PhD-FSTM-2025-124

The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 25 November 2025 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN MATHÉMATIQUES

by

Akram HEIDARI

PARAMETRIC ESTIMATION OF MCKEAN–VLASOV DIFFUSIONS

Dissertation defence committee

Dr Mark Podolskij, dissertation supervisor
Professor, University of Luxembourg

Dr Ivan Nourdin, chairman
Professor, University of Luxembourg

Dr Yannick Baraud
Professor, University of Luxembourg

Dr Ester Mariucci
Professor, Paris-Saclay University

Dr Almut Veraart
Professor, Imperial College London

Dedication

To my supervisor, Mark

for his kindness and his trust.

To my husband, Gholamreza

for his love and his unwavering support.

Acknowledgements

My deepest gratitude goes first and foremost to my supervisor, Mark Podolskij. I am sincerely thankful for his precise guidance, generous scientific support, patience, and invaluable assistance throughout every step of this challenging journey. His presence has been far more than an academic mentorship; it has been a source of human and friendly support that consistently brought me calmness and the motivation to move forward. Any progress I have made during these years is undoubtedly the result of his trust, guidance, and kindness.

I would also like to express my heartfelt appreciation to the members of my defence committee: Ivan Nourdin (chairman), Yannick Baraud, Christophe Ley, Ester Mariucci, and Almut Veraart. Thank you for dedicating your time to reviewing my thesis, for your valuable attention to this work, and for your participation in and contribution to my defence session.

My sincere thanks also go to all the people of DMath, especially the members of my research team — Chiara, Fran, Nicolas, Shiwi, and Vytauté — whose collaboration, scientific discussions, and friendly atmosphere have played an important role throughout these years.

I am likewise grateful to my sister, Nastaran, for her kindness and constant belief in me throughout this journey.

Finally, I wish to express my deepest gratitude to my dear husband, Gholamreza, whose love, patience, and unwavering support have made the difficulties of this path much easier to bear. Without his emotional support, deep understanding, and comforting presence, completing this journey would not have been possible. This achievement belongs as much to him as it does to me.

Contents

Chapter I

Introduction

This thesis is devoted to the statistical analysis of interacting particle systems and their mean-field limits, with a particular focus on McKean–Vlasov stochastic differential equations (SDEs), based on discrete observations over a fixed time interval. We study fundamental inference problems, including parameter estimation for drift and diffusion coefficients, local asymptotic normality of the associated likelihood functions, and goodness-of-fit testing for volatility structures. Overall, this work provides a comprehensive framework for statistical inference in interacting particle systems and nonlinear stochastic models.
In this chapter, we introduce the general framework of interacting particle systems and their mean-field counterparts. We present the class of SDEs arising in this context and discuss key mathematical results, such as the existence and uniqueness of solutions. We also describe the concept of propagation of chaos, which explains how the behaviour of individual particles becomes increasingly independent as the population size grows. Afterwards, we review recent developments in the statistical estimation of these models. The introduction concludes with an overview of the thesis structure and a summary of the main results.

I.1 Interacting Particle Systems and McKean-Vlasov Equations

Let us consider a system of N particles in $\mathbb{R}^d$, where the dynamics of each particle are described by the following SDE:

$$\begin{cases} dX_t^{i,N} = b\big(X_t^{i,N}, \mu_t^N\big)dt + a\big(X_t^{i,N}, \mu_t^N\big)dW_t^i, \qquad i = 1, ..., N, \quad t \in [0,T], \\ (X_0^i)_{1\le i\le N} \sim \mu_0^{\otimes N} \end{cases} \tag{I.1.1}$$

where $(W_t^i)_{1\le i\le N}$ denotes a family of independent d-dimensional Brownian motions, and the initial positions $(X_0^i)^{1\le i\le N}$ are independent and identically distributed (i.i.d.) random variables with common law μ_0, independent of Brownian motions. The functions b and a represent the drift and diffusion coefficients, respectively, and depend on both the current position of the particle and the empirical distribution of the system at time t, which is given by

$$\mu_t^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,N}}. \tag{I.1.2}$$

This model describes an **interacting particle system**, where each particle evolves not in isolation but under the influence of the whole system. The interaction appears through the dependence on the empirical measure. The study of such systems began with McKean's work in the context of plasma physics [76]. Since then, they have been widely explored and extended in the probabilistic literature; see for example [19, 45, 65, 78, 90].

As the number of particles N becomes large, one typically observes a limiting behaviour where the effect of any single particle becomes negligible, but the influence of the collective distribution remains. In this regime, the dynamics of the system are described by a **mean-field equation**, which is a d-dimensional SDE of the form

$$\begin{cases} dX_t = b(X_t, \mu_t)dt + a(X_t, \mu_t)dW_t, & t \in [0, T], \\ \mathcal{L}(X_t) = \mu_t \end{cases} \tag{I.1.3}$$

These equations are often referred to as **McKean–Vlasov SDEs**, distribution-dependent SDEs, or non-linear SDEs in the sense of McKean. Their distinctive feature is that the drift and diffusion coefficients depend on the law of the solution. This dependence on the evolving distribution introduces a non-linearity that sets them apart from classical SDEs and, in particular, means that the resulting processes are generally not Markovian.

To make this convergence precise, we introduce some basic assumptions on the coefficients.

IA1. *The drift $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and diffusion coefficient $a : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times d}$ satisfy:*

(i) Lipschitz continuity: there exists $C > 0$ such that for all $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$|b(x, \mu) - b(y, \nu)| + |a(x, \mu) - a(y, \nu)| \le C\big(|x - y| + W_2(\mu, \nu)\big)$$

where $W_2(\mu, \nu)$ denotes the 2-Wasserstein distance.

(ii) Linear growth: there exists $C > 0$ such that for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$|b(x, \mu)|^2 + |a(x, \mu)|^2 \le C\big(1 + |x|^2 + \int_{\mathbb{R}^d} |y|^2 \mu(dy)\big)$$

Under these conditions, both the particle system (I.1.1) and the limiting equation (I.1.3) are well-posed in the strong sense. These types of results can be found in standard references such as Sznitman's notes on propagation of chaos [90] or in more recent texts such as [79].

A central conceptual link between the N-particle system (I.1.1) and its corresponding mean-field counterpart (I.1.3) is provided by the concept of **propagation of chaos**, first introduced by [61] and later developed rigorously in the probabilistic setting in [76, 90]. This concept formalises the idea that, as $N \to \infty$, the joint behaviour of any fixed number of particles increasingly resembles that of independent copies of a single process. The common distribution of these limiting processes is precisely the law of the solution to the McKean–Vlasov equation. This idea forms the bridge between finite-particle models and their mean-field limit, and we recall its precise mathematical formulation below.

We say that the particle system exhibits pointwise propagation of chaos if the trajectory of any fixed particle converges, in the mean-square sense, to that of its mean-field counterpart uniformly on $[0,T]$. Formally, let $X_t^{i,N}$ denote the i-th component of the solution to (I.1.1), and let X_t^i solve the McKean–Vlasov SDE

$$\begin{cases} dX_t^i = b(X_t^i, \mu_t)dt + a(X_t^i, \mu_t)dW_t^i, & t \in [0,T], \\ \mathcal{L}(X_t^i) = \mu_t \end{cases}$$

driven by the same initial conditions and Brownian motions as in (I.1.1). The propagation of chaos holds if

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \mathbb{E}\left[\left|X_t^{i,N} - X_t^i\right|^2\right] = 0.$$

This property implies, in particular, the weak convergence $\mu_t^N \to \mu_t$, for each fixed $t \in [0,T]$. More precisely, the empirical measure μ_t^N converges in distribution to the deterministic law μ_t. By coupling the particle system with independent copies of the corresponding mean-field dynamics, one can establish that in the classical setting with Lipschitz coefficients, the system exhibits propagation of chaos. Furthermore, the existence and uniqueness of solutions to the mean-field equation follow from a fixed point argument (see [68], Theorem 3.3).

This not only justifies the mean-field approximation, but also underpins statistical procedures, since it allows one to replace large but finite systems with a tractable limit model. As a consequence, the particle system provides not only an approximation of the limiting dynamics, but also a practical tool for simulating or analysing such models numerically.

While propagation of chaos provides the fundamental probabilistic link between particle systems and their mean-field limits, the statistical analysis of such models relies on another cornerstone concept: the **local asymptotic normality (LAN)** property, which characterises the asymptotic behaviour of the log-likelihood ratio under local perturbations of the model parameters and underpins inference for estimators and tests.

LAN is a central notion in asymptotic statistics, originally introduced by Le Cam [70]. Intuitively, it describes the behaviour of a statistical model when the parameter of interest is perturbed in a neighbourhood of its true value while the sample size grows. More precisely, LAN provides a quadratic approximation of the log-likelihood ratio between two nearby parameter values, revealing an underlying Gaussian structure in the limit. This perspective is powerful: it characterises the asymptotic efficiency of estimators and tests, shows that the statistical experiment behaves locally like a Gaussian shift experiment, and forms the foundation for minimax theory in parametric inference [56, 71].

Formally, consider a sequence of statistical models $\{\mathbb{P}_{m,\theta} : \theta \in \Theta \in \mathbb{R}^d\}$. The LAN property at $\theta \in \Theta$ holds if, for every local perturbation of the form $\theta_m = \theta + r_m^{-1} h$ with $h \in \mathbb{R}^d$, the log-likelihood ratio admits the expansion

$$\log \frac{d\mathbb{P}_{m,\theta_m}}{d\mathbb{P}_{m,\theta}} = h^\top \mathcal{N} - \tfrac{1}{2} h^\top I_\theta h + o_{\mathbb{P}_{m,\theta}}(1), \qquad \text{as } m \to \infty,$$

where $\mathcal{N} \sim \mathcal{N}(0, I_\theta)$, and $I_\theta \in \mathbb{R}^{d\times d}$ denotes the Fisher information matrix. This expansion captures the idea that, locally, the model behaves like a Gaussian shift experiment with known information structure. The LAN property allows the application of minimax theorems to derive lower bounds on the asymptotic variance of estimators, providing a benchmark for optimal statistical procedures.

The LAN framework thus unifies the study of asymptotic efficiency and optimality, and it underlies likelihood-based estimation and testing in classical i.i.d. models, Markov processes, and diffusion processes [58, 67, 72]. Its generality makes it a natural tool for extending inference methods to more complex dependent settings. In particular, it motivates the investigation carried out in this thesis, where we aim to extend LAN-type results to McKean–Vlasov dynamics and develop statistical procedures for discretely observed particle systems.

While classical SDEs have been extensively studied in both theory and statistics, their inference techniques rely on the fact that the drift and diffusion coefficients depend only on the current state of the process. In contrast, McKean–Vlasov models incorporate the evolving law of the system into the dynamics. This fundamental difference means that many classical tools, such as likelihood expansions or quadratic variation methods, cannot be applied directly. The dependence on the unknown distribution μ_t introduces additional layers of complexity: one must account not only for the randomness of individual trajectories but also for the fluctuations of the empirical law that drives the dynamics. This dual source of randomness makes statistical inference in McKean–Vlasov models particularly delicate.

From the statistical viewpoint, interacting particle systems and their mean-field limits raise difficulties because standard approaches based on transition densities or Girsanov transformations become intractable. The empirical measure of the particle system appears as a high-dimensional, random object that must itself be estimated, further complicating likelihood-based arguments and asymptotic analysis.

For classical SDEs, a comprehensive toolkit for parametric inference has been established: likelihood-based methods, martingale estimating functions, and Girsanov-type changes of measure yield consistent and often efficient estimators under both low- and high-frequency sampling regimes (see, e.g., [58, 67, 72]). The theory of LAN provides the foundation for asymptotic efficiency and the construction of optimal statistical procedures [58, 67]. These approaches, however, rely crucially on transition densities and the Markov property—features that do not carry over to McKean–Vlasov models. Here the coefficients depend on the law of the solution, the transition structure is implicit and typically intractable, and the dynamics inherit mean-field effects that break simple Markovian arguments.

As a result, while a growing literature addresses propagation of chaos and estimation in specific mean-field models, rigorous results on joint estimation of drift and diffusion from discretely observed particle systems, on LAN in the mean-field regime, and on formal goodness-of-fit testing for volatility remain limited. This thesis develops asymptotic frameworks that address these inference challenges by establishing theoretical guarantees for estimation, extending LAN theory to mean-field settings, and proposing testing procedures for volatility structures. The approach combines probabilistic techniques for distribution-dependent dynamics with modern statistical theory, aiming to bridge the gap between mathematical modelling and practical inference in particle systems.

I.2 Literature Overview

The statistical inference of interacting particle systems and McKean–Vlasov equations has attracted increasing interest only in recent years. Originally introduced as models in plasma physics and first studied in [76], these systems have since inspired numerous contributions across a range of directions. In this section, we focus on those developments most relevant to the present thesis.

I.2.1 Parametric and Nonparametric Estimation for McKean–Vlasov Models

In this part, we review existing contributions on parameter estimation for McKean–Vlasov SDEs and closely related interacting particle systems. The focus lies primarily on parametric approaches, where the coefficients of the model are specified up to a finite-dimensional parameter, and maximum likelihood methods or related estimators are applied. We then conclude with a brief account of semi- and nonparametric results.

The statistical study of interacting diffusion processes can be traced back to the seminal contribution of Kasonga [63], who first analysed maximum likelihood estimation in systems of mean-field type. He considers a model of interacting diffusions where the drift coefficient depends linearly on an unknown parameter, and shows that, based on continuous observation of the system over a fixed interval $[0, T]$, the resulting estimator is both consistent and asymptotically normal as the number of particles $N \to \infty$.

Subsequent contributions extend this framework in several directions. Wen et al. [92] investigate maximum likelihood estimation for a class of McKean–Vlasov SDEs, they reformulate the model into a homogeneous diffusion of the form

$$dX_t = b(\theta, X_t, \mu_t)dt + dW_t,$$

where the unknown parameter θ enters through the drift coefficient, while the diffusion term is the standard Brownian motion. Based on continuous-time observations of the trajectory, they construct a likelihood function and establish estimation procedures for θ. The structure here is relatively simple: the diffusion coefficient is unit,

and the parametric dependence is confined to the drift. Later, Liu and Qiao [73] generalise this approach to path-dependent McKean–Vlasov SDEs with non-Lipschitz coefficients, constructing maximum likelihood estimators and establishing strong consistency.
Sharrock et al. [88] pursue a different approach by systematically analysing maximum likelihood estimation in both offline and online settings, providing a unified treatment of inference for distribution-dependent models. They consider a general family of McKean–Vlasov SDEs parametrised by $\theta \in \mathbb{R}^p$:

$$dX_t = b(\theta, X_t, \mu_t)\, dt + a(X_t)\, dW_t,$$

together with the corresponding particle system approximation. Their approach relies on a Girsanov-type representation of the likelihood, under the simplifying assumption that the diffusion coefficient is constant, where for convenience it is normalised to $a = 1$. They investigate two statistical scenarios: observing independent trajectories of the McKean–Vlasov SDE, and observing particles from the interacting system, and in both cases characterise the asymptotic behaviour of the maximum likelihood estimator as $t \to \infty$, as well as in the joint limit $t \to \infty$ and $N \to \infty$. In the offline setting, they establish consistency and asymptotic normality of the MLE, while in the online setting they propose a continuous-time stochastic gradient ascent algorithm that converges to the stationary points of the asymptotic log-likelihood. This dual perspective illustrates inference strategies both at the level of the mean-field limit and through finite-particle approximations.
In related directions, Bishwal [12] investigates parameter estimation in systems of interacting diffusions, when only discrete observations of the system are available and the parameter is a function of time. He studies both sieve and approximate maximum likelihood estimators of the drift parameters by analysing their asymptotic behaviour as the number of particles N increases, which contrasts with the large-time asymptotics considered in other works. While not formulated explicitly in McKean–Vlasov terms, these results provide important insight into drift estimation in interacting particle systems.
Another significant contribution is due to Chen [22], who focuses on the estimation of quadratic potential by maximum likelihood in interacting particle systems from continuous-time and single-trajectory data. The model is specified as

$$dX_t^{i,N} = \Theta\big(\bar{X}_t^N - X_t^{i,N}\big)\, dt + a\, dW_t^i, \qquad \bar{X}_t^N = \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N},$$

with an unknown positive-definite matrix parameter Θ. Remarkably, he establishes that the plain maximum likelihood estimator (without regularisation) achieves optimal convergence rates simultaneously in the mean-field limit and in long-time asymptotics, thereby overcoming the high dimensionality of the system through the symmetry of the interaction structure.
Recent advances in semiparametric and nonparametric inference have significantly enriched the range of tools available for analysing McKean–Vlasov models and interacting particle systems.

In the semiparametric realm, estimation strategies have been proposed that combine a finite-dimensional parameter of interest with an infinite-dimensional nuisance component, often arising from the interaction structure or law dependence of the system. Such frameworks allow efficient inference for the parametric part while retaining flexibility for the unspecified dynamics. Methods in this area typically build on profile likelihood, contrast minimisation, or estimating equations, and they adapt naturally to the mean-field setting where law-dependence complicates full parametric modeling. Recent advances include applications to McKean–Vlasov dynamics and related particle systems, highlighting both theoretical optimality and practical feasibility; see, for instance [10, 93].

On the nonparametric front, a variety of estimation strategies have been adapted to handle the complexity of mean-field dynamics. This includes kernel-based estimators, orthogonal projection methods, data-driven adaptive selection procedures, and deconvolution techniques. These approaches make it possible to estimate drift functions, interaction potentials, or law-dependent coefficients without imposing parametric assumptions, while still achieving optimal or near-optimal convergence rates. Noteworthy developments include [4, 29, 94].

I.2.2 Local Asymptotic Normality for Diffusions and McKean–Vlasov models

In this part, we review some of the key contributions to the study of the LAN property for diffusion models and their extensions to McKean–Vlasov processes. The literature in this direction begins with classical diffusion models and has more recently advanced to law-dependent systems, providing the methodological foundation for our work.

The study of the LAN property for discretely observed SDEs is developed extensively in the classical diffusion setting. Gobet [53] investigates the LAN property for a d-dimensional diffusion process

$$dX_t^{\alpha,\beta} = b(\alpha, X_t^{\alpha,\beta})dt + a(\beta, X_t^{\alpha,\beta})dW_t. \tag{I.2.1}$$

The process is assumed to be elliptic and ergodic under mild regularity conditions, ensuring the existence of a unique invariant distribution. The statistical experiment is based on discrete observations $(X_{k\Delta_n})_{0\leq k\leq n}$ with mesh size Δ_n. The asymptotic regime considered is $\Delta_n \to 0$ and $n\Delta_n \to \infty$, so that the horizon of observation increases to infinity while the sampling step decreases. He proves that the likelihood function associated with these discrete observations satisfies the LAN property, with different convergence rates for the drift and diffusion parameters: the rate for the drift parameter is $\sqrt{n\Delta_n}$, while for the diffusion parameter it is $\sqrt{n}$. This rate separation reflects the fact that drift parameters accumulate information only over time, whereas diffusion parameters can be estimated more efficiently from high-frequency data.

The proof relies on Malliavin calculus, which is used to transform the log-likelihood ratio into a form that admits a stochastic expansion. Since the transition density is typically not available in closed form, direct analysis of the likelihood becomes intractable. Malliavin calculus provides a way around this difficulty: by exploiting integration-by-parts formulas to express the derivatives of the log-likelihood in terms of conditional expectations of functionals of the diffusion. This representation yields Gaussian-type approximations for the score functions. The ergodicity assumption plays a crucial role in controlling long-time averages, ensuring that the information matrix stabilises as $n \to \infty$.

The result represents a foundation in the theory of LAN for discretely observed diffusions. By addressing both drift and diffusion parameters in a multidimensional and ergodic setting, and by quantifying their different asymptotic rates, it provides a methodological scheme for later studies–including extensions to non-ergodic or mean-field systems–where explicit likelihoods are unavailable and Malliavin calculus becomes the primary analytical tool.

Another cornerstone contribution is due to Della Maestra and Hoffmann [30], who extended the study of the LAN property to interacting particle systems and their mean-field limits. They consider a system of N interacting diffusions evolving according to

$$dX_t^{i,N} = b(\theta, X_t^{i,N}, \mu_t^N)dt + a(X_t^{i,N})dW_t^i. \tag{I.2.2}$$

The statistical experiment consists of observing the full trajectory of the system $(X_t^{i,N})_{t\in[0,T]}^{i=1,\dots,N}$ over a fixed time horizon $T > 0$, while letting $N \to \infty$.

A key simplification of this setting, compared to discretely observed diffusions, lies in the construction of the likelihood. Since the trajectory is observed continuously, the likelihood ratio with respect to a reference parameter can be expressed explicitly by Girsanov's theorem. This differs from the discrete-time case, where the transition densities of multidimensional diffusions are typically intractable and Malliavin calculus is required to approximate their derivatives. Thus, while the interacting structure of the model poses new challenges, the continuous-time observation framework yields a closed-form likelihood representation that facilitates the LAN analysis.

They prove that the LAN property holds for estimating a multidimensional parameter in the drift, in a mean-field regime $N \to \infty$. Their proof builds upon the classical Ibragimov and Hasminski theory of statistical experiments, adapted to the nonlinear dependencies of the McKean–Vlasov regime. A central feature of the result is that the maximum likelihood estimator enjoys not only asymptotic normality but also a sharper characterization via Hájek's convolution theorem. More precisely, they show that the MLE is asymptotically minimax optimal (up to constants), owing to strong probabilistic controls on the likelihood process. Furthermore, they derive explicit identifiability and non-degeneracy conditions for the Fisher information matrix, ensuring that the asymptotic variance is well-defined.

In addition to establishing LAN, the study provides structural insights into the associated nonlinear McKean–Vlasov model, since the Fisher information can be expressed in terms of the law of the limiting nonlinear diffusion. In this way, the work forms a bridge between the classical diffusion-based LAN results of Gobet and the modern theory of statistical inference for McKean–Vlasov dynamics.

I.2.3 Goodness-of-Fit Testing for Volatility in Diffusion Models

A central line of research on statistical testing for diffusion models has focused on the volatility function, which governs the local variability of the process. An important contribution in this direction is the paper by [35], which is motivated by the need to validate the functional form of the volatility in continuous-time financial models, where model misspecification can lead to substantial pricing and risk management errors. The authors consider a diffusion process of the form

$$dX_t = b(t, X_t)dt + a(t, X_t)dW_t, \qquad t \in [0, 1]$$

observed discretely over a fixed time horizon. The central problem is to test whether the variance function $a^2(t, x)$ belongs to a prescribed parametric family. They base their test on estimates of integrated volatility functionals $\int_0^1 a^2(t, X_t)dt$, and consider the more general case where the volatility depends on both time and the state variable X_t.
In this setting, the asymptotic behaviour of integrated volatility estimators becomes substantially more delicate. The authors prove that the estimators no longer converge to a normal distribution but instead converge stably in law to random variables with a non-standard limit distribution depending on the underlying process itself. Conditionally on the observed diffusion path, however, the limiting distribution is Gaussian. This work established a rigorous framework for linking high-frequency volatility estimation with formal model validation in diffusion processes.
Other related contributions to goodness-of-fit testing in diffusion models include the specification tests of [3, 26], and [34], as well as more recent extensions to fractional diffusions, such as [87]. Collectively, these studies provide the methodological foundation for volatility testing in diffusion models, although they remain confined to classical, non-interacting settings.

I.3 Contributions of the thesis

In this concluding section, we provide a summary of the principal findings of this thesis, as presented in Chapter II, Chapter III, and Chapter IV, and their main contributions.

I.3.1 Chapter II: Parameter estimation of discretely observed interacting particle systems

This subsection is devoted to presenting the principal result of Chapter II, derived from the paper:

- "Parameter estimation of discretely observed interacting particle systems",

  in collaboration with C. Amorino, V. Pilipauskaitė and M. Podolskij. *Stochastic Processes and Their Applications*, 163, 350–386, 2023.

We focus on the problem of joint parameter estimation in an interacting particle system of McKean–Vlasov type, given by

$$\begin{cases} dX_t^{\theta,i,N} = b\big(\theta_1, X_t^{\theta,i,N}, \mu_t^{\theta,N}\big)dt + a\big(\theta_2, X_t^{\theta,i,N}, \mu_t^{\theta,N}\big)dW_t^i, \qquad i = 1, ..., N, \quad t \in [0,T], \\ \mathcal{L}\big(X_0^{\theta,1,N}, ..., X_0^{\theta,N,N}\big) := \mu_0 \times ... \times \mu_0. \end{cases}$$

Since the transition densities of this system are not available in closed form, the standard maximum likelihood approach is infeasible. Instead, we construct an estimator by minimising a contrast function $S_n^N(\theta)$, defined in (II.2.1), which is based on an Euler-type approximation of the dynamics. The estimator $\hat{\theta}_n^N = (\theta_{n,1}^N, \theta_{n,2}^N)$ is then defined as

$$\hat{\theta}_n^N \in \arg\min_{\theta \in \Theta} S_n^N(\theta).$$

Our main results establish that $\hat{\theta}_n^N$ is consistent and asymptotically normal.

Theorem I.3.1. *(Consistency) Assume that IIA1-IIA5 hold, with only condition (I) in IIA4. Then the estimator $\hat{\theta}_n^N$ is consistent in probability:*

$$\hat{\theta}_n^N \xrightarrow{\mathbb{P}} \theta_0 \quad \textit{as } n, N \to \infty.$$

To establish the asymptotic normality of $\hat{\theta}_n^N$, we study the asymptotic behaviour of the first and second derivatives of the contrast function. After suitable normalisation, the first derivative converges in law to a Gaussian random variable, while the second derivative converges in probability to a deterministic matrix. This permits the use of a classical Taylor expansion around the true parameter to deduce the asymptotic distribution of the estimator. In this analysis, we impose an additional condition on the relative rates of N and Δ_n, requiring that $N\Delta_n \to 0$ as $N, n \to \infty$.

Theorem I.3.2. *(Asymptotic normality) Assume that IIA1-IIA7 hold. If $N\Delta_n \to 0$ then*

$$\big(\sqrt{N}(\hat{\theta}_{n,1}^N - \theta_{0,1}), \sqrt{N/\Delta_n}(\hat{\theta}_{n,2}^N - \theta_{0,2})\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, 2(\Sigma(\theta_0))^{-1}\big) \quad \textit{as } n, N \to \infty,$$

where

$$2(\Sigma(\theta_0))^{-1} \colon = 2\,\mathrm{diag}\,\big((\Sigma^{(1)}(\theta_0))^{-1}, (\Sigma^{(2)}(\theta_0))^{-1}\big)$$

with $\Sigma^{(j)}(\theta_0)$, $j = 1, 2$, being defined in IIA6.

Compared to the classical SDE setting, our analysis highlights the interplay between the number of particles and the observation frequency, and extends the contrast-based methodology developed by [40,64] and [95] to the McKean–Vlasov framework. The results show that, despite the additional complexity introduced by the interaction through the empirical measure, consistent and asymptotically normal estimation of both drift and diffusion parameters is achievable. Moreover, the analysis reveals the distinct convergence rates for drift and diffusion parameters, reflecting their different sensitivities to the discretisation step and particle number. This provides a precise characterisation of the asymptotic regime necessary for reliable parameter estimation in discretely observed interacting particle systems.

I.3.2 Chapter III: Local asymptotic normality for discretely observed McKean-Vlasov diffusions

In this subsection, we present the main result of Chapter III, which is based on the paper:

- "Local asymptotic normality for discretely observed McKean-Vlasov diffusions",

  in collaboration with M. Podolskij, 2025.

In this chapter, we establish the local asymptotic normality (LAN) property for the likelihood function arising from discretely observed d-dimensional McKean–Vlasov SDEs over a fixed time horizon. We consider an i.i.d. array of d-dimensional processes governed by

$$\begin{cases} dX_t^{i,\theta} = b_{\theta_1}\big(X_t^{i,\theta}, \mu_t^\theta\big)dt + a_{\theta_2}\big(X_t^{i,\theta}\big)dW_t^i \qquad i = 1, ..., N, \quad t \in [0, T] \\ \mathcal{L}(X_t^{i,\theta}) = \mu_t^\theta \end{cases}$$

The asymptotic regime of interest combines high-frequency sampling $\Delta_n := T/n \to 0$ with a growing number of particles $N \to \infty$, over a fixed time horizon T.
The main methodological difficulty stems from the absence of tractable transition densities, which prevents a direct likelihood expansion. To overcome this, we employ Malliavin calculus techniques, in particular an integration-by-parts representation, to derive an explicit expression for the logarithmic derivative of the transition density. This approach, inspired by earlier work of [52, 53] on classical diffusions, allows us to obtain a stochastic expansion of the log-likelihood ratio.
Under local perturbations of the form

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

we study the log-likelihood ratio between the measures $\mathbb{P}^{\theta^+}$ and $\mathbb{P}^{\theta^0}$, given by

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

where $p^\theta(s, t, x, y)$ denotes the transition density of the Euler scheme associated with parameter θ. The LAN property can now be stated as follows:

Theorem I.3.3. *Assume that Assumptions IIIA1–IIIA5 hold. Then,*

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

where $\mathcal{N}^{\theta^0}$ is a centred Gaussian vector with covariance matrix $\Sigma^{\theta^0} \in \mathbb{R}^{2\times 2}$ defined as

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

with

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

where

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

Compared with existing contributions, our analysis extends the LAN framework of [53] to the McKean–Vlasov setting with discrete-time observations, without requiring ergodicity. It complements recent contrast-based estimation methods for discrete particle systems [7], as well as the continuous-observation LAN results of [30], where likelihood representations are more explicit. More generally, our work can be seen as a mean-field extension of the classical LAN theory for discretely observed diffusion processes [40, 64, 95], addressing the additional challenges introduced by distributional dependence and the mean-field regime.

I.3.3 Chapter IV: On goodness-of-fit testing for volatility in McKean–Vlasov models

In this subsection, we state the principal result of Chapter IV, which is derived from the paper:

- "On goodness-of-fit testing for volatility in McKean–Vlasov models",
  in collaboration with M. Podolskij, 2025.

The central problem addressed in this chapter is the development of statistical goodness-of-fit tests for volatility in McKean–Vlasov particle systems based on empirical data. We consider a system of N independent particles $(X_t^i)^{i=1,\dots,N}$ evolving over a fixed time interval $[0, T]$. Each particle follows nonlinear dynamics governed by the McKean–Vlasov SDE:

$$\begin{cases} dX_t^i = b\big(X_t^i, \mu_t\big)dt + a\big(X_t^i, \mu_t\big)dW_t^i, & i = 1, \dots, N, \quad t \in [0, T] \\ \mathcal{L}(X_t^i) = \mu_t \end{cases}$$

The statistical question is whether the squared volatility function $a^2(x, \mu)$ belongs to a prescribed parametric family spanned by a collection of basis functions $a_1^2, \dots, a_d^2$. Formally, the null hypothesis can be expressed as

$$H_0 : L := \min_{(\lambda_1,\dots,\lambda_d)\in\mathbb{R}^d} \int_0^T \int_{\mathbb{R}} \Big(a^2(x, \mu_t) - \sum_{k=1}^d \lambda_k a_k^2(x, \mu_t)\Big)^2 \mu_t(dx)\, dt = 0.$$

This criterion provides a natural foundation for our test construction, as it measures the discrepancy between the true volatility function and its parametric approximation within a Hilbert space framework. To make the procedure operational, we introduce the empirical counterpart $\widehat{S}^N$ of the functional L, which is built from the observed particle trajectories and forms the basis of our goodness-of-fit procedure.
Our main theoretical contributions establish the asymptotic properties of $\widehat{S}^N$ in a joint high-frequency and large-population regime, where the number of particles $N \to \infty$ and the observation mesh $\Delta_n := T/n \to 0$, over a fixed horizon T. Within this framework, we prove consistency of the empirical estimators underlying $\widehat{S}^N$ and derive stochastic expansions that allow us to characterize their asymptotic distribution:

Corollary I.3.4. *If the Assumptions* (IVA1)-(IVA3) *are satisfied and* $N\Delta_n^2 \to 0$, *then*

$$\sqrt{N}(\widehat{S}^N - L) \xrightarrow{\mathcal{L}} \mathcal{G} \sim \mathcal{N}(0, \tau^2)$$

where τ^2 *denotes the asymptotic variance, which can be consistently estimated from the data (see Corollary IV.4.3).*

This result provides a valid testing procedure at any prescribed significance level α. Moreover, under the alternative $H_1 : L > 0$, the test statistic diverges to infinity in probability, guaranteeing consistency against fixed alternatives.
In comparison with existing literature, goodness-of-fit testing for volatility has been extensively investigated in classical diffusion models and their extensions (see [33–35, 87]), but all these contributions are confined to non-interacting settings. In contrast, no general methodology has been available for McKean–Vlasov particle systems, where the volatility depends on the evolving distribution of the process. Our approach provides a rigorous and practical framework for goodness-of-fit testing in interacting particle systems and fills this gap by establishing the first asymptotic theory for volatility testing in the mean-field regime, based on high-frequency and large-population asymptotics.

Chapter II

Parameter estimation of discretely observed interacting particle systems

Abstract: In this paper, we consider the problem of joint parameter estimation for drift and diffusion coefficients of a stochastic McKean-Vlasov equation and for the associated system of interacting particles. The analysis is provided in a general framework, as both coefficients depend on the solution and on the law of the solution itself. Starting from discrete observations of the interacting particle system over a fixed interval $[0, T]$, we propose a contrast function based on a pseudo likelihood approach. We show that the associated estimator is consistent when the discretization step (Δ_n) and the number of particles (N) satisfy $\Delta_n \to 0$ and $N \to \infty$, and asymptotically normal when additionally the condition $\Delta_n N \to 0$ holds.

