \|x\|)^q, \quad (4.23)$$

$$\mathbb{E}_{t_k, x}^{\bar{\theta}} \left[\left| \frac{\partial_{\theta_2} p^\theta}{p^\theta}(t_k, t_{k+1}, x, X_{t_{k+1}}^i) \right|^m \right] \leq L \exp(c\Delta_n\|x\|^2)(1 + \|x\|)^q. \quad (4.24)$$

As a consequence of Proposition 4.2 and Assumption A1, we conclude that all probability measures $(\mu_t)_{t \in [0, T]}$ are sub-Gaussian (cf. [5, Lemma 3.1]) and we conclude that

$$\mathbb{E}^\theta \left[\exp(r_0\|X_t^{\theta,i}\|^2) \right] \leq C \quad \text{for all } t \in [0, T], \quad (4.25)$$

for some $r_0 > 0$.

The following lemma will also be used throughout this section. Its proof is a direct consequence of the weak law of large numbers and integrability properties, and is therefore omitted.

Lemma 4.3. Assume that Assumptions A1–A2 are satisfied. Let $f : \mathbb{R}^d \times \mathcal{P}_l \rightarrow \mathbb{R}^p$ be a function such that for some $C > 0$, $k, l \in \mathbb{N}_0$, and all $(x, \mu), (y, \lambda) \in \mathbb{R}^d \times \mathcal{P}_l$, the following inequality holds:

$$\|f(x, \mu)\| \leq C \left(1 + \|x\|^k + W_l^l(\mu, \delta_0)\right). \quad (4.26)$$

Assume further that the map $(x, t) \mapsto f(x, \mu_t)$ is integrable with respect to $\mu_t(dx) dt$ on $\mathbb{R}^d \times [0, T]$. Then, as $n, N \rightarrow \infty$,

$$\frac{\Delta_n}{N} \sum_{i=1}^N \sum_{k=1}^n f(X_{t_k}^{i, \theta}, \mu_{t_k}^\theta) \xrightarrow{\mathbb{P}^\theta} \int_0^T \int_{\mathbb{R}^d} f(x, \mu_t^\theta) \mu_t^\theta(dx) dt.$$

The following proposition will be frequently used to establish the negligibility of certain stochastic terms in various settings.

Proposition 4.4. Suppose that Assumptions A1–A5 hold. Let $H_{t_{k+1}}^i$, $i = 1, \dots, N$, be a sequence of independent $\mathcal{F}_{t_{k+1}}$ -measurable random variables. Recall the definition (2.4).

(i) Assume that for any $\tau > 1$:

$$\mathbb{E}_{t_k, x}^{\theta_1(l), \theta_2^+} [H_{t_{k+1}}^i] = 0 \quad \text{and} \quad \left(\mathbb{E}_{t_k, x}^{\theta_1(l), \theta_2^+} |H_{t_{k+1}}^i|^\tau \right)^{1/\tau} = R_{t_k}^i(\Delta_n^2).$$

Then, as $N, n \rightarrow \infty$, we have

$$\frac{1}{\Delta_n \sqrt{N}} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^{\theta_1(l), \theta_2^+} \left[H^i |X_{t_{k+1}}^{i, \theta_1(l), \theta_2^+} = X_{t_{k+1}}^i \right] dl \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

(ii) Assume that for any $\tau > 1$:

$$\mathbb{E}_{t_k, x}^{\theta_1^0, \theta_2(l)} [H_{t_{k+1}}^i] = 0 \quad \text{and} \quad \left(\mathbb{E}_{t_k, x}^{\theta_1^0, \theta_2(l)} |H_{t_{k+1}}^i|^\tau \right)^{1/\tau} = R_{t_k}^i(\Delta_n^{3/2}).$$

Then, as $N, n \rightarrow \infty$, we have

$$\frac{1}{\sqrt{N} \Delta_n} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^{\theta_1^0, \theta_2(l)} \left[H^i |X_{t_{k+1}}^{i, \theta_1^0, \theta_2(l)} = X_{t_{k+1}}^i \right] dl \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

We note that the main challenge in Proposition 4.4 arises from the need to establish convergence under the original measure $\mathbb{P}^{\theta^0}$, which differs from the perturbed measures $\mathbb{P}^{\theta_1(l), \theta_2^+}$ and $\mathbb{P}^{\theta_1^0, \theta_2(l)}$. The proof presented below closely follows the methodology developed in [20].

