



HAL
open science

PARAMETRIC INFERENCE FOR ERGODIC MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS

Valentine Genon-Catalot, Catherine Larédo

► **To cite this version:**

Valentine Genon-Catalot, Catherine Larédo. PARAMETRIC INFERENCE FOR ERGODIC
MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS. 2023. hal-04071936

HAL Id: hal-04071936

<https://hal.science/hal-04071936>

Preprint submitted on 17 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARAMETRIC INFERENCE FOR ERGODIC MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS.

V. GENON-CATALOT⁽¹⁾, C. LARÉDO⁽²⁾

ABSTRACT. We consider a one-dimensional McKean-Vlasov stochastic differential equation with potential and interaction terms depending on unknown parameters. The sample path is continuously observed on a time interval $[0, 2T]$. We assume that the process is in stationary regime. As this distribution is not explicit, the exact likelihood does not lead to computable estimators. To overcome this difficulty, we consider a standard kernel estimator of the invariant density based on the sample path on $[0, T]$ and obtain original properties of this estimator. Then, we derive an explicit approximate likelihood using the sample path on $[T, 2T]$, including the kernel estimator of the invariant density and study the associated estimators of the unknown parameters. We prove their consistency and asymptotic normality with rate \sqrt{T} as T grows to infinity. Several classes of models illustrate the theory.

April 17, 2023

MSC2020 subject classifications: Primary 60J60, 60J99, 62F12, 62M05

Keywords and phrases: McKean-Vlasov stochastic differential equations, Continuous observations, Parametric and nonparametric inference, Invariant distribution, Asymptotic properties of estimators, Approximate likelihood, Long time asymptotics

1. INTRODUCTION

Stochastic systems of N interacting particles and their mean-field limits for large N , the McKean-Vlasov stochastic differential equations (SDEs), were first described in [38]. They appeared later in various areas of applications: for the modelling of granular media in statistical physics ([8]), in neurosciences (see *e.g.* [18], [3]), for population dynamics and ecology ([43], [9]), for epidemics dynamics ([4], [21]), in finance (see *e.g.* [26] and the references therein).

During the past decades, most contributions were devoted to their probabilistic properties (see *e.g.* among many references [50], [42], [6], [7], [27], [34]).

During the same period, statistical inference for these models remained unstudied except [32]. But recently, the interest for this topic is growing in two directions. Either, statistical studies are based on the direct observation of large interacting particle systems: see *e.g.* [19], [20], [26], [12], [5], [14], [2], [45]. Or, statistical inference is based on the observation of the mean-field limit, the McKean-Vlasov process. Indeed, observing the whole N -particles system might seem too

⁽¹⁾: Université Paris Cité, MAP5, UMR 8145 CNRS, F-75006, FRANCE,
email: valentine.genon-catalot@mi.parisdescartes.fr

⁽²⁾: Université Paris Cité, Laboratoire de Probabilités, Statistique & Modélisation, LPSM, France
email: catherine.laredo@inrae.fr

demanding or unrealistic for large N . Thus, inference for the limiting process of one typical particle for large N is now the subject of several contributions: see *e.g.* [48], [23],[24], [25], [46], [37]. This latter point of view is adopted in this paper. More precisely, consider the one-dimensional McKean-Vlasov stochastic differential equation (SDE)

$$(1) \quad dX_t = -b(\alpha, X_t)dt - \int_{\mathbb{R}} \Phi(\beta, X_t - y)\mu_t^\theta(dy)dt + \sigma dW_t, \quad X_0 = \eta$$

where (W_t) is a Wiener process, η is independent of (W_t) and μ_t^θ is the distribution of X_t . The functions $b(\alpha, x)$, $\Phi(\beta, x)$ depend on an unknown parameter $\theta = (\alpha, \beta)$ belonging to a convex set $\Theta_\alpha \times \Theta_\beta \subset \mathbb{R}^d \times \mathbb{R}^{d'}$. The potential term $b(\alpha, x)$ is an usual drift term describing the geometry of the space and $\Phi(\beta, x)$ is an interaction term describing the interaction between particles in the original system. A solution of (1) is a couple $((X_t, \mu_t^\theta), t \geq 0)$ composed with a process (X_t) and a family of distribution (μ_t^θ) satisfying (1). When defined, (X_t) is a time-*inhomogeneous* Markov process. Assumptions for existence and uniqueness of a solution to (1) and existence and uniqueness of an invariant distribution are known. Below, we assume that (1) admits an invariant distribution π_θ and that the process is in stationary regime, *i.e.* $\eta \sim \pi_\theta$.

Our aim is to estimate θ from a continuous observation of (X_t) throughout a time interval $[0, 2T]$ with asymptotic framework $T \rightarrow +\infty$.

A large number of contributions is concerned with statistical inference for ergodic diffusions (see *e.g.* the books [35], [31], [33], [30] and a lot of papers (see *e.g.* [1], [16], [17], [29], [44], [41]).

For what concerns McKean-Vlasov SDEs in stationary regime, papers are not so numerous: to our knowledge, we can refer to [48], [46] and [25] where the special McKean-Vlasov model with no potential term and with odd polynomial interaction term is considered. The McKean-Vlasov SDE studied here is much more general than in the latter paper and the approach quite different. Assuming that the initial variable η of process (1) follows the invariant distribution implies that, for all $t \geq 0$, the marginal distribution $\mu_t^\theta(dy)$ is constant and equal to the invariant distribution $\pi_\theta(dy)$. Therefore, the drift term of (1) does no longer depend on t and is given by

$$(2) \quad x \rightarrow S(\theta, x) := -b(\alpha, x) - \int_{\mathbb{R}} \Phi(\beta, x - y)\pi_\theta(dy) = -b(\alpha, x) - \Phi(\beta, \cdot) \star \pi_\theta(x).$$

Nevertheless, the invariant distribution is not explicitly known so that, although the exact likelihood can be theoretically studied, it does not lead to computable estimators. To overcome this difficulty, we start by studying a kernel estimator $\hat{\pi}_T(x)$ of the invariant density based on the sample path $(X_t, t \in [0, T])$. This estimator is studied in [10], [36], [35], [16], [49]. Its noteworthy property is that its variance rate is \sqrt{T} , *i.e.* the parametric rate which is important for our parametric setting. Nevertheless, the existing results concerning this estimator cannot be directly applied. For our purpose, we study a weighted mean integrated risk fitted to our problem and moreover we prove a convergence in distribution result which, up to our knowledge, are new results of intrinsic interest. In a second step, we propose an approximate likelihood based on the sample $(X_t, t \in [T, 2T])$ where the unknown invariant density of the drift $S(\theta, x)$ is replaced by the kernel estimator, thus leading to the approximate drift:

$$(3) \quad \hat{S}_T(\theta, x) = \hat{S}(\theta, x) = -b(\alpha, x) - \Phi(\beta, \cdot) \star \hat{\pi}_T(x),$$

Then, we study the asymptotic properties of the associated estimators of θ and prove their consistency and asymptotic normality with rate \sqrt{T} . Examples of classes of models illustrate the theory.