II.1 Introduction

In this paper we focus on parametric estimation of interacting particle system of the form

$$\begin{cases} dX_t^{\theta,i,N} = b\big(\theta_1, X_t^{\theta,i,N}, \mu_t^{\theta,N}\big)dt + a\big(\theta_2, X_t^{\theta,i,N}, \mu_t^{\theta,N}\big)dW_t^i, \qquad i = 1,\ldots,N, \quad t \in [0,T], \\ \mathcal{L}\big(X_0^{\theta,1,N},\ldots,X_0^{\theta,N,N}\big) := \mu_0 \times \ldots \times \mu_0. \end{cases} \tag{II.1.1}$$

Here the unknown parameter $\theta := (\theta_1, \theta_2)$ belongs to the set $\Theta := \Theta_1 \times \Theta_2$, where $\Theta_j \subset \mathbb{R}^{p_j}$, $j = 1, 2$, are compact and convex sets; we set $p := p_1 + p_2$. The processes $(W_t^i)_{t\in[0,T]}$, $i = 1,\ldots,N$, are independent $\mathbb{R}$-valued Brownian motions, independent of the initial value $(X_0^{\theta,1,N},\ldots,X_0^{\theta,N,N})$ of the system and $\mu_t^{\theta,N}$ is the empirical measure of the system at time t, i.e.

$$\mu_t^{\theta,N} := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{\theta,i,N}}.$$

The model coefficients are functions $b : U_1 \times \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ and $a : U_2 \times \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$, where U_1 and U_2 are two open sets containing Θ_1 and Θ_2, respectively, and $\mathcal{P}_2$ denotes the set of probability measures on $\mathbb{R}$ with a finite second moment, endowed with the Wasserstein 2-metric

$$W_2(\mu,\nu) := \Big(\inf_{m\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^2} |x-y|^2 m(dx,dy)\Big)^{\frac{1}{2}}, \tag{II.1.2}$$

and $\Gamma(\mu,\nu)$ denotes the set of probability measures on $\mathbb{R}^2$ with marginals μ and ν. The underlying observations are

$$\big(X_{t_{j,n}}^{\theta,i,N}\big)_{j=1,\ldots,n}^{i=1,\ldots,N},$$

where $t_{j,n} := Tj/n$ and $\Delta_n := T/n$ is the discretization step. We assume that the time horizon T is fixed, and $N, n \to \infty$.
The interacting particle system is naturally associated to its mean field equation as $N \to \infty$. The latter is described by the 1-dimensional McKean-Vlasov SDE

$$d\bar{X}_t^\theta = b\big(\theta_1, \bar{X}_t^\theta, \bar{\mu}_t^\theta\big)dt + a\big(\theta_2, \bar{X}_t^\theta, \bar{\mu}_t^\theta\big)dW_t, \quad t \in [0,T], \tag{II.1.3}$$

where $\bar{\mu}_t^\theta$ is the law of $\bar{X}_t^\theta$ and $(W_t)_{t\in[0,T]}$ is a standard Brownian motion, independent of the initial value $\bar{X}_0^\theta$ having the law $\bar{\mu}_0^\theta := \mu_0$. This equation is non-linear in the sense of McKean, see e.g. [76, 77, 90]. It means, in particular, that the coefficients depend not only on the current state but also on the current distribution of the solution. It is well known that, under appropriate assumptions on the coefficients a and b, it is possible to obtain a phenomenon commonly named *propagation of chaos* (see e.g. [90]). It implies that the empirical law $\mu_t^{\theta,N}$ weakly converges to $\bar{\mu}_t^\theta$ as $N \to \infty$. The McKean-Vlasov SDE in (II.1.3) links to a non-linear non-local partial differential equation on the space of probability measures (see e.g.

[19]), which naturally arises in several applications in statistical physics. Indeed, stochastic systems of interacting particles and the associated McKean non-linear Markov processes have been introduced in 1966 in [76] starting from statistical physics, to model the dynamics of plasma. Their importance has increased in time, and a huge number of probabilistic tools have been progressively developed in this context (see [19, 39, 74, 78], just to name a few).
On the other hand, however, statistical inference in this framework remained out of reach for many years (except for the early work of Kasonga in [63]), mainly as microscopic particle systems derived from statistical physics are not directly observable. Later on, McKean-Vlasov models found applications in several other fields, in which the data is observable. Nowadays, these models are used in finance (smile calibration in [54]; systemic risk in [42]) as well as social sciences (opinion dynamics in [21]) or mean-field games (see e.g. [15, 36, 50]). Moreover, some applications in neuroscience and population dynamics can be found respectively in [8] and [80]. At the same time, the interest in analysis of statistical models related to PDEs has gradually increased. A clear illustration of that is provided by the works on nonparametric Bayes and uncertainty quantification for inverse problems, as in [1, 83, 84].
Motivated by the increasing interest in statistical inference for McKean-Vlasov processes, we aim at estimating jointly the parameters θ_1, θ_2 starting from the discrete observations of the interacting particle systems (II.1.1) over a fixed time interval $[0, T]$. Despite recent interest in the study of the McKean-Vlasov SDEs, the problem of parameter estimation for this class has received relatively little attention. In [92] the authors established asymptotic consistency and normality of the maximum likelihood estimator for a class of McKean-Vlasov SDEs with constant diffusion coefficient, based on the continuous observation of the trajectory. This has been extended to the path dependent case in [73]. The mean field regime has been firstly considered by Kasonga in [63], who studied a system of interacting diffusion processes depending linearly in the drift coefficient on some unknown parameter. Starting from continuous observation of the system over a fixed time interval $[0, T]$, he showed that the MLE is consistent and asymptotically normal as $N \to \infty$. This has been extended in [88] to the case where the parametrisation is not linear, while Bishwal [12] extended it to the case where only discrete observations of the system are available and the parameter to be estimated is a function of time. In [50] the authors develop an asymptotic inference approach based on the approximation of the likelihood function for mean-fields models of large interacting financial systems. Moreover, Chen [22] has established the optimal convergence rate for the MLE in the large N and large T case. Even in this work the drift coefficient is linear and the diffusion coefficient is constant.
Let us also mention the works [48, 49], where parametric inference for a particular class of nonlinear self-stabilizing SDEs is studied, starting from continuous observation of the non-linear diffusion. Some different asymptotic regimes are considered, such as the small noise and the long time horizon. The problem of the semiparametric estimation of the drift coefficient starting from the observation of the particle system at time T, for $T \to \infty$ is studied in [10], while [29] considers non-parametric estimation of the drift term in a McKean-Vlasov SDE, based on the continuous observation of the associated interacting particle system over a fixed time horizon.

None of these works, however, consider the problem of the joint estimation of the drift and diffusion coefficients. Moreover, not only we are not aware of any work about parameter estimation for interacting particle system where the diffusion coefficient can depend on the solution and on the law of the solution itself, but in the majority of the above mentioned work the diffusion coefficient is directly assumed to be constant. We consider a more general model, as in (II.1.1), motivated by several applications in which the diffusion coefficient depends on the law. For example, this is the case in mathematical finance for the calibration of local and stochastic volatility models, with applications connected to the Dupire's local volatility function (see [13, 55, 69]). Moreover, they are used to capture the diversity of a financial market, as in [81].

We underline that the joint estimation of the two parameters introduces some significant difficulties: since the drift and the diffusion coefficient parameters are not estimated at the same rate, we have to deal with asymptotic properties in two different regimes. Another challenge comes from the fact that both coefficients depend on the empirical law of the process. This introduces some complexity compared to the case where a is constant.

A natural approach to estimation of unknown parameters in our context would be to use a maximum likelihood estimation. However, the likelihood function based on the discrete sample is not tractable in this setting, since it depends on the transition densities of the process, which are not explicitly known. To overcome this difficulty several methods have been developed, in the case of high frequency estimation for discretely observed classical SDEs. A widely-used method is to consider a pseudo likelihood function, for instance based on the high frequency approximation of the dynamic of the process by the dynamic of the Euler scheme, see for example [40, 64, 95].

Our statistical analysis is based upon minimisation of a contrast function, which is similar in spirit to the methods [40, 64, 95] that have been proposed in the setting of classical SDEs. The main result of the paper is the consistency and asymptotic normality of the resulting estimator, which is showed by using a central limit theorem for martingale difference triangular arrays. The convergence rates for estimation of the two parameters are different, which leads us to the study of the asymptotic properties of the contrast function in two different asymptotic schemes. Moreover, to illustrate our main results, we present numerical experiments for two models of interacting particle systems. Specifically, the first model is linear, while the second is a stochastic opinion dynamics model. While it is feasible to express the estimator explicitly for the linear model, the estimator for the stochastic opinion dynamics model is implicit and can only be obtained numerically. Our results show that the proposed estimators perform well in both cases.

We emphasize that our inference is made on the time horizon $[0, T]$ with T being fixed. It is well known that it is impossible to estimate the drift parameter of a classical SDE on a finite time horizon. However, due to increasing number of particles, we are able to consistently estimate the drift even when T is fixed. Moreover, it is worth remarking that our results apply to the system of N independent copies of a diffusion process as a special case. Non-parametric statistical inference for this type of system can be found for example in [23, 32, 75] (see also references therein). Closer

to the purpose of our work, [28, 31] discuss parameter estimation from discrete observations of independent copies of a diffusion process with mixed (or fixed) effects. Specifically, joint estimation of a fixed effect in the diffusion coefficient and parameters of the special distribution of a random effect (or a fixed effect) in the drift coefficient of the SDE is shown possible with the same rates of convergence in the same asymptotic framework as ours. Interested readers can find further references about SDEs with random effects in the aforementioned papers.
The outline of the paper is as follows. In Section II.2 we present the estimation approach, list the required assumptions and demonstrate some examples. Section II.3 is devoted to main results of the paper, which include consistency and asymptotic normality of the estimator. Section II.4 is devoted to numerical experiments. In Section II.5 we provide the technical lemmas we will use in order to show our main results. The proofs of the main results are collected in Section II.6 while the technical results are shown in Section II.7.

Notation

Throughout the paper all positive constants are denoted by C or C_q if they depend on an external parameter q. All vectors are row vectors, $\|\cdot\|$ denotes the Euclidean norm for vectors. We write $f(\theta) = f(\theta_1, \theta_2)$ for $\theta = (\theta_1, \theta_2)$. For $r = 0, 1, \dots,$ we denote by $C^r(X; \mathbb{R})$ the set of r times continuously differentiable functions $f : X \to \mathbb{R}$. We denote by $\partial_x f$ the partial derivative of a function $f(x, y, \dots)$ with respect to x. We denote by $\nabla_{\theta_j} f$ the vector $(\partial_{\theta_{j,1}} f, \dots, \partial_{\theta_{j,p_j}} f)$, $j = 1, 2$, and $\nabla_\theta f = (\nabla_{\theta_1} f, \nabla_{\theta_2} f)$. We say that a function $f : \mathbb{R} \times \mathcal{P}_l \to \mathbb{R}$ has *polynomial growth* if

$$|f(x, \mu)| \le C(1 + |x|^k + W_2^l(\mu, \delta_0)) \tag{II.1.4}$$

for some $k, l = 0, 1, \dots$ and all $(x, \mu) \in \mathbb{R} \times \mathcal{P}_l$, where $\mathcal{P}_l$ denotes the set of probability measures on $\mathbb{R}$ with a finite l-th absolute moment. For $p \in [1, \infty)$, the Wasserstein p-metric between two probability measures μ and ν in $\mathcal{P}_p$ is given as

$$W_p(\mu, \nu) := \Big(\inf_{m \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^2} |x - y|^p m(dx, dy) \Big)^{\frac{1}{p}};$$

where $\Gamma(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R}^2$ with marginals μ and ν. Finally, we suppress the dependence of several objects on the true parameter θ_0. In particular, we write $\mathbb{P} := \mathbb{P}^{\theta_0}$, $\mathbb{E} := \mathbb{E}^{\theta_0}$, $X_t^{i,N} := X_t^{\theta_0,i,N}$, $\bar{X}_t := \bar{X}_t^{\theta_0}$, $\mu_t^N := \mu_t^{\theta_0,N}$ and $\bar{\mu}_t := \bar{\mu}_t^{\theta_0}$. Furthermore, we denote by $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$, $\xrightarrow{L^p}$ the convergence in probability, in law, in L^p respectively. We also denote the value $a^2(\theta_2, x, \mu)$ as $c(\theta_2, x, \mu)$.

II.2 Minimal contrast estimator, assumptions and examples

We aim at estimating the unknown parameter $\theta_0 = (\theta_{0,1}, \theta_{0,2}) \in \Theta^\circ$ given equidistant discrete observations of the system introduced in (II.1.1). We study the asymptotic regime $N, n \to \infty$.

The estimator we propose is based upon a contrast function, which originates from the Gaussian quasi-likelihood. Starting from discrete observations of the model there are difficulties due to the fact that the transition density of the process is unknown. A common way to overcome this issue is to base the inference on a discretization of the continuous likelihood (see for example [46], [64] and [95] where classic SDEs are considered). This motivates us to consider the following contrast function:

$$S_n^N(\theta) := \sum_{i=1}^N \sum_{j=1}^n \left\{ \frac{\big(X_{t_{j,n}}^{i,N} - X_{t_{j-1,n}}^{i,N} - \Delta_n b\big(\theta_1, X_{t_{j-1,n}}^{i,N}, \mu_{t_{j-1,n}}^N\big)\big)^2}{\Delta_n c\big(\theta_2, X_{t_{j-1,n}}^{i,N}, \mu_{t_{j-1,n}}^N\big)} + \log c\big(\theta_2, X_{t_{j-1,n}}^{i,N}, \mu_{t_{j-1,n}}^N\big) \right\}, \tag{II.2.1}$$

for $\theta = (\theta_1, \theta_2)$. The estimator $\hat{\theta}_n^N = (\theta_{n,1}^N, \theta_{n,2}^N)$ of θ_0 is obtained as

$$\hat{\theta}_n^N \in \arg\min_{\theta \in \Theta} S_n^N(\theta).$$

Comparing $S_n^N(\theta)$ with the contrast function for parameter estimation for classical SDEs, the main difference consists in the fact that we have now an extra sum over the number of interacting diffusion processes. The interaction depends on the empirical measure of the system. The dependence of the drift and diffusion coefficients on the measure can take a general form. In order to meet this challenge and prove some asymptotic properties for $\hat{\theta}_n^N$ we need to introduce a set of assumptions. The first two assumptions ensure the system's existence and uniqueness, while the next two impose additional regularity conditions on the coefficients a and b.

IIA1. *(Boundedness of moments) For all $k \geq 1$,*

$$\int_{\mathbb{R}} |x|^k \mu_0(dx) \leq C_k.$$

IIA2. *(Lipschitz condition) The drift and diffusion coefficients are Lipschitz continuous in (x, μ), i.e. for all θ there exists C such that for all $(x,\mu), (y,\nu) \in \mathbb{R} \times \mathcal{P}_2$,*

$$|b(\theta_1, x, \mu) - b(\theta_1, y, \nu)| + |a(\theta_2, x, \mu) - a(\theta_2, y, \nu)| \leq C(|x-y| + W_2(\mu,\nu)).$$

IIA3. *(Regularity of the diffusion coefficient) The diffusion coefficient is uniformly bounded away from 0:*

$$\inf_{(\theta_2, x, \mu) \in \Theta_2 \times \mathbb{R} \times \mathcal{P}_2} c(\theta_2, x, \mu) > 0.$$

IIA4. *(Regularity of the derivatives) (I) For all (x,μ), the functions $b(\cdot, x, \mu)$, $a(\cdot, x, \mu)$ are in $C^3(U_1; \mathbb{R})$, $C^3(U_2; \mathbb{R})$ respectively. Furthermore, all their partial derivatives up to order three have polynomial growth, in the sense of* (II.1.4), *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 $C > 0$, $k, l = 0, 1, \ldots$ such that for all $r_1 + r_2 = 1, 2$, $h_1, h_2 = 1, \ldots, p_1$, $\tilde{h}_1, \tilde{h}_2 = 1, \ldots, p_2$, $(x,\mu), (y,\nu) \in \mathbb{R} \times \mathcal{P}_2$,

$$\begin{aligned}\big|\partial_{\theta_{1,h_1}}^{r_1} \partial_{\theta_{1,h_2}}^{r_2} b(\theta_1, x, \mu) - \partial_{\theta_{1,h_1}}^{r_1} \partial_{\theta_{1,h_2}}^{r_2} b(\theta_1, y, \nu)\big| + \big|\partial_{\theta_{2,\tilde{h}_1}}^{r_1} \partial_{\theta_{2,\tilde{h}_2}}^{r_2} a(\theta_2, x, \mu) - \partial_{\theta_{2,\tilde{h}_1}}^{r_1} \partial_{\theta_{2,\tilde{h}_2}}^{r_2} a(\theta_2, y, \nu)\big| \\ \leq C(|x-y| + W_2(\mu,\nu))\big(1 + |x|^k + |y|^k + W_2^l(\mu, \delta_0) + W_2^l(\nu, \delta_0)\big).\end{aligned}$$

Remark II.2.1. (i) It is possible to relax assumption IIA2 on the drift coefficient to allow for a locally Lipschitz condition in x with polynomial weights, cf. [37, Assumption 2.1]. In this setting the boundedness of moments shown in our Lemma II.5.1 can be replaced by [38, Theorem 3.3] and the propagation of chaos needed in order to prove Lemma II.5.2 would follow from [37, Proposition 3.1]. As a consequence the main results of this paper still hold.

(ii) IIA4(I) is sufficient to show consistency of the estimator $\hat{\theta}_n^N$. We require the additional condition (II) of IIA4 to prove the asymptotic normality. □

We now state an assumption on the identifiability of the model and some further conditions that are required to prove the asymptotic normality. For this purpose we define the functions $I : \Theta \to \mathbb{R}$, $J : \Theta_2 \to \mathbb{R}$ as

$$I(\theta) := \int_0^T \int_{\mathbb{R}} \frac{\left(b(\theta_1, x, \bar{\mu}_t) - b(\theta_{0,1}x, \bar{\mu}_t)\right)^2}{c(\theta_2, x, \bar{\mu}_t)} \bar{\mu}_t(dx)dt, \tag{II.2.2}$$

$$J(\theta_2) := \int_0^T \int_{\mathbb{R}} \Big(\frac{c(\theta_{0,2}, x, \bar{\mu}_t)}{c(\theta_2, x, \bar{\mu}_t)} + \log c(\theta_2, x, \bar{\mu}_t)\Big) \bar{\mu}_t(dx)dt, \tag{II.2.3}$$

where recall that $\bar{\mu}_t$ stands for $\bar{\mu}_t^{\theta_0}$. The next set of conditions are the following assumptions.

IIA5. *(Identifiability) The functions I, J defined above satisfy that for every $\varepsilon > 0$,*

$$\inf_{\theta\in\Theta:\|\theta_1-\theta_{0,1}\|\geq\varepsilon} I(\theta) > 0 \qquad \textit{and} \qquad \inf_{\theta_2\in\Theta_2:\|\theta_2-\theta_{0,2}\|\geq\varepsilon} (J(\theta_2) - J(\theta_{0,2})) > 0.$$

IIA6. *(Invertibility) We define a $p\times p$ block diagonal matrix $\Sigma(\theta_0) := \mathrm{diag}(\Sigma^{(1)}(\theta_0), \Sigma^{(2)}(\theta_0))$ whose main-diagonal blocks $\Sigma^{(j)}(\theta_0) = (\Sigma^{(j)}_{kl}(\theta_0))$ are defined via*

$$\Sigma^{(j)}_{kl}(\theta_0) := \begin{cases} 2\displaystyle\int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_{1,k}} b(\theta_{0,1}, x, \bar{\mu}_t)\, \partial_{\theta_{1,l}} b(\theta_{0,1}, x, \bar{\mu}_t)}{c(\theta_{0,2}, x, \bar{\mu}_t)} \bar{\mu}_t(dx)dt, & j = 1,\ k, l = 1, \dots, p_1, \\ \displaystyle\int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_{2,k}} c(\theta_{0,2}, x, \bar{\mu}_t)\, \partial_{\theta_{2,l}} c(\theta_{0,2}, x, \bar{\mu}_t)}{c^2(\theta_{0,2}, x, \bar{\mu}_t)} \bar{\mu}_t(dx)dt, & j = 2,\ k, l = 1, \dots, p_2. \end{cases}$$

We assume that $\det(\Sigma^{(j)}(\theta_0)) \neq 0$, $j = 1, 2$.

IIA7. *(Integral condition on the diffusion coefficient) At $\theta_{0,2}$ for all (x, μ) the diffusion coefficient takes the form*

$$a(\theta_{0,2}, x, \mu) := \tilde{a}\Big(x, \int_{\mathbb{R}} K(x, y)\mu(dy)\Big)$$

for some functions $\tilde{a}, K \in C^2(\mathbb{R}^2; \mathbb{R})$, which satisfy $|\partial_x^{r_1}\partial_y^{r_2}\tilde{a}(x, y)| + |\partial_x^{r_1}\partial_y^{r_2}K(x, y)| \leq C(1 + |x|^k + |y|^l)$ for some $k, l = 0, 1, \dots$ and all $r_1 + r_2 = 1, 2$, $(x, y) \in \mathbb{R}^2$.

Assumptions IIA1- IIA5 are required to prove the consistency of our estimator and are relatively standard in the literature for statistics of random processes. However,

Assumption IIA5 deserves some extra attention, as the quantities $I(\theta)$ and $J(\theta)$ are not at all explicit due to the presence of $\bar{\mu}_t$. Hence, it may be difficult to check Assumption IIA5 in practice and the identifiability of all parameters may not always be possible. In order to delve deeper into the topic, we refer to Section 2.4 in [30], where the authors have provided a thorough analysis. More specifically, for estimating the drift from continuous observations, they have identified explicit criteria that enable obtaining both identifiability and non-degeneracy of the Fisher information matrix. Notably, for a certain type of likelihood, they have established a connection between global identifiability and non-degeneracy of the Fisher information, which is highlighted in [30, Proposition 16]. It could be interesting to understand if it possible to prove an analogous proposition in our context, even if this is out of the purpose of the paper and it is therefore left for further investigation.
The additional conditions IIA6- IIA7 are needed to obtain the central limit theorem, even if they are not of the same type. Indeed, IIA6 is an invertibility condition which is always required when one wants to prove asymptotic normality. In IIA6, note that $\partial_{\theta_{1,k}} b(\theta_{0,1}, x, \bar{\mu}_t)$ and $\partial_{\theta_{2,k}} c(\theta_{0,2}, x, \bar{\mu}_t)$ are respectively $\partial_{\theta_{1,k}} b(\theta_{0,1}, x, \mu)|_{\mu=\bar{\mu}_t}$ and $\partial_{\theta_{2,k}} c(\theta_{0,2}, x, \mu)|_{\mu=\bar{\mu}_t}$, whereas $\bar{\mu}_t$ stands for $\bar{\mu}_t^{\theta_0}$. On the other hand, IIA7 is a technical condition needed in order to obtain the first statement of Lemma II.5.3. We shed light to the fact that the bounds in Lemma II.5.3 are stated for θ_0 and similarly we ask to IIA7 to be valid exclusively for the true parameter value $\theta_{0,2}$. Naturally, both $\tilde{a}$ and K in IIA7 can be functions on $\Theta_2 \times \mathbb{R}^2$ with the first argument fixed at $\theta_{0,2}$.
We also remark that, in the case where the unknown parameter θ appears only in the drift coefficient, there is no need to add a further assumption on the derivatives of the diffusion coefficient to estimate it, even if the diffusion coefficient still depends on the law of the process.

Example II.2.2. A number of interacting particle models (and associated mean field equations) have been analyzed in the literature. We highlight a few here to illustrate the scope of our paper.
We start by considering some examples where the diffusion coefficient is a constant on a compact set that does not include the origin. This case has several applications (see (i) and (ii)). After that, some more general examples are presented.

(i) The Kuramoto model is the most classical model for synchronization phenomena in large populations of coupled oscillators such as a clapping crowd, a population of fireflies or a system of neurons (see Section 5.2 of [20] and references therein). Let N oscillators be defined by N angles $X_t^{i,N}$, $i = 1, \dots, N$ (defined modulo 2π, in this way they can actually be considered as elements of the circle), evolving in $t \in [0, T]$ according to

$$dX_t^{i,N} = -\frac{\theta_{0,1}}{N} \sum_{j=1}^{N} \sin\left(X_t^{i,N} - X_t^{j,N}\right) dt + \theta_{0,2} dW_t^i.$$

This variant of the model satisfies our assumptions.

(ii) A popular model for opinion dynamics (see e.g. [21, 82]) takes the form

$$dX_t^{i,N} = -\frac{1}{N}\sum_{j=1}^{N}\varphi_{\theta_{0,1}}\left(\left|X_t^{i,N} - X_t^{j,N}\right|\right)\left(X_t^{i,N} - X_t^{j,N}\right)dt + \theta_{0,2}dW_t^i$$

for $i = 1, \ldots, N$, $t \in [0, T]$, where $\varphi_{\theta_{0,1}}(x) := \theta_{0,1,1}\mathbb{1}_{[0,\theta_{0,1,2}]}(x)$, $x \in \mathbb{R}$, is the influence function which acts on the "difference of opinions" between agents. To have our regularity assumptions hold true in practice we can replace the function $\varphi_{\theta_{0,1}}$ by its infinitely differentiable approximation as it is done in Section 5.2 of [88]. In [88] we also note that the proxy of $\varphi_{\theta_{0,1}}$ depends non-linearly on the parameter $\theta_{0,1,2}$.

(iii) Another example is

$$dX_t^{i,N} = \left(\theta_{0,1,1} + \frac{\theta_{0,1,2}}{N}\sum_{j=1}^{N} X_t^{j,N} - \theta_{0,1,3}X_t^{i,N}\right)dt + \theta_{0,2}\sqrt{1 + \left(X_t^{i,N}\right)^2}dW_t^i$$

for $i = 1, \ldots, N$, $t \in [0, T]$. We note that in the case $\theta_{0,1,2} = 0$ the interacting particle system reduces to N independent samples of a special case of the Pearson diffusion, which has applications in finance, see [41] and references therein.

(iv) We consider the dynamic of the system

$$dX_t^{i,N} = \left(\theta_{0,1,1} + \frac{\theta_{0,1,2}}{N}\sum_{j=1}^{N} X_t^{j,N} - \theta_{0,1,3}X_t^{i,N}\right)dt + \left(\theta_{0,2,1} + \theta_{0,2,2}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(X_t^{j,N}\right)^2}\right)dW_t^i$$

for $i = 1, \ldots, N$ in $t \in [0, T]$, where both the coefficients b and a depend on the law argument. We remark that the mean field limit of the above interacting particle system is a time-inhomogeneous Ornstein-Uhlenbeck process. See [63] for the case $\theta_{0,1,1} = \theta_{0,2,2} = 0$.

Some remarks are in order. Example (iv), where $\theta_{0,2,2} = 0$, has been thoroughly discussed in Section 4.1 of [30], specifically, with regard to the restrictions on μ_0 and $\theta_{0,1}$ that ensure the latter parameter satisfies A5, A6. In examples (i), (iii) and (iv), where either $\theta_{0,2,1}$ or $\theta_{0,2,2}$ is set to 0, it is obvious that A5, A6 hold for $\theta_{0,2} \neq 0$. Finally, we note that in examples (i), (iii), and (iv), where either $\theta_{0,2,1}$ or $\theta_{0,2,2}$ is set to 0, the drift and diffusion coefficients are respectively linear and multiplicative functions of θ, which allows us to solve our estimator in closed form.

II.3 Main results

Our main results demonstrate the consistency and the asymptotic normality of the estimator $\hat{\theta}_n^N$.

Theorem II.3.1. *(Consistency) Assume that IIA1- IIA5 hold, with only condition (I) in IIA4. Then the estimator $\hat{\theta}_n^N$ is consistent in probability:*

$$\hat{\theta}_n^N \xrightarrow{\mathbb{P}} \theta_0 \quad \textit{as } n, N \to \infty.$$

In order to obtain the asymptotic normality of our estimator we need to add an assumption on the relation between the rates N and Δ_n. In particular, we require that $N\Delta_n \to 0$ as $N, n \to \infty$.

Theorem II.3.2. *(Asymptotic normality) Assume that IIA1- IIA7 hold. If* $N\Delta_n \to 0$ *then*

$$\left(\sqrt{N}(\hat{\theta}^N_{n,1} - \theta_{0,1}), \sqrt{N/\Delta_n}(\hat{\theta}^N_{n,2} - \theta_{0,2})\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 2(\Sigma(\theta_0))^{-1}\right) \quad \text{as } n, N \to \infty,$$

where

$$2(\Sigma(\theta_0))^{-1} := 2\,\mathrm{diag}\left((\Sigma^{(1)}(\theta_0))^{-1}, (\Sigma^{(2)}(\theta_0))^{-1}\right)$$

with $\Sigma^{(j)}(\theta_0)$, $j = 1, 2$, *being defined in IIA6.*

As common in the literature on contrast function based methods, understanding the asymptotic behaviour of $S^N_n(\theta_1, \theta_2)$ and its derivatives is key to obtain the statements of Theorems II.3.1 and II.3.2. In particular, we show that, under proper normalisation, the first derivative of $S^N_n(\theta_1, \theta_2)$ converges to a Gaussian law with mean 0 and covariance matrix $2\Sigma(\theta_0)$ (see Proposition II.6.2), while the second derivative converges in probability to the matrix $\Sigma(\theta_0)$ defined in IIA6 (see Proposition II.6.3). These results lead to the statement of Theorem II.3.2.

The condition on the rate, at which the discretization step Δ_n converges to 0, has been discussed in detail in the framework of classical SDEs. In this context, one disposes discrete observations of the trajectory of only one particle up to a time $T := n\Delta_n \to \infty$. In [40] the corresponding condition was $T\Delta_n = n\Delta_n^2 \to 0$ as $n \to \infty$, which has been later improved to $n\Delta_n^3 \to 0$ in [95] thanks to a correction introduced in the contrast function. Finally, Kessler [64] proposed a contrast function based on a Gaussian approximation of the transition density, which allowed him to consider a weaker condition $n\Delta_n^p \to 0$ for an arbitrary integer p. Similar developments have been made in the setting of classical SDEs with jumps in [5, 6, 51, 89].

One may wonder if it possible to weaken the condition on the discretization step in the context of interacting particle systems. For a system of independent copies of a diffusion process with random and/or fixed effects, [27, 28, 31] require it in the same asymptotic framework as ours. In [28] also the rates of convergence of the estimators towards the parameters θ_1 of the distribution of a random effect in the drift coefficient, and the fixed effect θ_2 in the diffusion coefficient, are shown to be the same as ours. On the one hand, the condition $N\Delta_n \to 0$ allows us to approximate the derivative of the contrast function with a triangular array of martingale increments, as it is the case for classical SDEs. For this step, higher order approximations, similar to those in [64], could potentially help us relax this condition. On the other hand, we need it because of the correlation between particles and higher order approximations do not seem to solve this issue. Thus, we leave this investigation for future research.

A recent paper [30] establishes the *LAN property* for drift estimation in d-dimensional McKean-Vlasov models under continuous observations and with diffusion coefficient being a function of $(t, \bar{X}_t)$ only. The authors show that the Fisher information matrix is given as

$$\left(\int_0^T \int_{\mathbb{R}^d} \partial_{\theta_{1,k}}(c^{-\frac{1}{2}}b)(\theta_{0,1}, t, x, \bar{\mu}_t)^\top \partial_{\theta_{1,l}}(c^{-\frac{1}{2}}b)(\theta_{0,1}, t, x, \bar{\mu}_t)\bar{\mu}_t(dx)dt\right)_{1\le k,l\le p_1} \tag{II.3.1}$$

(cf. [88] where the diffusion coefficient is an identity matrix). This is consistent with our Theorem II.3.2 when restricted to drift estimation. In other words, our drift estimator is asymptotically efficient. When considering joint estimation of the drift and diffusion coefficients, the LAN property has not yet been shown, although the results of Gobet [53] in the classical diffusion setting give some hope. Indeed, Gobet [53] has shown that for classical SDEs, in the ergodic case, the Fisher information for the drift parameter is given by

$$(\Gamma_b^{\theta_0})_{k,l} = \int_{\mathbb{R}} \frac{\partial_{\theta_{1,k}} b(\theta_{0,1}, x)\, \partial_{\theta_{1,l}} b(\theta_{0,1}, x)}{c(\theta_{0,2}, x)} \pi(dx)$$

for $k, l = 1, \ldots, p_1$, while the one for the diffusion parameter is given by

$$(\Gamma_a^{\theta_0})_{k,l} = \int_{\mathbb{R}} \frac{\partial_{\theta_{2,k}} c(\theta_{0,2}, x)\, \partial_{\theta_{2,l}} c(\theta_{0,2}, x)}{c^2(\theta_{0,2}, x)} \pi(dx)$$

for $k, l = 1, \ldots, p_2$, where π is the invariant density associated to the diffusion. As $\Gamma_b^{\theta_0}$ modifies to (II.3.1) for McKean-Vlasov models, one could expect that $\Gamma_a^{\theta_0}$ modifies to our asymptotic variance as well. This is left for further investigation.

II.4 Numerical examples

We will now examine the finite-sample performance of the introduced estimator $\hat{\theta}_n^N$ on two examples of interacting particle systems.