Proof. We only show part (i) of Proposition 4.4 as the second part is shown similarly (cf. [20]). We define the random variable

$$\xi_k^N := \frac{1}{\Delta_n \sqrt{N}} \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^{\theta_1(l), \theta_2^+} \left[H_{t_{k+1}}^i |X_{t_{k+1}}^{i, \theta_1(l), \theta_2^+} = X_{t_{k+1}}^i \right] dl.$$

By martingale methods it suffices to prove the following convergence results:

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} [\xi_k^N] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \quad \text{and} \quad \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} [(\xi_k^N)^2] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

We start by handling the conditional expectation. We first observe the identity

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} [\xi_k^N] = \frac{1}{\Delta_n \sqrt{N}} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^{\theta_1(l), \theta_2^+} \left[H_{t_{k+1}}^i \frac{p^{\theta^0}}{p^{\theta_1(l), \theta_2^+}}(t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) \right] dl.$$

Consequently, we deduce the decomposition

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0}[\xi_k^N] =: V_{n,N}^1 + V_{n,N}^2,$$

with

$$V_{n,N}^1 := \frac{1}{\Delta_n \sqrt{N}} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}^i}^{\theta_1(l), \theta_2^+} \left[H_{t_{k+1}}^i \frac{p^{\theta_1, \theta_2^+} - p^{\theta_1(l), \theta_2^+}}{p^{\theta_1(l), \theta_2^+}}(t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) \right] dl,$$

$$V_{n,N}^2 := \frac{1}{\Delta_n \sqrt{N}} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \mathbb{E}_{t_k, X_{t_k}^i}^{\theta_1(l), \theta_2^+} \left[H_{t_{k+1}}^i \frac{p^{\theta^0} - p^{\theta_1(l), \theta_2^+}}{p^{\theta_1(l), \theta_2^+}}(t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) \right] dl.$$

For the first term $V_{n,N}^1$ we obtain the identity

$$V_{n,N}^1 = -\frac{u}{\Delta_n N} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 \int_0^1 \mathbb{E}_{t_k, X_{t_k}^i}^{\theta_1(l), \theta_2^+} \left[H_{t_{k+1}}^i \frac{\partial_{\theta_1} p^{\theta_1(r), \theta_2^+}}{p^{\theta_1(r), \theta_2^+}} \frac{p^{\theta_1(r), \theta_2^+}}{p^{\theta_1(l), \theta_2^+}}(t_k, t_{k+1}, X_{t_k}^i, X_{t_{k+1}}^i) \right] dr dl.$$

Now, we apply Hölder inequality with conjugates $q_1, q_2, q_3 > 1$, conditions of Proposition 4.4(i) and the inequalities of Proposition 4.2, which imply that

$$|V_{n,N}^1| \leq \frac{C}{N \sqrt{\Delta_n}} \sum_{k=1}^n \sum_{i=1}^N R_{t_k}^i(\Delta_n^2) \exp(c \Delta_n \|X_{t_k}^i\|^2 / q_2)$$

$$\times \left(\Delta_n^{-d/2} \int_{\mathbb{R}^d} \exp\left(-\frac{q_3 \|X_{t_k} - y\|^2}{c \Delta_n}\right) \exp\left(\frac{c(q_3 - 1) \|X_{t_k} - y\|^2}{\Delta_n}\right) \exp(c(2q_3 - 1) \Delta_n \|X_{t_k}^i\|^2) dy \right)^{1/q_3}.$$

We note that the above integral is finite when $(q_3 - 1)c - q_3/c < 0$, which can be achieved by choosing q_3 close enough to 1. We also note that, for any given constant $C > 0$, $\mathbb{E}[\exp(C \Delta_n \|X_{t_k}^i\|^2)]$ is uniformly bounded due to (4.25) if Δ_n is small enough. Hence, we conclude that

$$\mathbb{E}[|V_{n,N}^1|] \leq C \Delta_n^{1/2}.$$

The term $V_{n,N}^2$ is handled in exactly the same fashion and we deduce

$$\mathbb{E}[|V_{n,N}^2|] \leq C \Delta_n^{1/2}.$$

Consequently, we have proved that

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0}[\xi_k^N] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Concerning the conditional second moment we readily obtain the formula

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0}[(\xi_k^N)^2] = \frac{1}{\Delta_n^2 N} \sum_{k=1}^n \sum_{i_1, i_2=1}^N \int_0^1 \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^{\theta_1(l_1), \theta_2^+} \left[H_{t_{k+1}}^{i_1} |X_{t_{k+1}}^{i_1, \theta_1(l_1), \theta_2^+} = X_{t_{k+1}}^{i_1} \right]$$

$$\times \mathbb{E}_{t_k, X_{t_k}}^{\theta_1(l_2), \theta_2^+} \left[H_{t_{k+1}}^{i_2} |X_{t_{k+1}}^{i_2, \theta_1(l_2), \theta_2^+} = X_{t_{k+1}}^{i_2} \right] dl_1 dl_2$$

Applying Jensen's inequality and proceeding exactly as above, we conclude that

$$\sum_{k=1}^n \mathbb{E}^{\theta^0}[(\xi_k^N)^2 | \mathcal{F}_{t_k}] = O_{\mathbb{P}}(\Delta_n^2).$$