Section 2 presents our assumptions to ensure existence and uniqueness of a solution to (1) and existence of an invariant density. We prove a useful bound for the invariant density (Proposition 1) and obtain that the process (X_t) in stationary regime is identical to an ergodic diffusion

(Proposition 2). Section 3 is devoted to statistical results. The statistical assumptions are presented. Then, we study the theoretical exact maximum likelihood estimator (MLE) of θ , prove its consistency and asymptotic normality and exhibit the asymptotic Fisher information (Section 3.1). Section 3.2 concerns the kernel estimator $\hat{\pi}_T(x)$ based on $(X_t, t \in [0, T])$ (Theorem 1, Theorem 2 and Corollary 1). In Section 3.3, we define our approximate likelihood (contrast) which is the conditional likelihood of $(X_t, t \in [T, 2T])$ given X_T where the unknown invariant distribution is replaced by the kernel estimator $\hat{\pi}_T$. We study the associated pseudo-Hessian matrix (Proposition 6) and pseudo-score function (Theorem 3). We conclude by stating the consistency and asymptotic normality of our maximum contrast estimator (MCE). In Section 4, several classes of models are detailed where we compare the asymptotic variances of the MCE and the exact MLE. Section 5 is devoted to lemmas (Lemmas 1 and 2) which are the backbone of the proofs of Section 3. Lemma 2 is especially difficult to obtain. Section 6 gives concluding remarks. Section 7 is an appendix where we recall results from [22] concerning a central limit theorem for ergodic diffusions and properties of the infinitesimal generator, properties used all along the paper. Section 8, 9 and 10 contains all proofs.

2. ASSUMPTIONS AND PRELIMINARIES.

In the sequel, the notation \lesssim means \leq up to a constant. Let us set

$$(4) \quad V(\alpha, x) = \int_0^x b(\alpha, y) dy, \quad W(\beta, x) = \int_0^x \phi(\beta, y) dy.$$

We assume that, for all $\alpha \in \Theta_\alpha$ and all $\beta \in \Theta_\beta$, these functions satisfy:

- [H1] The function $W(\beta, \cdot)$ is even. The functions $x \rightarrow V(\alpha, x)$ and $x \rightarrow W(\beta, x)$ are C^2 and convex, one of the two being strictly convex: there exist constants $K(\alpha)$ and $\lambda(\beta)$ such that

$$\forall x, \frac{\partial^2 V}{\partial x^2}(\alpha, x) \geq K(\alpha) \geq 0, \quad \frac{\partial^2 W}{\partial x^2}(\beta, x) \geq \lambda(\beta) \geq 0 \quad \text{and} \quad K(\alpha) + \lambda(\beta) > 0.$$

- [H2] The functions $x \rightarrow \frac{\partial V}{\partial x}(\alpha, x)$, $\frac{\partial W}{\partial x}(\beta, x)$ are locally Lipschitz with polynomial growth, *i.e.* there exist $c = c(\alpha, \beta) > 0$, $\ell = \ell(\alpha, \beta) \in \mathbb{N}^*$ such that

$$\forall x, y \in \mathbb{R}, \left| \frac{\partial V}{\partial x}(\alpha, x) - \frac{\partial V}{\partial x}(\alpha, y) \right| + \left| \frac{\partial W}{\partial x}(\beta, x) - \frac{\partial W}{\partial x}(\beta, y) \right| \leq c|x - y|(1 + |x|^\ell + |y|^\ell).$$

- [H3] The functions $x \rightarrow \frac{\partial V}{\partial x}(\alpha, x)$, $\frac{\partial^2 V}{\partial x^2}(\alpha, x)$, $\frac{\partial W}{\partial x}(\beta, x)$, $\frac{\partial^2 W}{\partial x^2}(\beta, x)$ have ℓ polynomial growth: there exists a constant $c = c(\alpha, \beta)$ such that

$$\left| \frac{\partial V}{\partial x}(\alpha, x) \right| + \left| \frac{\partial^2 V}{\partial x^2}(\alpha, x) \right| + \left| \frac{\partial W}{\partial x}(\beta, x) \right| + \left| \frac{\partial^2 W}{\partial x^2}(\beta, x) \right| \leq c(1 + |x|^\ell).$$

According to [39], [6], [11],[27], under [H1]-[H3], equation (1) admits a unique solution $((X_t, \mu_t^\theta), t \geq 0)$. For what concerns invariant distributions, two cases are to be distinguished. If $V(\alpha, \cdot)$ is strictly convex ($K(\alpha) > 0$), model (1) admits a unique invariant distribution π_θ such that $\int_{\mathbb{R}} x^2 \pi_\theta(x) dx < +\infty$. If $V(\alpha, \cdot) \equiv 0$, (1) admits a one-parameter family of invariant distributions, the parameter being the expectation of the distribution. Thus, the invariant distribution is unique once its expectation is specified (see also more details in [25]).

If the initial variable X_0 of (1) follows an invariant distribution, then, for all t , $\mathcal{L}(X_t) = \mathcal{L}(X_0)$. This is why, in what follows, to cover all cases, we assume

- [H4] Either [H1]-[H2] hold with $K(\alpha) > 0$ and $X_0 = \eta$ follows the unique invariant distribution π_θ of (1) or $V(\alpha, \cdot) \equiv 0$, [H1]-[H2] hold with $\lambda(\beta) > 0$ and $X_0 = \eta \sim \pi_\theta$, where π_θ is the unique centered invariant distribution. In the latter case, π_θ is symmetric.

Under [H4], the invariant distribution π_θ has density $\pi_\theta(x)$ given as the solution of the implicit equation

$$(5) \quad \pi_\theta(x) = \frac{1}{Z_\theta} \exp \left[-\frac{2}{\sigma^2} (V(\alpha, x) + W(\beta, \cdot) \star \pi_\theta(x)) \right],$$

where $Z_\theta = \int_{\mathbb{R}} \exp \left[-\frac{2}{\sigma^2} V(\alpha, y) + W(\beta, \cdot) \star \pi_\theta(y) \right] dy$.

Contrary to classical SDEs, the invariant distribution is not explicit. Nevertheless, we have:

Proposition 1. *Under [H4], $V(\alpha, x) + W(\beta, \cdot) \star \pi_\theta(x) \geq (K(\alpha) + \lambda(\beta)) \frac{x^2}{2} + ax + b$ ($a = \frac{\partial V}{\partial x}(\alpha, 0) - \lambda(\beta) \int y \pi_\theta(y) dy$, $b = \frac{\lambda(\beta)}{2} \int y^2 \pi_\theta(y) dy$). Thus, the invariant distribution satisfies*

$$(6) \quad \pi_\theta(x) \lesssim \exp \left[-\frac{(K(\alpha) + \lambda(\beta))}{\sigma^2} \left(x + \frac{a}{K(\alpha) + \lambda(\beta)} \right)^2 \right].$$

As an obvious consequence, the invariant distribution has moments of any orders and (6) implies that, for all $k \in \mathbb{R}$ and all $\varepsilon > 0$,

$$(7) \quad \int_{\mathbb{R}} \exp(kx) (\pi_\theta(x))^\varepsilon dx < +\infty.$$

This property is crucially used in proofs for various values of ε .