II.4.1 Linear model

Consider an interacting particle system of the form:

$$dX_t^{i,N} = -\Big(\theta_{1,1} X_t^{i,N} + \frac{\theta_{1,2}}{N} \sum_{j=1}^N (X_t^{i,N} - X_t^{j,N})\Big)dt + \sqrt{\theta_2}dW_t^i, \tag{II.4.1}$$

where $i = 1, \ldots, N$, $t \in [0, T]$, for some $\theta_1 = (\theta_{1,1}, \theta_{1,2}) \in \mathbb{R}^2$, $\theta_{1,1} \neq 0$, $\theta_{1,1} + \theta_{1,2} \neq 0$, $\theta_2 > 0$ and $\int_{\mathbb{R}} x\mu_0(dx) \neq 0$. In this model, the parameter $\theta_{1,1}$ determines the intensity of attraction of each individual particle towards zero, while $\theta_{1,2}$ governs the degree of interaction, which is the attraction of each individual particle towards the empirical mean. Notably, for $\theta_{1,2} = 0$, the processes $(X_t^{i,N})_{t\in[0,T]}$, $i = 1, \ldots, N$, are independent.
Recall that for $\theta_2 = 1$, estimation of the parameter θ_1 from a continuous observation of the system has been studied in [63, 88]. Since the drift and squared diffusion

coefficients in (II.4.1) are linear in $\theta := (\theta_1, \theta_2)$, it is possible to find our estimator $\hat{\theta}_n^N$ in the closed form similarly as in [63, 88]:

$$\hat{\theta}_{n,1,1}^N = \frac{A_n^N - B_n^N}{D_n^N - C_n^N}, \qquad \hat{\theta}_{n,1,2}^N = \frac{A_n^N D_n^N - B_n^N C_n^N}{(C_n^N)^2 - C_n^N D_n^N}, \tag{II.4.2}$$

where

$$A_n^N := \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^n (X_{t_{j-1,n}}^{i,N} - \bar{X}_{t_{j-1,n}}^N)(X_{t_{j,n}}^{i,N} - X_{t_{j-1,n}}^{i,N}),\ B_n^N := \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^n X_{t_{j-1,n}}^{i,N}(X_{t_{j,n}}^{i,N} - X_{t_{j-1,n}}^{i,N}),$$

$$C_n^N := \frac{\Delta_n}{N}\sum_{i=1}^N \sum_{j=1}^n (X_{t_{j-1,n}}^{i,N} - \bar{X}_{t_{j-1,n}}^N)^2, \quad D_n^N := \frac{\Delta_n}{N}\sum_{i=1}^N \sum_{j=1}^n (X_{t_{j-1,n}}^{i,N})^2$$

with $\bar{X}_{t_{j-1,n}}^N := N^{-1}\sum_{k=1}^N X_{t_{j-1,n}}^{k,N}$, and then

$$\hat{\theta}_{n,2}^N = \frac{1}{NT}\sum_{i=1}^N \sum_{j=1}^n \left(X_{t_{j,n}}^{i,N} - X_{t_{j-1,n}}^{i,N} + \Delta_n\left(\hat{\theta}_{n,1,1}^N X_{t_{j-1,n}}^{i,N} + \frac{\hat{\theta}_{n,1,2}^N}{N}\sum_{j=1}^N (X_{t_{j-1,n}}^{i,N} - X_{t_{j-1,n}}^{j,N})\right)\right)^2. \tag{II.4.3}$$

To illustrate the finite sample performance of $\hat{\theta}_n^N$, we choose $\theta = (\theta_{1,1}, \theta_{1,2}, \theta_2) = (0.5, 1, 1)$ and $\mu_0 = \delta_1$ as in [88]. We simulate 1000 solutions of the system given by (II.4.1) using the Euler method with a step size of 0.01. We obtain observations of the system — data sets for all possible combinations of $T = 50, 100$, $\Delta_n = 0.1, 0.05, 0.01$ and $N = 50, 100$. Table II.3 presents the effect of N, Δ_n, T on the performance of $\hat{\theta}_n^N$. As N or T increases, the sample RMSE and bias of $\hat{\theta}_{n,1}^N$ decrease, whereas that of $\hat{\theta}_{n,2}^N$ do not change significantly. However, as Δ_n gets smaller, the performance of $\hat{\theta}_{n,2}$ improves, as well as that of $\hat{\theta}_{n,1,2}^N$.
We note that the numerical results presented above for $\Delta_n = 0.01$ can be viewed as the maximum likelihood estimation. Indeed, our contrast function up to a negative constant is the log-likelihood function for the Euler approximation with the same step Δ_n. Therefore, it is difficult to improve upon the estimation provided in the last lines of Table II.1. Interestingly, the performance of our estimator for $\Delta_n = 0.1$ and $\Delta_n = 0.05$ is quite similar to that of $\Delta_n = 0.01$, particularly with respect to the RMSE for the estimation of $\hat{\theta}_{n,1,1}^N$ and $\hat{\theta}_{n,1,2}^N$.

One possible application of our Theorem II.3.2 is to test the hypothesis of noninteraction of particles similarly as in [63]. Consider the null hypothesis $H_0 : \theta_{1,2} = 0$ and the alternative $H_1 : \theta_{1,2} \neq 0$. According to Theorem II.3.2, if $N\Delta_n \to 0$, then

$$\sqrt{N}(\hat{\theta}_{n,1,2}^N - \theta_{1,2}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\theta)),$$

where

$$V(\theta) := 2\Sigma_{11}^{(1)}(\theta)/(\Sigma_{11}^{(1)}(\theta)\Sigma_{22}^{(1)}(\theta) - \Sigma_{12}^{(1)}(\theta)\Sigma_{21}^{(1)}(\theta)),$$

and for all $i, j = 1, 2$,

$$\Sigma_{ij}^{(1)}(\theta) := \begin{cases} 2\theta_2^{-1}\int_0^T \int_{\mathbb{R}} x^2 \bar{\mu}_t(dx)dt, & i = j = 1, \\ 2\theta_2^{-1}\int_0^T \int_{\mathbb{R}} \left(x - \int_{\mathbb{R}} y\bar{\mu}_t(dy)\right)^2 \bar{\mu}_t(dx)dt, & \text{else}, \end{cases}$$

$N =$	50		100		50		100	
$(\Delta_n, T) =$	$(0.1, 50)$		$(0.1, 50)$		$(0.1, 100)$		$(0.1, 100)$	
$\hat{\theta}^N_{n,1,1}$	0.10	(0.00)	0.08	(0.00)	0.08	(0.00)	0.07	(0.00)
$\hat{\theta}^N_{n,1,2}$	0.15	(-0.10)	0.13	(-0.10)	0.13	(-0.10)	0.12	(-0.10)
$\hat{\theta}^N_{n,2}$	0.12	(-0.12)	0.12	(-0.12)	0.12	(-0.12)	0.12	(-0.12)
$(\Delta_n, T) =$	$(0.05, 50)$		$(0.05, 50)$		$(0.05, 100)$		$(0.05, 100)$	
$\hat{\theta}^N_{n,1,1}$	0.10	(0.01)	0.08	(0.01)	0.08	(0.01)	0.07	(0.00)
$\hat{\theta}^N_{n,1,2}$	0.12	(-0.05)	0.10	(-0.05)	0.10	(-0.05)	0.09	(-0.05)
$\hat{\theta}^N_{n,2}$	0.06	(-0.06)	0.06	(-0.06)	0.06	(-0.06)	0.06	(-0.06)
$(\Delta_n, T) =$	$(0.01, 50)$		$(0.01, 50)$		$(0.01, 100)$		$(0.01, 100)$	
$\hat{\theta}^N_{n,1,1}$	0.11	(0.01)	0.08	(0.01)	0.09	(0.01)	0.07	(0.01)
$\hat{\theta}^N_{n,1,2}$	0.11	(-0.02)	0.09	(-0.01)	0.09	(-0.01)	0.07	(-0.01)
$\hat{\theta}^N_{n,2}$	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)

Table II.1: Sample RMSE (and bias in brackets) of $\hat{\theta}^N_n$ for $\theta = (0.5, 1, 1)$ and different values of N, Δ_n, T. The number of replications is 1000.

can be explicitly computed in terms of the model parameters, see [63, 88]. By using Lemma II.5.2 and Theorem II.3.1, we have that

$$V^N_n := \hat{\theta}^N_{n,2} D^N_n / ((D^N_n - C^N_n) C^N_n) \xrightarrow{\mathbb{P}} V(\theta) \quad \text{as } n, N \to \infty.$$

Therefore, if $N\Delta_n \to 0$, under H_0, we can conclude that

$$Z^N_n := \hat{\theta}^N_{n,1,2} \sqrt{N/V^N_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{as } n, N \to \infty.$$

Thus, we reject H_0 if

$$|Z^N_n| > z_{\alpha/2},$$

where $\alpha \in (0, 1)$ is the chosen level of significance and z_α denotes the α-quantile of the standard normal distribution.
Next, we examine the performance of the test statistic Z^N_n. We simulate 1000 solutions of the system given by (II.4.1) with $\mu_0 = \delta_1$, using the Euler method with a step size of 0.01. Table II.2 reports the rejection rates of H_0 in favor of H_1 at a significance level of $\alpha = 5\%$ using Z^N_n for all possible combinations of $N, T = 50, 100$, $\Delta_n = 0.1$ and $\theta = (0.5, \theta_{1,2}, 1)$, where $\theta_{1,2} = 0, 0.1, 0.25, 0.5$, or 1. The empirical size is quite well observed. Rejection rates of incorrect H_0 increase with increasing $\theta_{1,2}$ or N and T.

II.4.2 Stochastic opinion dynamics model

We now consider an interacting particle system that can model opinion dynamics:

$$dX^{i,N}_t = -\frac{1}{N} \sum_{j=1}^{N} \varphi_{\theta_1}(|X^{i,N}_t - X^{j,N}_t|)(X^{i,N}_t - X^{j,N}_t)dt + \sqrt{\theta_2} dW^i_t, \qquad \text{(II.4.4)}$$

$\theta_{1,2}$	$(N,T) =$	$(50,50)$	$(100,50)$	$(50,100)$	$(100,100)$
0		4.8	4.6	4.2	4.1
0.1		17.8	22.5	21.4	28.9
0.25		61.3	78.2	75.6	87.0
0.5		97.2	99.7	99.8	99.9
1		100.0	100.0	100.0	100.0

Table II.2: Rejection rates (in %) of $H_0 : \theta_{1,2} = 0$ vs. $H_1 : \theta_{1,2} \neq 0$ at level $\alpha = 5\%$ with Z_n^N for $\theta = (0.5, \theta_{1,2}, 1)$, $\Delta_n = 0.1$ and different values of N, T. The number of replications is 1000.

where $i = 1, \dots, N$, $t \in [0,T]$, and

$$\varphi_{\theta_1}(x) := \theta_{1,2} \exp\Big(- \frac{0.01}{1 - (x - \theta_{1,1})^2} \Big) \mathbb{1}_{[\theta_{1,1}-1, \theta_{1,1}+1]}(x), \quad x \in \mathbb{R},$$

for some $-1 < \theta_{1,1} \leq 1$, $\theta_{1,2} > 0$, $\theta_2 > 0$. The interaction kernel $\varphi_{\theta_1}(x)$ provides an infinitely differentiable approximation to the scaled indicator function $\theta_{1,2}\mathbb{1}_{[0,\theta_{1,1}+1]}(x)$, $x \geq 0$. We interpret that $\theta_{1,1}$ governs the intensity of attraction of each individual particle towards the scaled empirical mean of all the others within a distance $\theta_{1,1} + 1$. The position of each particle represents its opinion, and over time, the opinions of particles merge into metastable "soft clusters". For further information on this stochastic opinion dynamics model, see [88] and references therein.

Note that the squared diffusion coefficient is a multiplicative function of θ_2 which enables us to express $\hat{\theta}_{n,2}^N$ in terms of $(\hat{\theta}_{n,1,1}^N, \hat{\theta}_{n,1,2}^N)$. However, the latter estimator is implicit and can only be found using a numerical method. To illustrate the performance of $\hat{\theta}_n^N = (\hat{\theta}_{n,1,1}^N, \hat{\theta}_{n,1,2}^N, \hat{\theta}_{n,2}^N)$ we choose the parameter $\theta = (\theta_{1,1}, \theta_{1,2}, \theta_2) = (-0.5, 2, 0.04)$ as in [88], and the initial distribution $\mu_0 = \mathcal{N}(0,1)$ for each individual particle. We simulate 1000 solutions of the system given by (II.4.4) using the Euler method with a step size of 0.01. We obtain 1000 data sets for $\Delta_n = 0.1$ and all possible combinations of $N, T = 50, 100$ as in the previous subsection. Table II.3 presents the effect of N, T on the performance of $\hat{\theta}_n^N$. As N increases, the sample RMSE and bias of $\hat{\theta}_n^N$ decrease, whereas they do not change that much with increasing T. We can also see that $\hat{\theta}_{n,1,1}^N$ is more accurate than $\hat{\theta}_{n,1,2}^N$.

$(N,T) =$	$(50,50)$		$(100,50)$		$(50,100)$		$(100,100)$	
$\hat{\theta}_{n,1,1}^N$	0.0340	(0.0159)	0.0263	(0.0145)	0.0280	(0.0154)	0.0206	(0.0137)
$\hat{\theta}_{n,1,2}^N$	0.1652	(-0.1378)	0.1503	(-0.1347)	0.1526	(-0.1420)	0.1472	(-0.1416)
$\hat{\theta}_{n,2}^N$	0.0027	(-0.0026)	0.0026	(-0.0025)	0.0033	(-0.0032)	0.0033	(-0.0033)

Table II.3: Sample RMSE (and bias in brackets) of $\hat{\theta}_n^N$ for $\theta = (-0.5, 2, 0.04)$, $\Delta_n = 0.1$ and different values of N, T. The number of replications is 1000.

II.5 Technical lemmas

Before proving the main statistical results stated in previous section, we need to introduce some additional notations and to state some lemmas which will be useful in the sequel.

Define $\mathcal{F}_t^N := \sigma\{(W_u^k)_{u\in[0,t]}, X_0^{k,N}; k = 1, \ldots, N\}$ and $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_t^N]$. For a set $(Y_{t,n}^{i,N})$ of random variables and $\delta \geq 0$, the notation

$$Y_{t,n}^{i,N} = R_t^i(\Delta_n^\delta)$$

means that $Y_{t,n}^{i,N}$ is $\mathcal{F}_t^N$-measurable and the set $(Y_{t,n}^{i,N}/\Delta_n^\delta)$ is bounded in L^q for all $q \geq 1$, uniformly in t, i, n, N. That is

$$\mathbb{E}\big[\big|Y_{t,n}^{i,N}/\Delta_n^\delta\big|^q\big]^{1/q} \leq C_q$$

for all t, i, n, N, $q \geq 1$.

We will repeatedly use some moment inequalities gathered in the following lemma.

Lemma II.5.1. *Assume IIA1-IIA2. Then, for all $p \geq 1$, $0 \leq s < t \leq T$ such that $t - s \leq 1$, $i \in \{1, \ldots, N\}$, $N \in \mathbb{N}$, the following hold true.*

1. $\sup_{t\in[0,T]} \mathbb{E}[|X_t^{i,N}|^p] < C$, *moreover,* $\sup_{t\in[0,T]} \mathbb{E}[W_p^q(\mu_t^N, \delta_0)] < C$ *for* $p \leq q$.
2. $\mathbb{E}[|X_t^{i,N} - X_s^{i,N}|^p] \leq C(t-s)^{\frac{p}{2}}$.
3. $\mathbb{E}_s[|X_t^{i,N} - X_s^{i,N}|^p] \leq C(t-s)^{\frac{p}{2}} R_s^i(1)$.
4. $\mathbb{E}[W_2^p(\mu_t^N, \mu_s^N)] \leq C(t-s)^{\frac{p}{2}}$.
5. $\mathbb{E}_s[W_2^p(\mu_t^N, \mu_s^N)|] \leq C(t-s)^{\frac{p}{2}} R_s(1)$.

The asymptotic properties of the estimator are deduced by the asymptotic behaviour of our contrast function. To study it, the following lemma will be useful.

Lemma II.5.2. *Assume IIA1-IIA2. Let $f : \mathbb{R} \times \mathcal{P}_l \to \mathbb{R}$ satisfy for some $C > 0$, $k, l = 0, 1, \ldots$ and all $(x, \mu), (y, \nu) \in \mathbb{R} \times \mathcal{P}_l$,*

$$|f(x,\mu) - f(y,\nu)| \leq C(|x-y| + W_2(\mu,\nu))(1 + |x|^k + |y|^k + W_l^l(\mu,\delta_0) + W_l^l(\nu,\delta_0)). \tag{II.5.1}$$

Moreover, let the mapping $(x,t) \mapsto f(x, \bar{\mu}_t)$ be integrable with respect to $\bar{\mu}_t(dx)dt$ over $\mathbb{R} \times [0,T]$. Then

$$\frac{\Delta_n}{N} \sum_{i=1}^N \sum_{j=1}^n f(X_{t_{j-1,n}}^{i,N}, \mu_{t_{j-1,n}}^N) \xrightarrow{\mathbb{P}} \int_0^T \int_{\mathbb{R}} f(x,\bar{\mu}_t)\bar{\mu}_t(dx)dt \quad \text{as } n, N \to \infty.$$

It is worth underlining that the boundedness of the moments and the convergence of the Riemann sums, which are obtained almost for free in the classical SDE case, are more complex in our setting. In particular, the proof of Lemma II.5.2 consists now in three steps, the first deals with the convergence of the proper Riemann sums, in the second step we move from the interacting particle system to the iid system though the propagation of chaos property, while the third step is an application of the law of large numbers.
Another challenge compared to the classical SDE case is gathered in next lemma. Indeed, our main results heavily rely on the study of derivatives of our contrast function and so on the moment bounds of its numerator. To accomplish this, we need to use Itô's lemma on the squared diffusion coefficient as a function of the particle system's state. Therefore, we must understand how to express derivatives of a with respect to the measure argument. That is the purpose of the extra hypothesis IIA7, thanks to which the problem reduces to study the derivatives of K.
We recall that, in the sequel, we will denote by $c(\theta_2, x, \mu)$ the value $a^2(\theta_2, x, \mu)$.

Lemma II.5.3. *Assume IIA1-IIA2. Then, the following hold true.*

1. *If also IIA7 is satisfied, then* $\mathbb{E}_{t_{j,n}}[(X^{i,N}_{t_{j+1,n}} - X^{i,N}_{t_{j,n}} - \Delta_n b(\theta_{0,1}, X^{i,N}_{t_{j,n}}, \mu^N_{t_{j,n}}))^2] = \Delta_n c(\theta_{0,2}, X^{i,N}_{t_{j,n}}, \mu^N_{t_{j,n}}) + R^i_{t_{j,n}}(\Delta_n^2).$
2. $\mathbb{E}_{t_{j,n}}[(X^{i,N}_{t_{j+1,n}} - X^{i,N}_{t_{j,n}} - \Delta_n b(\theta_{0,1}, X^{i,N}_{t_{j,n}}, \mu^N_{t_{j,n}}))^4] = 3\Delta_n^2 c^2(\theta_{0,2}, X^{i,N}_{t_{j,n}}, \mu^N_{t_{j,n}}) + R^i_{t_{j,n}}(\Delta_n^{\frac{5}{2}}).$
3. $|\mathbb{E}_{t_{j,n}}[X^{i,N}_{t_{j+1,n}} - X^{i,N}_{t_{j,n}} - \Delta_n b(\theta_{0,1}, X^{i,N}_{t_{j,n}}, \mu^N_{t_{j,n}})]| = R^i_{t_{j,n}}(\Delta_n^{\frac{3}{2}}).$

We underline that IIA7 is needed in order to prove that the size of the remainder function in the first point is Δ_n^2. Without it, the size of the rest function would have been $\Delta_n^{\frac{3}{2}}$, which would not have been enough to obtain the asymptotic normality as in Proposition II.6.2 (see the proof of (II.6.23)). The proof of the lemmas stated in this section can be found in Section II.7.

II.6 Proofs

II.6.1 Consistency

Let us prove the (asymptotic) consistency of $\hat{\theta}^N_n = (\hat{\theta}^N_{n,1}, \hat{\theta}^N_{n,2})$ component-wise. Our approach is similar to that taken in the proof of [91, Theorem 5.7]. In particular, we consider a criterion function $\theta \mapsto S^N_n(\theta)$ as a random element taking values in $(C(\Theta; \mathbb{R}), \|\cdot\|_\infty)$. The uniform convergence of criterion functions is proved in the following lemma.

Lemma II.6.1. *Assume IIA1- IIA3, IIA4(I), IIA5. Then as $N, n \to \infty$,*

$$\sup_{(\theta_1,\theta_2)\in\Theta} \left| \frac{\Delta_n}{N} S^N_n(\theta_1, \theta_2) - J(\theta_2) \right| \xrightarrow{\mathbb{P}} 0, \tag{II.6.1}$$

$$\sup_{(\theta_1,\theta_2)\in\Theta}\left|\frac{1}{N}(S_n^N(\theta_1,\theta_2)-S_n^N(\theta_{0,1},\theta_2))-I(\theta_1,\theta_2)\right|\xrightarrow{\mathbb{P}}0, \tag{II.6.2}$$

where the functions I, J are defined in (II.2.2), (II.2.3) *respectively.*

Proof. It suffices to show the following steps:

1. $\frac{\Delta_n}{N}S_n^N(\theta_1,\theta_2)\xrightarrow{\mathbb{P}}J(\theta_2)$ for every $(\theta_1,\theta_2)\in\Theta$.
2. The sequence $(\theta_1,\theta_2)\mapsto\frac{\Delta_n}{N}S_n^N(\theta_1,\theta_2)$ is tight in $(C(\Theta;\mathbb{R}),\|\cdot\|_\infty)$.
3. $\frac{1}{N}(S_n^N(\theta_1,\theta_2)-S_n^N(\theta_{0,1},\theta_2))\xrightarrow{\mathbb{P}}I(\theta_1,\theta_2)$ for every $(\theta_1,\theta_2)\in\Theta$,
4. The sequence $(\theta_1,\theta_2)\mapsto\frac{1}{N}(S_n^N(\theta_1,\theta_2)-S_n^N(\theta_{0,1},\theta_2))$ is tight in $(C(\Theta;\mathbb{R}),\|\cdot\|_\infty)$.

Let us omit the notation for dependence on N,n, in particular, write X_t^i for $X_t^{i,N}$, μ_t for μ_t^N, t_j for $t_{j,n}$. Denote $f(\cdot,X_t^i,\mu_t)$ by $f_t^i(\cdot)$ for a function f, for example equal to h or g defined as

$$h(\theta,x,\mu)=\frac{(b(\theta_{0,1},x,\mu)-b(\theta_1,x,\mu))^2}{c(\theta_2,x,\mu)},\qquad g(\theta,x,\mu)=\frac{b(\theta_{0,1},x,\mu)-b(\theta_1,x,\mu)}{c(\theta_2,x,\mu)} \tag{II.6.3}$$

for all $\theta=(\theta_1,\theta_2)\in\Theta_1\times\Theta_2=\Theta$, $x\in\mathbb{R}$, $\mu\in\mathcal{P}_2$.

• Step 3. We start proving that for every $\theta=(\theta_1,\theta_2)\in\Theta_1\times\Theta_2=\Theta$,

$$\frac{1}{N}(S_n^N(\theta_1,\theta_2)-S_n^N(\theta_{0,1},\theta_2))\xrightarrow{\mathbb{P}}I(\theta)=\int_0^T\int_{\mathbb{R}}h(\theta,x,\bar{\mu}_t)\bar{\mu}_t(dx)dt.$$

Let us first decompose the left hand side as a sum of a main term and remainder. We have

$$S_n^N(\theta_1,\theta_2)=\sum_{i=1}^N\sum_{j=1}^n\frac{(H_j^i+\Delta_n(b_{t_{j-1}}^i(\theta_{0,1})-b_{t_{j-1}}^i(\theta_1)))^2}{\Delta_n c_{t_{j-1}}^i(\theta_2)}+(\log c)_{t_{j-1}}^i(\theta_2),$$

where $H_j^i=X_{t_{j-1}}^i-X_{t_{j-1}}^i-\Delta_n b_{t_{j-1}}^i(\theta_{0,1})$ for all i,j. We decompose

$$\frac{1}{N}(S_n^N(\theta_1,\theta_2)-S_n^N(\theta_{0,1},\theta_2))=I_n^N(\theta)+2\rho_n^N(\theta), \tag{II.6.4}$$

where

$$I_n^N(\theta)=\frac{\Delta_n}{N}\sum_{i=1}^N\sum_{j=1}^n h_{t_{j-1}}^i(\theta),\qquad \rho_n^N(\theta)=\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^n g_{t_{j-1}}^i(\theta)H_j^i. \tag{II.6.5}$$

Then

$$I_n^N(\theta)\xrightarrow{\mathbb{P}}I(\theta)$$

follows from Lemma II.5.2 if the function $h(\theta,\cdot)$ is locally Lipschitz with polynomial growth. To check this assumption we note that the functions $b(\theta_{0,1},\cdot)-b(\theta_1,\cdot)$,

$a(\theta_2,\cdot)$ are Lipschitz continuous and have linear growth by IIA2. We also recall that $\inf_{x,\mu} c(\theta_2,x,\mu) > 0$ by IIA3. Hence, $h(\theta,\cdot)$ satisfies the assumption of Lemma II.5.2. It remains to show that

$$\rho_n^N(\theta) \xrightarrow{\mathbb{P}} 0. \tag{II.6.6}$$

With $H_j^i = B_j^i + A_j^i$, where

$$B_j^i = \int_{t_{j-1}}^{t_j} (b_s^i(\theta_{0,1}) - b_{t_{j-1}}^i(\theta_{0,1}))ds, \qquad A_j^i = \int_{t_{j-1}}^{t_j} a_s^i(\theta_{0,2})dW_s^i,$$

for all i,j, let us further decompose

$$\rho_n^N(\theta) = \rho_{n,1}^N(\theta) + \rho_{n,2}^N(\theta), \tag{II.6.7}$$

where

$$\rho_{n,1}^N(\theta) = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^n g_{t_{j-1}}^i(\theta)B_j^i, \qquad \rho_{n,2}^N(\theta) = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^n g_{t_{j-1}}^i(\theta)A_j^i.$$

It is enough to show that

$$\rho_{n,k}^N(\theta) \xrightarrow{L^k} 0, \qquad k=1,2. \tag{II.6.8}$$

First, let us show (II.6.8) in case $k=2$. Note that for all $i_1 = i_2$ and $j_1 \neq j_2$,

$$\mathbb{E}[g_{t_{j_1-1}}^{i_1}(\theta)A_{j_1}^{i_1} g_{t_{j_2-1}}^{i_2}(\theta)A_{j_2}^{i_2}] = 0 \tag{II.6.9}$$

follows from $\mathbb{E}_{t_{j_1-1}}[A_{j_1}^{i_1}] = 0$, whereas independence of Brownian motions implies (II.6.9) for all $i_1 \neq i_2$ and j_1, j_2. We conclude that

$$\mathbb{E}[(\rho_{n,2}^N(\theta))^2] = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^n \mathbb{E}[(g_{t_{j-1}}^i(\theta)A_j^i)^2]. \tag{II.6.10}$$

Next, the Itô isometry gives

$$\mathbb{E}[(g_{t_{j-1}}^i(\theta)A_j^i)^2] = \int_{t_{j-1}}^{t_j} \mathbb{E}[(g^2)_{t_{j-1}}^i(\theta)c_s^i(\theta_{0,2})]ds,$$

where $\mathbb{E}[(g^2)_{t_{j-1}}^i(\theta)c_s^i(\theta_{0,2})] = O(1)$ uniformly in $t_{j-1} \leq s \leq t_j, j, i$ thanks to $\inf_{x,\mu} c(\theta_2,x,\mu) > 0$ by IIA3, linear growth of $a(\theta_{0,2},\cdot)$, $b(\theta_1,\cdot)$ by IIA2 and moment bounds in Lemma II.5.1(1). We conclude that $\mathbb{E}[(g_{t_{j-1}}^i(\theta)A_j^i)^2] = O(\Delta_n)$ uniformly in i,j, which in turn implies

$$\mathbb{E}[(\rho_{n,2}^N(\theta))^2] = O\big(N^{-1}\big).$$

Finally, let us show (II.6.8) in case $k=1$. For this purpose, use

$$\mathbb{E}[|g_{t_{j-1}}^i(\theta)B_j^i|] \leq \int_{t_{j-1}}^{t_j} \mathbb{E}[|g_{t_{j-1}}^i(\theta)(b_s^i(\theta_{0,1}) - b_{t_{j-1}}^i(\theta_{0,1}))|]ds$$

and then the Cauchy–Schwarz inequality. Note $\mathbb{E}[(g^2)^i_{t_{j-1}}(\theta)] = O(1)$ uniformly in j, i follows in the same way as above. Lipschitz continuity of $b(\theta_{0,1}, \cdot)$ by IIA2 and moment bounds in Lemma II.5.1(2) and (4) imply $\mathbb{E}[(b^i_s(\theta_{0,1}) - b^i_{t_{j-1}}(\theta_{0,1}))^2] = O(\Delta_n)$ uniformly in $t_{j-1} \le s \le t_j, j, i$. We conclude that

$$\mathbb{E}[|\rho^N_{n,1}|] = O(\Delta_n^{\frac{1}{2}}).$$

This completes the proof of Step 3.

• Step 4. Recall the decomposition (II.6.4), (II.6.7). It is enough to show tightness of

$$\theta \mapsto I^N_n(\theta), \qquad \theta \mapsto \rho^N_{n,k}(\theta), \qquad k = 1, 2.$$

Our approach to showing tightness of both sequences are based upon [62, Theorem 14.5]. We need to show that for all N, n:

$$\mathbb{E}\big[\sup_\theta \|\nabla_\theta I^N_n(\theta)\|\big] \le C, \qquad \mathbb{E}\big[\sup_\theta \|\nabla_\theta \rho^N_{n,1}(\theta)\|\big] \le C. \tag{II.6.11}$$

The above bounds follow if for all N, n, and $i, j, t_{j-1} \le s \le t_j$,

$$\mathbb{E}\big[\sup_\theta \|\nabla_\theta h^i_{t_{j-1}}(\theta)\|\big] \le C, \qquad \mathbb{E}\big[|b^i_s(\theta_{0,1})| \sup_\theta \|\nabla_\theta g^i_{t_{j-1}}(\theta)\|\big] \le C, \tag{II.6.12}$$

where $h, g : \Theta \times \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ are defined by (II.6.3). In $\nabla_{\theta_k} h, \nabla_{\theta_k} g,\ k = 1, 2$, we note $\nabla_{\theta_1}(b(\theta_{0,1}, \cdot) - b(\theta_1, \cdot)) = -\nabla_{\theta_1} b(\theta_1, \cdot)$. Moreover, by the mean value theorem, $|b(\theta_{0,1}, \cdot) - b(\theta_1, \cdot)| \le C \sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\|$ for all $\theta_1 \in \Theta_1$, since Θ_1 is convex, bounded. Additionally using $\inf_{\theta_2, x, \mu} c(\theta_2, x, \mu) > 0$ by IIA3, we get

$$\|\nabla_{\theta_1} g(\theta, \cdot)\| \le C \sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\|, \qquad \|\nabla_{\theta_2} g(\theta, \cdot)\| \le C \sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\| \sup_{\theta_2} \|\nabla_{\theta_2} a(\theta_2, \cdot)\|,$$

and

$$\|\nabla_{\theta_1} h(\theta, \cdot)\| \le C \sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\|^2, \qquad \|\nabla_{\theta_2} h(\theta, \cdot)\| \le C \sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\|^2 \sup_{\theta_2} \|\nabla_{\theta_2} a(\theta_2, \cdot)\|$$

for all θ. We have the polynomial growth of $\sup_{\theta_1} \|\nabla_{\theta_1} b(\theta_1, \cdot)\|$, $\sup_{\theta_2} \|\nabla_{\theta_2} a(\theta_2, \cdot)\|$ thanks to assumption IIA4 and linear growth of $b(\theta_{0,1}, \cdot)$ thanks to IIA2. The Cauchy-Schwarz inequality and moment bounds in Lemma II.5.1(1) yield (II.6.12) and so (II.6.11).
Following the approach of [58, Theorem 20 in Appendix 1], we want to show that for all N, n and $\theta, \theta' \in \Theta$,

$$\mathbb{E}[|\rho^N_{n,2}(\theta)|^2] \le C, \qquad \mathbb{E}[|\rho^N_{n,2}(\theta) - \rho^N_{n,2}(\theta')|^2] \le C\|\theta - \theta'\|_2^2.$$

We note that the second relation implies the first one because $\rho^N_{n,2}(\theta) = 0$ with $\theta_1 = \theta_{0,1}$ and Θ_2 is bounded. In the same way as in (II.6.10) we get

$$\mathbb{E}[|\rho^N_{n,2}(\theta) - \rho^N_{n,2}(\theta')|^2] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^n \mathbb{E}[|(g^i_{t_{j-1}}(\theta) - g^i_{t_{j-1}}(\theta'))A^i_j|^2],$$

where the Itô isometry gives

$$\mathbb{E}[|(g^i_{t_{j-1}}(\theta) - g^i_{t_{j-1}}(\theta'))A^i_j|^2] = \int_{t_{j-1}}^{t_j} \mathbb{E}[(g^i_{t_{j-1}}(\theta) - g^i_{t_{j-1}}(\theta'))^2 c^i_s(\theta_{0,2})]ds.$$

By the mean value theorem,

$$|g(\theta,\cdot) - g(\theta',\cdot)| \leq \|\theta - \theta'\| \sup_{\theta} \|\nabla_\theta g(\theta,\cdot)\|$$

since Θ is convex. Then

$$\mathbb{E}\big[\sup_{\theta} \|\nabla_\theta g^i_{t_{j-1}}(\theta)\|^2 c^i_s(\theta_{0,2})\big] \leq C$$

for all $t_{j-1} \leq s \leq t_j, j, i$ and N, n follows in a similar way as the second bound in (II.6.12) does using, in addition, linear growth of $a(\theta_{0,2},\cdot)$, which follows from its Lipschitz continuity by IIA2.