This completes the proof of Proposition 4.4. $\square$

4.2 Proof of Proposition 3.2

4.2.1 The drift term

We start by considering the difference $\zeta_k^{i,\theta_1} - \widehat{\zeta}_k^{i,\theta_1}$. Throughout this subsection we set for simplicity

$$\theta = (\theta_1, \theta_2) := (\theta_1(l), \theta_2^+).$$

According to Proposition 3.1 we have the identity

$$\zeta_k^{i,\theta_1} = \frac{u}{\Delta_n \sqrt{N}} \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^\theta \left[\sum_{r=1}^d \delta \left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} U_r^i \right) \middle| X_{t_{k+1}}^{i,\theta} = X_{t_{k+1}}^i \right] dl.$$

In the following discussion we will study the decomposition

$$\begin{aligned} \delta \left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} U_r^i \right) &= \Delta_n z_{r,\theta_1}^\theta \left(X_{t_k}^{i,\theta} \right) \\ &\times \left[a_{\theta_2}^{-2} \left(X_{t_k}^{i,\theta} \right) \left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \right]_r + H_{t_{k+1}}^i. \end{aligned} \quad (4.27)$$

Here the random variable $H_{t_{k+1}}^i$ is a reminder term, which necessarily satisfies $\mathbb{E}_{t_k, x}^{\theta_1(l), \theta_2^+} [H_{t_{k+1}}^i] = 0$ since the other two terms in (4.27) obviously satisfy this identity. Hence, if we prove that

$$\left(\mathbb{E}_{t_k, x}^\theta |H_{t_{k+1}}^i|^\tau \right)^{1/\tau} = R_{t_k}^i(\Delta_n^2) \quad \text{for any } \tau > 1,$$

the proof of Proposition 3.2 is completed for ζ_k^{i,θ_1} due to Proposition 4.4(i).

We apply the product formula (2.7) and conclude the identity

$$\delta \left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} U_r^i \right) = \partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} \delta \left(U_r^i \right) - \int_0^{\Delta_n} (\mathcal{D}_t \partial_{\theta_1} X_{r,\Delta_n}^{i,\theta})^\top U_{r,t}^i dt.$$

In the next step, we recall the definition $U_s^i = a_{\theta_2}^{-1}(X_{t_k+s}^{i,\theta}) Y_s^{i,\theta} (Y_{\Delta_n}^{i,\theta})^{-1}$, $s \in [0, \Delta_n]$, and define its approximation via

$$\widehat{U}_s^i := a_{\theta_2}^{-1}(X_{t_k+s}^{i,\theta}).$$

With this notation at hand, we finally obtain the decomposition

$$\delta \left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} U_r^i \right) = M_n^i + H_n^{i,1} + H_n^{i,2} + H_n^{i,3}$$

with

$$\begin{aligned} M_n^i &:= \Delta_n z_{r,\theta_1}^\theta \left(X_{t_k}^{i,\theta} \right) \delta \left(\widehat{U}_r^i \right) \\ H_n^{i,1} &:= - \int_0^{\Delta_n} (\mathcal{D}_t \partial_{\theta_1} X_{r,\Delta_n}^{i,\theta})^\top U_{r,t}^i dt \\ H_n^{i,2} &:= \left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} - \Delta_n z_{r,\theta_1}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \delta \left(\widehat{U}_r^i \right) \\ H_n^{i,3} &:= \partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} \delta \left(U_r^i - \widehat{U}_r^i \right) \end{aligned}$$

The term M_n^i represents the main contribution while $H_n^{i,j}$, $j = 1, 2, 3$, turn out to be negligible.

We start with the term $H_n^{i,1}$. Due to the formula (3.13), we readily obtain the statement

$$\sup_{t \in [0, \Delta_n]} \left| \partial_{\theta_1} X_{r,t}^{i,\theta} \right| = R_{t_{k+1}}^i(\Delta_n). \quad (4.28)$$

This implies the same statement for the Malliavin derivative:

$$\left\| \mathcal{D}_t \partial_{\theta_1} X_{r, \Delta_n}^{i, \theta} \right\| = R_{t_{k+1}}^i(\Delta_n).$$

As a consequence, we deduce that

$$H_n^{i,1} = R_{t_{k+1}}^i(\Delta_n^2).$$

Now, we handle the term $H_n^{i,2}$. We note that

$$\delta \left(\widehat{U}_r^i \right) = R_{t_{k+1}}^i(\Delta_n^{1/2}).$$

On the other hand, we deduce that

$$\partial_{\theta_1} X_{r, \Delta_n}^{i, \theta} - \Delta_n z_{r, \theta_1}^\theta \left(X_{t_k}^{i, \theta} \right) = R_{t_{k+1}}^i(\Delta_n^{3/2}),$$

which readily implies that $H_n^{i,2} = R_{t_{k+1}}^i(\Delta_n^2)$.

Next, we treat the term $H_n^{i,3}$. Due to (2.6), we conclude that

$$\delta \left(U_r^i - \widehat{U}_r^i \right) = R_{t_{k+1}}^i(\Delta_n).$$

Hence, the application of (4.28), implies that $H_n^{i,3} = R_{t_{k+1}}^i(\Delta_n^2)$.