Under [H4], the initial variable η follows an invariant distribution $\pi_\theta(x)dx$. Then, the distribution $\mathcal{L}(X_t) = \mu_t^\theta(dy)$ satisfies that

$$\forall t \geq 0, \mu_t^\theta(dy) = \pi_\theta(y)dy.$$

Therefore, the following holds (see (2)):

$$(8) \quad dX_t = -(b(\alpha, X_t) + \Phi(\beta, \cdot) \star \pi_\theta(X_t))dt + \sigma dW_t = S(\theta, X_t)dt + \sigma dW_t, \quad X_0 = \eta \sim \pi_\theta(x)dx.$$

Proposition 2. *Assume [H4] and consider the stochastic differential equation (see (2))*

$$(9) \quad dY_t = S(\theta, Y_t)dt + \sigma dW_t.$$

Then (Y_t) is a positive recurrent diffusion with stationary density given by (5). If $Y_0 \sim \pi_\theta(x)dx$, it is ergodic. Moreover,

- *If $Y_0 \neq X_0$, $(Y_t) \not\equiv (X_t)$.*
- *If $Y_0 = X_0 = \eta \sim \pi_\theta(x)dx$, then $X_t = Y_t$ for all $t \geq 0$.*

The result simply follows from the uniqueness of solutions. Thus, under [H4], (X_t) is equal to the solution of a classical SDE in stationary regime and is ergodic. This result has important consequences. One can apply to (X_t) results for classical ergodic SDEs. If f satisfies $\int |f(x)|\pi_\theta(x)dx < +\infty$, the ergodic theorem (see *e.g.* [47], Chap. V-7), yields

$$(10) \quad \frac{1}{T} \int_0^T f(X_s)ds \rightarrow_{a.s.} \int f(x)\pi_\theta(x)dx.$$

More results are given in Section 7. They rely on the infinitesimal generator of the SDEs (8):

$$(11) \quad Lg = (\sigma^2/2)g'' - (b(\alpha, \cdot) + \Phi(\beta, \cdot) \star \pi_\theta(\cdot))g'.$$

The operator L acts on $\mathbb{L}^2(\pi_\theta(x)dx)$ and is defined on the domain \mathcal{D} ,

$$(12) \quad \mathcal{D} = \{g \in \mathbb{L}^2(\pi_\theta(x)dx), g' \text{ abs. continuous}, Lg \in \mathbb{L}^2(\pi_\theta(x)dx), \lim_{|x| \rightarrow \infty} g'(x)\pi_\theta(x) = 0\}.$$

3. ESTIMATION RESULTS.

From now on, we assume that [H4] holds for all θ and study the estimation of $\theta = (\alpha, \beta)$ from the continuous observation of $(X_t, t \in [0, 2T])$ satisfying (8).

Throughout this study, we assume that σ^2 is fixed, known or unknown, but since it is identifiable from the continuous observation of (X_t) on $[0, 2T]$, we are not concerned by its estimation. Clearly, the invariant density of (X_t) defined by the implicit equation (5) also depends on σ^2 . Since it is fixed, we omit for sake of clarity, its dependence with respect to σ^2 and just write $\pi_\theta(\cdot)$, as well as in all statistical notations used in the sequel.

We detail first a maximum likelihood approach (Section 3.1) which is theoretical. Then, we propose a tractable contrast method which relies on a nonparametric estimator of the invariant density whose properties are studied in Section 3.2. The contrast is studied in Section 3.3.

We denote by \mathbb{P}_θ the distribution on $C([0, 2T])$ of $(X_t, t \in [0, 2T])$ and by θ_0 the true value of the parameter. For M a matrix or a vector, we denote by M' its transpose. Recall the notation (2) and set

$$(13) \quad K(\theta_0, \theta) = \int_{\mathbb{R}} (S(\theta, x) - S(\theta_0, x))^2 \pi_{\theta_0}(x) dx,$$

$$(14) \quad \tilde{K}(\theta_0, \theta) = \int_{\mathbb{R}} \{b(\alpha, x) - b(\alpha_0, x) + [\Phi(\beta, \cdot) - \Phi(\beta_0, \cdot)] \star \pi_{\theta_0}(x)\}^2 \pi_{\theta_0}(x) dx.$$

These two quantities are well defined since, under [S1], $b(\alpha, \cdot)$ and $\Phi(\beta, \cdot) \star \pi_\theta$ have polynomial growth and, by (6) π_θ has moments of any order. Statistical assumptions are required.

- [S0] The parameter space $\Theta = \Theta_\alpha \times \Theta_\beta$ is compact and the true value θ_0 belongs to $\mathring{\Theta}$.
- [S1] The functions $(\alpha, x) \rightarrow V(\alpha, x)$ and $(\beta, x) \rightarrow W(\beta, x)$ are respectively defined on $U_\alpha \times \mathbb{R}$ and $U_\beta \times \mathbb{R}$ where U_α (resp. U_β) is an open set containing Θ_α (resp. Θ_β) and are such that all the derivatives $(\alpha, x) \rightarrow \frac{\partial^{i+j} V}{\partial x^i \partial \alpha^j}(\alpha, x)$, $(\beta, x) \rightarrow \frac{\partial^{i+j} W}{\partial x^i \partial \beta^j}(\beta, x)$ exist, are continuous respectively on $U_\alpha \times \mathbb{R}$ (resp. $U_\beta \times \mathbb{R}$) and have polynomial ℓ -growth with respect to x , uniformly in α, β .
- [S2] $\{K(\theta_0, \theta) = 0\} \Rightarrow \{\theta = \theta_0\}$.
- [S3] For all α, β , and for $z_1, z_2 \in \mathbb{R}^d \times \mathbb{R}^{d'}$,

$$\{z_1 \cdot \nabla_\alpha S(\theta, x) + z_2 \cdot \nabla_\beta S(\theta, x) \equiv 0\} \Rightarrow \{z_1 = z_2 = 0\}$$

where ∇_α (resp. ∇_β) denotes the gradient vector w.r.t. α (resp. β) and $x \cdot y$ denotes the Euclidian scalar product of two vectors x, y .

- [S4] $\{\tilde{K}(\theta_0, \theta) = 0\} \Rightarrow \{\theta = \theta_0\}$.
- [S5] For all α, β , and for $z_1, z_2 \in \mathbb{R}^d \times \mathbb{R}^{d'}$,

$$\{z_1 \cdot \nabla_\alpha b(\alpha, x) + z_2 \cdot [\nabla_\beta \Phi(\beta, \cdot)] \star \pi_\theta(x) \equiv 0\} \Rightarrow \{z_1 = z_2 = 0\}.$$

Assumptions [S0]-[S1] are standard for MLEs or MCEs. As we can see from the expression (15), the likelihood is not explicitly known so that the MLE is not computable. Therefore, we introduce an explicit contrast (pseudo-likelihood) leading to computable estimators. Assumptions [S2]-[S3] are the identifiability conditions for the MLE (resp. [S4]-[S5] for the MCE). Assumption [S3] (resp. [S5]) concerns the invertibility of the Fisher (resp. pseudo-Fisher) information matrix required for the asymptotic normality of the MLE (resp. MCE). By direct computation, comparing (13) and (14) is not straightforward, except on specific examples (see Section 4). Below, Proposition 3 (resp. Proposition 5) states that [S2] ensures the consistency of the MLE (resp. [S4] of the MCE). Therefore, the statistical theory ensures that if [S4] holds, [S2] holds.

3.1. Maximum Likelihood Estimator based on $(X_t, t \in [0, T])$ and $(X_t, t \in [0, 2T])$. To estimate θ from $(X_t, t \in [0, T])$, let us first look at the conditional log-likelihood of $(X_t, t \in [0, T])$ given X_0 . The Girsanov formula holds for (X_t) and using (2) yields

$$(15) \quad \ell_T(\theta) = \frac{1}{\sigma^2} \left(\int_0^T S(\theta, X_t) dX_t - \frac{1}{2} \int_0^T S^2(\theta, X_t) dt \right).$$

From (2), we see that $S(\theta, x)$ depends on π_θ given by the implicit equation (see (5)) and is not explicit. Therefore, (15) cannot lead to computable estimators of θ since the derivatives of the log-likelihood depend on the derivatives w.r.t. θ of π_θ . Nevertheless, a theoretical study of the MLE is possible. A first step is the following proposition which determines the rate of the problem and the parameters that can be identifiable on this model.