• Step 1. We want to prove that for every $\theta \in \Theta$,

$$\frac{\Delta_n}{N} S^N_n(\theta) \xrightarrow{\mathbb{P}} J(\theta_2) = \int_0^T \int_{\mathbb{R}} f(\theta_2, x, \bar{\mu}_t)\bar{\mu}_t(dx)dt, \tag{II.6.13}$$

where

$$f(\theta_2, x, \mu) = \frac{c(\theta_{0,2}, x, \mu)}{c(\theta_2, x, \mu)} + \log c(\theta_2, x, \mu)$$

for every $(\theta_2, x, \mu) \in \Theta_2 \times \mathbb{R} \times \mathcal{P}_2$. For this purpose, in $\Delta_n S^N_n(\theta)$ let us decompose every term as

$$\frac{(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1))^2}{c^i_{t_{j-1}}(\theta_2)} + \Delta_n(\log c)^i_{t_{j-1}}(\theta_2) = \Delta_n f^i_{t_{j-1}}(\theta_2) + r^i_j. \tag{II.6.14}$$

We can decompose r^i_j further with

$$X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1) = B^i_j(\theta_1) + A^i_j, \tag{II.6.15}$$

where

$$B^i_j(\theta_1) = \int_{t_{j-1}}^{t_j} b^i_s(\theta_{0,1})ds - \Delta_n b^i_{t_{j-1}}(\theta_1), \qquad A^i_j = \int_{t_{j-1}}^{t_j} a^i_s(\theta_{0,2})dW^i_s, \tag{II.6.16}$$

note

$$\mathbb{E}_{t_{j-1}}[(A^i_j)^2] = \int_{t_{j-1}}^{t_j} c^i_s(\theta_{0,2})ds.$$

We get

$$r^i_j = \sum_{k=0}^{2} r^i_{j,k}, \qquad \text{where } r^i_{j,k} = \frac{H^i_{j,k}}{c^i_{t_{j-1}}(\theta_2)}, \qquad k = 0, 1, 2, \tag{II.6.17}$$

and

$$H^i_{j,2} = (A^i_j)^2 - \mathbb{E}_{t_{j-1}}[(A^i_j)^2], \qquad H^i_{j,1} = 2A^i_j B^i_j(\theta_1) + (B^i_j(\theta_1))^2,$$
$$H^i_{j,0} = \mathbb{E}_{t_{j-1}}[(A^i_j)^2] - \Delta_n c^i_{t_{j-1}}(\theta_{0,2}).$$

Our proof of (II.6.13) consists of the following steps:

$$\frac{\Delta_n}{N}\sum_{i=1}^N\sum_{j=1}^n f^i_{t_{j-1}}(\theta_2) \xrightarrow{\mathbb{P}} J(\theta_2), \qquad \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^n r^i_{j,k} \xrightarrow{L^1} 0, \quad k=0,1,2. \tag{II.6.18}$$

Let us start from the convergence in (II.6.18) for $k = 2$. It is enough to show that $\sup_i \mathbb{E}[(\sum_j r^i_{j,2})^2] = o(1)$. We note that $\mathbb{E}[r^i_{j_1,2} r^i_{j_2,2}] = 0$, $j_1 \neq j_2$, since $\mathbb{E}_{t_{j-1}}[r^i_{j,2}] = 0$. We are left to show that $\sup_i \sum_j \mathbb{E}[(r^i_{j,2})^2] = o(1)$. Thanks to assumption IIA3 it reduces to showing $\sup_i \sum_j \mathbb{E}[(H^i_{j,2})^2] = o(1)$, where $\mathbb{E}_{t_{j-1}}[(H^i_{j,2})^2] = \mathbb{E}_{t_{j-1}}[(A^i_j)^4] - (\mathbb{E}_{t_{j-1}}[(A^i_j)^2])^2$ leads to $\mathbb{E}[(H^i_{j,2})^2] \le \mathbb{E}[(A^i_j)^4]$ for all i, j. Furthermore, by the Burkholder-Davis-Gundy inequality and Jensen's inequality,

$$\mathbb{E}[(A^i_j)^4] \le C\mathbb{E}\Big[\Big(\int_{t_{j-1}}^{t_j} c^i_s(\theta_{0,2})ds\Big)^2\Big] \le C\Delta_n \int_{t_{j-1}}^{t_j} \mathbb{E}[(c^2)^i_s(\theta_{0,2})]ds = O(\Delta_n^2) \tag{II.6.19}$$

uniformly in i, j, where the last relation follows thanks to linear growth of $a(\theta_{0,2}, \cdot)$ by IIA2 and moment bounds in Lemma II.5.1(1). We conclude that $\sup_{i,j} \mathbb{E}[(R^i_{j,2})^2] = O(\Delta_n^2)$.

We now turn to the convergence in (II.6.18) for $k = 1$. It is enough to show that $n \sup_{i,j} \mathbb{E}[|r^i_{j,1}|] = o(1)$. Assumption IIA3 implies $\mathbb{E}[|r^i_{j,1}|] \le C\mathbb{E}[|H^i_{j,1}|]$ for all i, j, where $\sup_{i,j} \mathbb{E}[(A^i_j)^2] = O(\Delta_n)$ follows from (II.6.19). Moreover, by Jensen's inequality,

$$\mathbb{E}[(B^i_j(\theta_1))^2] \le 2\Delta_n \int_{t_{j-1}}^{t_j} \mathbb{E}[(b^i_s(\theta_{0,1}))^2]ds + 2\Delta_n^2 \mathbb{E}[(b^i_{t_{j-1}}(\theta_1))^2] = O(\Delta_n^2)$$

uniformly in i, j, where the last relation follows thanks to linear growth of $b(\theta_1, \cdot)$ for every θ_1 by IIA2 and moment bounds in Lemma II.5.1(1). We conclude that $\sup_{i,j} \mathbb{E}[|r^i_{j,1}|] = O(\Delta_n^{\frac{3}{2}})$.

Next, we consider the convergence in (II.6.18) for $k = 0$. It is enough to show that $n \sup_{i,j} \mathbb{E}[|r^i_{j,0}|] = o(1)$. Assumption IIA3 implies $\mathbb{E}[|r^i_{j,0}|] \le C\mathbb{E}[|H^i_{j,0}|]$, where

$$\mathbb{E}[|H^i_{j,0}|] \le \int_{t_{j-1}}^{t_j} \mathbb{E}[|c^i_s(\theta_{0,2}) - c^i_{t_{j-1}}(\theta_{0,2})|]ds.$$

Lipschitz continuity of $a(\theta_{0,2}, \cdot)$ and Lemma II.5.1(2) and (4) imply $\mathbb{E}[(a^i_s(\theta_{0,2}) - a^i_{t_{j-1}}(\theta_{0,2}))^2] = O(\Delta_n)$ uniformly in $t_{j-1} \le s \le t_j, j, i$. Finally, linear growth of $a(\theta_{0,2}, \cdot)$ and moment bounds in Lemma II.5.1(1) guarantee $\mathbb{E}[(a^i_s(\theta_{0,2}) + a^i_{t_{j-1}}(\theta_{0,2}))^2] = O(1)$ uniformly in $t_{j-1} \le s \le t_j, j, i$. We conclude by Cauchy-Schwarz inequality that $\mathbb{E}[|c^i_s(\theta_{0,2}) - c^i_{t_{j-1}}(\theta_{0,2})|] = O(\Delta_n^{\frac{1}{2}})$ uniformly in $t_{j-1} \le s \le t_j, j, i$, whence $\sup_{i,j} \mathbb{E}[|r^i_{j,0}|] = O(\Delta_n^{\frac{3}{2}})$.

The first relation in (II.6.18) follows from Lemma II.5.2 if the function $f(\theta_2,\cdot)$ is locally Lipschitz with polynomial growth. To check this assumption, use $|\log y_1 - \log y_2| \le |y_1-y_2|/\min(y_1,y_2)$ for $y_1,y_2>0$ and assumption IIA3. Note $b(\theta_1,\cdot)$, $a(\theta_2,\cdot)$ are Lipschitz continuous and have linear growth by IIA2. Hence, the function $f(\theta_2,\cdot)$ satisfies the assumption of Lemma II.5.2.

• Step 2. We want to prove that the sequence $\frac{\Delta_n}{N}S_n^N(\theta)$ in $(C(\Theta;\mathbb{R}),\|\cdot\|_\infty)$ is tight. So we have to show that for all N,n,

$$\frac{\Delta_n}{N}\mathbb{E}\Big[\sup_\theta \sum_{k=1}^2 \|\nabla_{\theta_k}S_n^N(\theta)\|\Big] \le C.$$

We have

$$\nabla_{\theta_k}S_n^N(\theta) = \sum_{i=1}^N\sum_{j=1}^n \zeta_{j,k}^i(\theta), \qquad k=1,2,$$

where

$$\zeta_{j,1}^i(\theta) = -\frac{2(X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1))}{c_{t_{j-1}}^i(\theta_2)}\nabla_{\theta_1}b_{t_{j-1}}^i(\theta_1),$$

$$\zeta_{j,2}^i(\theta) = -\frac{(X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1))^2}{\Delta_n (c^2)_{t_{j-1}}^i(\theta_2)}\nabla_{\theta_2}c_{t_{j-1}}^i(\theta_2) + \frac{1}{c_{t_{j-1}}^i(\theta_2)}\nabla_{\theta_2}c_{t_{j-1}}^i(\theta_2).$$

It suffices to show that for all N,n and i,j,

$$\mathbb{E}\big[\sup_\theta \|\zeta_{j,k}^i(\theta)\|\big] \le C, \qquad k=1,2. \tag{II.6.20}$$

Using IIA3 and the Cauchy-Schwarz inequality, we get

$$\mathbb{E}\big[\sup_\theta \|\zeta_{j,1}^i(\theta)\|\big] \le C\big(\mathbb{E}\big[\sup_{\theta_1}|X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1)|^2\big]\big)^{\frac12}\big(\mathbb{E}\big[\sup_{\theta_1}\|\nabla_{\theta_1}b_{t_{j-1}}^i(\theta_1)\|^2\big]\big)^{\frac12},$$

$$\begin{aligned}\mathbb{E}\big[\sup_\theta \|\zeta_{j,2}^i(\theta)\|\big] \le{}& \frac{C}{\Delta_n}\big(\mathbb{E}\big[\sup_{\theta_1}|X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1)|^4\big]\big)^{\frac12}\big(\mathbb{E}\big[\sup_{\theta_2}\|\nabla_{\theta_2}a_{t_{j-1}}^i(\theta_2)\|^2\big]\big)^{\frac12}\\ &+ C\mathbb{E}\big[\sup_{\theta_2}\|\nabla_{\theta_2}a_{t_{j-1}}^i(\theta_2)\|\big].\end{aligned}$$

We use polynomial growth of $\sup_{\theta_1}\|\nabla_{\theta_1}b(\theta_1,\cdot)\|$, $\sup_{\theta_2}\|\nabla_{\theta_2}a(\theta_2,\cdot)\|$ and moment bounds in Lemma II.5.1(1). Moreover, Lemma II.5.1(2) gives $\sup_{i,j}\mathbb{E}[|X_{t_j}^i - X_{t_{j-1}}^i|^4] = O(\Delta_n^2)$. Finally, $b(\theta_{0,1},\cdot)$ has a linear growth and the mean value theorem implies $b(\theta_1,\cdot)-b(\theta_{0,1},\cdot) = \int_0^1 \nabla_{\theta_1}b(\theta_{0,1}+(\theta_1-\theta_{0,1})u,\cdot)du\cdot(\theta_1-\theta_{0,1})$ for all θ_1 in Θ_1, where Θ_1 is convex, bounded and we recall that $\sup_{\theta_1}\|\nabla_{\theta_1}b(\theta_1,\cdot)\|$ has polynomial growth. The moment bounds in Lemma II.5.1(1) imply $\mathbb{E}[\sup_{\theta_1}|b_{t_{j-1}}^i(\theta_1)|^4] \le C$, completing the proof of (II.6.20). □

II.6.1.1 Proof of Theorem II.3.1

Proof. Assumption IIA5 implies that for every $\varepsilon>0$ there exists $\eta>0$ such that $J(\theta_2)-J(\theta_{0,2})>\eta$ for every θ_2 with $\|\theta_2-\theta_{0,2}\|\ge\varepsilon$. Thus $\{\|\hat\theta_{n,2}^N-\theta_{0,2}\|\ge\varepsilon\} \subseteq \{J(\hat\theta_{n,2}^N)-J(\theta_{0,2})>\eta\}$. The probability of the latter event converges to 0 in view of

$$J(\hat\theta_{n,2}^N)-J(\theta_{0,2}) = J_{n,0}^N + J_{n,1}^N,$$

where the definition of $\hat{\theta}_n^N$ and (II.6.1) imply respectively

$$J_{n,0}^N := \frac{\Delta_n}{N}(S_n^N(\hat{\theta}_{n,1}^N, \hat{\theta}_{n,2}^N) - S_n^N(\hat{\theta}_{n,1}^N, \theta_{0,2})) \leq 0,$$
$$J_{n,1}^N := J(\hat{\theta}_{n,2}^N) - J(\theta_{0,2}) - J_{n,0}^N \leq 2 \sup_{(\theta_1,\theta_2)\in\Theta} \left| \frac{\Delta_n}{N} S_n^N(\theta_1, \theta_2) - J(\theta_2) \right| = o_{\mathbb{P}}(1).$$

Consistency of $\hat{\theta}_{n,1}^N$ follows in a similar way. Assumption IIA5 implies that for every $\varepsilon > 0$ there exists $\eta > 0$ such that $I(\theta_1, \theta_2) > \eta$ for every (θ_1, θ_2) with $\|\theta_1 - \theta_{0,1}\| \geq \varepsilon$. Thus $\{\|\hat{\theta}_{n,1}^N - \theta_{0,1}\| \geq \varepsilon\} \subseteq \{I(\hat{\theta}_{n,1}^N, \hat{\theta}_{n,2}^N) > \eta\}$. The probability of the latter event converges to 0 because

$$I(\hat{\theta}_{n,1}^N, \hat{\theta}_{n,2}^N) = I_{n,0}^N + I_{n,1}^N,$$

where the definition of $\hat{\theta}_n^N$ and (II.6.2) imply respectively

$$I_{n,0}^N := \frac{1}{N}(S_n^N(\hat{\theta}_{n,1}^N, \hat{\theta}_{n,2}^N) - S_n^N(\theta_{0,1}, \hat{\theta}_{n,2}^N)) \leq 0,$$
$$I_{n,1}^N := I(\hat{\theta}_{n,1}^N, \hat{\theta}_{n,2}^N) - I_{n,0}^N = o_{\mathbb{P}}(1).$$

□

II.6.2 Asymptotic normality

The proof of the asymptotic normality of our estimator is obtained following a classical route. It consists in proving the asymptotic normality of the first derivative of the contrast function (II.2.1) (see for example [47, Section 5a]). We introduce in particular the appropriate normalization matrix

$$M_n^N := \operatorname{diag}\Big(\underbrace{\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}}}_{p_1 \text{ times}}, \underbrace{\sqrt{\frac{\Delta_n}{N}}, \ldots, \sqrt{\frac{\Delta_n}{N}}}_{p_2 \text{ times}}\Big).$$

The proof of Theorem II.3.2 is based on the following proposition.

Proposition II.6.2. *Assume IIA1- IIA4(I) and (II), IIA7. If $N\Delta_n \to 0$ then as $N, n \to \infty$,*

$$\nabla_\theta S_n^N(\theta_0) M_n^N \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\Sigma(\theta_0)),$$

where $\Sigma(\theta_0)$ is a $p \times p$ matrix defined in IIA6.

We observe that, as $\nabla_\theta S_n^N(\hat{\theta}_n^N) = 0$, by Taylor's formula we obtain

$$(\hat{\theta}_n^N - \theta_0) \int_0^1 \nabla_\theta^2 S_n^N(\theta_0 + s(\hat{\theta}_n^N - \theta_0))ds = -\nabla_\theta S_n^N(\theta_0). \tag{II.6.21}$$

Multiplying the equation (II.6.21) by M_n^N, we obtain

$$(\hat{\theta}_n^N - \theta_0)(M_n^N)^{-1} \int_0^1 \Sigma_n^N(\theta_0 + s(\hat{\theta}_n^N - \theta_0))ds = -\nabla_\theta S_n^N(\theta_0) M_n^N, \tag{II.6.22}$$

where

$$\Sigma_n^N(\theta) := M_n^N \nabla_\theta^2 S_n^N(\theta) M_n^N = \begin{pmatrix} \Sigma_n^{N,(1)}(\theta) & \Sigma_n^{N,(12)}(\theta) \\ \Sigma_n^{N,(21)}(\theta) & \Sigma_n^{N,(2)}(\theta) \end{pmatrix}$$

with

$$\Sigma_n^{N,(1)}(\theta) = (1/N)\nabla_{\theta_1}^2 S_n^N(\theta), \qquad \Sigma_n^{N,(12)}(\theta) = (\sqrt{\Delta_n}/N)\nabla_{\theta_1}\nabla_{\theta_2} S_n^N(\theta),$$
$$\Sigma_n^{N,(21)}(\theta) = (\sqrt{\Delta_n}/N)\nabla_{\theta_2}\nabla_{\theta_1} S_n^N(\theta), \qquad \Sigma_n^{N,(2)}(\theta) = (\Delta_n/N)\nabla_{\theta_2}^2 S_n^N(\theta).$$

The analysis of the second derivatives of the contrast function is gathered in the following proposition, which will be proven at the end of this section.

Proposition II.6.3. *Assume IIA1- IIA5 with both (I) and (II) in IIA4. Then as $N, n \to \infty$,*

1. $\Sigma_n^N(\theta_0) \xrightarrow{\mathbb{P}} \Sigma(\theta_0)$,
2. $\sup_{s\in[0,1]} \|\Sigma_n^N(\theta_0 + s(\hat{\theta}_n^N - \theta_0)) - \Sigma_n^N(\theta_0)\| \xrightarrow{\mathbb{P}} 0$, *where $\|\cdot\|$ refers to the operator norm on the space of $p \times p$ matrices induced by the Euclidean norm for vectors.*

By Proposition II.6.3 assumption IIA6 implies that the probability that $\int_0^1 \Sigma_n^N(\theta_0 + s(\hat{\theta}_n^N - \theta_0))ds$ is invertible tends to 1. Applying its inverse to the equation (II.6.22), by Proposition II.6.2 and the continuous mapping theorem, we get

$$\big(\sqrt{N}(\hat{\theta}_{n,1}^N - \theta_{0,1}), \sqrt{N/\Delta_n}(\hat{\theta}_{n,2}^N - \theta_{0,2})\big) = (\hat{\theta}_n^N - \theta_0)(M_n^N)^{-1} \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, 2(\Sigma(\theta_0))^{-1}\big).$$

II.6.3 Proof of Proposition II.6.2

Proof. As in the proof of consistency, we omit the notation for dependence on N, n. In particular, we write X_t^i for $X_t^{i,N}$, μ_t for μ_t^N, t_j for $t_{j,n}$. Denote by $f_{t_{j-1}}^i(\theta)$ the values of $f(\theta, X_{t_{j-1}}^i, \mu_{t_{j-1}})$. We note that $-\nabla_\theta S_n^N(\theta) M_n^N$ consists of $-\partial_{\theta_{1,h}} S_n^N(\theta)/\sqrt{N} =: \sum_{j=1}^n \xi_{j,h}^{(1)}(\theta)$ and $-\sqrt{\Delta_n/N}\partial_{\theta_{2,\tilde{h}}} S_n^N(\theta) =: \sum_{j=1}^n \xi_{j,\tilde{h}}^{(2)}(\theta)$, where

$$\xi_{j,h}^{(1)}(\theta) := \frac{1}{\sqrt{N}} \sum_{i=1}^N 2\frac{\partial_{\theta_{1,h}} b_{t_{j-1}}^i(\theta_1)}{c_{t_{j-1}}^i(\theta_2)} (X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1)),$$
$$\xi_{j,\tilde{h}}^{(2)}(\theta) := \sqrt{\frac{\Delta_n}{N}} \sum_{i=1}^N \frac{\partial_{\theta_{2,\tilde{h}}} c_{t_{j-1}}^i(\theta_2)}{\Delta_n (c_{t_{j-1}}^i(\theta_2))^2} (X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_1))^2 - \frac{\partial_{\theta_{2,\tilde{h}}} c_{t_{j-1}}^i(\theta_2)}{c_{t_{j-1}}^i(\theta_2)}$$

for $h = 1, \dots, p_1$, $\tilde{h} = 1, \dots, p_2$. To prove the asymptotic normality of $-\nabla_\theta S_n^N(\theta_0) M_n^N$ we want to use a central limit theorem for martingale difference arrays, in accordance with Theorems 3.2 and 3.4 of [57]. Approximation of $-\nabla_\theta S_n^N(\theta_0) M_n^N$ by a martingale array follows from

$$\sum_{j=1}^n \mathbb{E}_{t_{j-1}}[\xi_{j,h}^{(1)}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \qquad \sum_{j=1}^n \mathbb{E}_{t_{j-1}}[\xi_{j,\tilde{h}}^{(2)}(\theta_0)] \xrightarrow{\mathbb{P}} 0 \tag{II.6.23}$$

for $h = 1, \dots, p_1$, $\tilde{h} = 1, \dots, p_2$. Moreover, application of the central limit theorem requires that for some $r > 0$ the following convergences hold:

$$\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[\xi^{(1)}_{j,h_1}(\theta_0)\xi^{(1)}_{j,h_2}(\theta_0)] \xrightarrow{\mathbb{P}} 4\int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_{1,h_1}} b(\theta_{0,1}, x, \bar{\mu}_t)\partial_{\theta_{1,h_2}} b(\theta_{0,1}, x, \bar{\mu}_t)}{c(\theta_{0,2}, x, \bar{\mu}_t)} \bar{\mu}_t(dx)dt, \tag{II.6.24}$$

$$\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[\xi^{(2)}_{j,\tilde{h}_1}(\theta_0)\xi^{(2)}_{j,\tilde{h}_2}(\theta_0)] \xrightarrow{\mathbb{P}} 2\int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_{2,\tilde{h}_1}} c(\theta_{0,2}, x, \bar{\mu}_t)\partial_{\theta_{2,\tilde{h}_2}} c(\theta_{0,2}, x, \bar{\mu}_t)}{c^2(\theta_{0,2}, x, \bar{\mu}_t)} \bar{\mu}_t(dx)dt, \tag{II.6.25}$$

$$\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[\xi^{(1)}_{j,h}(\theta_0)\xi^{(2)}_{j,\tilde{h}}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{II.6.26}$$

$$\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[|\xi^{(1)}_{j,h}(\theta_0)|^{2+r}] \xrightarrow{\mathbb{P}} 0, \qquad \sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[|\xi^{(2)}_{j,\tilde{h}}(\theta_0)|^{2+r}] \xrightarrow{\mathbb{P}} 0, \tag{II.6.27}$$

where $h, h_1, h_2 = 1, \dots, p_1$, $\tilde{h}, \tilde{h}_1, \tilde{h}_2 = 1, \dots, p_2$.

• Proof of (II.6.23).
Assumptions IIA3 and IIA4(I) imply that $F^i_{j,h} := 2\partial_{\theta_{1,h}} b^i_{t_{j-1}}(\theta_{0,1})(c^i_{t_{j-1}}(\theta_{0,2}))^{-1}$ satisfies $|F^i_{j,h}| \le C(1 + |X^i_{t_{j-1}}|^{k_1} + W_2^{l_1}(\mu_{t_{j-1}}, \delta_0))$. Hence, from Lemma II.5.1(1) it is easy to see that $F^i_{j,h} = R^i_{t_{j-1}}(1)$. If $N\Delta_n \to 0$ then Lemma II.5.3(3) implies

$$\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[\xi^{(1)}_{j,h}(\theta_0)] = \frac{1}{N^{\frac{1}{2}}} \sum_{i=1}^{N}\sum_{j=1}^{n} R^i_{t_{j-1}}(1) R^i_{t_{j-1}}(\Delta_n^{\frac{3}{2}}) \xrightarrow{L^1} 0$$

and so the convergence in probability. In a similar way, using Lemma II.5.3(1), we obtain

$$\begin{aligned}
\sum_{j=1}^{n} \mathbb{E}_{t_{j-1}}[\xi^{(2)}_{j,\tilde{h}}(\theta_0)] &= \Big(\frac{\Delta_n}{N}\Big)^{\frac{1}{2}} \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{\partial_{\theta_{2,\tilde{h}}} c^i_{t_{j-1}}(\theta_{0,2})}{\Delta_n (c^i_{t_{j-1}}(\theta_{0,2}))^2}(\Delta_n c^i_{t_{j-1}}(\theta_{0,2}) + R^i_{t_{j-1}}(\Delta_n^2)) - \frac{\partial_{\theta_{2,\tilde{h}}} c^i_{t_{j-1}}(\theta_{0,2})}{c^i_{t_{j-1}}(\theta_{0,2})} \\
&= \Big(\frac{\Delta_n}{N}\Big)^{\frac{1}{2}} \sum_{i=1}^{N}\sum_{j=1}^{n} \frac{\partial_{\theta_{2,\tilde{h}}} c^i_{t_{j-1}}(\theta_{0,2})}{\Delta_n (c^i_{t_{j-1}}(\theta_{0,2}))^2} R^i_{t_{j-1}}(\Delta_n^2) \\
&= \Big(\frac{\Delta_n}{N}\Big)^{\frac{1}{2}} \sum_{i=1}^{N}\sum_{j=1}^{n} R^i_{t_{j-1}}(\Delta_n),
\end{aligned}$$

which converges to 0 in L^1 and so in probability if $N\Delta_n \to 0$.

• Proof of (II.6.24).
We have

$$\mathbb{E}_{t_{j-1}}[\xi^{(1)}_{j,h_1}(\theta_0)\xi^{(1)}_{j,h_2}(\theta_0)] = \frac{1}{N}\sum_{i_1,i_2=1}^{N} \mathbb{E}_{t_{j-1}}[(A^{i_1}_j + B^{i_1}_j)(A^{i_2}_j + B^{i_2}_j)] F^{i_1}_{j,h_1} F^{i_2}_{j,h_2}, \tag{II.6.28}$$

where

$$F^i_{j,h} := 2\frac{\partial_{\theta_{1,h}} b^i_{t_{j-1}}(\theta_{0,1})}{c^i_{t_{j-1}}(\theta_{0,2})} = R^i_{t_{j-1}}(1),$$

and

$$B^i_j := \int_{t_{j-1}}^{t_j} (b^i_s(\theta_{0,1}) - b^i_{t_{j-1}}(\theta_{0,1}))ds, \qquad A^i_j := \int_{t_{j-1}}^{t_j} a^i_s(\theta_{0,2})dW^i_s. \tag{II.6.29}$$

We have $\mathbb{E}_{t_{j-1}}[(B^i_j)^2] = R^i_{t_{j-1}}(\Delta_n^3)$ and $\mathbb{E}_{t_{j-1}}[(A^i_j)^2] = R^i_{t_{j-1}}(\Delta_n)$, whereas if $i_1 \neq i_2$ then $\mathbb{E}_{t_{j-1}}[A^{i_1}_j A^{i_2}_j] = 0$ because of the independence of Brownian motions. Hence, by the Cauchy-Schwarz inequality,

$$\mathbb{E}_{t_{j-1}}[(A^{i_1}_j + B^{i_1}_j)(A^{i_2}_j + B^{i_2}_j)] = \mathbb{E}_{t_{j-1}}[(A^{i_1}_j)^2]\mathbf{1}(i_1 = i_2) + R^{i_1,i_2}_{t_{j-1}}(\Delta_n^2).$$

We get

$$\sum_{j=1}^n \mathbb{E}_{t_{j-1}}[\xi^{(1)}_{j,h_1}(\theta_0)\xi^{(1)}_{j,h_2}(\theta_0)] = \frac{1}{N}\sum_{j=1}^n\sum_{i=1}^N \mathbb{E}_{t_{j-1}}[(A^i_j)^2]F^i_{j,h_1}F^i_{j,h_2} + \frac{1}{N}\sum_{j=1}^n\sum_{i_1,i_2=1}^N R^{i_1,i_2}_{t_{j-1}}(\Delta_n^2),$$

where the last sum converges to 0 in L^1 and so in probability if $N\Delta_n \to 0$. We can therefore focus on the first sum. We decompose the term $\mathbb{E}_{t_{j-1}}[(A^i_j)^2]$ into $\Delta_n c^i_{t_{j-1}}(\theta_{0,2})$ and

$$\mathbb{E}_{t_{j-1}}[(A^i_j)^2] - \Delta_n c^i_{t_{j-1}}(\theta_{0,2}) = \int_{t_{j-1}}^{t_j} \mathbb{E}_{t_{j-1}}[c^i_s(\theta_{0,2}) - c^i_{t_{j-1}}(\theta_{0,2})]ds = R^i_{t_{j-1}}(\Delta_n^{\frac{3}{2}}).$$

The result follows from $\Delta_n \to 0$ and application of Lemma II.5.2.

• Proof of (II.6.27), first convergence.
We want to show (II.6.27) with $r = 2$. We use the same notation as in (II.6.28) and consider the terms

$$\mathbb{E}_{t_{j-1}}[(A^{i_1}_j + B^{i_1}_j)(A^{i_2}_j + B^{i_2}_j)(A^{i_3}_j + B^{i_3}_j)(A^{i_4}_j + B^{i_4}_j)]F^{i_1}_{j,h}F^{i_2}_{j,h}F^{i_3}_{j,h}F^{i_4}_{j,h}. \tag{II.6.30}$$

We have $F^i_j = R^i_{t_{j-1}}(1)$, moreover, $\mathbb{E}_{t_{j-1}}[(A^i_j)^4] = R^i_{t_{j-1}}(\Delta_n^2)$, $\mathbb{E}_{t_{j-1}}[(B^i_j)^4] = R^i_{t_{j-1}}(\Delta_n^6)$ and so $\mathbb{E}_{t_{j-1}}[(A^i_j + B^i_j)^4] = R^i_{t_{j-1}}(\Delta_n^2)$. Application of the Cauchy-Schwarz inequality shows that the term in (II.6.30) is also $R^{i_1,i_2,i_3,i_4}_{t_{j-1}}(\Delta_n^2)$. In case where i_1, i_2, i_3, i_4 are pairwise distinct we decompose A^i_j into

$$A^i_{j,2} := \int_{t_{j-1}}^{t_j} (a^i_s(\theta_{0,2}) - a^i_{t_{j-1}}(\theta_{0,2}))dW^i_s, \qquad A^i_{j,1} := \int_{t_{j-1}}^{t_j} a^i_{t_{j-1}}(\theta_{0,2})dW^i_s, \tag{II.6.31}$$

which satisfy $\mathbb{E}_{t_{j-1}}[(A^i_{j,k})^4] = R^i_{t_{j-1}}(\Delta_n^{2k})$, $k = 1, 2$. In particular the independence of the Brownian motions implies

$$\mathbb{E}_{t_{j-1}}[A^{i_1}_{j,1}A^{i_2}_{j,1}A^{i_3}_{j,1}A^{i_4}_{j,k}]F^{i_1}_{j,h}F^{i_2}_{j,h}F^{i_3}_{j,h}F^{i_4}_{j,h} = 0$$

for $k = 1, 2$. The term converging to 0 at the slowest rate in (II.6.30) is then, up to a permutation of the indices i_1, i_2, i_3, i_4,

$$\mathbb{E}_{t_{j-1}}[A^{i_1}_{j,1}A^{i_2}_{j,1}A^{i_3}_{j,2}A^{i_4}_{j,2} + A^{i_1}_{j,1}A^{i_2}_{j,1}A^{i_3}_{j,1}B^{i_4}_j]F^{i_1}_{j,h}F^{i_2}_{j,h}F^{i_3}_{j,h}F^{i_4}_{j,h} = R^{i_1,i_2,i_3,i_4}_{t_{j-1}}(\Delta_n^3).$$

We get

$$\sum_{j=1}^{n}\mathbb{E}_{t_{j-1}}[(\xi^{(1)}_{j,h}(\theta_0))^4] = \frac{1}{N^2}\sum_{j=1}^{n}\Big(\sum_{i\in I}R^i_{t_{j-1}}(\Delta_n^3) + \sum_{i\in I^c}R^i_{t_{j-1}}(\Delta_n^2)\Big), \tag{II.6.32}$$

where I denotes a set of all $i = (i_1, i_2, i_3, i_4) \in \{1, \ldots, N\}^4$ such that i_1, i_2, i_3, i_4 are pairwise distinct. We note that $\mathrm{card}(I) = O(N^4)$ and $\mathrm{card}(I^c) = O(N^3)$. We conclude that (II.6.32) converges to 0 in L^1 and so in probability if $N\Delta_n \to 0$.