Finally, we treat the main term M_n^i . First note that $\delta(\widehat{U}_r^i)$ is a regular Itô integral with respect to B^i . For the Brownian motion B^i we obtain the following representation:

$$\begin{aligned} dB_t^i &= a_{\theta_2}^{-1}(X_{t_k+t}^{i, \theta}) dX_{t_k+t}^{i, \theta} - a_{\theta_2}^{-1}(X_{t_k+t}^{i, \theta}) b_{\theta_1}(X_{t_k+t}^{i, \theta}, \mu_{t_k+t}^\theta) dt \\ &= a_{\theta_2}^{-1}(X_{t_k}^{i, \theta}) dX_{t_k+t}^{i, \theta} + \left(\mathbb{I}_d - a_{\theta_2}^{-1}(X_{t_k}^{i, \theta}) a_{\theta_2}(X_{t_k+t}^{i, \theta}) \right) dB_t^i \\ &\quad + a_{\theta_2}^{-1}(X_{t_k}^{i, \theta}) b_{\theta_1}(X_{t_k+t}^{i, \theta}, \mu_{t_k+t}^\theta) dt. \end{aligned}$$

Using the above representation we conclude that

$$\begin{aligned} \delta \left(\widehat{U}_r^i \right) &= \sum_{m=1}^d \int_0^{\Delta_n} \left(a_{\theta_2}^{-1} \right)_{rm} \left(X_{t_k+t}^{i, \theta} \right) dB_t^{m,i} \\ &= \sum_{m=1}^d \int_0^{\Delta_n} \left(a_{\theta_2}^{-1} \right)_{rm} \left(X_{t_k}^{i, \theta} \right) dB_t^{m,i} + R_{t_{k+1}}^i(\Delta_n) \\ &= \sum_{m=1}^d \left(a_{\theta_2}^{-2} \right)_{rm} \left(X_{t_k}^{i, \theta} \right) \int_0^{\Delta_n} dX_{t_k+t}^{m,i, \theta} + R_{t_{k+1}}^i(\Delta_n) \\ &= \left[a_{\theta_2}^{-2}(X_{t_k}^{i, \theta}) \left(X_{t_{k+1}}^{i, \theta} - m_{t_k, t_{k+1}}^\theta(X_{t_k}^{i, \theta}) \right) \right]_r + R_{t_{k+1}}^i(\Delta_n). \end{aligned}$$

This completes the proof of Proposition 3.2 for the drift component.

4.2.2 The diffusion term

We proceed by considering the difference $\zeta_k^{i, \theta_2} - \widehat{\zeta}_k^{i, \theta_2}$. In this subsection we set for simplicity of notation

$$\theta = (\theta_1, \theta_2) := (\theta_1^0, \theta_2(l)).$$

As in Section 4.2.1, according to Proposition 3.1 we have the identity

$$\zeta_k^{i,\theta_2} = \frac{v}{\sqrt{\Delta_n N}} \int_0^1 \mathbb{E}_{t_k, X_{t_k}}^\theta \left[\sum_{r=1}^d \delta \left(\partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} U_r^i \right) \middle| X_{t_{k+1}}^{i,\theta} = X_{t_{k+1}}^i \right] dl.$$

We consider the following decomposition:

$$\begin{aligned} \delta \left(\partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} U_r^i \right) &= \left[\partial_{\theta_2} a_{\theta_2}(X_{t_k}^{i,\theta}) a_{\theta_2}^{-1}(X_{t_k}^{i,\theta}) \left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \right]_r \\ &\times \left[a_{\theta_2}^{-2}(X_{t_k}^{i,\theta}) \left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \right]_r \\ &- \left[\partial_{\theta_2} a_{\theta_2}(X_{t_k}^{i,\theta}) a_{\theta_2}^{-1}(X_{t_k}^{i,\theta}) V_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) a_{\theta_2}^{-2}(X_{t_k}^{i,\theta}) \right]_{r,r} + H_{t_{k+1}}^i. \end{aligned} \quad (4.29)$$

As in the previous section, we only need to show that

$$\left(\mathbb{E}_{t_k, x}^\theta |H_{t_{k+1}}^i|^\tau \right)^{1/\tau} = R_{t_k}^i(\Delta_n^{3/2}) \quad \text{for any } \tau > 1,$$

which completes the proof of Proposition 3.2 is completed for ζ_k^{i,θ_2} due to Proposition 4.4(ii).