Proposition 3. *Assume [H4], [S1]. Then, as T tends to infinity, under \mathbb{P}_{θ_0} , using (13),*

$$\frac{1}{T} [\ell_T(\theta) - \ell_T(\theta_0)] \rightarrow_{a.s.} -\frac{1}{2\sigma^2} \int_{\mathbb{R}} [S(\theta, x) - S(\theta_0, x)]^2 \pi_{\theta_0}(x) dx = -\frac{1}{2\sigma^2} K(\theta_0, \theta).$$

The proof of Proposition 3 is standard as (X_t) is identical to (Y_t) an ergodic diffusion process, and therefore satisfies (10). Now, let us define the MLE as

$$(16) \quad \widehat{\theta}_T = \arg \sup_{\theta \in \Theta} \ell_T(\theta).$$

Under [S0]-[S2], $K(\theta_0, \theta)$ possesses a unique minimum at $\theta = \theta_0$, which leads to the consistency of $\widehat{\theta}_T$. Moreover, Proposition 3 implies that the estimation rate of θ is \sqrt{T} as for ergodic diffusions.

We can define the $(d + d')$ matrix

$$(17) \quad I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ [I_{12}(\theta)]' & I_{22}(\theta) \end{pmatrix}, \quad \text{where}$$

$$I_{11}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial S}{\partial \alpha_i}(\theta, x) \frac{\partial S}{\partial \alpha_k}(\theta, x) \pi_\theta(x) dx \right)_{i,k=1,\dots,d},$$

$$I_{12}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial S}{\partial \alpha_i}(\theta, x) \frac{\partial S}{\partial \beta_j}(\theta, x) \pi_\theta(x) dx \right)_{i=1,\dots,d; j=1\dots d'}$$

$$I_{22}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial S}{\partial \beta_j}(\theta, x) \frac{\partial S}{\partial \beta_\ell}(\theta, x) \pi_\theta(x) dx \right)_{j,\ell=1,\dots,d'}, \quad \text{and}$$

$$\frac{\partial S}{\partial \alpha_i}(\theta, x) = \frac{\partial b}{\partial \alpha_i}(\alpha, x) + \Phi(\beta, \cdot) \star \frac{\partial \pi_\theta}{\partial \alpha_i}(\theta, x), \quad \frac{\partial S}{\partial \beta_j}(\theta, x) = \left[\frac{\partial \Phi}{\partial \beta_j}(\beta, \cdot) \star \pi_\theta \right](\theta, x) + \left[\Phi(\beta, \cdot) \star \frac{\partial \pi_\theta}{\partial \beta_j} \right](\theta, x).$$

Then, under [S1], the Fisher information matrix associated with $\ell_T(\theta)$ is $\sigma^{-2} I(\theta)$ (see Proposition 4). As usual, [S2] implies [S3] as $K(\theta_0, \theta_0 + h) = h' I(\theta_0) h + o(\|h\|^2)$ but the reverse is not true.

Note that σ^2 implicitly appears in the $I_{ij}(\theta)$, $i, j = 1, 2$ through π_θ . As $S(\theta, \cdot)$ depends on π_θ , the derivation of the MLE requires, additionally to the ability of differentiating π_θ w.r.t θ , the knowledge of σ^2 . This is specific to McKean-Vlasov diffusions. We can state:

Proposition 4. *Assume [H4], [S0]-[S2]-[S3]. Then, the MLE $\widehat{\theta}_T$ is consistent and $\sqrt{T}(\widehat{\theta}_T - \theta_0)$ converges in distribution under \mathbb{P}_{θ_0} to the Gaussian law $\mathcal{N}_{d+d'}(0, \sigma^2 I^{-1}(\theta_0))$.*

The proof of this result is standard. However, the result remains theoretical as the MLE is not computable. Besides, one cannot easily check either [S2] or [S3] (see Section 4).

In order to compare the estimators derived from the two methods (the MLE and the MCE below), assume now that the observation is $(X_t, t \in [0, 2T])$. We know that, under the same assumptions as in Proposition 4, $\sqrt{2T}(\widehat{\theta}_{2T} - \theta_0)$ converges in distribution under \mathbb{P}_{θ_0} to the Gaussian law $\mathcal{N}_{d+d'}(0, \sigma^2 I^{-1}(\theta_0))$. Therefore, $\sqrt{T}(\widehat{\theta}_{2T} - \theta_0)$ converges in distribution under \mathbb{P}_{θ_0} to the Gaussian law $\mathcal{N}_{d+d'}(0, \frac{\sigma^2}{2} I^{-1}(\theta_0))$.

3.2. Nonparametric estimator of the invariant density. When there is no ambiguity, we drop in this section the index 0 and write θ instead of θ_0 .

Before studying the second estimation method, we need properties for the nonparametric estimator of $\pi_\theta(\cdot)$ that we plug in $S(\theta, x)$.

Let K a nonnegative continuous even function with support $[-A, A]$, $A > 0$ such that

$$(18) \quad \int K(v)dv = 1, \quad \int v^2 K(v)dv = \sigma_K^2 < +\infty, \quad \text{and set } K_h(x) = (1/h)K(x/h).$$

Then, a kernel density estimator of $\pi_\theta(x)$ is, if h_T is a bandwidth satisfying $h_T = o(1)$,

$$(19) \quad \widehat{\pi}_T(x) = \frac{1}{T} \int_0^T K_{h_T}(X_t - x) dt.$$

Note that the function $\widehat{\pi}_T(x)$ is a probability density.

For n *i.i.d.* observations, the mean integrated risk of a kernel density estimator is well known (see *e.g.* [51], [13]). The rate of the integrated square-bias term is a power of the bandwidth and the rate of the integrated variance term is $(nh)^{-1}$, lower than the parametric rate. For *i.i.d.* observations on $[0, T]$ of diffusions paths with fixed T , similar results on the estimation of $\pi_T(x) = T^{-1} \int_0^T p_t(x_0, x) dt$ where $p_t(x_0, x)$ is the transition density and x_0 the initial value, have been obtained by [40].

For a kernel estimator such as $\widehat{\pi}_T$ based on the continuous observation of a one-dimensional stationary process, this risk was firstly investigated in [10], see also for the case of ergodic diffusion processes, [36], [35], [16] for pointwise risk or [49] for sup-norm risk or [15]. The results are different from the *i.i.d.* case. Indeed, under appropriate assumptions, the variance rate of the kernel estimator is the parametric rate T^{-1} which is unusual for nonparametric density estimation. This is why it is interesting in our model to replace the unknown density π_θ by the kernel estimator $\widehat{\pi}_T$ as this substitution will not affect the rate of convergence of our parametric estimators. However, these results are not sufficient for our purpose and we need to study a weighted mean integrated risk fitted to our problem together with a convergence in distribution result which, up to our knowledge, have not yet been studied.

Theorem 1. *Assume [H4]. We have, for all $p \geq 0$, using (18), (19),*

(1) *There exists a continuous function $f_\theta : \mathbb{R} \rightarrow \mathbb{R}^+$ such that*

$$B_T^2(p, \theta) = B_T^2(p) = \int_{\mathbb{R}} (1 + |x|^p) (\mathbb{E}_\theta \widehat{\pi}_T(x) - \pi_\theta(x))^2 dx \leq h_T^4 \sigma_K^4 \sup_{|t| \leq A} f_\theta(t).$$

(2) *There exists a constant $C_\theta(p)$ such that*

$$V_T(p, \theta) = V_T(p) = \int_{\mathbb{R}} \text{Var}_\theta(\widehat{\pi}_T(x)) (1 + |x|^p) dx \leq \frac{C_\theta(p)}{T}.$$

The invariant density is twice continuously differentiable by [H1] so that the rate of the bias term is consistent with previous results. The second result on the variance term is new and difficult to obtain. It strongly relies on property (7) of the invariant distribution.