• Proof of (II.6.25).
We rewrite the left hand side of (II.6.25) as

$$\frac{\Delta_n}{N}\sum_{j=1}^{n}\sum_{i_1,i_2=1}^{N}\Delta_n^{-2}C^{i_1}_{j,\tilde{h}_1}C^{i_2}_{j,\tilde{h}_2}\mathbb{E}_{t_{j-1}}[D^{i_1}_jD^{i_2}_j], \tag{II.6.33}$$

where

$$C^i_{j,\tilde{h}} := \frac{\partial_{\theta_{2,\tilde{h}}}c^i_{t_{j-1}}(\theta_{0,2})}{(c^i_{t_{j-1}}(\theta_{0,2}))^2} = R^i_{t_{j-1}}(1), \qquad D^i_j := (X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1}))^2 - \Delta_n c^i_{t_{j-1}}(\theta_{0,2}).$$

We consider the term $\mathbb{E}_{t_{j-1}}[D^{i_1}_jD^{i_2}_j]$ in (II.6.33). By Lemma II.5.3(1) it equals

$$\begin{aligned}&\mathbb{E}_{t_{j-1}}[(X^{i_1}_{t_j} - X^{i_1}_{t_{j-1}} - \Delta_n b^{i_1}_{t_{j-1}}(\theta_{0,1}))^2(X^{i_2}_{t_j} - X^{i_2}_{t_{j-1}} - \Delta_n b^{i_2}_{t_{j-1}}(\theta_{0,1}))^2] \\ &\quad - \Delta_n c^{i_1}_{t_{j-1}}(\theta_{0,2})\Delta_n c^{i_2}_{t_{j-1}}(\theta_{0,2}) + R^{i_1,i_2}_{t_{j-1}}(\Delta_n^3).\end{aligned} \tag{II.6.34}$$

If $i_1 = i_2$ then Lemma II.5.3(2) implies

$$\mathbb{E}_{t_{j-1}}[(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1}))^4] = 3\Delta_n^2(c^i_{t_{j-1}}(\theta_{0,2}))^2 + R^i_{t_{j-1}}(\Delta_n^{\frac{5}{2}}),$$

whence

$$\mathbb{E}_{t_{j-1}}[(D^i_j)^2] = 2\Delta_n^2(c^i_{t_{j-1}}(\theta_{0,2}))^2 + R^i_{t_{j-1}}(\Delta_n^{\frac{5}{2}}). \tag{II.6.35}$$

If $i_1 \neq i_2$ then to deal with the term in (II.6.34) we decompose

$$X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1}) = A^i_{j,1} + A^i_{j,2} + B^i_j$$

as in (II.6.29), (II.6.31), where $\mathbb{E}_{t_{j-1}}[(A^i_{j,k})^4] = R^i_{t_{j-1}}(\Delta_n^{2k})$, $k = 1, 2$, and $\mathbb{E}_{t_{j-1}}[(B^i_j)^4] = R^i_{t_{j-1}}(\Delta_n^6)$. We note that

$$\mathbb{E}_{t_{j-1}}[(A^{i_1}_{j,1})^2(A^{i_2}_{j,1})^2] = \Delta_n c^{i_1}_j(\theta_{0,2})\Delta_n c^{i_2}_j(\theta_{0,2}).$$

Moreover, we have

$$\mathbb{E}_{t_{j-1}}[(A^{i_1}_{j,1})^2 A^{i_2}_{j,1} A^{i_2}_{j,2}] = c^{i_1}_{t_{j-1}}(\theta_{0,2}) a^{i_2}_{t_{j-1}}(\theta_{0,2}) V^{i_1,i_2}_j,$$

where independence of Brownian motions together with Itô isometry implies

$$\begin{aligned} V^{i_1,i_2}_j &:= \mathbb{E}_{t_{j-1}}\Big[(W^{i_1}_{t_j} - W^{i_1}_{t_{j-1}})^2 \int_{t_{j-1}}^{t_j} dW^{i_2}_s \int_{t_{j-1}}^{t_j} (a^{i_2}_s(\theta_{0,2}) - a^{i_2}_{t_{j-1}}(\theta_{0,2})) dW^{i_2}_s\Big] \\ &= \int_{t_{j-1}}^{t_j} \mathbb{E}_{t_{j-1}}[(W^{i_1}_{t_j} - W^{i_1}_{t_{j-1}})^2 (a^{i_2}_t(\theta_{0,2}) - a^{i_2}_{t_{j-1}}(\theta_{0,2}))] dt. \end{aligned} \tag{II.6.36}$$

Assumption IIA7 allows us to apply Itô's lemma to $a^{i_2}_t(\theta_{0,2})$. We get that the conditional expectation in (II.6.36) equals

$$\begin{aligned} &\mathbb{E}_{t_{j-1}}\Big[(W^{i_1}_{t_j} - W^{i_1}_{t_{j-1}})^2 \int_{t_{j-1}}^{t} \sum_{k=1}^{N} \Big(b^k_s(\theta_{0,1}) \partial_{x_k} a^{i_2}_s(\theta_{0,2}) + \frac{1}{2} c^k_s(\theta_{0,2}) \partial^2_{x_k} a^{i_2}_s(\theta_{0,2})\Big) ds\Big] \\ &\quad + \mathbb{E}_{t_{j-1}}\Big[(W^{i_1}_{t_j} - W^{i_1}_{t_{j-1}})^2 \int_{t_{j-1}}^{t} \sum_{k=1}^{N} a^k_s(\theta_{0,2}) \partial_{x_k} a^{i_2}_s(\theta_{0,2}) dW^k_s\Big]. \end{aligned}$$

The first term is clearly a $R^{i_1,i_2}_{t_{j-1}}(\Delta_n^2)$ function. Regarding the second one, for $k \neq i_1$, the independence of the Brownian motions makes it directly equal to 0. For $k = i_1$, instead, we have

$$\mathbb{E}_{t_{j-1}}\Big[(W^{i_1}_{t_j} - W^{i_1}_{t_{j-1}})^2 \int_{t_{j-1}}^{t} a^{i_1}_s(\theta_{0,2}) \partial_{x_{i_1}} a^{i_2}_s(\theta_{0,2}) dW^{i_1}_s\Big],$$

where under IIA7 we obtain

$$\partial_{x_{i_1}} a^{i_2}_s(\theta_{0,2}) := \partial_y \tilde{a}\Big(X^{i_2}_s, \frac{1}{N} \sum_{l=1}^{N} K(X^{i_2}_s, X^l_s)\Big) \frac{1}{N} \partial_y K(X^{i_2}_s, X^{i_1}_s)$$

with $\partial_y \tilde{a}$, $\partial_y K$ having polynomial growth. Using the Cauchy-Schwarz inequality, it follows that the above quantity is upper bounded by

$$\begin{aligned} &\Big(3\Delta_n^2 \mathbb{E}_{t_{j-1}}\Big[\Big(\int_{t_{j-1}}^{t} a^{i_1}_s(\theta_{0,2}) \partial_{x_{i_1}} a^{i_2}_s(\theta_{0,2}) dW^{i_1}_s\Big)^2\Big]\Big)^{\frac{1}{2}} \\ &\quad = \Big(3\Delta_n^2 \int_{t_{j-1}}^{t} \mathbb{E}_{t_{j-1}}[(a^{i_1}_s(\theta_{0,2}) \partial_{x_{i_1}} a^{i_2}_s(\theta_{0,2}))^2] ds\Big)^{\frac{1}{2}} = \frac{1}{N} R^{i_1,i_2}_{t_{j-1}}(\Delta_n^{\frac{3}{2}}). \end{aligned}$$

It implies

$$\mathbb{E}_{t_{j-1}}[(A^{i_1}_{j,1})^2 A^{i_2}_{j,1} A^{i_2}_{j,2}] = R^{i_1,i_2}_{t_{j-1}}(\Delta_n^3) + \frac{1}{N} R^{i_1,i_2}_{t_{j-1}}(\Delta_n^{\frac{5}{2}}). \tag{II.6.37}$$

We conclude that

$$\mathbb{E}_{t_{j-1}}[(X^{i_1}_{t_j} - X^{i_1}_{t_{j-1}} - \Delta_n b^{i_1}_{t_{j-1}}(\theta_{0,1}))^2 (X^{i_2}_{t_j} - X^{i_2}_{t_{j-1}} - \Delta_n b^{i_2}_{t_{j-1}}(\theta_{0,1}))^2]$$

$$= \Delta_n c_j^{i_1}(\theta_{0,2})\Delta_n c_j^{i_2}(\theta_{0,2}) + R_{t_{j-1}}^{i_1,i_2}(\Delta_n^3) + \frac{1}{N} R_{t_{j-1}}^{i_1,i_2}(\Delta_n^{\frac{5}{2}}),$$

whence

$$\mathbb{E}_{t_{j-1}}[D_j^{i_1} D_j^{i_2}] = R_{t_{j-1}}^{i_1,i_2}(\Delta_n^3) + \frac{1}{N} R_{t_{j-1}}^{i_1,i_2}(\Delta_n^{\frac{5}{2}}) \tag{II.6.38}$$

if $i_1 \neq i_2$. Finally, we plug (II.6.35), (II.6.38) back into (II.6.33), where use of the conditions $N\Delta_n \to 0$, $\Delta_n \to 0$ and Lemma II.5.2 completes the proof of the convergence in (II.6.25).

• Proof of (II.6.27), second convergence.
We prove it for $r = 2$. We use the same notation as in (II.6.33) and rewrite the left hand side of (II.6.27) as

$$\frac{\Delta_n^2}{N^2} \sum_{j=1}^{n} \sum_{i_1,i_2,i_3,i_4=1}^{N} \Delta_n^{-4} C_{j,\tilde{h}}^{i_1} C_{j,\tilde{h}}^{i_2} C_{j,\tilde{h}}^{i_3} C_{j,\tilde{h}}^{i_4} \mathbb{E}_{t_{j-1}}[D_j^{i_1} D_j^{i_2} D_j^{i_3} D_j^{i_4}]. \tag{II.6.39}$$

We have $\mathbb{E}_{t_{j-1}}[(D_j^i)^4] = R_{t_{j-1}}^i(\Delta_n^4)$ and $\mathrm{card}(I^c) = O(N^3)$, where I denotes a set of all $i = (i_1, i_2, i_3, i_4) \in \{1, \dots, N\}^4$ such that i_1, i_2, i_3, i_4 are pairwise distinct. In (II.6.39) the sum over $i \in I^c$ converges to 0 in L^1 and so in probability since $N\Delta_n \to 0$. In case $i \in I$ we use the decomposition

$$\begin{aligned} D_j^i &= (A_{j,1}^i + A_{j,2}^i + B_j^i)^2 - \Delta_n c_{t_{j-1}}^i(\theta_{0,2}) \\ &= (A_{j,2}^i + B_j^i)(2A_{j,1}^i + A_{j,2}^i + B_j^i) + (A_{j,1}^i)^2 - \Delta_n c_{t_{j-1}}^i(\theta_{0,2}). \end{aligned}$$

We note that

$$\mathbb{E}_{t_{j-1}}[((A_{j,1}^i)^2 - \Delta_n c_{t_{j-1}}^i(\theta_{0,2}))^4] = R_{t_{j-1}}^i(\Delta_n^4),$$
$$\mathbb{E}_{t_{j-1}}[(A_{j,k}^i)^8] = R_{t_{j-1}}^i(\Delta_n^{4k}),\ k = 1, 2, \qquad \mathbb{E}_{t_{j-1}}[(B_j^i)^8] = R_{t_{j-1}}^i(\Delta_n^{12}).$$

Moreover, because of the independence of Brownian motions, we have

$$\mathbb{E}_{t_{j-1}}\Big[\prod_{k=1}^{4}((A_{j,1}^{i_k})^2 - \Delta_n c_{t_{j-1}}^{i_k}(\theta_{0,2}))\Big] = 0$$

and in a similar manner as in (II.6.37) under IIA7 we have

$$\begin{aligned} &\mathbb{E}_{t_{j-1}}\Big[A_{j,2}^{i_1} A_{j,1}^{i_1} \prod_{k=2}^{4}((A_{j,1}^{i_k})^2 - \Delta_n c_{t_{j-1}}^{i_k}(\theta_{0,2}))\Big] \\ &= a_{t_{j-1}}^{i_1}(\theta_{0,2}) \prod_{k=2}^{4} c_{t_{j-1}}^{i_k}(\theta_{0,2}) \int_{t_{j-1}}^{t_j} \mathbb{E}_{t_{j-1}}\Big[(a_s^{i_1}(\theta_{0,2}) - a_{t_{j-1}}^{i_1}(\theta_{0,2})) \prod_{l=2}^{4}((W_{t_j}^{i_l} - W_{t_{j-1}}^{i_l})^2 - \Delta_n)\Big] ds \\ &= R_{t_{j-1}}^{i_1,i_2,i_3,i_4}(\Delta_n^5) + \frac{1}{N} R_{t_{j-1}}^{i_1,i_2,i_3,i_4}(\Delta_n^{\frac{9}{2}}), \end{aligned}$$

whence it follows

$$\mathbb{E}_{t_{j-1}}[D_j^{i_1} D_j^{i_2} D_j^{i_3} D_j^{i_4}] = R_{t_{j-1}}^{i_1,i_2,i_3,i_4}(\Delta_n^5) + \frac{1}{N} R_{t_{j-1}}^{i_1,i_2,i_3,i_4}(\Delta_n^{\frac{9}{2}}).$$

We recall that $\text{card}(I) = O(N^4)$. Since $N\Delta_n \to 0$, $\Delta_n \to 0$, the sum over $i \in I$ in (II.6.39) converges to 0 in L^1 and so in probability.

• Proof of (II.6.26).
We rewrite the left hand side of (II.6.26) as

$$\frac{\Delta_n^{\frac{1}{2}}}{N}\sum_{j=1}^{n}\sum_{i_1,i_2=1}^{N}\mathbb{E}_{t_{j-1}}[(A_{j,1}^{i_1}+A_{j,2}^{i_1}+B_j^{i_1})D_j^{i_2}]\Delta_n^{-1}C_{j,\tilde{h}}^{i_2}F_{j,h}^{i_1}, \tag{II.6.40}$$

where

$$D_j^i = (A_{j,2}^i + B_j^i)(2A_{j,1}^i + A_{j,2}^i + B_j^i) + (A_{j,1}^i)^2 - \Delta_n c_{t_{j-1}}^i(\theta_{0,2})$$

with the notations introduced above. We recall that $F_{j,h}^i = R_{t_{j-1}}^i(1)$, $C_{j,\tilde{h}}^i = R_{t_{j-1}}^i(1)$, $\mathbb{E}_{t_{j-1}}[(B_j^i)^4] = R_{t_{j-1}}^i(\Delta_n^6)$, $\mathbb{E}_{t_{j-1}}[(A_{j,k}^i)^4] = R_{t_{j-1}}^i(\Delta_n^{2k})$, $k = 1,2$, and so $\mathbb{E}_{t_{j-1}}[((A_{j,1}^i)^2 - \Delta_n c_{t_{j-1}}^i(\theta_{0,2}))^2] = R_{t_{j-1}}^i(\Delta_n^2)$. We note that

$$\mathbb{E}_{t_{j-1}}[A_{j,1}^{i_1}((A_{j,1}^{i_2})^2 - \Delta_n c_{t_{j-1}}^{i_2}(\theta_{0,2}))] = 0$$

for all i_1, i_2. This is a consequence of the independence of the Brownian motions for $i_1 \neq i_2$, while for $i_1 = i_2$ it derives from the fact that the odd moments are centered. Hence, in case $i_1 = i_2 = i$ the term $\mathbb{E}_{t_{j-1}}[(A_{j,1}^i)^2 A_{j,2}^i]$ makes the main contribution to

$$\mathbb{E}_{t_{j-1}}[(A_{j,1}^i + A_{j,2}^i + B_j^i)D_j^i] = R_{t_{j-1}}^i(\Delta_n^2).$$

Now we can see that the sum over $i_1 = i_2$ in (II.6.40) converges to 0 in L^1 and so in probability. In case $i_1 \neq i_2$ we have

$$\mathbb{E}_{t_{j-1}}[A_{j,2}^{i_1}((A_{j,1}^{i_2})^2 - \Delta_n c_{t_{j-1}}^{i_2}(\theta_{0,2}))] = 0.$$

Moreover,

$$\mathbb{E}_{t_{j-1}}[A_{j,1}^{i_1}A_{j,1}^{i_2}A_{j,2}^{i_2}] = a_{t_{j-1}}^{i_1}(\theta_{0,2})a_{t_{j-1}}^{i_2}(\theta_{0,2})\int_{t_{j-1}}^{t_j}\mathbb{E}_{t_{j-1}}[(W_{t_j}^{i_1} - W_{t_{j-1}}^{i_1})(a_s^{i_2}(\theta_{0,2}) - a_{t_{j-1}}^{i_2}(\theta_{0,2})]ds.$$

The application of Itô's lemma to $a_s^{i_2}(\theta_{0,2})$ under IIA7 similarly as in the proof of (II.6.37) provides

$$\mathbb{E}_{t_{j-1}}[A_{j,1}^{i_1}A_{j,1}^{i_2}A_{j,2}^{i_2}] = R_{t_{j-1}}^{i_1,i_2}(\Delta_n^{\frac{5}{2}}) + \frac{1}{N}R_{t_{j-1}}^{i_1,i_2}(\Delta_n^2).$$

We conclude that

$$\mathbb{E}_{t_{j-1}}[(A_{j,1}^{i_1} + A_{j,2}^{i_1} + B_j^{i_1})D_j^{i_2}] = R_{t_{j-1}}^{i_1,i_2}(\Delta_n^{\frac{5}{2}}) + \frac{1}{N}R_{t_{j-1}}^{i_1,i_2}(\Delta_n^2).$$

in case $i_1 \neq i_2$. Hence, the sum over $i_1 \neq i_2$ in (II.6.40) converges to 0 in L^1 and so in probability when $N\Delta_n \to 0$, $\Delta_n \to 0$. This concludes the proof of the asymptotic normality of $-\nabla_\theta S_n^N(\theta_0)M_n^N$. □

II.6.4 Proof of Proposition II.6.3

Proof. The proof relies on the computation of the second derivatives of the contrast function. We have that, for any $k, l = 1, \dots, p_1$,

$$\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} S_n^N(\theta) = 2\sum_{i=1}^N \sum_{j=1}^n \Big\{ \Delta_n \frac{\partial_{\theta_{1,k}} b^i_{t_{j-1}}(\theta_1) \partial_{\theta_{1,l}} b^i_{t_{j-1}}(\theta_1)}{c^i_{t_{j-1}}(\theta_2)} - \frac{\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} b^i_{t_{j-1}}(\theta_1)}{c^i_{t_{j-1}}(\theta_2)} (X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1)) \Big\},$$

where the last factor can further be decomposed into $\Delta_n(b^i_{t_{j-1}}(\theta_{0,1}) - b^i_{t_{j-1}}(\theta_1))$ and $X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1})$. We can see that $\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} S_n^N(\theta)/N$ converges to

$$\Sigma^{(1)}_{kl}(\theta) := 2\int_0^1 \int_{\mathbb{R}} \Big\{ \frac{\partial_{\theta_{1,k}} b(\theta_1, x, \bar\mu_t) \partial_{\theta_{1,l}} b(\theta_1, x, \bar\mu_t)}{c(\theta_2, x, \bar\mu_t)} - \frac{\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} b(\theta_1, x, \bar\mu_t)}{c(\theta_2, x, \bar\mu_t)} (b(\theta_{0,1}, x, \bar\mu_t) - b(\theta_1, x, \bar\mu_t)) \Big\} \bar\mu_t(dx)dt \tag{II.6.41}$$

uniformly in θ in probability. Indeed, the proof follows along the lines of the proof of (II.6.1). We refer to Steps 3, 4 of the proof of Lemma II.6.1, where in (II.6.5) in $I_n^N(\theta)$, $\rho_n^N(\theta)$ it is enough to replace the functions $h(\theta, \cdot)$ and $g(\theta, \cdot)$ with the integrand of (II.6.41) and $\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} b(\theta_1, \cdot)/c(\theta_2, \cdot)$ respectively, and to check them for the respective conditions. We note that both functions have polynomial growth. Moreover, the integrand in (II.6.41) is locally Lipschitz continuous, which allows us to apply Lemma II.5.2 and yields the convergence in probability of the sequence $\partial_{\theta_{1,k}}\partial_{\theta_{1,l}} S_n^N(\theta)/N$ for every θ. To get tightness in $(C(\Theta; \mathbb{R}), \|\cdot\|_\infty)$, we use that uniformly in θ the partial derivatives with respect to $\theta_{i',j'}$, $j' = 1, \dots, p_{i'}$, $i' = 1, 2$, of the two functions have polynomial growth.

In the same way as above we get that for any $k = 1, \dots, p_1$, $l = 1, \dots, p_2$, once multiplied by $\sqrt{\Delta_n}/N$,

$$\partial_{\theta_{1,k}}\partial_{\theta_{2,l}} S_n^N(\theta) = 2\sum_{i=1}^N \sum_{j=1}^n \frac{\partial_{\theta_{1,k}} b^i_{t_{j-1}}(\theta_1) \partial_{\theta_{2,l}} c^i_{t_{j-1}}(\theta_2)}{(c^i_{t_{j-1}}(\theta_2))^2} (X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1)),$$

converges to 0 uniformly in θ in probability.

Finally, we have that for any $k, l = 1, \dots, p_2$,

$$\partial_{\theta_{2,k}}\partial_{\theta_{2,l}} S_n^N(\theta) = \sum_{i=1}^N \sum_{j=1}^n \Big\{ \frac{\partial_{\theta_{2,k}}\partial_{\theta_{2,l}} c^i_{t_{j-1}}(\theta_2) c^i_{t_{j-1}}(\theta_2) - \partial_{\theta_{2,k}} c^i_{t_{j-1}}(\theta_2) \partial_{\theta_{2,l}} c^i_{t_{j-1}}(\theta_2)}{(c^i_{t_{j-1}}(\theta_2))^2} + \frac{2\partial_{\theta_{2,k}} c^i_{t_{j-1}}(\theta_2) \partial_{\theta_{2,l}} c^i_{t_{j-1}}(\theta_2) - \partial_{\theta_{2,k}}\partial_{\theta_{2,l}} c^i_{t_{j-1}}(\theta_2) c^i_{t_{j-1}}(\theta_2)}{\Delta_n (c^i_{t_{j-1}}(\theta_2))^3} \times (X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1))^2 \Big\},$$

where the last factor can further be decomposed into $(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1))^2 - \Delta_n c^i_{t_{j-1}}(\theta_{0,2})$ and $\Delta_n c^i_{t_{j-1}}(\theta_{0,2})$. We note that $(\Delta_n/N)\partial_{\theta_{2,k}}\partial_{\theta_{2,l}} S^N_n(\theta)$ converges to

$$\begin{aligned}\Sigma^{(2)}_{kl}(\theta) := \int_0^T \int_{\mathbb{R}} \Big\{ & \frac{\partial_{\theta_{2,k}}\partial_{\theta_{2,l}} c(\theta_2, x, \bar{\mu}_t) c(\theta_2, x, \bar{\mu}_t) - \partial_{\theta_{2,k}} c(\theta_2, x, \bar{\mu}_t)\partial_{\theta_{2,l}} c(\theta_2, x, \bar{\mu}_t)}{c(\theta_2, x, \bar{\mu}_t)^2} \\ & + \frac{2\partial_{\theta_{2,k}} c(\theta_2, x, \bar{\mu}_t)\partial_{\theta_{2,l}} c(\theta_2, x, \bar{\mu}_t) - \partial_{\theta_{2,k}}\partial_{\theta_{2,l}} c(\theta_2, x, \bar{\mu}_t) c(\theta_2, x, \bar{\mu}_t)}{c(\theta_2, x, \bar{\mu}_t)^3} \\ & \times c(\theta_{0,2}, x, \bar{\mu}_t) \Big\} \bar{\mu}_t(dx)dt\end{aligned}$$

uniformly in θ in probability. We will prove the uniform in θ convergence to the second term of $\Sigma^{(2)}_{kl}(\theta)$ only:

$$\sum_{j=1}^n \chi^N_{n,j}(\theta) \xrightarrow{\mathbb{P}} \tilde{\Sigma}^{(2)}_{kl}(\theta) := \int_0^T \int_{\mathbb{R}} \tilde{f}(\theta_2, x, \bar{\mu}_t) c(\theta_{0,2}, x, \bar{\mu}_t)\bar{\mu}_t(dx)dt, \tag{II.6.42}$$

where

$$\chi^N_{n,j}(\theta) = \frac{1}{N}\sum_{i=1}^N \tilde{f}^i_{t_{j-1}}(\theta_2)(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1))^2$$

and function $\tilde{f} : \Theta_2 \times \mathbb{R} \times \mathcal{P} \to \mathbb{R}$ is given by $(2(\partial_{\theta_{2,k}} c)(\partial_{\theta_{2,l}} c) - (\partial_{\theta_{2,k}}\partial_{\theta_{2,l}} c)c)/c^3$. For every θ the convergence in (II.6.42) follows from

$$\sum_{j=1}^n \mathbb{E}_{t_{j-1}}[\chi^N_{n,j}(\theta)] \xrightarrow{\mathbb{P}} \tilde{\Sigma}^{(2)}_{kl}(\theta), \qquad \sum_{j=1}^n \mathbb{E}_{t_{j-1}}[(\chi^N_{n,j}(\theta))^2] \xrightarrow{\mathbb{P}} 0$$

by [47, Lemma 9]. Indeed, the above relations hold, because by Lemma II.5.3(1),

$$\mathbb{E}_{t_{j-1}}[\chi^N_{n,j}(\theta)] = \frac{1}{N}\sum_{i=1}^N \tilde{f}^i_{t_{j-1}}(\theta_2)(\Delta_n c^i_{t_{j-1}}(\theta_{0,2}) + R^i_{t_{j-1}}(\Delta_n^{3/2})),$$

by Jensen's inequality and Lemma II.5.3(2),

$$\mathbb{E}_{t_{j-1}}[(\chi^N_{n,j}(\theta))^2] \le \frac{1}{N}\sum_{i=1}^N (\tilde{f}^i_{t_{j-1}}(\theta_2))^2 R^i_{t_{j-1}}(\Delta_n^2),$$

by polynomial growth of $\partial^{i'}_{\theta_{2,j'}} c(\theta_2, \cdot)$, $i' = 0, 1, 2$, $j' = 1, \ldots, p_2$, IIA3 and Point 1. of Lemma II.5.1,

$$(\tilde{f}^i_{t_{j-1}}(\theta_2))^2 = R^i_{t_{j-1}}(1).$$

The tightness in $(C(\Theta;\mathbb{R}), \|\cdot\|_\infty)$ follows from $\mathbb{E}[\sup_\theta \|\nabla_\theta \sum_{j=1}^n \chi^N_{n,j}(\theta)\|] = O(1)$. Indeed, we have

$$\nabla_{\theta_1}\chi^N_{n,j}(\theta) = -2\frac{\Delta_n}{N}\sum_{i=1}^N \nabla_{\theta_1} b^i_{t_{j-1}}(\theta_1)\tilde{f}^i_{t_{j-1}}(\theta_2)(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1)),$$

$$\nabla_{\theta_2}\chi_{n,j}^N(\theta) = \frac{1}{N}\sum_{i=1}^N \nabla_{\theta_2}\tilde{f}^i_{t_{j-1}}(\theta_2)(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1))^2,$$

where by polynomial growth of $\sup_{\theta_1}\|\nabla_{\theta_1} b(\theta_1,\cdot)\|$, $\sup_{\theta_2}|\partial^{i'}_{\theta_{2,j'}} c(\theta_2,\cdot)|$, $i'=0,1,2,3$, $j'=1,\dots,p_2$, and IIA3,

$$\sup_\theta \|\nabla_{\theta_1} b^i_{t_{j-1}}(\theta_1)\tilde{f}^i_{t_{j-1}}(\theta_2)\| = R^i_{t_{j-1}}(1), \qquad \sup_{\theta_2}\|\nabla_{\theta_2}\tilde{f}^i_{t_{j-1}}(\theta_2)\| = R^i_{t_{j-1}}(1)$$

and

$$\sup_{\theta_1}|X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_1)| \le |X^i_{t_j} - X^i_{t_{j-1}}| + \Delta_n \sup_{\theta_1}|b^i_{t_{j-1}}(\theta_1)|$$

with $\sup_{\theta_1}|b^i_{t_{j-1}}(\theta_1)| = R^i_{t_{j-1}}(1)$. Finally, we have $\mathbb{E}[|X^i_{t_j} - X^i_{t_{j-1}}|^4] \le C\Delta_n^2$ uniformly in i,j and N,n by Lemma II.5.1(2).
We conclude that the matrix $\Sigma_n^N(\theta)$ converges to $\Sigma(\theta) = \operatorname{diag}(\Sigma^{(1)}(\theta), \Sigma^{(2)}(\theta))$ uniformly in θ and so at $\theta=\theta_0$ in probability. Hence,

$$\|\Sigma_n^N(\theta_0 + s(\hat\theta_n^N - \theta_0)) - \Sigma_n^N(\theta_0)\| \le o_{\mathbb{P}}(1) + \|\Sigma(\theta_0 + s(\hat\theta_n^N - \theta_0)) - \Sigma(\theta_0)\|,$$

where the uniform convergence in probability (in s) of the last term to 0 follows from continuity of $\Sigma(\theta)$ at $\theta=\theta_0$ and consistency of the estimator sequence $\hat\theta_n^N$. □

II.7 Proof of technical results

II.7.1 Proof of Lemma II.5.1

Proof. Proof of Lemma II.5.1(1).
We have, for any $i=1,\dots,N$, $0\le t\le T$, $p\ge 2$,

$$\begin{aligned}\mathbb{E}[|X^i_t|^p] &\le \mathbb{E}\Big[\Big|X^i_0 + \int_0^t b^i_u(\theta_{0,1})du + \int_0^t a^i_u(\theta_{0,2})dW^i_u\Big|^p\Big] \\ &\le C\Big(\mathbb{E}[|X^i_0|^p] + t^{p-1}\int_0^t \mathbb{E}[|b^i_u(\theta_{0,1})|^p]du + t^{\frac{p}{2}-1}\int_0^t \mathbb{E}[|a^i_u(\theta_{0,2})|^p]du\Big),\end{aligned}$$

where we have used the Burkholder-Davis-Gundy and Jensen inequalities. We observe that, as a consequence of the lipschitzianity gathered in IIA2, for the true value of the parameter both coefficients are upper bounded by $C(1+|X^i_u| + W_2(\mu_u,\delta_0))$. Due to Jensen's inequality, we have

$$\mathbb{E}[W_2^p(\mu_u,\delta_0)] \le \frac{1}{N}\sum_{j=1}^N \mathbb{E}[|X^j_u|^p] = \mathbb{E}[|X^i_u|^p].$$

The last identity follows from the fact that the particles are equally distributed. We obtain

$$\mathbb{E}[|X^i_t|^p] \le C\Big(\mathbb{E}[|X^i_0|^p] + (t^{p-1} + t^{\frac{p}{2}-1})\Big(t + 2\int_0^t \mathbb{E}[|X^i_u|^p]du\Big)\Big). \qquad \text{(II.7.1)}$$

We infer by Gronwall's lemma that

$$\mathbb{E}[|X_t^i|^p] \leq C(\mathbb{E}[|X_0^i|^p] + T^p + T^{\frac{p}{2}}) \exp(C'(T^p + T^{\frac{p}{2}})).$$

As the constants do not depend on $t \leq T$ and $\mathbb{E}[|X_0^i|^p] < \infty$ by IIA1, we have the wanted result for $p \geq 2$. Then, by a Jensen argument and the boundedness of the moments for $p \geq 2$, it follows the result also for $p < 2$.

Proof of Lemma II.5.1(2).
We have for any $0 \leq s < t \leq T$, $p \geq 2$,

$$\begin{aligned}\mathbb{E}[|X_t^i - X_s^i|^p] &= \mathbb{E}\Big[\Big|\int_s^t b_u^i(\theta_{0,1})du + \int_s^t a_u^i(\theta_{0,2})dW_u^i\Big|^p\Big] \\ &\leq C\Big((t-s)^{p-1}\int_s^t \mathbb{E}[|b_u^i(\theta_{0,1})|^p]du + (t-s)^{\frac{p}{2}-1}\int_s^t \mathbb{E}[|a_u^i(\theta_{0,2})|^p]ds\Big),\end{aligned}$$

where we have used the Jensen and Burkholder-Davis-Gundy inequalities. Because of (II.1.4) and the just shown Lemma II.5.1(1), the result follows letting $t - s \leq 1$.

Proof of Lemma II.5.1(3).
According to the definition of $R_s^i(1)$, we want to evaluate the L^q norm of $\mathbb{E}_s[|X_t^i - X_s^i|^p]$. For any $0 \leq s < t \leq T$ such that $t - s \leq 1$ and $p \geq 2$, $q \geq 1$,

$$\mathbb{E}\big[\big|\mathbb{E}_s[|X_t^i - X_s^i|^p]\big|^q\big]^{\frac{1}{q}} \leq \mathbb{E}[|X_t^i - X_s^i|^{pq}]^{\frac{1}{q}} \leq C(t-s)^{\frac{p}{2}}$$

follows by conditional Jensen's inequality and Lemma II.5.1(2).

Proof of Lemma II.5.1(4).
This is a straightforward consequence of

$$W_2^p(\mu_t, \mu_s) \leq \Big(\frac{1}{N}\sum_{j=1}^N |X_t^j - X_s^j|^2\Big)^{\frac{p}{2}} \leq \frac{1}{N}\sum_{j=1}^N |X_t^j - X_s^j|^p \tag{II.7.2}$$

by Jensen's inequality for any $0 \leq s < t \leq T$ such that $t - s \leq 1$, $p \geq 2$ and Lemma II.5.1(2).

Proof of Lemma II.5.1(5).
It follows directly from (II.7.2), where we use Minkowski's inequality as follows:

$$\mathbb{E}\big[\big|\mathbb{E}_s[W_2^p(\mu_t, \mu_s)]\big|^q\big]^{\frac{1}{q}} \leq \frac{1}{N}\sum_{j=1}^N \mathbb{E}\big[\big|\mathbb{E}_s[|X_t^j - X_s^j|^{pq}]\big|\big]^{\frac{1}{q}},$$

and then Lemma II.5.1(3). □

II.7.2 Proof of Lemma II.5.2

Proof. Step 1. We prove that

$$\frac{\Delta_n}{N}\sum_{i=1}^{N}\sum_{j=1}^{n} f(X^{i,N}_{t_{j-1,n}},\mu^N_{t_{j-1,n}}) - \frac{1}{N}\sum_{i=1}^{N}\int_0^T f(X^{i,N}_s,\mu^N_s)ds \xrightarrow{L^1} 0.$$

Here we note $\Delta_n = t_{j,n} - t_{j-1,n}$ and decompose the above integral into integrals over $[t_{j-1,n}, t_{j,n})$. We can see that the above convergence follows from

$$\sum_{j=1}^{n}\int_{t_{j-1,n}}^{t_{j,n}} \mathbb{E}[|f(X^{i,N}_{t_{j-1,n}},\mu^N_{t_{j-1,n}}) - f(X^{i,N}_s,\mu^N_s)|]ds \to 0, \qquad N,n\to\infty,$$

for fixed i, which in turn follows using the condition (II.5.1), Cauchy-Schwarz inequality and moment bounds in Lemma II.5.1(1), (2) and (4). In particular, $\mathbb{E}[|X^{i,N}_{t_{j-1,n}} - X^{i,N}_s|^2] \le C\Delta_n$ for all $t_{j-1,n}\le s\le t_{j,n}$, j and n,N.