We apply the product formula (2.7) and conclude the identity

$$\delta \left(\partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} U_r^i \right) = \partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} \delta(U_r^i) - \int_0^{\Delta_n} (\mathcal{D}_t \partial_{\theta_2} X_{r, \Delta_n}^{i,\theta})^\top U_{r,t}^i dt.$$

Substituting U_r^i by $\widehat{U}_r^i$ as in the previous proof, we readily deduce the approximation

$$\partial_{\theta_2} X_{r, \Delta_n}^{i,\theta} = \left[\partial_{\theta_2} a_{\theta_2}(X_{t_k}^i) a_{\theta_2}^{-1}(X_{t_k}^{i,\theta}) \left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \right]_r + R_{t_{k+1}}^i(\Delta_n)$$

and

$$\delta(U_r^i) = \left[a_{\theta_2}^{-2}(X_{t_k}^{i,\theta}) \left(X_{t_{k+1}}^{i,\theta} - m_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) \right) \right]_r + R_{t_{k+1}}^i(\Delta_n).$$

Furthermore, since $V_{t_k, t_{k+1}}^\theta(X_{t_k}^{i,\theta}) = \Delta_n a_{\theta_2}^2(X_{t_k}^{i,\theta}) + R_{t_k}^i(\Delta_n^{3/2})$, we conclude that

$$\int_0^{\Delta_n} (\mathcal{D}_t \partial_{\theta_2} X_{r, \Delta_n}^{i,\theta})^\top U_{r,t}^i dt = \left[\partial_{\theta_2} a_{\theta_2}(X_{t_k}^{i,\theta}) a_{\theta_2}^{-1}(X_{t_k}^{i,\theta}) V_{t_k, t_{k+1}}^\theta \left(X_{t_k}^{i,\theta} \right) a_{\theta_2}^{-2}(X_{t_k}^{i,\theta}) \right]_{r,r} + R_{t_{k+1}}^i(\Delta_n^{3/2}).$$

This concludes the proof of Proposition 3.2.

4.3 Proof of Theorem 3.4

In view of Proposition 3.2 and [23, Theorem VII-5-2], it suffices to check the following conditions:

$$\sum_{k=1}^n \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}] \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2} u^2 \Sigma_b^{\theta^0} \quad (4.30)$$

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left(\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right)^2 \right] - \left(\mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right] \right)^2 \xrightarrow{\mathbb{P}^{\theta^0}} u^2 \Sigma_b^{\theta^0} \quad (4.31)$$

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left| \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right|^4 \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \quad (4.32)$$

$$\sum_{k=1}^n \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_2}] \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2} v^2 \Sigma_a^{\theta^0} \quad (4.33)$$

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left(\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right)^2 \right] - \left(\mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] \right)^2 \xrightarrow{\mathbb{P}^{\theta^0}} v^2 \Sigma_a^{\theta^0} \quad (4.34)$$

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left| \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right|^4 \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \quad (4.35)$$

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] - \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right] \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \quad (4.36)$$

Proof of (4.30): We first observe that $m_{t_k, t_{k+1}}^{\theta}(x) = x + \int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k, x}^{\theta} [b_{\theta_1}(X_s^{i, \theta}, \mu_s^{\theta})] ds$. Applying the Taylor expansion and recalling (3.17), we deduce that

$$m_{t_k, t_{k+1}}^{\theta^0}(X_{t_k}^i) - m_{t_k, t_{k+1}}^{\theta_1(l), \theta_2^+}(X_{t_k}^i) = -\frac{lu\Delta_n}{\sqrt{N}} z_{t_k}^{\theta^0}(X_{t_k}^i) + R_{t_k}^i(\varepsilon_{n, N} \Delta_n / \sqrt{N}) \quad (4.37)$$

with $\varepsilon_{n, N} \rightarrow 0$. We introduce the notation

$$\bar{\zeta}_k^{i, \theta_1} := \frac{u}{\sqrt{N}} \int_0^1 z_{t_k}^{\theta_1(l), \theta_2^+}(X_{t_k}^i)^{\top} \left[a_{\theta_2^+}^{-2}(X_{t_k}^i) \left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_0}(X_{t_k}^i) \right) \right] dl,$$

which is a direct analogue of $\widehat{\zeta}_k^{i, \theta_1}$ with $m^{\theta_1(l), \theta_2^+}$ replaced by m^{θ_0} . By definition it holds that $\mathbb{E}_{t_k}^{\theta^0}[\bar{\zeta}_k^{i, \theta_1}] = 0$. Hence, we conclude the statement

$$\begin{aligned} \sum_{k=1}^n \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}] &= \frac{u}{\sqrt{N}} \sum_{k=1}^n \sum_{i=1}^N \int_0^1 z_{t_k}^{\theta_1(l), \theta_2^+}(X_{t_k}^i)^{\top} \\ &\quad \times \left[a_{\theta_2^+}^{-2}(X_{t_k}^i) \left(m_{t_k, t_{k+1}}^{\theta^0}(X_{t_k}^i) - m_{t_k, t_{k+1}}^{\theta_1(l), \theta_2^+}(X_{t_k}^i) \right) \right] dl \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2} u^2 \Sigma_b^{\theta^0}, \end{aligned}$$

which follows directly from (4.37) and Lemma 4.3.