Theorem 2. Assume [H4], [S1]. Then, if $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is a C^1 function such that ψ and ψ' have polynomial growth, under \mathbb{P}_θ ,

$$(20) \quad \sqrt{T} \int_{\mathbb{R}} \psi(y)(\widehat{\pi}_T(y) - \mathbb{E}_\theta \widehat{\pi}_T(y)) dy = \frac{\sigma}{\sqrt{T}} \int_0^T g'_\psi(X_s) dW_s + O_P\left(\frac{1}{\sqrt{T}}\right) + O_P(\sqrt{T}h_T),$$

where g_ψ satisfies $Lg_\psi(y) = -(\psi(y) - \int \psi(y)\pi_\theta(y)dy) := -\psi_c(y)$, L is the infinitesimal generator defined in (11), so that, using Remark 1 of the Appendix (Section 7),

$$(21) \quad g'_\psi(x) = 2\sigma^{-2}\pi_\theta^{-1}(x) \int_{-\infty}^x \psi_c(y)\pi_\theta(y)dy = -2\sigma^{-2}\pi_\theta^{-1}(x) \int_x^{+\infty} \psi_c(y)\pi_\theta(y)dy.$$

The above result is original and useful for the study of the asymptotic properties of our MCE.

Corollary 1. Under the assumptions of Theorem 2, if $\sqrt{T}h_T = o(1)$, under \mathbb{P}_θ , as T tends to infinity,

$$\sqrt{T} \int_{\mathbb{R}} \psi(y)(\widehat{\pi}_T(y) - \pi_\theta(y)) dy \rightarrow_{\mathcal{L}} \mathcal{N}\left(0, \sigma^2 \int (g'_\psi(x))^2 \pi_\theta(x) dx\right).$$

As our aim is parametric inference, we do not seek a square bias-variance compromise. But, we need a sufficient condition to erase the bias and obtain a \sqrt{T} global rate.

3.3. Maximum contrast estimator. To derive this estimator, we use the interval $[0, T]$ to build the nonparametric estimator $\widehat{\pi}_T(x)$ of $\pi_\theta(x)$ and substitute, in the conditional likelihood of $(X_t, t \in [T, 2T])$, given X_T , $\pi_\theta(x)$ by $\widehat{\pi}_T(x)$. Recall (see (3)):

$$\widehat{S}_T(\theta, x) = \widehat{S}(\theta, x) = -b(\alpha, x) - \Phi(\beta, \cdot) \star \widehat{\pi}_T(x),$$

and consider the pseudo-likelihood or contrast

$$(22) \quad \Lambda_T(\theta) = \frac{1}{\sigma^2} \left(\int_T^{2T} \widehat{S}(\theta, X_s) dX_s - \frac{1}{2} \int_T^{2T} (\widehat{S}(\theta, X_s))^2 ds \right).$$

The stochastic integral is well defined as $\widehat{\pi}_T(x)$ is computed using $(X_t, 0 \leq t \leq T)$. The following proposition clarifies the identifiability assumption [S3] associated with the contrast (22).

Proposition 5. Assume [H4], [S1]. Then, as T tends to infinity, under \mathbb{P}_{θ_0} , the contrast defined in (22) with $\widehat{\pi}_T$ given in (19), satisfies (see (14)),

$$\frac{1}{T} [\Lambda_T(\theta) - \Lambda_T(\theta_0)] \rightarrow_{a.s.} -\frac{1}{2\sigma^2} \widetilde{K}(\theta_0, \theta).$$

Analogously, Assumption [S4] implies [S5]. The identifiability assumption for $\Lambda_T(\theta)$ is not the same as for $\ell_T(\theta)$ as $\widetilde{K}(\theta_0, \theta) \neq K(\theta_0, \theta)$. The comparison of $\widetilde{K}(\theta_0, \theta)$ and $K(\theta_0, \theta)$ is not straightforward. In Section 4, examples are given where it is possible.

Now define the maximum contrast estimator (MCE) associated with (22) by:

$$(23) \quad \widehat{\theta}_T = \arg \sup_{\theta \in \Theta} \Lambda_T(\theta)$$

Contrary to the MLE, this MCE does not require the knowledge of σ^2 and is explicit.

Under [S0]-[S1] and [S4], we deduce standardly from Proposition 5 that the MCE is consistent.

To obtain the limiting distribution of the normalized MCE $\sqrt{T}(\widehat{\theta}_T - \theta_0)$, we have to study the derivatives of the contrast with respect to the $\alpha = (\alpha_j, j = 1, \dots, d)$, $\beta = (\beta_k, k = 1, \dots, d')$. This requires properties of the difference

$$(24) \quad \widehat{S}(\theta, x) - S(\theta, x) = -\Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(x).$$

Let us define the $(d + d') \times (d + d')$ pseudo Hessian matrix $\mathcal{I}_T(\theta) = \begin{pmatrix} \mathcal{I}_{T,1,1} & \mathcal{I}_{T,1,2} \\ [\mathcal{I}_{T,1,2}]' & \mathcal{I}_{T,2,2} \end{pmatrix}$ where

$$\mathcal{I}_{T,1,1} = \left(\frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \alpha_{j'}}(\theta) \right)_{1 \leq j, j' \leq d}, \quad \mathcal{I}_{T,2,2} = \left(\frac{\partial^2 \Lambda_T}{\partial \beta_k \partial \beta_{k'}}(\theta) \right)_{1 \leq k, k' \leq d'},$$

$$\mathcal{I}_{T,1,2} = \left(\frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \beta_k}(\theta) \right)_{1 \leq j \leq d, 1 \leq k \leq d'}.$$

Proposition 6. *Assume [H4], [S1]. As T tends to infinity, under \mathbb{P}_θ , the normalized pseudo Hessian matrix $\frac{1}{T} \mathcal{I}_T(\theta)$ converges to $-\sigma^{-2} \mathcal{I}(\theta)$ where*

$$(25) \quad \mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{1,1}(\theta) & \mathcal{I}_{1,2}(\theta) \\ [\mathcal{I}_{1,2}(\theta)]' & \mathcal{I}_{2,2}(\theta) \end{pmatrix}, \quad \text{with}$$

$$\mathcal{I}_{1,1}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial b}{\partial \alpha_j}(\alpha, x) \frac{\partial b}{\partial \alpha_{j'}}(\alpha, x) \pi_\theta(x) dx \right)_{1 \leq j, j' \leq d},$$

$$\mathcal{I}_{1,2}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial b}{\partial \alpha_j}(\alpha, x) \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(x) \pi_\theta(x) dx \right)_{1 \leq j \leq d, 1 \leq k \leq d'}$$

$$\mathcal{I}_{2,2}(\theta) = \left(\int_{\mathbb{R}} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(x) \frac{\partial \Phi}{\partial \beta_{k'}}(\beta, \cdot) \star \pi_\theta(x) \pi_\theta(x) dx \right)_{1 \leq k, k' \leq d'}.$$

Note that $\mathcal{I}(\theta)$ is invertible if and only if, for $z = (z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^{d'}$,

$$z' \mathcal{I}(\theta) z = 0 \Rightarrow \{z_1 = z_2 = 0\}.$$

Now

$$z' \mathcal{I}(\theta) z = \int (z_1 \cdot \nabla_\alpha b(\alpha, x) + z_2 \cdot \nabla_\beta \phi(\beta, \cdot) \star \pi_\theta(x))^2 \pi_\theta(x) dx.$$

Therefore, $z' \mathcal{I}(\theta) z = 0 \Leftrightarrow z_1 \cdot \nabla_\alpha b(\alpha, x) + z_2 \cdot \nabla_\beta \phi(\beta, \cdot) \star \pi_\theta(x) \equiv 0$. Thus, [S5].