Step 2. Next, let us prove that

$$\frac{1}{N}\sum_{i=1}^{N}\int_0^T f(X^{i,N}_s,\mu^N_s)ds - \frac{1}{N}\sum_{i=1}^{N}\int_0^T f(\bar{X}^i_s,\bar{\mu}_s)ds \xrightarrow{L^1} 0, \qquad N\to\infty,$$

where each $(\bar{X}^i_t)_{t\in[0,T]}$ satisfies (II.1.3) with $(W_t)_{t\in[0,T]} = (W^i_t)_{t\in[0,T]}$ and $\bar{X}^i_0 = X^{i,N}_0$. It suffices to prove

$$\int_0^T \mathbb{E}[|f(X^{i,N}_s,\mu^N_s) - f(\bar{X}^i_s,\bar{\mu}_s)|]ds \to 0,$$

where i is fixed and the integral is over a bounded interval. For this purpose, let us use again the condition (II.5.1) and the Cauchy-Schwarz inequality. Following the same arguments as in the proof of Lemma II.5.1(1) and Gronwall lemma, it is easy to show that for all $p>0$ there exists $C_p>0$ such that for all s,i,N it holds $\mathbb{E}[|\bar{X}^i_s|^p] < C_p$. Moreover we have

$$\mathbb{E}[|X^{i,N}_s - \bar{X}^i_s|^2] \le \frac{C}{\sqrt{N}}$$

for all $0\le s\le T$ and i,N, thanks to Theorem 3.20 in [20], based on Theorem 1 of [44]. We remark that, from the boundedness of the moments, the quantity q appearing in the statement of Theorem 3.20 in [20] is larger than 4. Hence, the rate $N^{-(q-2)/q}$ is negligible compared to $N^{-1/2}$. The propagation of chaos stated above implies

$$\mathbb{E}[W_2^2(\mu^N_s,\bar{\mu}_s)] \le \frac{C}{\sqrt{N}}.$$

Indeed, to get the last relation, we introduce the empirical measure $\bar{\mu}^N_s = N^{-1}\sum_{i=1}^N \delta_{\bar{X}^i_s}$ of the independent particle system at time s and use the triangle inequality for W_2. Then

$$\mathbb{E}[W_2^2(\mu^N_s,\bar{\mu}^N_s)] \le \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[|X^{i,N}_s - \bar{X}^i_s|^2] \le \frac{C}{\sqrt{N}},$$

whereas Theorem 1 of [44] implies

$$\mathbb{E}[W_2^2(\bar{\mu}_s^N, \bar{\mu}_s)] \leq \frac{C}{\sqrt{N}}.$$

Step 3. Finally, the law of large numbers gives

$$\frac{1}{N}\sum_{i=1}^{N}\int_0^T f(\bar{X}_s^i, \bar{\mu}_s)ds \xrightarrow{\mathbb{P}} \mathbb{E}\Big[\int_0^T f(\bar{X}_s, \bar{\mu}_s)ds\Big], \qquad N \to \infty.$$

□

II.7.3 Proof of Lemma II.5.3

Proof. We use the same notation as before.

Proof of Lemma II.5.3(2). We decompose $X_{t_j}^i - X_{t_{j-1}}^i - \Delta_n b_{t_{j-1}}^i(\theta_{0,1})$ into $A_{j,1}^i$ and $H_{j,2}^i := A_{j,2}^i + B_j^i$, where

$$\begin{gathered} A_{j,1}^i := \int_{t_{j-1}}^{t_j} a_{t_{j-1}}^i(\theta_{0,2})dW_s^i, \qquad A_{j,2}^i := \int_{t_{j-1}}^{t_j} (a_s^i(\theta_{0,2}) - a_{t_{j-1}}^i(\theta_{0,2}))dW_s^i, \\ B_j^i := \int_{t_{j-1}}^{t_j} (b_s^i(\theta_{0,1}) - b_{t_{j-1}}^i(\theta_{0,1}))ds, \end{gathered} \tag{II.7.3}$$

are the same as in (II.6.29), (II.6.31).
Firstly, we will show that for any $p \geq 2$,

$$\mathbb{E}[|H_{j,2}^i|^p] \leq C\Delta_n^p. \tag{II.7.4}$$

Using Jensen's inequality and Lipschitz continuity of $b(\theta_1, \cdot)$ we get

$$\begin{aligned} \mathbb{E}[|B_j^i|^p] &\leq \mathbb{E}\Big[\Delta_n^{p-1}\int_{t_{j-1}}^{t_j} |b_s^i(\theta_{0,1}) - b_{t_{j-1}}^i(\theta_{0,1})|^p ds\Big] \\ &\leq C\Delta_n^{p-1}\int_{t_{j-1}}^{t_j} (\mathbb{E}[|X_s^i - X_{t_{j-1}}^i|^p] + \mathbb{E}[W_2^p(\mu_s, \mu_{t_{j-1}})])ds \\ &\leq C\Delta_n^{p-1}\int_{t_{j-1}}^{t_j} (s - t_{j-1})^{\frac{p}{2}}ds = C\Delta_n^{\frac{3}{2}p}, \end{aligned} \tag{II.7.5}$$

where the last inequality follows from Lemma II.5.1(2) and (4). Further use of the Burkholder-Davis-Gundy and Jensen inequalities gives

$$\begin{aligned} \mathbb{E}[|A_{j,2}^i|^p] &\leq C\mathbb{E}\Big[\Big(\int_{t_{j-1}}^{t_j} |a_s^i(\theta_{0,2}) - a_{t_{j-1}}^i(\theta_{0,2})|^2 ds\Big)^{\frac{p}{2}}\Big] \\ &\leq C\Delta_n^{\frac{p}{2}-1}\int_{t_{j-1}}^{t_j} \mathbb{E}[|a_s^i(\theta_{0,2}) - a_{t_{j-1}}^i(\theta_{0,2})|^p]ds \\ &\leq C\Delta_n^p, \end{aligned} \tag{II.7.6}$$

where the last inequality follows from Lipschitz continuity of $a(\theta_2, \cdot)$ and Lemma II.5.1(2) and (4) as so does (II.7.5). Hence, we have shown (II.7.4).
Next, we have

$$\mathbb{E}[|A^i_{j,1}|^p] = C\Delta_n^{\frac{p}{2}}\mathbb{E}[|a^i_{t_{j-1}}(\theta_{0,2})|^p] \leq C\Delta_n^{\frac{p}{2}} \tag{II.7.7}$$

since we know the absolute moments of a centered normal distribution and have linear growth of $a(\theta_{0,2}, \cdot)$, moment bounds in Lemma II.5.1(1). In particular, we note

$$\mathbb{E}_{t_{j-1}}[(A^i_{j,1})^4] = 3\Delta_n^2(c^2)^i_{t_{j-1}}(\theta_{0,2}).$$

Finally, we have

$$\mathbb{E}_{t_{j-1}}[(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1}))^4] = 3\Delta_n^2(c^2)^i_{t_{j-1}}(\theta_{0,2}) + \sum_{k=0}^{3}\binom{4}{k}\mathbb{E}_{t_{j-1}}[(A^i_{j,1})^k(H^i_{j,2})^{4-k}]. \tag{II.7.8}$$

For any $k = 0, 1, 2, 3$ and $q \geq 1$, using Jensen's inequality for conditional expectation, we get

$$\mathbb{E}\big[\big|\mathbb{E}_{t_{j-1}}[(A^i_{j,1})^k(H^i_{j,2})^{4-k}]\big|^q\big] \leq \mathbb{E}\big[\big|(A^i_{j,1})^k(H^i_{j,2})^{4-k}\big|^q\big] \leq C\Delta_n^{(4-\frac{k}{2})q},$$

where the last inequality follows from (II.7.6), (II.7.4) using Cauchy-Schwarz inequality. Hence, the term converging to 0 in L^q at the slowest rate is the one for which $k = 3$. We therefore obtain that the remaining sum on the right hand side of (II.7.8) is an $R^i_{t_{j-1}}(\Delta_n^{\frac{5}{2}})$ function.

Proof of Lemma II.5.3(3). This follows directly from (II.7.5) by decomposing the dynamics of X^i as in (II.7.3) and remarking that the stochastic integral is centered.

Proof of Lemma II.5.3(1). We decompose $X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1})$ into

$$A^i_j := A^i_{j,1} + A^i_{j,2} = \int_{t_{j-1}}^{t_j} a^i_s(\theta_{0,2})dW^i_s,$$

and B^i_j satisfying respectively $\mathbb{E}[|A^i_j|^{2p}] \leq C\Delta_n^p$ and $\mathbb{E}[|B^i_j|^{2p}] \leq C\Delta_n^{3p}$, whence $\mathbb{E}[|A^i_jB^i_j|^p] \leq C\Delta_n^{2p}$ for any $p \geq 1$, see (II.7.5)-(II.7.7). We conclude that

$$\mathbb{E}_{t_{j-1}}[(X^i_{t_j} - X^i_{t_{j-1}} - \Delta_n b^i_{t_{j-1}}(\theta_{0,1}))^2] = \int_{t_{j-1}}^{t_j}\mathbb{E}_{t_{j-1}}[c^i_s(\theta_{0,2})]ds + R^i_{t_{j-1}}(\Delta_n^2).$$

We are left to show that we can replace $\mathbb{E}_{t_{j-1}}[c^i_s(\theta_{0,2})]$ with $c^i_{t_{j-1}}(\theta_{0,2})$ and that the remaining integral is an $R^i_{t_{j-1}}(\Delta_n^2)$ function.
Under IIA7 we have that for any i,

$$(x_1, \ldots, x_N) \mapsto c\Big(\theta_{0,2}, x_i, \frac{1}{N}\sum_{j=1}^{N}\delta_{x_j}\Big) = \tilde{a}^2\Big(x_i, \frac{1}{N}\sum_{j=1}^{N}K(x_i, x_j)\Big) =: g^i(x_1, \ldots, x_N)$$

is a twice continuously differentiable function from $\mathbb{R}^N$ to $\mathbb{R}$. Given a vector $(X_s^1, \dots, X_s^N)_{s\in[0,T]}$ of processes, we denote

$$(\partial_{x_k}^l c)_s^i(\theta_{0,2}) := \partial_{x_k}^l g^i(X_s^1, \dots, X_s^N).$$

We apply the multidimensional Itô's formula to $g^i(X_s^1, \dots, X_s^N) = c_s^i(\theta_{0,2})$ as follows:

$$\begin{aligned} c_s^i(\theta_{0,2}) - c_{t_{j-1}}^i(\theta_{0,2}) = &\sum_{k=1}^N \int_{t_{j-1}}^s \Big((\partial_{x_k} c)_u^i(\theta_{0,2}) b_u^k(\theta_{0,1}) + \frac{1}{2}(\partial_{x_k}^2 c)_u^i(\theta_{0,2}) c_u^k(\theta_{0,2})\Big) du \\ &+ \sum_{k=1}^N \int_{t_{j-1}}^s (\partial_{x_k} c)_u^i(\theta_{0,2}) a_u^k(\theta_{0,2}) dW_u^k. \end{aligned}$$

Since the driving $(W_u^1, \dots, W_u^N)_{u\in[t_{j-1},s]}$ is independent of $\mathcal{F}_{t_{j-1}}^N$, it follows that

$$\begin{aligned} &\mathbb{E}_{t_{j-1}}[c_s^i(\theta_{0,2})] - c_{t_{j-1}}^i(\theta_{0,2}) \\ &= \mathbb{E}_{t_{j-1}}\Big[\sum_{k=1}^N \int_{t_{j-1}}^s \Big((\partial_{x_k} c)_u^i(\theta_{0,2}) b_u^k(\theta_{0,1}) + \frac{1}{2}(\partial_{x_k}^2 c)_u^i(\theta_{0,2}) c_u^k(\theta_{0,2})\Big) du\Big]. \qquad \text{(II.7.9)} \end{aligned}$$

To conclude, we need to bound each $(\partial_{x_k}^l c)_u^i(\theta_{0,2})$, $l = 1, 2$. To do that, we rely on the assumption about the dependence of the diffusion coefficient on the convolution with a probability measure gathered in IIA7. To compute the derivatives with respect to x_k we need to consider two different cases, depending on whether $k \neq i$ or $k = i$. When $k \neq i$ we have $(\partial_{x_k} c)_u^i(\theta_{0,2}) = 2a_u^i(\theta_{0,2})(\partial_{x_k} a)_u^i(\theta_{0,2})$, where

$$(\partial_{x_k} a)_u^i(\theta_{0,2}) := \partial_y \tilde{a}\Big(X_u^i, \frac{1}{N}\sum_{j=1}^N K(X_u^i, X_u^j)\Big)\frac{1}{N}\partial_y K(X_u^i, X_u^k), \qquad \text{(II.7.10)}$$

while for $k = i$ we have $(\partial_{x_i} c)_u^i(\theta_{0,2}) = 2a_u^i(\theta_{0,2})(\partial_{x_i} a)_u^i(\theta_{0,2})$, where

$$\begin{aligned} (\partial_{x_i} a)_u^i(\theta_{0,2}) := &\partial_x \tilde{a}\Big(X_u^i, \frac{1}{N}\sum_{j=1}^N K(X_u^i, X_u^j)\Big) + \partial_y \tilde{a}\Big(X_u^i, \frac{1}{N}\sum_{j=1}^N K(X_u^i, X_u^j)\Big) \\ &\times\Big(\frac{1}{N}\sum_{j=1}^N \partial_x K(X_u^i, X_u^j) + \frac{1}{N}\partial_y K(X_u^i, X_u^i)\Big). \end{aligned}$$

From polynomial growth of the l-th order partial derivatives of $K, \tilde{a}$ for $l = 0, 1$, that of $b(\theta_{0,1}, \cdot)$, moment bounds in Lemma II.5.1(1) applying Jensen's inequality it follows that $\sum_{k=1}^N (\partial_{x_k} c)_u^i(\theta_{0,2}) b_u^k(\theta_{0,1})$ is bounded in L^p for any $p \geq 1$ uniformly in u, i. We proceed similarly to compute $(\partial_{x_k}^2 c)_u^i(\theta_{0,2})$. Then from polynomial growth of the l-th order partial derivatives of $K, \tilde{a}$ for $l = 0, 1, 2$, moment bounds in Lemma II.5.1(1) applying Jensen's inequality it follows that $\sum_{k=1}^N (\partial_{x_k}^2 c)_u^i(\theta_{0,2}) c_u^k(\theta_{0,2})$ is bounded in L^p for any $p \geq 1$ uniformly in u, i. For any $p \geq 1$, $t_{j-1} \leq s \leq t_j$, repeatedly applying Jensen's inequality to (II.7.9) we get

$$\mathbb{E}\big[\big|\mathbb{E}_{t_{j-1}}[c_s^i(\theta_{0,2})] - c_{t_{j-1}}^i(\theta_{0,2})\big|^p\big] \leq C(s - t_{j-1})^p,$$

whence

$$\mathbb{E}\Big[\Big|\int_{t_{j-1}}^{t_j}\big(\mathbb{E}_{t_{j-1}}[c_s^i(\theta_{0,2})] - c_{t_{j-1}}^i(\theta_{0,2})\big)ds\Big|^p\Big] \leq C\Delta_n^{2p}.$$

which completes the proof. □

Chapter III

Local asymptotic normality for discretely observed McKean-Vlasov diffusions

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.

III.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 [8,15,21,36,42,50,54,80]. 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:

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

where the unknown parameter $\theta := (\theta_1, \theta_2)$ is an element of the set $\Theta = \Theta_1 \times \Theta_2$, and $\Theta_k \subset \mathbb{R}, k = 1, 2$ 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

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

are the drift and diffusion coefficients, 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

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

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

$$\Big(X_{t_j}^{i,\theta}\Big)_{j=1,...,n}^{i=1,...,N}, \tag{III.1.2}$$

where $t_j := Tj/n$ and $\Delta_n := T/n$ denotes the discretization step. We consider the asymptotic regime $\Delta_n \to 0$, $N \to \infty$, 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 $h \in \mathbb{R}^p$:

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

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. [56, 70, 71]).

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 [7, 12, 22, 24, 48, 49, 63, 73, 88, 92]. Non-parametric approaches have also gained increasing attention, with notable contributions including [4, 10, 25, 29, 85].
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 [52,53], enables a stochastic expansion of the log-likelihood ratio and ultimately yields the desired LAN result. In contrast to Gobet's framework [53], 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 [7], 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 [30], 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 III.2 we introduce some notations, formulate the assumptions and state some technical lemmas we will use in order to show our main results. Section III.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 III.4.

III.2 Notation and assumptions

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

III.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 $\mathrm{tr}(A)$. For a function $f : \mathbb{R}^d \times \Theta_k \to \mathbb{R}$, we denote by $\nabla_x f$ (resp. $\partial_{\theta_k} f$) the derivative with respect to x (resp. with respect to θ_k). A function $f : \mathbb{R}^d \times \mathcal{P}_l \to \mathbb{R}^d$ 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

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

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

$$\begin{aligned} m^\theta_{t_j,t_{j+1}}(x) &:= \mathbb{E}^\theta_{t_j,x}\left[X^{i,\theta}_{t_{j+1}}\right] \in \mathbb{R}^d, \\ V^\theta_{t_j,t_{j+1}}(x) &:= \mathbb{E}^\theta_{t_j,x}\left[\left(X^{i,\theta}_{t_{j+1}} - m^\theta_{t_j,t_{j+1}}(x)\right)^{\otimes 2}\right] \in \mathbb{R}^{d\times d}. \end{aligned} \tag{III.2.1}$$

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

$$Y_t^i = R_t^i(\varepsilon), \qquad \varepsilon = \varepsilon(n, N), \tag{III.2.2}$$

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).

III.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 (III.1.1):

IIIA1. *The initial distribution μ_0 is sub-Gaussian, i.e. there exists a constant $\sigma > 0$ such that*

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

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

IIIA2. *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 IIIA2 guarantees existence and uniqueness of the McKean-Vlasov SDE (III.1.1). As for the regularity of the drift $b : \Theta_1 \times \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d$ 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 $\mathcal{P}_2 \to \mathbb{R}$, will be useful in our setting. We refer in particular to [29, Section 2] and the references therein.

Definition III.2.1. A mapping $f : \mathcal{P}_2 \to \mathbb{R}^d$ is said to have a linear functional derivative, if there exists $\partial_\mu f : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d$ 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 $g : \mathbb{R}^d \to \mathbb{R}^d$ is a μ-integrable function, we simply have $\partial_\mu f(y, \mu) = g(y)$.

IIIA3. *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 $C > 0$, $k, l = 0, 1, \ldots$ such that for all $r_1 + r_2 = 1, 2$ $(x, \mu), (x', \mu') \in \mathbb{R}^d \times \mathcal{P}_2$,

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

(III) The map $\mu \mapsto b_{\theta_1}(x, \mu)$ admits a functional derivative in the sense of Definition III.2.1. There exists $k, k', l \geq 1$ such that

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

IIIA4. *For every point θ in the interior of Θ, the coefficient functions are twice differentiable in x. Furthermore, the following estimates hold:*

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

(b) $\|g(\cdot,x)\| \leq C(1+\|x\|^q)$ *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.

IIIA5. *(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 IIIA3-IIIA5 are direct analogues of conditions imposed in [52,53], which in particular imply uniform lower and upper bounds for the transition density established in Proposition III.4.2.

III.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

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

where $h_i \in \mathbb{H}$ and $f : \mathbb{R}^k \to \mathbb{R}$ 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

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

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

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

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 $\mathrm{Dom}(\delta)$, consists of all processes $u \in L^2([0,T]\times\Omega;\mathbb{H})$ such that
$$\forall F \in \mathbb{D}^{1,2}, \quad |\mathbb{E}[\langle \mathcal{D}F, u\rangle_{\mathbb{H}}]| \le c(u)\|F\|_{L^2(\Omega)}.$$

(ii) For $u \in \mathrm{Dom}(\delta)$, the Skorohod integral $\delta(u) \in L^2(\Omega)$ is defined by the *integration by parts formula*
$$\forall F \in \mathbb{D}^{1,2}, \quad \mathbb{E}[F\,\delta(u)] = \mathbb{E}[\langle \mathcal{D}F, u\rangle_{\mathbb{H}}]. \tag{III.2.3}$$

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 III.2.2. *Let $p > 1$. Then the following assertions hold:*

(i) If $u \in \mathbb{D}^{1,p}(\mathbb{H})$, then $u \in \mathrm{Dom}(\delta)$ and
$$\|\delta(u)\|_p \le c_p \left(\|u\|_{L^p(\Omega;\mathbb{H})} + \|\mathcal{D}u\|_{L^p(\Omega;\mathbb{H}\otimes\mathbb{H})}\right). \tag{III.2.4}$$

(ii) Let $F \in \mathbb{D}^{1,2}$ and $u \in \mathrm{Dom}(\delta)$ 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
$$\delta(Fu) = F\delta(u) - \langle \mathcal{D}F, u\rangle_{\mathbb{H}}. \tag{III.2.5}$$

III.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 III.4.2 and [53, 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
$$(\theta_1^+, \theta_2^+) := \left(\theta_1^0 + \frac{u}{\sqrt{N}},\ \theta_2^0 + \frac{v}{\sqrt{N/\Delta_n}}\right), \qquad \theta^+ := (\theta_1^+, \theta_2^+), \tag{III.3.1}$$

and exploit the Markov property of the processes $(X_t^{\theta,i})_{t\ge 0}$, $i = 1,\dots,N$, to express the log-likelihood ratio between the measures $\mathbb{P}^{\theta^+}$ and $\mathbb{P}^{\theta^0}$ as
$$z(\theta^0,\theta^+) := \log \frac{d\mathbb{P}^{\theta^+}}{d\mathbb{P}^{\theta^0}}(X_{t_j})_{j=1,\dots,n} = \sum_{j=1}^{n}\sum_{i=1}^{N} \log\left(\frac{p^{\theta^+}}{p^{\theta^0}}\right)\left(t_j, t_{j+1}, X_{t_j}^i, X_{t_{j+1}}^i\right). \tag{III.3.2}$$

From the expansion (III.3.2), we obtain the identity
$$z(\theta^0,\theta^+) = \sum_{j=1}^{n}\sum_{i=1}^{N} \left(\zeta_j^{i,\theta_1} + \zeta_j^{i,\theta_2}\right), \tag{III.3.3}$$

where the individual components are given by

$$\zeta_j^{i,\theta_1} := \frac{u}{\sqrt{N}} \int_0^1 \frac{\partial_{\theta_1} p^{\theta_1(l),\theta_2^+}}{p^{\theta_1(l),\theta_2^+}} \left(t_j, t_{j+1}, X^i_{t_j}, X^i_{t_{j+1}}\right) dl,$$
$$\zeta_j^{i,\theta_2} := \frac{v}{\sqrt{N/\Delta_n}} \int_0^1 \frac{\partial_{\theta_2} p^{\theta_1^0,\theta_2(l)}}{p^{\theta_1^0,\theta_2(l)}} \left(t_j, t_{j+1}, X^i_{t_j}, X^i_{t_{j+1}}\right) dl. \tag{III.3.4}$$

Here, we use the notation

$$\theta_1(l) := \theta_1^0 + \frac{lu}{\sqrt{N}}, \qquad \theta_2(l) := \theta_2^0 + \frac{lv}{\sqrt{N/\Delta_n}}. \tag{III.3.5}$$

In the next step, we present an expansion formula for the transition density expressions appearing in (III.3.4). To this end, we introduce the notation ($t \in [0, \Delta_n]$)

$$X^{i,\theta}_{t_j+t} = X^{i,\theta}_{t_j} + \int_0^t b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, \mu^\theta_{t_j+s}\big) ds + \int_0^t a_{\theta_2}\big(X^{i,\theta}_{t_j+s}\big) dB^i_s,$$

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

$$\begin{aligned}
\partial_{\theta_1} X_t^{i,\theta} &= \int_0^t \Big(\partial_{\theta_1} b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, \mu^\theta_{t_j+s}\big) + \nabla_x b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, \mu^\theta_{t_j+s}\big)\, \partial_{\theta_1} X_s^{i,\theta} \\
&\quad + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, y, \mu^\theta_{t_j+s}\big)\, \partial_{\theta_1} \mu^\theta_{t_j+s}(dy)\Big)\, ds + \sum_{r=1}^d \int_0^t \nabla_x a^r_{\theta_2}\big(X^{i,\theta}_{t_j+s}\big)\, \partial_{\theta_1} X_s^{i,\theta}\, dB^i_{r,s}, \\
\partial_{\theta_2} X_t^{i,\theta} &= \int_0^t \Big(\nabla_x b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, \mu^\theta_{t_j+s}\big)\, \partial_{\theta_2} X_s^{i,\theta} + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, y, \mu^\theta_{t_j+s}\big)\, \partial_{\theta_2} \mu^\theta_{t_j+s}(dy)\Big)\, ds \\
&\quad + \sum_{r=1}^d \int_0^t \Big(\partial_{\theta_2} a^r_{\theta_2}\big(X^{i,\theta}_{t_j+s}\big) + \nabla_x a^r_{\theta_2}\big(X^{i,\theta}_{t_j+s}\big)\, \partial_{\theta_2} X_s^{i,\theta}\Big)\, dB^i_{r,s}.
\end{aligned} \tag{III.3.6}$$

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 III.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

$$Y_t^{i,\theta} := \mathbb{I}_d + \int_0^t \nabla_x b_{\theta_1}\big(X^{i,\theta}_{t_j+s}, \mu^\theta_{t_j+s}\big) Y_s^{i,\theta}\, ds + \sum_{r=1}^d \int_0^t \nabla_x a^r_{\theta_2}\big(X^{i,\theta}_{t_j+s}\big) Y_s^{i,\theta}\, dB_s^{r,i}, \tag{III.3.7}$$

where $a^r_{\theta_2}$ denotes the rth column of the matrix-valued function a_{θ_2}. It can be shown that $Y^{i,\theta}_t$ is invertible for all $t \geq 0$, and the Malliavin derivative satisfies the following identity (see [86, Eq. (2.59)]):

$$\mathcal{D}_s X^{i,\theta}_{t_j+t} = Y^{i,\theta}_t (Y^{i,\theta}_s)^{-1} a_{\theta_2}(X^{i,\theta}_{t_j+s})\ \mathbb{1}_{[0,t]}(s). \tag{III.3.8}$$

Our first result provides an identity for the terms ζ^{i,θ_1}_j and ζ^{i,θ_2}_j introduced in (III.3.4).

Proposition III.3.1. *Assume that Assumptions IIIA1–IIIA5 hold. Then, for $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$, the transition density $p^\theta(t_j, t_{j+1}, x, y)$ is absolutely continuous with respect to θ, and we have the representation*

$$\begin{aligned} \frac{\partial_{\theta_1} p^\theta}{p^\theta}(t_j, t_{j+1}, x, y) &= \frac{1}{\Delta_n} \mathbb{E}^\theta_{t_j,x}\left[\sum_{r=1}^d \delta\left(\partial_{\theta_1} X^{i,\theta}_{r,\Delta_n} U^i_r\right) \middle| X^{i,\theta}_{t_{j+1}} = y \right], \\ \frac{\partial_{\theta_2} p^\theta}{p^\theta}(t_j, t_{j+1}, x, y) &= \frac{1}{\Delta_n} \mathbb{E}^\theta_{t_j,x}\left[\sum_{r=1}^d \delta\left(\partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} U^i_r\right) \middle| X^{i,\theta}_{t_{j+1}} = y \right], \end{aligned} \tag{III.3.9}$$

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

$$U^i_s := \left(\mathcal{D}_s X^{i,\theta}_{t_{j+1}}\right)^{-1} = a^{-1}_{\theta_2}(X^{i,\theta}_{t_j+s})\, Y^{i,\theta}_s \left(Y^{i,\theta}_{\Delta_n}\right)^{-1}, \qquad s \in [0, \Delta_n].$$

The application of Malliavin calculus to derive the transformation in (III.3.9) is essential for establishing the LAN property of the statistical model. This formula was originally shown by Gobet (see, e.g., [52, 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^θ_t on the parameter θ.
The next result applies Proposition III.3.1 to derive an asymptotic expansion of the log-likelihood function $z(\theta^0, \theta^+)$.

Proposition III.3.2. *Assume that Assumptions IIIA1–IIIA5 hold. Define the quantities*

$$z^\theta_t(x) := \partial_{\theta_1} b_{\theta_1}(x, \mu^\theta_t) + \int_{\mathbb{R}^d} \partial_\mu b_{\theta_1}(x, y, \mu^\theta_t)\, \partial_{\theta_1} \mu^\theta_t(dy) \tag{III.3.10}$$

and

$$\widehat{\zeta}^{i,\theta_1}_j := \frac{u}{\sqrt{N}} \int_0^1 z^{\theta_1(l),\theta_2^+}_{t_j}(X^i_{t_j})^\top \left[a^{-2}_{\theta_2^+}(X^i_{t_j}) \left(X^i_{t_{j+1}} - m^{\theta_1(l),\theta_2^+}_{t_j,t_{j+1}}(X^i_{t_j}) \right) \right] dl,$$

$$\begin{aligned} \widehat{\zeta}^{i,\theta_2}_j := \frac{v}{\sqrt{N\Delta_n}} \int_0^1 \operatorname{tr}\Big[&\partial_{\theta_2} a_{\theta_2(l)}(X^i_{t_j})\, a^{-1}_{\theta_2(l)}(X^i_{t_j}) \\ &\times \left(\left(X^i_{t_{j+1}} - m^{\theta_1^0,\theta_2(l)}_{t_j,t_{j+1}}(X^i_{t_j}) \right)^{\otimes 2} - V^{\theta_1^0,\theta_2(l)}_{t_j,t_{j+1}}(X^i_{t_j}) \right) a^{-2}_{\theta_2(l)}(X^i_{t_j}) \Big]\, dl. \end{aligned}$$

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

$$\sum_{j=1}^{n} \sum_{i=1}^{N} \left(\zeta_j^{i,\theta_k} - \widehat{\zeta}_j^{i,\theta_k} \right) \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

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

$$\begin{cases} dX_t^{i,N,\theta} = b_{\theta_1}\left(X_t^{i,N,\theta}, \mu_t^N\right)dt + a_{\theta_2}\left(X_t^{i,N,\theta}\right)dW_t^i, \qquad i = 1, \dots, N, \quad t \in [0, T], \\ \mathrm{Law}\left(X_0^{1,N,\theta}, \dots, X_0^{N,N,\theta}\right) := \mu_0^{\otimes N}, \end{cases} \tag{III.3.11}$$

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 [7, 30, 88]. In many cases, statistical inference methods developed for the particle system also extend to its mean-field limit (III.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 III.3.2—a key step in the LAN analysis.

In the case of model (III.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 III.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 [30].

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

Theorem III.3.4. *Assume that Assumptions IIIA1–IIIA5 hold. Then, as $N, n \to \infty$,*

$$z(\theta^0, \theta^+) \xrightarrow{\mathbb{P}^{\theta^0}-\text{law}} \begin{pmatrix} u \\ v \end{pmatrix}^\top \mathcal{N}^{\theta^0} - \frac{1}{2} \begin{pmatrix} u \\ v \end{pmatrix}^\top \Sigma^{\theta^0} \begin{pmatrix} u \\ v \end{pmatrix}, \tag{III.3.12}$$

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

$$\Sigma_b^{\theta^0} = \int_0^T \int_{\mathbb{R}^d} z_s^{\theta^0}(x)^\top a_{\theta_2^0}^{-2}(x) z_s^{\theta^0}(x)\, \mu_s^{\theta^0}(dx)\, ds,$$

$$\Sigma_a^{\theta^0} = 2\int_0^T \int_{\mathbb{R}^d} \operatorname{tr}\left(\partial_{\theta_2} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \partial_{\theta_2} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x)\right) \mu_s^{\theta^0}(dx)\, ds.$$

Let us now comment on the implications of Theorem III.3.4. The article [7] investigates parametric estimation of the drift and diffusion coefficients in one-dimensional particle systems of the form (III.3.11), observed at discrete time points. The proposed estimator, based on minimising a contrast function, is shown to be asymptotically normal. Under the additional condition $N\Delta_n \to 0$, the asymptotic covariance matrix takes the same form as Σ^{θ^0}.
However, a key distinction arises when comparing the particle system model (III.3.11) to the i.i.d. observation scheme of the McKean–Vlasov system (III.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 [30] in the context of continuously observed particle systems.
For ease of exposition, we have considered the case of univariate parameters. Nevertheless, as shown in [53] 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

$$\left(\Sigma_b^{\theta^0}\right)_{kl} = \int_0^T \int_{\mathbb{R}^d} z_{k,s}^{\theta^0}(x)^\top a_{\theta_2^0}^{-2}(x) z_{l,s}^{\theta^0}(x)\, \mu_s^{\theta^0}(dx)\, ds,$$

$$\left(\Sigma_a^{\theta^0}\right)_{kl} = 2\int_0^T \int_{\mathbb{R}^d} \operatorname{tr}\left(\partial_{\theta_{k,2}} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x) \partial_{\theta_{l,2}} a_{\theta_2^0}(x) a_{\theta_2^0}^{-1}(x)\right) \mu_s^{\theta^0}(dx)\, ds,$$

and

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

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

III.4 Proofs

III.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 IIIA3 and IIIA4, 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 [86, Section 2.2]). Moreover, the following uniform estimates hold:

$$\mathbb{E}^\theta \left[\sup_{0 \leq t \leq T} \|Z_t\|^p \right] + \sup_{r \in [0,T]} \mathbb{E}^\theta \left[\sup_{r \leq t \leq T} \|\mathcal{D}_r Z_t\|^p \right] \leq C, \tag{III.4.1}$$

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 [7].