Proof of (4.31): We observe that

$$\mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right] = \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N (\widehat{\zeta}_k^{i, \theta_1} - \bar{\zeta}_k^{i, \theta_1}) \right] = R_{t_k}^i(\Delta_n)$$

due to (4.37). Hence, the second term in (4.31) is asymptotically negligible. As for the first term, we obtain the decomposition

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left(\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right)^2 \right] = \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i_1 \neq i_2} \widehat{\zeta}_k^{i_1, \theta_1} \widehat{\zeta}_k^{i_2, \theta_1} + \sum_{i=1}^N (\widehat{\zeta}_k^{i, \theta_1})^2 \right].$$

Again due to (4.37), we deduce the estimate

$$\mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i_1 \neq i_2} \widehat{\zeta}_k^{i_1, \theta_1} \widehat{\zeta}_k^{i_2, \theta_1} \right] = \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i_1 \neq i_2} (\widehat{\zeta}_k^{i_1, \theta_1} - \bar{\zeta}_k^{i_1, \theta_1})(\widehat{\zeta}_k^{i_2, \theta_1} - \bar{\zeta}_k^{i_2, \theta_1}) \right] = R_{t_k}^i(\Delta_n^2).$$

In other words,

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i_1 \neq i_2} \widehat{\zeta}_k^{i_1, \theta_1} \widehat{\zeta}_k^{i_2, \theta_1} \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Finally, we have that

$$(\widehat{\zeta}_k^{i, \theta_1})^2 = \frac{u^2}{N} \left[z_{t_k}^{\theta^0} (X_{t_k}^i)^\top a_{\theta_2^0}^{-2} (X_{t_k}^i) \left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta^0} (X_{t_k}^i) \right) \right]^2 + R_{t_{k+1}}^i (\varepsilon_{n, N} \Delta_n / N)$$

with $\varepsilon_{n, N} \rightarrow 0$. Since $V_{t_k, t_{k+1}}^\theta (X_{t_k}^{i, \theta}) = \Delta_n a_{\theta_2}^2 (X_{t_k}^{i, \theta}) + R_{t_k}^i(\Delta_n^{3/2})$, we conclude that

$$\sum_{k=1}^n \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} \left[(\widehat{\zeta}_k^{i, \theta_1})^2 \right] \xrightarrow{\mathbb{P}^{\theta^0}} u^2 \Sigma_b^{\theta^0}.$$

Proof of (4.32): Observe that

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left| \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right|^4 \right] &\leq C \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left| \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} - \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}] \right|^4 \right] \\ &\quad + C \sum_{k=1}^n \left| \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}] \right|^4. \end{aligned}$$

As in the proof of (4.31), we immediately deduce the convergence

$$\sum_{k=1}^n \left| \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}] \right|^4 \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Setting $y_k^i := \widehat{\zeta}_k^{i, \theta_1} - \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_1}]$, we obtain the decomposition

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left| \sum_{i=1}^N y_k^i \right|^4 \right] = \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{(i_1, i_2, i_3, i_4) \in A_N \cup B_N \cup C_N} y_k^{i_1} \cdots y_k^{i_4} \right],$$

where

$$A_N := \{\text{all indices are distinct}\},$$

$$B_N := \{\text{the indices contain two pairs } (m_1, m_1) \text{ and } (m_2, m_2) \text{ with } m_1 \neq m_2\},$$

$$C_N := \{i_1 = i_2 = i_3 = i_4\}.$$

Due to independence and (4.37) we have

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{(i_1, i_2, i_3, i_4) \in A_N} y_k^{i_1} \cdots y_k^{i_4} \right] = O_{\mathbb{P}^{\theta^0}}(\Delta_n^3).$$

We also obtain

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{(i_1, i_2, i_3, i_4) \in B_N} y_k^{i_1} \cdots y_k^{i_4} \right] = O_{\mathbb{P}^{\theta^0}}(\Delta_n)$$

and

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{(i_1, i_2, i_3, i_4) \in C_N} y_k^{i_1} \cdots y_k^{i_4} \right] = O_{\mathbb{P}^{\theta^0}}(\Delta_n/N).$$

This completes the proof of (4.32).

Proof of (4.33): Similarly to the proof of (4.30), we obtain the identity:

$$\begin{aligned} \mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_2}] &= \frac{v}{\sqrt{N\Delta_n}} \int_0^1 \text{tr} \left[\partial_{\theta_2} a_{\theta_2(l)}(X_{t_k}^i) a_{\theta_2(l)}^{-1}(X_{t_k}^i) \right. \\ &\quad \times \left. \left(\left(m_{t_k, t_{k+1}}^{\theta^0}(X_{t_k}^i) - m_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) \right)^{\otimes 2} + \left(V_{t_k, t_{k+1}}^{\theta^0}(X_{t_k}^i) - V_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) \right) a_{\theta_2(l)}^{-2}(X_{t_k}^i) \right) \right] dl. \end{aligned}$$