Let us study now the pseudo-score function. For this, define $j = 1, \dots, d, k = 1, \dots, d'$,

$$(26) \quad \Psi_j(\theta, y) = \int \frac{\partial b}{\partial \alpha_j}(\alpha, x) \Phi(\beta, x - y) \pi_\theta(x) dx; \quad \Psi_{j,c}(\theta, y) = \Psi_j(\theta, y) - \int \Psi_j(\theta, z) \pi_\theta(z) dz,$$

(27)

$$\Xi_k(\theta, y) = \int \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(x) \Phi(\beta, x - y) \pi_\theta(x) dx; \quad \Xi_{k,c}(\theta, y) = \Xi_k(\theta, y) - \int \Xi_k(\theta, z) \pi_\theta(z) dz.$$

By Proposition 9, there exist functions g_j, h_k satisfying

$$(28) \quad Lg_j(\theta, \cdot)(y) = -\Psi_{j,c}(\theta, y); \quad Lh_k(\theta, \cdot)(y) = -\Xi_{k,c}(\theta, y), \quad j = 1, \dots, d; \quad k = 1, \dots, d'.$$

Next, define the square $d + d'$ matrix

$$(29) \quad \mathcal{H}(\theta) = \begin{pmatrix} \mathcal{H}_{11}(\theta) & \mathcal{H}_{12}(\theta) \\ [\mathcal{H}_{12}(\theta)]' & \mathcal{H}_{22}(\theta) \end{pmatrix},$$

where $\mathcal{H}_{11}, \mathcal{H}_{12}, \mathcal{H}_{22}$ are respectively the $d \times d, d \times d'$ and $d' \times d'$ matrices for $j, j' = 1, \dots, d, k, k' = 1, \dots, d'$,

$$\mathcal{H}_{11}(j, j') = \int g'_j(\theta, x) g'_{j'}(\theta, x) \pi_\theta(x) dx, \quad \mathcal{H}_{12}(j, k) = \int g'_j(\theta, x) h'_k(\theta, x) \pi_\theta(x) dx,$$

$$\mathcal{H}_{22}(k, k') = \int h'_k(\theta, x) h'_{k'}(\theta, x) \pi_\theta(x) dx.$$

Here, $g'_j(\theta, x), h'_k(\theta, x)$ denote the derivatives of $g_j(\theta, x), h_k(\theta, x)$ w.r.t. x (see (26), (27), (28)).

Theorem 3. *Assume [H4], [S1] and that h_T tends to 0 in such a way that $h_T\sqrt{T} = o(1)$. Then, $\frac{\sigma^2}{\sqrt{T}} \begin{pmatrix} \nabla_\alpha \Lambda_T(\theta) \\ \nabla_\beta \Lambda_T(\theta) \end{pmatrix}$ satisfies that, under \mathbb{P}_θ , for $j = 1, \dots, d$; $k = 1, \dots, d'$,*

$$\begin{aligned} \frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \alpha_j}(\theta) &= -\frac{\sigma}{\sqrt{T}} \left(\int_T^{2T} \frac{\partial b}{\partial \alpha_j}(\alpha, X_t) dW_t + \int_0^T g'_j(\theta, X_t) dW_t \right) + o_P(1), \\ \frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \beta_k}(\theta) &= -\frac{\sigma}{\sqrt{T}} \left(\int_T^{2T} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(X_t) dW_t + \int_0^T h'_k(\theta, X_t) dW_t \right) + o_P(1). \end{aligned}$$

Therefore, $\frac{\sigma^2}{\sqrt{T}} \begin{pmatrix} \nabla_\alpha \Lambda_T(\theta) \\ \nabla_\beta \Lambda_T(\theta) \end{pmatrix}$ converges in distribution to $\mathcal{N}_{d+d'}(0, \Sigma(\theta))$ with, using (25), (29),

$$(30) \quad \Sigma(\theta) = \begin{pmatrix} \Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ [\Sigma_{12}(\theta)]' & \Sigma_{22}(\theta) \end{pmatrix}, \text{ where } \Sigma_{ij}(\theta) = \sigma^2(\mathcal{I}_{ij}(\theta) + \mathcal{H}_{ij}(\theta)) \text{ for } i, j = 1, 2.$$

To clarify the previous results, let us describe the main terms occurring in $\frac{\partial \Lambda_T}{\partial \alpha}(\theta)$ assuming $d = d' = 1$. We have, under \mathbb{P}_θ ,

$$\begin{aligned} \frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \alpha}(\theta) &= -\frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) dW_t - A_{2,T} + o_P(1), \\ A_{2,T} &= \frac{1}{T} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) \sqrt{T} (\Phi(\beta, \cdot) \star (\hat{\pi}_T - \pi_\theta))(X_t) \\ &= \sqrt{T} \int_{\mathbb{R}} (\hat{\pi}_T(y) - \pi_\theta(y)) \frac{1}{T} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) \Phi(\beta, X_t - y) dy. \end{aligned}$$

Substituting $\frac{1}{T} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) \Phi(\beta, X_t - y) dy$ by its limit $\Psi(\theta, y) = \int \frac{\partial b}{\partial \alpha}(\alpha, x) \Phi(\beta, x - y) \pi_\theta(x) dx$ as $T \rightarrow \infty$ yields that the main term of $A_{2,T}$ is equal to

$$A_{2,T} = \sqrt{T} \int_{\mathbb{R}} \Psi(\theta, y) (\hat{\pi}_T(y) - \pi_\theta(y)) dy + o_P(1) = \frac{\sigma}{\sqrt{T}} \int_0^T g'(X_s) dW_s + o_P(1),$$

where, applying Theorem 2, g satisfies $Lg = -\Psi_c(\theta, \cdot)$. Finally $\frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \alpha}(\theta)$ is the sum of two uncorrelated terms and converges in distribution to a Gaussian random variable with variance $\Sigma_{11} = \sigma^2 \int [(\frac{\partial b}{\partial \alpha}(\alpha, x))^2 + (g'(x))^2] \pi_\theta(x) dx$ using the central limit theorem (32).

Now, the convergence in distribution of the MCE derives classically from the previous results.

Proposition 7. *Assume [H4], [S0]-[S1], [S4]-[S5]. The MCE defined in (23) is consistent and satisfies that, under \mathbb{P}_{θ_0} , using (25), (30),*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow_{\mathcal{L}} \mathcal{N}_{d+d'}(0, \mathcal{I}^{-1}(\theta_0) \Sigma(\theta_0) \mathcal{I}^{-1}(\theta_0)).$$

Note that this variance satisfies, using (29), $\sigma^{-2}(\mathcal{I}^{-1}(\theta) + \mathcal{I}^{-1}(\theta) \mathcal{H}(\theta) \mathcal{I}^{-1}(\theta))$.