Lemma III.4.1. *Assume that Assumptions IIIA1–IIIA2 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 (III.1.1) to standard SDEs with time-varying coefficients. Consider the SDE

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

obtained by freezing the laws $(\mu_t)_{t \in [0,T]}$ in the original model (III.1.1). Under Assumption IIIA2, the stochastic process $\widetilde{X}^\theta$ follows an SDE with bounded coefficients. We also note that the processes $\widetilde{X}^\theta$ and $X^{\theta,i}$ have the same law due to uniqueness of the solution of (III.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 $p^\theta(t_j, t_{j+1}, x, y)$ (see [53, Proposition 1.2]).

Proposition III.4.2. *Assume that Assumptions IIIA2–IIIA5 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 $t_j, t_{j+1} \in [0,T]$, the transition density $p^\theta(t_j, t_{j+1}, x, y)$ satisfies the following Gaussian-type bounds:*

$$p^\theta(t_j, t_{j+1}, x, y) \leq \frac{L}{\Delta_n^{d/2}} \exp\left(- \frac{\|x-y\|^2}{c\Delta_n} \right) \exp(c\Delta_n \|x\|^2), \tag{III.4.2}$$

$$p^\theta(t_j, t_{j+1}, x, y) \geq \frac{1}{L\Delta_n^{d/2}} \exp\left(-c \frac{\|x-y\|^2}{\Delta_n} \right) \exp(-c\Delta_n \|x\|^2). \tag{III.4.3}$$

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

$$\mathbb{E}_{t_j,x}^{\bar{\theta}} \left[\left| \frac{\partial_{\theta_1} p^\theta}{p^\theta}(t_j, t_{j+1}, x, X_{t_{j+1}}^i) \right|^m \right] \leq L\Delta_n^{m/2} \exp(c\Delta_n \|x\|^2)(1 + \|x\|)^q, \tag{III.4.4}$$

$$\mathbb{E}_{t_j,x}^{\bar{\theta}} \left[\left| \frac{\partial_{\theta_2} p^\theta}{p^\theta}(t_j, t_{j+1}, x, X_{t_{j+1}}^i) \right|^m \right] \leq L \exp(c\Delta_n \|x\|^2)(1 + \|x\|)^q. \tag{III.4.5}$$

As a consequence of Proposition III.4.2 and Assumption IIIA1, we conclude that all probability measures $(\mu_t)_{t\in[0,T]}$ are sub-Gaussian (cf. [11, Lemma 3.1]) and we conclude that

$$\mathbb{E}^{\theta}\left[\exp(r_0\|X_t^{\theta,i}\|^2)\right] \leq C \qquad \text{for all } t\in[0,T], \tag{III.4.6}$$

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 III.4.3. *Assume that Assumptions IIIA1–IIIA2 are satisfied. Let $f:\mathbb{R}^d\times\mathcal{P}_l\to\mathbb{R}^p$ 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:*

$$\|f(x,\mu)\| \leq C\left(1+\|x\|^k + W_l^l(\mu,\delta_0)\right). \tag{III.4.7}$$

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 $n,N\to\infty$,

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

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

Proposition III.4.4. *Suppose that Assumptions IIIA1–IIIA5 hold. Let $H_{t_{j+1}}^i$, $i=1,\dots,N$, be a sequence of independent $\mathcal{F}_{t_{j+1}}$-measurable random variables. Recall the definition* (III.2.2)*.*

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

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

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

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

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

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

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

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

We note that the main challenge in Proposition III.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 [53].

Proof. We only show part (i) of Proposition III.4.4 as the second part is shown similarly (cf. [53]). We define the random variable

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

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

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}[\xi_j^N] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \qquad \text{and} \qquad \sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}[(\xi_j^N)^2] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

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

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

Consequently, we deduce the decomposition

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}[\xi_j^N] =: V_{n,N}^1 + V_{n,N}^2,$$

with

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

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

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

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

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

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

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

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 $C > 0$, $\mathbb{E}[\exp(C\Delta_n \|X_{t_j}^i\|^2)]$ is uniformly bounded due to (III.4.6) if Δ_n is small enough. Hence, we conclude that

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

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

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

Consequently, we have proved that

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}[\xi_j^N] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Concerning the conditional second moment we readily obtain the formula

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

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

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

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

This completes the proof of Proposition III.4.4. □

III.4.2 Proof of Proposition III.3.2

III.4.2.1 The drift term

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

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

According to Proposition III.3.1 we have the identity

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

In the following discussion we will study the decomposition

$$\delta\left(\partial_{\theta_1} X_{r,\Delta_n}^{i,\theta} U_r^i \right) = \Delta_n z_{r,\theta_1}^{\theta} \left(X_{t_j}^{i,\theta} \right) \tag{III.4.8}$$

$$\times \left[a_{\theta_2}^{-2}\left(X_{t_j}^{i,\theta}\right)\left(X_{t_{j+1}}^{i,\theta} - m_{t_j,t_{j+1}}^{\theta}\left(X_{t_j}^{i,\theta}\right)\right)\right]_r + H_{t_{j+1}}^i.$$

Here the random variable $H_{t_{j+1}}^i$ is a reminder term, which necessarily satisfies $\mathbb{E}_{t_j,x}^{\theta_1(l),\theta_2^+}[H_{t_{j+1}}^i] = 0$ since the other two terms in (III.4.8) obviously satisfy this identity. Hence, if we prove that

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

the proof of Proposition III.3.2 is completed for ζ_j^{i,θ_1} due to Proposition III.4.4(i). We apply the product formula (III.2.5) and conclude the identity

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

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

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

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

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

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 (III.3.6), we readily obtain the statement

$$\sup_{t\in[0,\Delta_n]}\left|\partial_{\theta_1} X_{r,t}^{i,\theta}\right| = R_{t_{j+1}}^i(\Delta_n). \tag{III.4.9}$$

This implies the same statement for the Malliavin derivative:

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

As a consequence, we deduce that

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

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

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

On the other hand, we deduce that

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

which readily implies that $H_n^{i,2} = R_{t_{j+1}}^i(\Delta_n^2)$.
Next, we treat the term $H_n^{i,3}$. Due to (III.2.4), we conclude that

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

Hence, the application of (III.4.9), implies that $H_n^{i,3} = R_{t_{j+1}}^i(\Delta_n^2)$.
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:

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

Using the above representation we conclude that

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

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

III.4.2.2 The diffusion term

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

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

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

$$\zeta_j^{i,\theta_2} = \frac{v}{\sqrt{\Delta_n N}} \int_0^1 \mathbb{E}^\theta_{t_j, X_{t_j}} \left[\sum_{r=1}^d \delta\left(\partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} U^i_r\right) \middle| X^{i,\theta}_{t_{j+1}} = X^i_{t_{j+1}} \right] dl.$$

We consider the following decomposition:

$$\begin{aligned} \delta\left(\partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} U^i_r\right) &= \left[\partial_{\theta_2} a_{\theta_2}(X^{i,\theta}_{t_j}) a^{-1}_{\theta_2}(X^{i,\theta}_{t_j}) \left(X^{i,\theta}_{t_{j+1}} - m^\theta_{t_j,t_j}\left(X^{i,\theta}_{t_j}\right)\right)\right]_r \\ &\times \left[a^{-2}_{\theta_2}(X^{i,\theta}_{t_j}) \left(X^{i,\theta}_{t_{j+1}} - m^\theta_{t_j,t_{j+1}}\left(X^{i,\theta}_{t_j}\right)\right)\right]_r \\ &- \left[\partial_{\theta_2} a_{\theta_2}(X^{i,\theta}_{t_j}) a^{-1}_{\theta_2}(X^{i,\theta}_{t_j}) V^\theta_{t_j,t_{j+1}}\left(X^{i,\theta}_{t_j}\right) a^{-2}_{\theta_2}(X^{i,\theta}_{t_j})\right]_{r,r} + H^i_{t_{j+1}}. \end{aligned} \tag{III.4.10}$$

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

$$\left(\mathbb{E}^\theta_{t_j,x} |H^i_{t_{j+1}}|^\tau\right)^{1/\tau} = R^i_{t_j}(\Delta_n^{3/2}) \qquad \text{for any } \tau > 1,$$

which completes the proof of Proposition III.3.2 is completed for ζ_j^{i,θ_2} due to Proposition III.4.4(ii).
We apply the product formula (III.2.5) and conclude the identity

$$\delta\left(\partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} U^i_r\right) = \partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} \delta\left(U^i_r\right) - \int_0^{\Delta_n} \left(\mathcal{D}_t \partial_{\theta_2} X^{i,\theta}_{r,\Delta_n}\right)^\top U^i_{r,t} dt.$$

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

$$\partial_{\theta_2} X^{i,\theta}_{r,\Delta_n} = \left[\partial_{\theta_2} a_{\theta_2}(X^i_{t_j}) a^{-1}_{\theta_2}(X^{i,\theta}_{t_j}) \left(X^{i,\theta}_{t_{j+1}} - m^\theta_{t_j,t_j}\left(X^{i,\theta}_{t_j}\right)\right)\right]_r + R^i_{t_{j+1}}(\Delta_n)$$

and

$$\delta\left(U^i_r\right) = \left[a^{-2}_{\theta_2}(X^{i,\theta}_{t_j}) \left(X^{i,\theta}_{t_{j+1}} - m^\theta_{t_j,t_{j+1}}\left(X^{i,\theta}_{t_j}\right)\right)\right]_r + R^i_{t_{j+1}}(\Delta_n).$$

Furthermore, since $V^\theta_{t_j,t_{j+1}}(X^{i,\theta}_{t_j}) = \Delta_n a^2_{\theta_2}(X^{i,\theta}_{t_j}) + R^i_{t_j}(\Delta_n^{3/2})$, we conclude that

$$\int_0^{\Delta_n} \left(\mathcal{D}_t \partial_{\theta_2} X^{i,\theta}_{r,\Delta_n}\right)^\top U^i_{r,t} dt = \left[\partial_{\theta_2} a_{\theta_2}(X^{i,\theta}_{t_j}) a^{-1}_{\theta_2}(X^{i,\theta}_{t_j}) V^\theta_{t_j,t_{j+1}}\left(X^{i,\theta}_{t_j}\right) a^{-2}_{\theta_2}(X^{i,\theta}_{t_j})\right]_{r,r} + R^i_{t_{j+1}}(\Delta_n^{3/2}).$$

This concludes the proof of Proposition III.3.2.

III.4.3 Proof of Theorem III.3.4

In view of Proposition III.3.2 and [59, Theorem VII-5-2], it suffices to check the following conditions:

$$\sum_{j=1}^{n}\sum_{i=1}^{N}\mathbb{E}_{t_j}^{\theta^0}\Big[\widehat{\zeta}_j^{i,\theta_1}\Big] \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2}u^2\Sigma_b^{\theta^0} \tag{III.4.11}$$

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\Big(\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\Big)^2\Big] - \Bigg(\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\Big]\Bigg)^2 \xrightarrow{\mathbb{P}^{\theta^0}} u^2\Sigma_b^{\theta^0} \tag{III.4.12}$$

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\Big|\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\Big|^4\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \tag{III.4.13}$$

$$\sum_{j=1}^{n}\sum_{i=1}^{N}\mathbb{E}_{t_j}^{\theta^0}\Big[\widehat{\zeta}_j^{i,\theta_2}\Big] \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2}v^2\Sigma_a^{\theta^0} \tag{III.4.14}$$

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\Big(\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_2}\Big)^2\Big] - \Bigg(\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_2}\Big]\Bigg)^2 \xrightarrow{\mathbb{P}^{\theta^0}} v^2\Sigma_a^{\theta^0} \tag{III.4.15}$$

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\Big|\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_2}\Big|^4\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \tag{III.4.16}$$

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_2}\Big] - \mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\Big]\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_2}\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0 \tag{III.4.17}$$

Proof of (III.4.11)*:* We first observe that $m_{t_j,t_{j+1}}^{\theta}(x) = x + \int_{t_j}^{t_{j+1}}\mathbb{E}_{t_j,x}^{\theta}[b_{\theta_1}(X_s^{i,\theta},\mu_s^{\theta})]ds$. Applying the Taylor expansion and recalling (III.3.10), we deduce that

$$m_{t_j,t_{j+1}}^{\theta^0}(X_{t_j}^i) - m_{t_j,t_{j+1}}^{\theta_1(l),\theta_2^+}(X_{t_j}^i) = -\frac{lu\Delta_n}{\sqrt{N}}z_{t_j}^{\theta^0}(X_{t_j}^i) + R_{t_j}^i(\varepsilon_{n,N}\Delta_n/\sqrt{N}) \tag{III.4.18}$$

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

$$\overline{\zeta}_j^{i,\theta_1} := \frac{u}{\sqrt{N}}\int_0^1 z_{t_j}^{\theta_1(l),\theta_2^+}(X_{t_j}^i)^\top\Big[a_{\theta_2^+}^{-2}(X_{t_j}^i)\Big(X_{t_{j+1}}^i - m_{t_j,t_{j+1}}^{\theta_0}(X_{t_j}^i)\Big)\Big]dl,$$

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

$$\sum_{j=1}^{n}\sum_{i=1}^{N}\mathbb{E}_{t_j}^{\theta^0}\Big[\widehat{\zeta}_j^{i,\theta_1}\Big] = \frac{u}{\sqrt{N}}\sum_{j=1}^{n}\sum_{i=1}^{N}\int_0^1 z_{t_j}^{\theta_1(l),\theta_2^+}(X_{t_j}^i)^\top \times\Big[a_{\theta_2^+}^{-2}(X_{t_j}^i)\Big(m_{t_j,t_{j+1}}^{\theta^0}(X_{t_j}^i) - m_{t_j,t_{j+1}}^{\theta_1(l),\theta_2^+}(X_{t_j}^i)\Big)\Big]dl \xrightarrow{\mathbb{P}^{\theta^0}} -\frac{1}{2}u^2\Sigma_b^{\theta^0},$$

which follows directly from (III.4.18) and Lemma III.4.3.

Proof of (III.4.12)*:* We observe that

$$\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^N \widehat{\zeta}_j^{i,\theta_1}\Big] = \mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^N (\widehat{\zeta}_j^{i,\theta_1} - \overline{\zeta}_j^{i,\theta_1})\Big] = R_{t_j}^i(\Delta_n)$$

due to (III.4.18). Hence, the second term in (III.4.12) is asymptotically negligible. As for the first term, we obtain the decomposition

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}\Big[\Big(\sum_{i=1}^N \widehat{\zeta}_j^{i,\theta_1}\Big)^2\Big] = \sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i_1\neq i_2} \widehat{\zeta}_j^{i_1,\theta_1}\widehat{\zeta}_j^{i_2,\theta_1} + \sum_{i=1}^N (\widehat{\zeta}_j^{i,\theta_1})^2\Big].$$

Again due to (III.4.18), we deduce the estimate

$$\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i_1\neq i_2} \widehat{\zeta}_j^{i_1,\theta_1}\widehat{\zeta}_j^{i_2,\theta_1}\Big] = \mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i_1\neq i_2} (\widehat{\zeta}_j^{i_1,\theta_1} - \overline{\zeta}_j^{i_1,\theta_1})(\widehat{\zeta}_j^{i_2,\theta_1} - \overline{\zeta}_j^{i_2,\theta_1})\Big] = R_{t_j}^i(\Delta_n^2).$$

In other words,

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i_1\neq i_2} \widehat{\zeta}_j^{i_1,\theta_1}\widehat{\zeta}_j^{i_2,\theta_1}\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Finally, we have that

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

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

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

Proof of (III.4.13)*:* Observe that

$$\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}\Big[\Big|\sum_{i=1}^N \widehat{\zeta}_j^{i,\theta_1}\Big|^4\Big] \le C\sum_{j=1}^n \mathbb{E}_{t_j}^{\theta^0}\Big[\Big|\sum_{i=1}^N \widehat{\zeta}_j^{i,\theta_1} - \mathbb{E}_{t_j}^{\theta^0}\big[\widehat{\zeta}_j^{i,\theta_1}\big]\Big|^4\Big]$$

$$+ C\sum_{j=1}^n\Big|\sum_{i=1}^N \mathbb{E}_{t_j}^{\theta^0}\big[\widehat{\zeta}_j^{i,\theta_1}\big]\Big|^4.$$

As in the proof of (III.4.12), we immediately deduce the convergence

$$\sum_{j=1}^n\Big|\sum_{i=1}^N \mathbb{E}_{t_j}^{\theta^0}\big[\widehat{\zeta}_j^{i,\theta_1}\big]\Big|^4 \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

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

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

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 (III.4.18) we have

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

We also obtain

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

and

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

This completes the proof of (III.4.13).

Proof of (III.4.14)*:* Similarly to the proof of (III.4.11), we obtain the identity:

$$\mathbb{E}_{t_j}^{\theta^0}\left[\widehat{\zeta}_j^{i,\theta_2}\right] = \frac{v}{\sqrt{N\Delta_n}} \int_0^1 \operatorname{tr}\Big[\partial_{\theta_2} a_{\theta_2(l)}(X_{t_j}^i)\, a_{\theta_2(l)}^{-1}(X_{t_j}^i)$$
$$\times \left(\left(m_{t_j,t_{j+1}}^{\theta^0}(X_{t_j}^i) - m_{t_j,t_{j+1}}^{\theta_1^0,\theta_2(l)}(X_{t_j}^i) \right)^{\otimes 2} + \left(V_{t_j,t_{j+1}}^{\theta^0}(X_{t_j}^i) - V_{t_j,t_{j+1}}^{\theta_1^0,\theta_2(l)}(X_{t_j}^i) \right) \right) a_{\theta_2(l)}^{-2}(X_{t_j}^i) \Big]\, dl.$$

Due to (III.4.18) 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

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

By mean value theorem we deduce

$$V_{t_j,t_{j+1}}^{\theta^0}(x) - V_{t_j,t_{j+1}}^{\theta_1^0,\theta_2(l)}(x) = (\theta_2^0 - \theta_2(l))\partial_{\theta_2} V_{t_j,t_{j+1}}^{\theta_1^0,\widetilde{\theta}_2}(x)$$

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

$$V^{\theta^0}_{t_j,t_{j+1}}(X^i_{t_j}) - V^{\theta_1^0,\theta_2(l)}_{t_j,t_{j+1}}(X^i_{t_j}) = -\frac{2lv\Delta_n^{3/2}}{\sqrt{N}}\partial_{\theta_2} a_{\theta_2^0}(X^i_{t_j}) a_{\theta_2^0}(X^i_{t_j}) + R^i_{t_j}\left(\varepsilon_{n,N}\Delta_n^{3/2}/\sqrt{N}\right) \tag{III.4.19}$$

with $\varepsilon_{n,N} \to 0$. Hence, (III.4.19) implies the convergence in (III.4.14).

Proof of (III.4.15)*:* In the previous proof we have shown that

$$\mathbb{E}^{\theta^0}_{t_j}\left[\widehat{\zeta}^{i,\theta_2}_j\right] = R^i_{t_j}(\Delta_n/N).$$

Hence, we immediately conclude that

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

Furthermore, following the same arguments as in the proof of (III.4.12), we deduce the estimate

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

Last but not least, a direct computation shows that

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

This completes the proof of (III.4.15).

Proof of (III.4.16)*:* This statement is shown in exactly the same way as (III.4.13).

Proof of (III.4.17)*:* Applying the estimates from the proof of (III.4.11) and (III.4.14), we immediately obtain the convergence

$$\sum_{j=1}^n \mathbb{E}^{\theta^0}_{t_j}\Big[\sum_{i=1}^N \widehat{\zeta}^{i,\theta_1}_j\Big]\mathbb{E}^{\theta^0}_{t_j}\Big[\sum_{i=1}^N \widehat{\zeta}^{i,\theta_2}_j\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0.$$

Also, using the same arguments as in the proof of (III.4.12), we deduce that

$$\sum_{j=1}^n \mathbb{E}^{\theta^0}_{t_j}\Big[\sum_{i=1}^N \widehat{\zeta}^{i,\theta_1}_j \sum_{i=1}^N \widehat{\zeta}^{i,\theta_2}_j\Big] = \sum_{j=1}^n \mathbb{E}^{\theta^0}_{t_j}\Big[\sum_{i=1}^N \widehat{\zeta}^{i,\theta_1}_j \widehat{\zeta}^{i,\theta_2}_j\Big] + o_{\mathbb{P}^{\theta^0}}(1).$$

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

$$\mathbb{E}^{\theta^0}_{t_j}\left[\left(X^i_{t_{j+1}} - m^{\theta_1(l),\theta_2^+}_{t_j,t_{j+1}}(X^i_{t_j})\right)_{r_1} V^{\theta_1^0,\theta_2(l)}_{t_j,t_{j+1}}(X^i_{t_j})_{r_2,r_3}\right] = R^i_{t_j}(\Delta_n^2/\sqrt{N}),$$

$$\mathbb{E}_{t_j}^{\theta^0}\left[\left(X_{t_{j+1}}^i - m_{t_j,t_{j+1}}^{\theta_1(l),\theta_2^+}(X_{t_j}^i)\right)_{r_1}\left(X_{t_{j+1}}^i - m_{t_j,t_{j+1}}^{\theta_1^0,\theta_2(l)}(X_{t_j}^i)\right)_{r_2,r_3}^{\otimes 2}\right] = R_{t_j}^i(\Delta_n^2)$$

This implies the convergence

$$\sum_{j=1}^{n}\mathbb{E}_{t_j}^{\theta^0}\Big[\sum_{i=1}^{N}\widehat{\zeta}_j^{i,\theta_1}\widehat{\zeta}_j^{i,\theta_2}\Big] \xrightarrow{\mathbb{P}^{\theta^0}} 0,$$

and the proof of Theorem III.3.4 is complete.

Chapter IV

On goodness-of-fit testing for volatility in McKean–Vlasov models

Abstract: This paper develops a statistical framework for goodness-of-fit testing of volatility functions in McKean–Vlasov stochastic differential equations, which describe large systems of interacting particles with distribution-dependent dynamics. While integrated volatility estimation in classical SDEs is now well established, formal model validation and goodness-of-fit testing for McKean–Vlasov systems remain largely unexplored, particularly in regimes with both large particle limits and high-frequency sampling. We propose a test statistic based on discrete observations of particle systems, analysed in a joint regime where both the number of particles and the sampling frequency increase. The estimators involved are proven to be consistent, and the test statistic is shown to satisfy a central limit theorem, converging in distribution to a centred Gaussian law.

IV.1 Introduction

McKean–Vlasov stochastic differential equations (SDEs) have gained increasing prominence as a widely used modeling framework for complex systems consisting of large populations of interacting agents. Unlike classical SDEs, the dynamics of each particle depend not only on its individual state but also on the statistical distribution of the entire system. This distinctive feature makes McKean–Vlasov models particularly well-suited for capturing systemic interactions and emergent behavior in diverse applications ranging from economics and finance to physics and engineering [16, 36, 60, 90].

In the context of financial modeling, McKean–Vlasov dynamics have been used to study systemic risk [43], mean-field interactions in portfolio management [17], and the evolution of agent-based financial markets. More broadly, their use has expanded to problems involving optimal control, equilibrium analysis, and the dynamics of large-scale interacting systems, as surveyed in [18, 68]. These models allow for a nuanced representation of endogenous feedback effects in financial systems, where local decisions and aggregate dynamics are tightly coupled through distributional dependencies.

Given the growing importance of McKean–Vlasov models in applications, there is an increasing need for statistical methods that can rigorously validate their structural components, particularly the volatility function, which governs the system's stochastic fluctuations. While classical diffusion models often rely on volatility functions depending solely on the state or time variables, this assumption breaks down in systems where interaction between agents drives the evolution. In such settings, volatility may depend on the entire population distribution, and any mis-specification can significantly affect downstream predictions and risk measures.

Despite the importance of model validation, the literature on goodness-of-fit testing for McKean–Vlasov equations remains scarce. Existing parametric testing procedures for volatility, such as those developed for standard SDEs [3, 26, 34, 35] and fractional SDEs [87], are tailored to non-interacting systems only. Parametric and non-parametric estimation methods for McKean-Vlasov SDEs have been investigated in [4,7,10,12,22,24,25,29,48,49,63,73,85,88,92], but they mostly focus on the drift function. This creates a critical methodological gap: there is currently no general method for assessing whether a given volatility structure adequately captures the behavior of a McKean–Vlasov system based on empirical data.

We address this gap by developing a statistical testing procedure for McKean–Vlasov particle systems. Our focus lies on constructing a goodness-of-fit test for the volatility function under high-frequency and large-population asymptotics. In this work, we consider a system of N interacting particles $(X^i)^{i=1,\dots,N}$, defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and evolving over a fixed time interval $[0,T]$. The particles are modeled as independent copies of a non-linear process satisfying the McKean–Vlasov SDE:

$$\begin{cases} dX_t^i = b\big(X_t^i, \mu_t\big)dt + a\big(X_t^i, \mu_t\big)dW_t^i \qquad i = 1, \dots, N, \quad t \in [0,T] \\ \text{Law}\big(X_0^1, \dots, X_0^N\big) := \mu_0 \times \dots \times \mu_0 \end{cases} \tag{IV.1.1}$$

where the processes $(W_t^i)_{t\in[0,T]}$, $i = 1, \ldots, N$, are independent Brownian motions, and μ_t denotes the law of X_t^i. The model coefficients

$$b : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}, \qquad a : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$$

are measurable functions that depend on the current state and the current distribution of the solution. Here, $\mathcal{P}_2$ denotes the space of probability measures on $\mathbb{R}$ with finite second moments. This space is equipped with the Wasserstein 2-metric, defined by

$$W_2(\mu, \nu) = \Big(\inf_{m\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^2} |x - y|^2 m(dx, dy) \Big)^{\frac{1}{2}},$$

where $\Gamma(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R}^2$ with marginals μ and ν.

Our primary objective is to develop a goodness-of-fit testing framework for the volatility function $a(x, \mu)$, based on discrete-time observations of the system. We consider observations of the form

$$\Big(X_{t_j}^i \Big)_{j=1,\ldots,n}^{i=1,\ldots,N}, \qquad \text{with } t_j = Tj/n, \tag{IV.1.2}$$

and study the regime where both the observation frequency increases ($\Delta_n := T/n \to 0$) and the number of particles grows ($N \to \infty$), with a fixed time horizon $T > 0$.

The main statistical goal of this work is to develop a goodness-of-fit test for the volatility function $a(x, \mu)$, under the null hypothesis that it belongs to a given parametric family. To this end, we introduce a test statistic based on discrete-time observations of an interacting particle system, constructed through an appropriate distance measure. We prove consistency of the underlying estimators and establish a central limit theorem for the proposed statistic. This yields a testing procedure that maintains the correct asymptotic level and is consistent against any fixed alternative. The main methodological challenge lies in the measure dependence of $a(x, \mu)$, which generates non-linear and path-dependent effects that render standard techniques from classical SDE analysis inapplicable.

To the best of our knowledge, this work provides the first rigorous statistical testing framework for volatility structures in McKean–Vlasov models based on discrete-time observations of interacting particle systems. By combining high-frequency asymptotics with the mean-field structure, our approach extends the scope of model validation to complex stochastic systems with distribution-dependent dynamics and lays the theoretical groundwork for hypothesis testing in nonlinear diffusion models.

The structure of the paper is as follows. Section IV.2 introduces the framework and sets out the standing assumptions for model (IV.1.1). In Section IV.3, we develop the proposed goodness-of-fit testing procedure, detailing the construction of the test statistic. Section IV.4 presents the main theoretical results, establishing the consistency and asymptotic normality of the estimators, and deriving the limiting distribution of the test statistic under the null hypothesis. All proofs and supporting technical arguments are collected in Section IV.5.

IV.2 Assumptions

In this section, we introduce the main assumptions associated with the model (IV.1.1), which are satisfied by a wide class of stochastic volatility models.

IVA1. *For all $k \geq 1$,*

$$\int_{\mathbb{R}} |x|^k \mu_0(dx) \leq C_k.$$

The following assumption ensures the existence and uniqueness of a strong solution to (IV.1.1), guaranteeing well-posedness of the model.

IVA2. *The drift and diffusion coefficients satisfy Lipschitz continuity and a linear growth condition. Specifically, there exists a constant $C > 0$ such that for all $(x, \mu), (y, \lambda) \in \mathbb{R} \times \mathcal{P}_2$:*

$$|b(x,\mu) - b(y,\lambda)| + |a(x,\mu) - a(y,\lambda)| \leq C\big(|x-y| + W_2(\mu,\lambda)\big),$$
$$|b(x,\mu)|^2 + |a(x,\mu)|^2 \leq C\big(1 + |x|^2 + W_2^2(\mu,\delta_0)\big),$$

Regarding the regularity of the diffusion function $a : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$, we adopt the notion of linear differentiability, which is widely used in the literature on McKean–Vlasov equations and mean-field games to characterize the smoothness of the mapping $\mu \mapsto a(x,\mu)$ from $\mathcal{P}_2 \to \mathbb{R}$. This concept is particularly well-suited to our framework, and we refer the reader to Section 2 of [29] and the references therein for a detailed exposition.

Definition IV.2.1. A mapping $f : \mathcal{P}_2 \to \mathbb{R}$ is said to have a linear functional derivative, if there exists $\partial_\mu f : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}$ such that

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

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

IVA3. *The map $\mu \mapsto a(x,\mu)$ admits a functional derivative in the sense of Definition IV.2.1. Furthermore, there exists a constant $C > 0$ such that for all $(x,\mu), (x',\mu') \in \mathbb{R} \times \mathcal{P}_2$,*

$$\left|\partial_\mu a(x,y,\mu) - \partial_\mu a(x',y',\mu')\right| \leq C(|x-x'| + |y-y'| + W_2(\mu,\mu')).$$

Additionally it holds that

$$|\partial_y \partial_\mu a(x,y,\mu)| \leq C \qquad \forall (x,y,\mu),$$
$$|\partial_y \partial_\mu a(x,y,\mu) - \partial_y \partial_\mu a(x,y,\mu')| \leq C W_2(\mu,\mu').$$

Finally, the function $a(x,t) := a(x,\mu_t)$ is in $C^{2,1}(\mathbb{R} \times [0,T])$.

We note that Assumption IVA3 ensures, in particular, that the process $(a(X_t^i, t))_{t\in[0,T]}$ is a continuous semimartingale. Indeed, by Itô's formula one obtains

$$a(X_t^i, t) = a(X_0, 0) + \int_0^t \Big(\partial_t a(X_s^i, s) + b(X_s^i, \mu_s)\,\partial_x a(X_s^i, s) + \tfrac{1}{2}\,\partial_{xx} a(X_s^i, s)\,a^2(X_s^i, s)\Big)\,ds$$
$$+ \int_0^t \partial_x a(X_s^i, s) a(X_s^i, s)\,dW_s^i. \qquad \text{(IV.2.1)}$$

This representation plays a key role in deriving error estimates for high-frequency statistics of the process $(X_t^i)_{t\in[0,T]}$ (cf. [9]).

IV.3 Testing parametric hypotheses for the volatility

In this section, we develop a goodness-of-fit testing framework for the volatility structure in McKean–Vlasov SDEs. Our goal is to assess whether a given parametric form of the volatility function is consistent with the observed behavior of a discretely sampled particle system. We begin by formally stating the parametric hypothesis. Then, we introduce our proposed test statistic $\widehat{S}^N$ and describe its construction under high-frequency and large-population asymptotics.
Let

$$a_1^2, \dots, a_d^2 : \mathbb{R} \times \mathcal{P}_2 \to \mathbb{R}^+$$

be a collection of known functions, assumed to be linearly independent and to satisfy the same regularity conditions as the volatility function $a(x, \mu)$. Our objective is to test whether the squared volatility function $a^2(x, \mu)$ belongs to the linear span of the basis functions $a_1^2, \dots, a_d^2$. More precisely, the null hypothesis is given by

$$H_0 : L := \min_{(\lambda_1,\dots,\lambda_d)\in\mathbb{R}^d} \int_0^T \int_{\mathbb{R}} \Big(a^2(x, \mu_t) - \sum_{k=1}^d \lambda_k a_k^2(x, \mu_t)\Big)^2 \mu_t(dx)\,dt = 0, \quad \text{(IV.3.1)}$$

with the alternative hypothesis $H_1 : L > 0$. Here, μ_t denotes the distribution of the underlying particle system $(X_t^1, \dots, X_t^N)$, which is not directly observable. In practice, we approximate μ_t by the empirical distribution of the observed particles. The criterion in (IV.3.1) serves as a natural foundation for our test construction, since it directly measures model discrepancy in a Hilbert space framework. Moreover, it admits a discretized version that can be readily implemented using particle observations.