Due to (4.37) the term containing the difference of conditional expectations is negligible, and we only need to study the difference of conditional variances. Via Itô formula we deduce the decomposition

$$\begin{aligned} \left(V_{t_k, t_{k+1}}^{\theta}(x) \right)_{r_1, r_2} &= x_{r_1} x_{r_2} + \int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k, x}^{\theta} \left[a_{\theta_2}^2(X_s^{i, \theta})_{r_1, r_2} + b_{r_1, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta}) X_{r_2, s}^{i, \theta} + b_{r_2, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta}) X_{r_1, s}^{i, \theta} \right] ds \\ &\quad - m_{t_k, t_{k+1}}^{\theta}(x)_{r_1} m_{t_k, t_{k+1}}^{\theta}(x)_{r_2} \\ &= \int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k, x}^{\theta} \left[a_{\theta_2}^2(X_s^{i, \theta})_{r_1, r_2} + b_{r_1, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta})(X_{r_2, s}^{i, \theta} - x_{r_2}) + b_{r_2, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta})(X_{r_1, s}^{i, \theta} - x_{r_1}) \right] ds \\ &\quad + \left(\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k, x}^{\theta} \left[b_{r_1, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta}) \right] ds \right) \left(\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k, x}^{\theta} \left[b_{r_2, \theta_1}(X_s^{i, \theta}, \mu_s^{\theta}) \right] ds \right). \end{aligned}$$

By mean value theorem we deduce

$$V_{t_k, t_{k+1}}^{\theta^0}(x) - V_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(x) = (\theta_2^0 - \theta_2(l)) \partial_{\theta_2} V_{t_k, t_{k+1}}^{\theta_1^0, \tilde{\theta}_2}(x)$$

for some $\tilde{\theta}_2$ satisfying $|\theta_2^0 - \tilde{\theta}_2| \leq |\theta_2^0 - \theta_2(l)|$. Consequently, we conclude that

$$V_{t_k, t_{k+1}}^{\theta^0}(X_{t_k}^i) - V_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) = -\frac{2lv\Delta_n^{3/2}}{\sqrt{N}} \partial_{\theta_2} a_{\theta_2^0}(X_{t_k}^i) a_{\theta_2^0}(X_{t_k}^i) + R_{t_k}^i \left(\varepsilon_{n, N} \Delta_n^{3/2} / \sqrt{N} \right) \quad (4.38)$$

with $\varepsilon_{n, N} \rightarrow 0$. Hence, (4.38) implies the convergence in (4.33).

Proof of (4.34): In the previous proof we have shown that

$$\mathbb{E}_{t_k}^{\theta^0} [\widehat{\zeta}_k^{i, \theta_2}] = R_{t_k}^i(\Delta_n/N).$$

Hence, we immediately conclude that

$$\sum_{k=1}^n \left(\mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] \right)^2 \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Furthermore, following the same arguments as in the proof of (4.31), we deduce the estimate

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\left(\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right)^2 \right] = \sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} \left[(\widehat{\zeta}_k^{i, \theta_2})^2 \right] + o_{\mathbb{P}^{\theta^0}}(1).$$

Last but not least, a direct computation shows that

$$\sum_{i=1}^N \mathbb{E}_{t_k}^{\theta^0} \left[(\widehat{\zeta}_k^{i, \theta_2})^2 \right] \xrightarrow{\mathbb{P}^{\theta^0}} v^2 \Sigma_a^{\theta^0}.$$

This completes the proof of (4.34).

Proof of (4.35): This statement is shown in exactly the same way as (4.32).

Proof of (4.36): Applying the estimates from the proof of (4.30) and (4.33), we immediately obtain the convergence

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \right] \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Also, using the same arguments as in the proof of (4.31), we deduce that

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_2} \right] = \sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \widehat{\zeta}_k^{i, \theta_2} \right] + o_{\mathbb{P}^{\theta^0}}(1).$$

Finally, using the representations from Proposition 3.2, we obtain for any $1 \leq r_1, r_2, r_3 \leq d$:

$$\begin{aligned} \mathbb{E}_{t_k}^{\theta^0} \left[\left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_1(l), \theta_2^+}(X_{t_k}^i) \right)_{r_1} V_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i)_{r_2, r_3} \right] &= R_{t_k}^i(\Delta_n^2 / \sqrt{N}), \\ \mathbb{E}_{t_k}^{\theta^0} \left[\left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_1(l), \theta_2^+}(X_{t_k}^i) \right)_{r_1} \left(X_{t_{k+1}}^i - m_{t_k, t_{k+1}}^{\theta_1^0, \theta_2(l)}(X_{t_k}^i) \right)_{r_2, r_3}^{\otimes 2} \right] &= R_{t_k}^i(\Delta_n^2) \end{aligned}$$

This implies the convergence

$$\sum_{k=1}^n \mathbb{E}_{t_k}^{\theta^0} \left[\sum_{i=1}^N \widehat{\zeta}_k^{i, \theta_1} \widehat{\zeta}_k^{i, \theta_2} \right] \xrightarrow{\mathbb{P}^{\theta^0}} 0,$$

and the proof of Theorem 3.4 is complete.