4. EXAMPLES

Example 1 is the Ornstein-Uhlenbeck McKean-Vlasov process. Example 2 deals with the case where $b(\alpha, x) = \alpha b(x)$ is linear in the parameter α and $\Phi(\beta, x) = \beta x$, $\alpha, \beta \in \mathbb{R}$. For this family of models, we can compare $K(\theta_0, \theta)$, $\tilde{K}(\theta_0, \theta)$ and $I(\theta)$, $\mathcal{I}(\theta)$. In Example 3, we consider $b(\alpha, x) \equiv 0$ and $x \rightarrow \Phi(\beta, x)$ an odd polynomial. Thus, we compare the present approach with the one developed in [25], which is devoted to the study of McKean-Vlasov models with null potential term and odd polynomial interaction term. For sake of simplicity, we focus on the

comparison of the estimators $\widehat{\theta}_T$ (16) (instead of $\widehat{\theta}_{2T}$) and $\widehat{\theta}_T$ (23).

Example 1. $V(\alpha, x) = \alpha x^2/2, W(\beta, x) = \beta x^2/2, \alpha > 0, \beta > 0$. Equation (1) is $dX_t = -\alpha X_t dt - \beta(X_t - \mathbb{E}_\theta X_t)dt + \sigma dW_t$, ((23) so that $\mathbb{E}_\theta X_t = \mathbb{E}_\theta X_0 e^{-\alpha t}$. Therefore, the invariant distribution is necessarily centered, and

$$dX_t = -(\alpha + \beta)X_t dt + \sigma dW_t.$$

The invariant distribution π_θ is the Gaussian distribution $\mathcal{N}(0, \sigma^2/2(\alpha + \beta))$ and $S(\theta, x) = -(\alpha + \beta)x$. Here, $K(\theta_0, \theta) = [\alpha + \beta - (\alpha_0 + \beta_0)]^2 \int x^2 \pi_{\theta_0}(x) dx = \frac{\sigma^2}{2(\alpha_0 + \beta_0)} [\alpha + \beta - (\alpha_0 + \beta_0)]^2$. The only identifiable parameter is $\tau = \tau(\theta) = \alpha + \beta$ ([S2], [S3] not satisfied). The MLE of τ is

$$\widehat{\tau} = -\frac{\int_0^T X_t dX_t}{\int_0^T X_t^2 dt} \quad \text{and} \quad \sqrt{T}(\widehat{\tau} - \tau) \rightarrow_{\mathcal{L}} \mathcal{N}(0, 2\tau).$$

Let us look at the contrast $\Lambda_T(\theta)$. We have $\widehat{S}(\theta, x) = -\alpha x - \beta \int (x - y) \widehat{\pi}_T(y) dy = -(\alpha + \beta)x - \beta \int y \widehat{\pi}_T(y) dy = S(\theta, x) - \frac{\beta}{T} \int_0^T X_t dt$ since

$$\int y \widehat{\pi}_T(y) dy = \frac{1}{Th_T} \int_0^T \int_{\mathbb{R}} y K\left(\frac{y - X_t}{h_T}\right) dy = \frac{1}{T} \int_0^T \int_{\mathbb{R}} (X_t + vh_T) K(v) dv = \frac{1}{T} \int_0^T X_t dt.$$

We have $\widetilde{K}(\theta_0, \theta) = K(\theta_0, \theta)$. With $\tau = \tau(\theta) = \alpha + \beta$,

$$\sigma^2 \Lambda_T(\tau, \beta) = \int_T^{2T} [-\tau X_t - \beta \frac{1}{T} \int_0^T X_t dt] dX_t - \frac{1}{2} \int_T^{2T} [-\tau X_t - \beta \frac{1}{T} \int_0^T X_t dt]^2 dt.$$

Thus, the MCE satisfies

$$\frac{1}{T} \mathcal{I}_T \begin{pmatrix} \widehat{\tau}_T \\ \widehat{\beta}_T \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \int_T^{2T} X_s dX_s \\ -\frac{1}{T} \int_0^T X_t dt \int_T^{2T} dX_s \end{pmatrix} \quad \text{with} \quad \mathcal{I}_T = \begin{pmatrix} -\int_T^{2T} X_s^2 ds & \frac{1}{T} \int_0^T X_t dt \int_T^{2T} X_s ds \\ \frac{1}{T} \int_0^T X_t dt \int_T^{2T} X_s ds & -(\frac{1}{T} \int_0^T X_t dt)^2 \end{pmatrix}.$$

For given T , \mathcal{I}_T/T is invertible but, as T grows to infinity, it converges to the non invertible matrix $\begin{pmatrix} -\sigma^2/2\tau & 0 \\ 0 & 0 \end{pmatrix}$. After some computations, we find

$$\widehat{\tau}_T = \frac{1}{\frac{1}{T} \int_T^{2T} X_s^2 ds - (\frac{1}{T} \int_0^T X_t dt)^2} \left(-\frac{1}{T} \int_T^{2T} X_t dX_t + \frac{1}{T} \int_T^{2T} dX_s \frac{1}{T} \int_0^T X_s ds \right).$$

By the ergodic theorem, $\frac{1}{T} \int_0^T X_t dt$ converges almost surely to $\int y \pi_\theta(y) dy = 0$. Then, we apply the result of Proposition 8 to $f_c(x) = x$ with $Lg(x) = (\sigma^2/2)g'' - (\alpha + \beta)xg'$. The function $g(x) = x/(\alpha + \beta)$ satisfies $Lg(x) = -x$. Therefore, $\sigma^2(f_c) = \sigma^2/(\alpha + \beta)^2$. Thus,

$$\sqrt{T} \left(\frac{1}{T} \int_0^T X_t dt \right) \rightarrow_{\mathcal{L}} X \sim \mathcal{N}(\sigma^2/(\alpha + \beta)^2).$$

This allows to study $\widehat{\tau}_T$ and we can prove that $\sqrt{T}(\widehat{\tau}_T - \tau) = o_P(1)$. We can compute $\widehat{\beta}_T$, but there is no result for it. Note that $z_1 \frac{\partial b}{\partial \alpha}(\alpha, x) + z_2 \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(x) = (z_1 + z_2)x \equiv 0$ does not imply $z_1 = z_2 = 0$. So [S5] is not satisfied.

Example 2. Consider the class of models $b(\alpha, x) = \alpha b(x), \alpha > 0$, with b twice continuously differentiable, $\Phi(\beta, x) = \beta x$ with $\alpha b'(x) + \beta \geq K > 0, b''(x) \neq 0$. Assumption [H4] is satisfied

and the model has a unique invariant density π_θ ,

$$\pi_\theta(x) = \frac{1}{Z_\theta} \exp \left[-\frac{2}{\sigma^2} \left(\alpha \int_0^x b(y) dy + \beta \left(\frac{x^2}{2} - x\gamma_1(\theta) \right) \right) \right],$$

where $\gamma_1(\theta) = \int y\pi_\theta(y)dy$ is not explicitly known and Z_θ is the unknown normalization constant. Now, we compare $K(\theta_0, \theta)$, $\tilde{K}(\theta_0, \theta)$ and $I(\theta)$, $\mathcal{I}(\theta)$.