Remark IV.3.1. The distance measure L introduced in (IV.3.1) is conceptually related to the distance proposed in [34, 35] for classical SDEs, although the two approaches differ in essential aspects. To clarify this, recall that [34, 35] study the one-dimensional diffusion model

$$dX_t = b(X_t)\,dt + a(X_t)\,dW_t,$$

observed at discrete time points t_j. They introduce the *random* distance measure

$$M^2 := \min_{(\lambda_1,\dots,\lambda_d)\in\mathbb{R}^d} \int_0^T \Big(a^2(X_t) - \sum_{k=1}^d \lambda_k a_k^2(X_t)\Big)^2 dt,$$

and consider the hypothesis test $H_0 : M^2 = 0$ versus $H_1 : M^2 > 0$. A key point is that, under high-frequency observations of a single trajectory of $(X_t)_{t\in[0,T]}$, one can only verify whether

$$a^2(X_t) = \sum_{k=1}^d \lambda_k a_k^2(X_t)$$

holds for some choice of λ_k, *along the realized path* $(X_t(\omega))_{t\in[0,T]}$. There is no possibility to test this identity outside the observed trajectory, i.e., for $x \notin (X_t(\omega))_{t\in[0,T]}$. In contrast, in our setting with N independent trajectories as in (IV.1.1), the condition $L = 0$ entails that

$$a^2(x, \mu_t) = \sum_{k=1}^d \lambda_k a_k^2(x, \mu_t)$$

for some λ_k, holding for μ_t-almost every $x \in \mathbb{R}$, $t \in [0, T]$. Nevertheless, this identity is testable only with respect to the distributions μ_t of the observed particles (X_t^i), and not for arbitrary distributions. □

Standard arguments (cf. [2]) show that this L^2-distance admits the closed-form expression:

$$L = g(\Gamma_1, \dots, \Gamma_d, \mathcal{B}, \Lambda) = \mathcal{B} - (\Gamma_1, \dots, \Gamma_d)\Lambda^{-1}(\Gamma_1, \dots, \Gamma_d)^\top \tag{IV.3.2}$$

where the quantities $\mathcal{B}, \Gamma_1, \dots, \Gamma_d$ and the matrix $\Lambda = (\Lambda_{k,l})_{1\le k,l\le d}$ are given by

$$\begin{aligned}
\mathcal{B} &:= \int_0^T \int_{\mathbb{R}} a^4(x, \mu_t)\mu_t(dx)dt, \\
\Gamma_k &:= \int_0^T \int_{\mathbb{R}} a_k^2(x, \mu_t)a^2(x, \mu_t)\mu_t(dx)dt, \qquad k = 1, \dots, d \\
\Lambda_{k,l} &:= \int_0^T \int_{\mathbb{R}} a_k^2(x, \mu_t)a_l^2(x, \mu_t)\mu_t(dx)dt, \qquad k, l = 1, \dots, d
\end{aligned} \tag{IV.3.3}$$

In order to construct a consistent estimator of L in (IV.3.2), we replace the quantities in (IV.3.3) with their empirical counterparts, based on the discrete-time observations introduced in (IV.1.2). Accordingly, we define the following estimators:

$$\begin{aligned}
\widehat{\mathcal{B}} &:= \frac{1}{3N\Delta_n} \sum_{i=1}^N \sum_{j=1}^n |X_{t_{j+1}}^i - X_{t_j}^i|^4 \\
\widehat{\Gamma}_k &:= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n a_k^2(X_{t_j}^i, \mu_{t_j}^N)|X_{t_{j+1}}^i - X_{t_j}^i|^2 \qquad k = 1, \dots, d \\
\widehat{\Lambda}_{k,l} &:= \frac{\Delta_n}{N} \sum_{i=1}^N \sum_{j=1}^n a_k^2(X_{t_j}^i, \mu_{t_j}^N)a_l^2(X_{t_j}^i, \mu_{t_j}^N) \qquad k, l = 1, \dots, d
\end{aligned} \tag{IV.3.4}$$

(For simplicity of notation we suppress the dependence of estimators on n and N). Here, μ_t^N denotes the empirical measure of the system at time t, i.e.

$$\mu_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^i}.$$

We then introduce the following test statistic as the empirical analogue of the quantity (IV.3.2):

$$\widehat{S}^N = g(\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_d, \widehat{\mathcal{B}}, \widehat{\Lambda}) = \widehat{\mathcal{B}} - \widehat{\mathbf{\Gamma}}^\top \widehat{\Lambda}^{-1} \widehat{\mathbf{\Gamma}}$$

where the vector $\widehat{\mathbf{\Gamma}} = (\widehat{\Gamma}_1, \dots, \widehat{\Gamma}_d)^\top$ is defined by the components in (IV.3.4). This statistic captures the deviation from the null hypothesis and forms the basis of our goodness-of-fit test.

In the following section, we briefly summarise the key statistical properties of our proposed estimator $\widehat{S}^N$. Specifically, building on the consistent estimation of $\mathcal{B}$, Γ_k, and $\Lambda_{k,l}$ by $\widehat{\mathcal{B}}$, $\widehat{\Gamma}_k$, and $\widehat{\Lambda}_{k,l}$ (as established in Theorems IV.4.1 and IV.4.2, respectively), we show that $\widehat{S}^N$ is a consistent estimator of L as $N \to \infty$ and $\Delta_n \to 0$, and provide the associated central limit theorem.

IV.4 Main results

In this section, we present the core theoretical contributions of this paper, focusing on the asymptotic properties of our proposed estimators and their associated test statistics. We begin by demonstrating the consistency and convergence rates of our estimators. Specifically, Theorem IV.4.1 establishes the consistent approximation of the limiting matrix Λ by its empirical counterpart $\widehat{\Lambda}$. Similarly, Theorem IV.4.2 provides stochastic expansions for the quantities $\widehat{\mathcal{B}}$ and $\widehat{\Gamma}_k$ (for $k = 1, \dots, d$), which are essential for deriving the limiting distribution of our test statistic. Corollary IV.4.3 then establishes the joint asymptotic normality of these key components. These results together characterize the limiting distribution and its asymptotic covariance structure. Finally, combining these foundational results, we present the main asymptotic normality result for our proposed estimator $\widehat{S}^N$.

In what follows, we introduce two theorems that provide stochastic expansions for the quantities $\widehat{\Lambda}$, $\widehat{\mathcal{B}}$, and $\widehat{\Gamma}_k$. Throughout the sequel, we will frequently use the notation $o_{\mathbb{P}}(1)$ to denote terms that converge to 0 in probability.

Theorem IV.4.1. *Assume that Assumptions IVA1-IVA3 hold and $N\Delta_n^2 \to 0$. Then*

$$\sqrt{N}\big(\widehat{\Lambda} - \Lambda\big) = \sqrt{N} M_\Lambda + o_{\mathbb{P}}(1)$$

where M_Λ is a $(d \times d)$-matrix with elements

$$M_{\Lambda,k,l} := \frac{1}{N}\sum_{i=1}^{N} \big(Z^i_{\Lambda,k,l} - \mathbb{E}(Z^1_{\Lambda,k,l})\big)$$

$$Z^i_{\Lambda,k,l} := \int_0^T a_k^2(X_s^i, \mu_s) a_l^2(X_s^i, \mu_s) ds \qquad k, l = 1, \dots, d.$$

Theorem IV.4.2. *Assume that Assumptions IVA1-IVA3 hold and $N\Delta_n^2 \to 0$. Then*

$$\sqrt{N}\left(\widehat{\Gamma}_k - \Gamma_k\right) = \sqrt{N}M_k + o_{\mathbb{P}}(1)$$
$$\sqrt{N}\left(\widehat{\mathcal{B}} - \mathcal{B}\right) = \sqrt{N}M_{\mathcal{B}} + o_{\mathbb{P}}(1)$$

with

$$M_k := \frac{1}{N}\sum_{i=1}^{N}\left(Z_k^i - \mathbb{E}(Z_k^1)\right), \qquad Z_k^i := \int_0^T a_k^2(X_s^i, \mu_s)a^2(X_s^i, \mu_s)ds$$
$$M_{\mathcal{B}} := \frac{1}{N}\sum_{i=1}^{N}\left(Z_{\mathcal{B}}^i - \mathbb{E}(Z_{\mathcal{B}}^1)\right), \qquad Z_{\mathcal{B}}^i := \int_0^T a^4(X_s^i, \mu_s)ds$$

Building upon the individual asymptotic properties established in Theorems IV.4.1 and IV.4.2, we are now in a position to derive the joint asymptotic distribution of the involved components $(M_1, \ldots, M_d, M_{\mathcal{B}}, M_\Lambda)$. The following statement is a simple consequence of the standard central limit theorem and the δ-method.

Corollary IV.4.3. *Assume that Assumptions IVA1-IVA3 are satisfied and $N\Delta_n^2 \to 0$.*

(i) It holds that

$$\widehat{Z} := \sqrt{N}\begin{pmatrix} M_1 \\ \vdots \\ M_d \\ M_{\mathcal{B}} \\ \text{vec }(M_\Lambda)_{k,l} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathbf{Z}^*, \qquad \mathbf{Z}^* \stackrel{d}{=} \mathcal{N}_{d^2+d+1}(0, \Sigma)$$

where the components of the covariance matrix Σ are given by

$$\Sigma_{p,q} = \text{Cov}\left(\widehat{Z}_p, \widehat{Z}_q\right).$$

(ii) It holds that

$$\sqrt{N}(\widehat{S}^N - L) \xrightarrow{\mathcal{L}} \mathcal{G} \sim \mathcal{N}(0, \tau^2)$$

where the asymptotic variance τ^2 is defined as

$$\tau^2 = \nabla g^\top(\Gamma_1, \ldots, \Gamma_d, \mathcal{B}, \Lambda)\Sigma\nabla g(\Gamma_1, \ldots, \Gamma_d, \mathcal{B}, \Lambda).$$

The asymptotic result of Corollary IV.4.3 forms the theoretical basis of our goodness-of-fit test. Under the null hypothesis $H_0 : L = 0$, it yields

$$\sqrt{N}\,\widehat{S}^N \xrightarrow{\mathcal{L}} \mathcal{N}(0, \tau^2).$$

Suppose we can construct a consistent estimator $\widehat{\tau}^2$ of the asymptotic variance τ^2, that is,

$$\widehat{\tau}^2 \xrightarrow{\mathbb{P}} \tau^2.$$

Then, for a given significance level $\alpha \in (0,1)$, the null hypothesis $H_0 : L = 0$ is rejected whenever

$$\frac{\sqrt{N}\,\widehat{S}^N}{\widehat{\tau}} > z_{1-\alpha},$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. By construction, this test attains the correct asymptotic size α. Moreover, under the alternative $H_1 : L > 0$, we have $\sqrt{N}\,\widehat{S}^N \xrightarrow{\mathbb{P}} +\infty$, which ensures that the procedure is consistent against any fixed alternative.

In the final step we construct a consistent estimator $\widehat{\tau}^2$ of τ^2. We introduce the vector

$$\widehat{V}^i := \left(\widehat{Z}^i_1, \dots, \widehat{Z}^i_d,\ \widehat{Z}^i_{\mathcal{B}},\ \mathrm{vec}\,(\widehat{Z}^i_\Lambda)_{k,l}\right)^\top$$

with the estimators given by

$$\widehat{Z}^i_k := \sum_{j=1}^n a_k^2(X^i_{t_j}, \mu^N_{t_j})|X^i_{t_{j+1}} - X^i_{t_j}|^2 \qquad k = 1, \dots, d$$

$$\widehat{Z}^i_{\mathcal{B}} := \frac{1}{3\Delta_n}\sum_{j=1}^n |X^i_{t_{j+1}} - X^i_{t_j}|^4$$

$$(\widehat{Z}^i_\Lambda)_{k,l} := \Delta_n \sum_{j=1}^n a_k^2(X^i_{t_j}, \mu^N_{t_j}) a_l^2(X^i_{t_j}, \mu^N_{t_j}) \qquad k, l = 1, \dots, d$$

Then the empirical covariance estimator of the covariance matrix Σ is defined as

$$\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^N \left(\widehat{V}^i - \overline{V}\right)\left(\widehat{V}^i - \overline{V}\right)^\top, \qquad \text{with } \overline{V} = \frac{1}{N}\sum_{i=1}^N \widehat{V}^i.$$

Applying similar methods as in Theorem IV.4.1, we conclude that $\widehat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$ and consequently the estimator

$$\widehat{\tau}^2 := \nabla g^\top \left(\widehat{\Gamma_1}, \dots, \widehat{\Gamma_d}, \widehat{\mathcal{B}}, \widehat{\Lambda}\right) \widehat{\Sigma} \nabla g \left(\widehat{\Gamma_1}, \dots, \widehat{\Gamma_d}, \widehat{\mathcal{B}}, \widehat{\Lambda}\right)$$

satisfies $\widehat{\tau}^2 \xrightarrow{\mathbb{P}} \tau^2$.

Remark IV.4.4. For practical applications, the null hypothesis will almost never hold exactly. It is therefore natural to ask how well the linear span of the functions $a_1^2, \dots, a_d^2$ can approximate the true squared volatility coefficient a^2. The distance measure L is not ideal in this context, since its numerical size is difficult to interpret. A more convenient criterion was introduced in [87] for fractional diffusion models and, in our setting, takes the form

$$G := \frac{L}{\mathcal{B}}.$$

In contrast to L, the statistic G enjoys the appealing property $G \in [0,1]$, which follows directly from Pythagoras' theorem. This normalization allows deviations from the null hypothesis to be expressed in relative terms rather than in absolute units. Moreover, for any fixed $\delta \in (0,1)$, one can test

$$H_0: \ G \in [0,\delta] \qquad \text{vs.} \qquad H_1: \ G \in (\delta, 1],$$

and the asymptotic normality of G follows directly from Corollary IV.4.3. □

IV.5 Proofs

As a preliminary step, we recall a collection of moment bounds, adapted from [7], that will serve as a foundation for establishing the main results of this paper.

Lemma IV.5.1. *Assumptions IVA1–IVA3 hold. Then, for any $p \geq 1$, there exists a constant $C > 0$ such that the following bounds hold uniformly over all particles $i \in \{1, \dots, N\}$, for all $N \in \mathbb{N}$, and for all times $t \in [0,T]$:*

(i) $\sup_{t\in[0,T]} \mathbb{E}[|X_t^i|^p] < C$, *and moreover,* $\sup_{t\in[0,T]} \mathbb{E}[W_p^q(\mu_t^N, \delta_0)] < C$ *for* $p \leq q$.

(ii) $\mathbb{E}[|X_{t_{j+1}}^i - X_{t_j}^i|^p] \leq C\Delta_n^{p/2}$.

(iii) $\mathbb{E}[W_2^2(\mu_t^N, \mu_t)] \leq CN^{-1}$.

IV.5.1 Proof of Theorem IV.4.1

We restrict our attention to the case $d = 1$ and set $f := a_1^4$. The extension to higher dimensions $d > 1$ is straightforward and involves only additional notational complexity. We consider the following decomposition

$$\begin{aligned}
\widehat{\Lambda} - \Lambda &= \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} f(X_{t_j}^i, \mu_{t_j}^N) - \int_0^T \int_{\mathbb{R}} f(x, \mu_s)\mu_s(dx)ds \\
&= \frac{1}{N} \sum_{i=1}^{N} \int_0^T f(X_s^i, \mu_s)ds - \int_0^T \int_{\mathbb{R}} f(x, \mu_s)\mu_s(dx)ds \\
&+ \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} f(X_{t_j}^i, \mu_{t_j}) - \frac{1}{N} \sum_{i=1}^{N} \int_0^T f(X_s^i, \mu_s)ds \\
&+ \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \Big[f(X_{t_j}^i, \mu_{t_j}^N) - f(X_{t_j}^i, \mu_{t_j}) \Big] \\
&=: M_\Lambda + H_{(1)} + H_{(2)}
\end{aligned}$$

The term M_Λ corresponds to the deviation between the empirical mean and its deterministic counterpart. We thus need to show that $\sqrt{N}(|H_{(1)}| + |H_{(2)}|) = o_{\mathbb{P}}(1)$. First of all, we note that the term $H_{(1)}$ corresponds to the Riemann sum approximation error associated with stochastic process $f_t := f(X_t^i, \mu_t)$. According to (IV.2.1), this stochastic process is a continuous Itô semimartingale, we know from [9, Section 7 and 8] that

$$\Delta_n \sum_{j=1}^{n} f(X^i_{t_j}, \mu_{t_j}) - \int_0^T f(X^i_s, \mu_s) ds = \Delta_n A^n_i$$

with

$$\frac{1}{N} \sum_{i=1}^{N} A^n_i = O_{\mathbb{P}}(1).$$

In other words, $\sqrt{N} H_{(1)} = O_{\mathbb{P}}(\Delta_n \sqrt{N})$ and the latter is negligible as $N\Delta_n^2 \to 0$. Now, we focus on the term $H_{(2)}$. Here we use the notion of the linear functional derivative and apply it to the function f:

$$H_{(2)} = H_{(2.1)} + H_{(2.2)},$$

where the terms $H_{(2.1)}$ and $H_{(2.2)}$ are defined via

$$H_{(2.1)} := \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \int_{\mathbb{R}} \partial_\mu f(X^i_{t_j}, y, \mu_{t_j})(\mu^N_{t_j} - \mu_{t_j})(dy)$$

$$H_{(2.2)} := \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \int_0^1 \int_{\mathbb{R}} \Big\{ \partial_\mu f(X^i_{t_j}, y, \lambda \mu^N_{t_j} + (1-\lambda)\mu_{t_j}) - \partial_\mu f(X^i_{t_j}, y, \mu_{t_j}) \Big\} \times (\mu^N_{t_j} - \mu_{t_j})(dy) d\lambda$$

For the term $H_{(2.1)}$ we obtain the identity

$$H_{(2.1)} = \frac{\Delta_n}{N^2} \sum_{i,k=1}^{N} \sum_{j=1}^{n} \left(\partial_\mu f(X^i_{t_j}, X^k_{t_j}, \mu_{t_j}) - \int_{\mathbb{R}} \partial_\mu f(X^i_{t_j}, y, \mu_{t_j}) \mu_{t_j}(dy) \right).$$

If we write $H_{(2.1)} = N^{-2} \sum_{i,k=1}^{N} R_n(i,k)$ and use the arguments from Hoeffding decomposition for U-statistics to compute the variance of $H_{(2.1)}$, we immediately conclude that

$$H_{(2.1)} = O_{\mathbb{P}}(1/N).$$

Now, we move on to handling the term $H_{(2.2)}$. For a function $g : \mathbb{R} \to \mathbb{R}$ with bounded derivative we know that

$$\left| \int_{\mathbb{R}} g(x)(\mu - \nu)(dx) \right| \le \|g'\|_\infty W_1(\mu, \nu).$$

Applying this inequality, Lemma IV.5.1(iii) and Assumption IVA3, we obtain for the term $H_{(2.2)}$:

$$|H_{(2.2)}| \le \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} \int_0^1 \sup_{y \in \mathbb{R}} \left| \partial_y \partial_\mu f(X^i_{t_j}, y, \lambda \mu^N_{t_j} + (1-\lambda)\mu_{t_j}) - \partial_y \partial_\mu f(X^i_{t_j}, y, \mu_{t_j}) \right| d\lambda \times W_1(\mu^N_{t_j}, \mu_{t_j})$$

$$\leq C\Delta_n \sum_{j=1}^{n} W_2(\mu_{t_j}^N, \mu_{t_j}) W_1(\mu_{t_j}^N, \mu_{t_j}) = O_{\mathbb{P}}(N^{-1}).$$

Hence, we conclude that $H_{(2)} = O_{\mathbb{P}}(N^{-1})$, which completes the proof of Theorem IV.4.1.

IV.5.2 Proof of Theorem IV.4.2

We start with the term $\widehat{\Gamma}_k$. We obtain the decomposition

$$\widehat{\Gamma}_k - \Gamma_k = \Gamma_{k.1} + \Gamma_{k.2}$$

where the terms $\Gamma_{k.1}$ and $\Gamma_{k.2}$ are defined as

$$\Gamma_{k.1} := \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} a_k^2(X_{t_j}^i, \mu_{t_j}^N) a^2(X_{t_j}^i, \mu_{t_j}^N) - \int_0^T \int_{\mathbb{R}} a_k^2(x, \mu_t) a^2(x, \mu_t) \mu_t(dx) dt,$$

$$\Gamma_{k.2} := \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} a_k^2(X_{t_j}^i, \mu_{t_j}^N) |X_{t_{j+1}}^i - X_{t_j}^i|^2 - \frac{\Delta_n}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} a_k^2(X_{t_j}^i, \mu_{t_j}^N) a^2(X_{t_j}^i, \mu_{t_j}^N).$$

Analogously to the proof of Theorem IV.4.1 we conclude that

$$\Gamma_{k.1} = M_k + o_{\mathbb{P}}(N^{-1}).$$

Applying the methods of [9, Section 7 and 8], we deduce the decomposition

$$\Gamma_{k.2} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} a_k^2(X_{t_j}^i, \mu_{t_j}^N) a^2(X_{t_j}^i, \mu_{t_j}^N) \left((W_{t_{j+1}}^i - W_{t_j}^i)^2 - \Delta_n \right) + O_{\mathbb{P}}(N^{-1}).$$

By martingale methods we immediately obtain that

$$\mathrm{var}(\Gamma_{k.2}) = N^{-1}\Delta_n.$$

This implies that $\sqrt{N}\Gamma_{k.2} = o_{\mathbb{P}}(1)$.
The term $\widehat{\mathcal{B}}$ is handled the same way as $\widehat{\Gamma}_k$, which completes the proof of Theorem IV.4.2.

□

Bibliography

[1] K. Abraham and R. Nickl. On statistical Calderón problems. *Mathematical Statistics and Learning*, 2(2), 165–216, 2020.

[2] N. J. Achieser. *Theory of approximation*. Dover Publications Inc., New York, 1956.

[3] Y. Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9, 385–426, 1996.

[4] C. Amorino, D. Belomestny, V. Pilipauskaitė, M. Podolskij, and S. Y. Zhou. Polynomial rates via deconvolution for nonparametric estimation in McKean–Vlasov SDEs. *Probability Theory and Related Fields*, 193, 539–584, 2024.

[5] C. Amorino and A. Gloter. Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scandinavian Journal of Statistics*, 47(2), 279-346, 2020.

[6] C. Amorino and A. Gloter. Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function. *Statistical Inference for Stochastic Processes*, 24(1), 61–148, 2021.

[7] C. Amorino, A. Heidari, V. Pilipauskaitė, and M. Podolskij. Parameter estimation of discretely observed interacting particle systems. *Stochastic processes and their Applications*, 163, 350–386, 2023.

[8] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *Journal of Mathematical Neuroscience*, 2(1), 1–50, 2012.

[9] O. E. Barndorff-Nielsen, S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard. A central limit theorem for realised power and bipower variations of continuous semimartingales. *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, Springer, 2006.

[10] D. Belomestny, V. Pilipauskaitė, and M. Podolskij. Semiparametric estimation of McKean-Vlasov SDEs. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 59(1), 79–96, 2022.

[11] D. Belomestny, M. Podolskij and S.-Y. Zhou. On nonparametric estimation of the interaction function in particle system models. *Working paper*, arXiv preprint arxiv: 2402.14419, 2025.

[12] J. P. N. Bishwal. Estimation in interacting diffusions: Continuous and discrete sampling. *Applied Mathematics*, 2(9), 1154–1158, 2011.

[13] M. Bossy and J.F. Jabir. On the wellposedness of some McKean models with moderated or singular diffusion coefficient. *International Symposium on BSDEs*, Springer, 43–87, 2017.

[14] C. Brugna and G. Toscani. Kinetic models of opinion formation in the presence of personal conviction. *Physical Review E*, 92(5), 052818, 2015.

[15] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions. *The master equation and the convergence problem in mean field games*. Princeton University Press, Princeton, 2019.

[16] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications II: Mean field games with common noise and stochastic control.* Springer, 2018.

[17] R. Carmona and B. Webster. A mean field game approach to competitive electricity markets. *Applied Mathematical Finance*, 24(2), 150–170, 2017.

[18] R. Carmona and A. Webster. *Mean field games in financial economics.* Springer, 2020.

[19] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability Theory and Related Fields*, 140, 19–40, 2008.

[20] L. P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. arXiv preprint arXiv:2106.14812, 2022.

[21] B. Chazelle, Q. Jiu, Q. Li and C. Wang. Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. *Journal of Differential Equations*, 263(1), 365–397, 2017.

[22] X. Chen. Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electronic Communications in Probability*, 26, 1–13, 2021.

[23] F. Comte and V. Genon-Catalot. Nonparametric drift estimation for i.i.d. paths of stochastic differential equations. *Annals of Statistics*, 48(6), 3336–3365, 2020.

[24] F. Comte and V. Genon-Catalot. Parametric inference for ergodic McKean-Vlasov stochastic differential equations. *Bernoulli*, 30(3), 1971–1997, 2024.

[25] F. Comte, V. Genon-Catalot, and C. Laredo. Nonparametric moment method for McKean–Vlasov stochastic differential equations. *Working paper*, 2024.

[26] V. Corradi and H. White. Specification tests for the variance of a diffusion. *J. Time Ser. Anal.*, 20, 253–270, 1999.

[27] M. Delattre, V. Genon-Catalot, V., and C. Larédo. Approximate maximum likelihood estimation for stochastic differential equations with random effects in the drift and the diffusion. *Metrika*, 81, 953–983, 2018.

[28] M. Delattre, V. Genon-Catalot, V., and C. Larédo. Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic Processes and their Applications*, 128(6), 1929–1957, 2018.

[29] L. Della Maestra and M. Hoffmann. Nonparametric estimation for interacting particle systems: McKean–Vlasov models. *Probability Theory and Related Fields*, 182(1), 551–613, 2022.

[30] L. Della Maestra and M. Hoffmann. The LAN property for McKean–Vlasov models in a mean-field regime. *Stochastic Processes and their Applications*, 155, 109–146, 2023.

[31] C. Denis, C. Dion, and M. Martinez. Consistent procedures for multiclass classification of discrete diffusion paths. *Scandinavian Journal of Statistics*, 2020.

[32] C. Denis, C. Dion-Blanc, and M. Martinez. A ridge estimator of the drift from discrete repeated observations of the solutions of a stochastic differential equation. *Bernoulli*, 27(4), 2675–2713, 2021.

[33] H. Dette and C. von Lieres und Wilkau. On a test for a parametric form of volatility in continuous time financial models. *Finance and Stochastics*, 7, 363–384, 2003.

[34] H. Dette and M. Podolskij. Testing the parametric form of the volatility in continuous time diffusion models—a stochastic process approach. *Journal of Econometrics*, 143, 56–73, 2008.

[35] H. Dette, M. Podolskij, and M. Vetter. Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scandinavian Journal of Statistics*, 33, 259–278, 2006.

[36] B. Djehiche, F. Gozzi, G. Zanco, and M. Zanella. Optimal portfolio choice with path dependent benchmarked labor income: a mean field model. *Stochastic Processes and their Applications*, 145, 48–85, 2022.

[37] G. Dos Reis, S. Engelhardt, and G. Smith. Simulation of McKean-Vlasov SDEs with super-linear growth. *IMA Journal of Numerical Analysis*, 42(1), 874–922, 2022.

[38] G. Dos Reis, W. Salkeld, and J. Tugaut. Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law. *The Annals of Applied Probability*, 29(3), 1487–1540, 2019.

[39] B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Processes and Their Applications*, 71(1), 33–53, 1997.

[40] D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4), 547–557, 1989.

[41] J. L. Forman and M. Sørensen. The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3), 438–465, 2008.

[42] J.P. Fouque and L.H. Sun. *Systemic risk illustrated*. Handbook on Systemic Risk, Eds J.P Fouque and J Langsam, 2013.

[43] J. P. Fouque and X. Sun. Systemic risk with feedback. *Quantitative Finance*, 13(12), 1845–1856, 2013.

[44] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3), 707–738, 2015.

[45] J. Gärtner. On the McKean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, 137.1, 197–248, 1988.

[46] V. Genon-Catalot. Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1), 99–116, 1990.

[47] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29, 119–151, 1993.

[48] V. Genon-Catalot and C. Larédo. Parametric inference for small variance and long time horizon McKean-Vlasov diffusion models. *Electronic Journal of Statistics*, 15(2), 5811–5854, 2021.

[49] V. Genon-Catalot and C. Larédo. Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. *Stochastic Processes and their Applications*, 142, 513–548, 2021.

[50] K. Giesecke, G. Schwenkler, and J. Sirignano. Inference for large financial systems. *Mathematical Finance*, 30(1), 3–46, 2019.

[51] A. Gloter, D. Loukianova, and H. Mai. Jump filtering and efficient drift estimation for Lévy-driven SDEs. *The Annals of Statistics*, 46(4), 1445, 2018.

[52] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion. *Bernouilli*, 7 (6), 899–912, 2001.

[53] E. Gobet. LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 38(5), 711–737, 2002.

[54] J. Guyon and P. Henry-Labordere. The smile calibration problem solved. *Available at SSRN 1885032*, 2011.

[55] I. Gyongy. Mimicking the one-dimensional marginal distributions of processes having an Ito Differential. *Probability Theory and Related Fields*, 71, 501–516, 1986.

[56] J. Hajek. Local asymptotic minimax admissibility in estimation. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles*, University of California Press, 175–194, 1971.

[57] P. Hall and C. Heyde. *Martingale limit theory and its application.* Academic Press, New York, 1980.

[58] I. A. Ibragimov and R. Z. Has'minskii. *Statistical estimation: Asymptotic theory.* Springer Science & Business Media, Vol. 16, 2013.

[59] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes.* Springer, Berlin, 1987.

[60] B. Jourdain, F. Malrieu, and W. A. Woyczynski. Poisson approximation for the solutions of McKean-Vlasov equations. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 45(4), 865–882, 2009.

[61] M. Kac. Foundations of kinetic theory. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, vol. III, 171–197. University of California Press, 1956.

[62] O. Kallenberg. *Foundations of modern probability* (Vol. 2). Springer-Verlag, New York, 1997.

[63] R.A. Kasonga. Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, 50(3), 865–875, 1990.

[64] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2), 211–229, 1997.

[65] V. N. Kolokoltsov. *Nonlinear Markov processes and kinetic equations.* Cambridge University Press, vol. 182, 2010.

[66] Y. A. Kutoyants. *Parameter estimation for stochastic processes.* Heldermann Verlag, vol. 6, 1988.

[67] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer, 2004.

[68] H. Lacker. *Mean field games and control in economics*. Cambridge University Press, 2020.

[69] D. Lacker, M. Shkolnikov, and J. Zhang. Inverting the Markovian projection, with an application to local stochastic volatility models. *The Annals of Probability*, 48(5), 2189–2211, 2020.

[70] L. Le Cam. Limits of experiments. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles*, University of California Press, 245–261, 1971.

[71] L. Le Cam and G.L. Yang. *Asymptotics in statistics*. Springer, 1990.

[72] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes I: General theory; II: Applications*. Springer, vol. 5–6, 2000.

[73] M. Liu and H. Qiao. Parameter estimation of path-dependent McKean-Vlasov stochastic differential equations. *Acta Mathematica Scientia*, 42, 876–886, 2022.

[74] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Processes and Their Applications*, 95(1), 109–132, 2001.

[75] N. Marie and A. Rosier. Nadaraya–Watson estimator for I.I.D. paths of diffusion processes. *Scandinavian Journal of Statistics*, 2022.

[76] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56.6, 1907–1911, 1966.

[77] H.P. McKean. Propagation of chaos for a class of non-linear parabolic equations. *Stochastic Differential Equations* (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), 41–57, 1967.

[78] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. *Probabilistic Models for Nonlinear Partial Differential Equations*, Springer Berlin Heidelberg, 42–95, 1996.

[79] Y. Mishura and A. Veretennikov. *Theory of stochastic processes with applications to financial mathematics and risk theory*. Springer, 2016.

[80] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *Journal of Mathematical Biology*, 38(6), 534–570, 1999.

[81] S. A. A. Monter, M. Shkolnikov, and J. Zhang. Dynamics of observables in rank-based models and performance of functionally generated portfolios. *The Annals of Applied Probability*, 29(5), 2849–2883, 2019.

[82] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM review*, 56(4), 577–621, 2014.

[83] R. Nickl. Bernstein-von Mises theorems for statistical inverse problems I: Schrodinger equation. *Journal of the European Mathematical Society*, 22(8), 2697–2750, 2020.

[84] R. Nickl. On Bayesian inference for some statistical inverse problems with partial differential equations. *Bernoulli News*, 24(2), 5–9, 2017.

[85] R. Nickl, G.A. Pavliotis, and K. Ray. Bayesian nonparametric inference in McKean-Vlasov models. *Annals of Statistics*, 53(1), 170–193, 2025.

[86] D. Nualart. *The Malliavin calculus and related topics*. Springer-Verlag, Berlin Heidelberg, New York, 1995.

[87] M. Podolskij and K. Wasmuth. Goodness-of-fit testing for fractional diffusions. *Statistical Inference for Stochastic Processes*, 16(2), 147–159, 2013.

[88] L. Sharrock, N. Kantas, P. Parpas, and G.A. Pavliotis. Online parameter estimation for the McKean-Vlasov stochastic differential equation. *Stochastic Processes and their Applications*, 162, 481–546, 2021.

[89] Y. Shimizu. M-Estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Statistical Inference for Stochastic Processes*, 9, 179–225, 2006.

[90] A.-S. Sznitman. Topics in propagation of chaos. *Ecole d'Eté de Probabilités de Saint-Flour XIX*, Springer Berlin Heidelberg, 165–251, 1991.

[91] A. W. Van der Vaart. *Asymptotic statistics*. Cambridge University Press, Cambridge, 1998.

[92] J. Wen, X. Wang, S. Mao, and X. Xiao. Maximum likelihood estimation of McKean-Vlasov stochastic differential equation and its application. *Applied Mathematics and Computation*, 274, 237–246, 2016.

[93] H. Yang, A. Hasan, Y. Ng, and V. Tarokh. Neural McKean-Vlasov processes: Distributional dependence in diffusion processes. *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, vol 238, 262–270, 2024.

[94] R. Yao, X. Chen, and Y. Yang. Mean-Field nonparametric estimation of interacting particle systems. *Proceedings of the 35th Conference on Learning Theory*, vol 178, 2242–2275, 2022.

[95] N. Yoshida. Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2), 220–242, 1992.