References

- [1] C. Amorino, A. Heidari, V. Pilipauskaitė and M. Podolskij (2023): Parameter estimation of discretely observed interacting particle systems. *Stochastic processes and their Applications*, 163, 350–386.
- [2] C. Amorino, D. Belomestny, V. Pilipauskaitė, M. Podolskij and S.-Y. Zhou (2024): Polynomial rates via deconvolution for nonparametric estimation in McKean–Vlasov SDEs. *Probability Theory and Related Fields* 193, 539–584.
- [3] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul (2012): Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *Journal of Mathematical Neuroscience*, 2(1), 1–50.
- [4] D. Belomestny, V. Pilipauskaitė and M. Podolskij (2022): Semiparametric estimation of McKean-Vlasov SDEs. *Annales de l’Institut Henri Poincaré*, 59(1), 79–96.
- [5] D. Belomestny, M. Podolskij and S.-Y. Zhou (2025): On nonparametric estimation of the interaction function in particle system models. *Working paper*, arXiv preprint arxiv: 2402.14419.
- [6] J.P.N. Bishwal (2011): Estimation in interacting diffusions: Continuous and discrete sampling. *Applied Mathematics*, 2(9), 1154–1158.
- [7] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions (2019): *The master equation and the convergence problem in mean field games*. Princeton University Press, Princeton.
- [8] B. Chazelle, Q. Jiu, Q. Li and C. Wang (2017): Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. *Journal of Differential Equations*, 263(1), 365–397.
- [9] X. Chen (2021): Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electronic Communications in Probability*, 26, 1–13.
- [10] F. Comte and V. Genon-Catalot (2024): Parametric inference for ergodic McKean-Vlasov stochastic differential equations. *Bernoulli*, 30(3), 1971–1997.
- [11] F. Comte, V. Genon-Catalot and C. Laredo (2024): Nonparametric moment method for McKean-Vlasov stochastic differential equations. *Working paper*.
- [12] L. Della Maestra and M. Hoffmann (2022): Nonparametric estimation for interacting particle systems: McKean–Vlasov models. *Probability Theory and Related Fields*, 182(1), 551–613.
- [13] L. Della Maestra and M. Hoffmann (2023): The LAN property for McKean–Vlasov models in a mean-field regime. *Stochastic Processes and their Applications*, 155, 109–146.
- [14] B. Djehiche, F. Gozzi, G. Zanco and M. Zanella (2022): Optimal portfolio choice with path dependent benchmarked labor income: a mean field model. *Stochastic Processes and their Applications*, 145, 48–85.
- [15] J.P. Fouque and L.H. Sun (2013): *Systemic risk illustrated*. Handbook on Systemic Risk, Eds J.P. Fouque and J. Langsam.
- [16] V. Genon-Catalot and C. Larédo (2021): Parametric inference for small variance and long time horizon McKean-Vlasov diffusion models. *Electronic Journal of Statistics*, 15(2), 5811–5854.

- [17] V. Genon-Catalot and C. Larédo (2021): Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. *Stochastic Processes and their Applications*, 142, 513–548.
- [18] K. Giesecke, G. Schwenkler and J. Sirignano (2019): Inference for large financial systems. *Mathematical Finance*, 30(1), 3–46.
- [19] E. Gobet (2001): Local asymptotic mixed normality property for elliptic diffusion. *Bernoulli*, 7 (6), 899–912.
- [20] E. Gobet (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 38(5), 711–737.
- [21] J. Guyon and P. Henry-Labordere (2011): The smile calibration problem solved. *Available at SSRN 1885032*.
- [22] J. Hajek (1971). Local asymptotic minimax admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles*, University of California Press, 175–194.
- [23] J. Jacod and A.N. Shiryaev (1987): *Limit theorems for stochastic processes*. Springer, Berlin.
- [24] R.A. Kasonga (1990): Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, 50(3), 865–875.
- [25] L. Le Cam (1971): Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles*, University of California Press, 245–261.
- [26] L. Le Cam and G.L. Yang (1990). *Asymptotics in statistics*. Springer.
- [27] M. Liu and H. Qiao (2022): Parameter estimation of path-dependent McKean-Vlasov stochastic differential equations. *Acta Mathematica Scientia*, 42, 876–886.
- [28] A. Mogilner and L. Edelstein-Keshet (1999): A non-local model for a swarm. *Journal of Mathematical Biology*, 38(6), 534–570.
- [29] R. Nickl, G.A. Pavliotis and K. Ray (2025): Bayesian nonparametric inference in McKean-Vlasov models. *Annals of Statistics*, 53(1), 170–193.
- [30] D. Nualart (1995): *The Malliavin calculus and related topics*. Springer-Verlag, Berlin Heidelberg, New York.
- [31] L. Sharrock, N. Kantas, P. Parpas and G.A. Pavliotis (2021): Online parameter estimation for the McKean-Vlasov stochastic differential equation. *Stochastic Processes and their Applications*, 162, 481–546.
- [32] J. Wen, X. Wang, S. Mao and X. Xiao (2016): Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application. *Applied Mathematics and Computation*, 274, 237–246.