The equation for X_t is given by

$$X_t = X_0 + \alpha \int_0^t [b(X_s) + \beta(X_s - \mathbb{E}_\theta(X_s))] ds + \sigma W_t.$$

When $X_0 \sim \pi_\theta$, the process is in stationary regime. Thus, taking expectations yields

$$(31) \quad \int b(x)\pi_\theta(x)dx = 0.$$

This implies,

$$\begin{aligned} K(\theta_0, \theta) &= \int [(\alpha - \alpha_0)b(x) + (\beta - \beta_0)(x - \gamma_1(\theta_0)) + \beta(\gamma_1(\theta_0) - \gamma_1(\theta))]^2 \pi_{\theta_0}(x) dx \\ &= \tilde{K}(\theta_0, \theta) + \beta^2(\gamma_1(\theta) - \gamma_1(\theta_0))^2. \end{aligned}$$

Assume that $\tilde{K}(\theta_0, \theta) = 0$. Then, for all x , $(\alpha - \alpha_0)b(x) + (\beta - \beta_0)(x - \gamma_1(\theta_0)) = 0$. Since b'' is not identically nul, this implies $\alpha = \alpha_0$ and $\beta = \beta_0$. Thus, by the contrast $\Lambda_T(\theta)$, both parameters are identifiable. One sees that, if [S4] holds, [S2] holds. So, both parameters are also identifiable by the likelihood method. We have

$$\begin{aligned} \sigma^2 \ell_T(\theta) &= - \int_0^T [\alpha b(X_t) + \beta(X_t - \gamma_1(\theta))] dX_t - \frac{1}{2} \int_T^{2T} [\alpha b(X_t) + \beta(X_t - \gamma_1(\theta))]^2 dt, \\ \sigma^2 \Lambda_T(\theta) &= - \int_T^{2T} [\alpha b(X_t) + \beta(X_t - \hat{m}_T)] dX_t - \frac{1}{2} \int_0^T [\alpha b(X_t) + \beta(X_t - \hat{m}_T)]^2 dt, \end{aligned}$$

with $\hat{m}_T = T^{-1} \int_0^T X_s ds$. Here, we again see that the MLE of θ is not explicitly computable as there is no explicit expression of the function $\theta \rightarrow \gamma_1(\theta)$ and its derivatives w.r.t. $\theta = (\alpha, \beta)$. Nevertheless, we can compute the Fisher information matrix which is $\sigma^{-2}I(\theta)$, where, using (31) and (17), $I(\theta) = (I_{ij}(\theta))_{1 \leq i, j \leq 2}$, with

$$\begin{aligned} I_{11}(\theta) &= \int b^2(x)\pi_\theta(x)dx + \left(\beta \frac{\partial \gamma_1}{\partial \alpha}(\theta)\right)^2, \quad I_{22}(\theta) = \int (x - \gamma_1(\theta))^2 \pi_\theta(x)dx + \left(\beta \frac{\partial \gamma_1}{\partial \beta}(\theta)\right)^2, \\ I_{12}(\theta) &= \int b(x)(x - \gamma_1(\theta))\pi_\theta(x)dx + \beta^2 \frac{\partial \gamma_1}{\partial \alpha}(\theta) \frac{\partial \gamma_1}{\partial \beta}(\theta). \end{aligned}$$

Note that $I(\theta) = \int \begin{pmatrix} b(x) \\ x - \gamma_1(\theta) \end{pmatrix} \begin{pmatrix} b(x), & x - \gamma_1(\theta) \end{pmatrix} \pi_\theta(x) dx + \begin{pmatrix} \beta \frac{\partial \gamma_1}{\partial \alpha}(\theta) \\ \beta \frac{\partial \gamma_1}{\partial \beta}(\theta) \end{pmatrix} \begin{pmatrix} \beta \frac{\partial \gamma_1}{\partial \alpha}(\theta), & \beta \frac{\partial \gamma_1}{\partial \beta}(\theta) \end{pmatrix}$.

Hence, for $z = (z_1, z_2)'$, $z'I(\theta)z = 0$ is equivalent to for all x , $z_1(b(x) + \beta \frac{\partial \gamma_1}{\partial \alpha}(\theta)) + z_2(x - \gamma_1(\theta) + \beta \frac{\partial \gamma_1}{\partial \beta}(\theta)) = 0$. This, in turn implies, since $b''(x) \neq 0$, $z_1 = z_2 = 0$. So the matrix $I(\theta)$ is invertible. The MLE converges to $\mathcal{N}(0, \sigma^2 I^{-1}(\theta))$.

Contrary to the MLE, the MCE of θ is explicit and solution of the linear system:

$$-\mathcal{I}_T \begin{pmatrix} \hat{\alpha}_T \\ \hat{\beta}_T \end{pmatrix} = \begin{pmatrix} \int_T^{2T} b(X_t) dX_t \\ \int_T^{2T} (X_t - \hat{m}_T) dX_t \end{pmatrix} \text{ with } \mathcal{I}_T = \begin{pmatrix} \int_T^{2T} b^2(X_t) dt & \int_T^{2T} b(X_t)(X_t - \hat{m}_T) dt \\ \int_T^{2T} b(X_t)(X_t - \hat{m}_T) dt & \int_T^{2T} (X_t - \hat{m}_T)^2 dt \end{pmatrix}.$$

The matrix \mathcal{I}_T/T converges as T tends to infinity to (see (25))

$$\mathcal{I}(\theta) = \begin{pmatrix} \int b^2(x)\pi_\theta(x)dx & \int (x - \gamma_1(\theta))b(x)\pi_\theta(x)dx \\ \int (x - \gamma_1(\theta))b(x)\pi_\theta(x)dx & \int (x - \gamma_1(\theta))^2\pi_\theta(x)dx \end{pmatrix}.$$

The relation $z_1b(x) + z_2(x - \gamma_1(\theta)) \equiv 0$ implies $z_1 = z_2 = 0$: [S4] is satisfied and $\mathcal{I}(\theta)$ is invertible. In this model, we can compute the functions g, h of Theorem 3. Indeed, we have, using (31),

$$\begin{aligned} \Psi(\theta, y) &= \beta \int b(x)(x - y)\pi_\theta(x)dx = \beta \int xb(x)\pi_\theta(x)dx, \\ \Xi(\theta, y) &= \int (x - \gamma_1(\theta))\beta(x - y)\pi_\theta(x)dx = \beta \int x(x - \gamma_1(\theta))\pi_\theta(x)dx \end{aligned}$$

Thus, $y \rightarrow \Psi(\theta, y)$ and $y \rightarrow \Xi(\theta, y)$ are constant so that $g \equiv h \equiv 0$. Therefore, $\Sigma(\theta) = \sigma^2\mathcal{I}(\theta)$. The asymptotic variance of the MCE is equal to $(\sigma^2\mathcal{I}(\theta))^{-1}$. Note that

$$I(\theta) = \mathcal{I}(\theta) + \begin{pmatrix} \beta \frac{\partial \gamma_1}{\partial \alpha}(\theta) \\ \beta \frac{\partial \gamma_1}{\partial \beta}(\theta) \end{pmatrix} \begin{pmatrix} \beta \frac{\partial \gamma_1}{\partial \alpha}(\theta), & \beta \frac{\partial \gamma_1}{\partial \beta}(\theta) \end{pmatrix}.$$

This shows that there is a loss of information when using the contrast $\Lambda_T(\theta)$ instead of the exact log-likelihood which is not surprising.

Example 3. $b(\alpha, x) \equiv 0$, $\Phi(\beta, x) = \beta_1x + \beta_3x^3$ with $\beta = (\beta_1, \beta_3)$, $\beta_1 > 0, \beta_3 \geq 0$. Assume that (X_t) is in centered stationary regime. The invariant distribution π_β is not only centered by also symmetric (see [H4]). Using this property, we obtain, with $\gamma_i(\beta) = \int y^i \pi_\beta(y)dy$,

$$\Phi(\beta, \cdot) \star \pi_\beta(x) = \int [\beta_1(x - y) + \beta_3(x - y)^3] \pi_\beta(x)dx = (\beta_1 + 3\beta_3\gamma_2(\beta))x + \beta_3x^3.$$

Let us first study the MLE. If (β_{10}, β_{30}) denotes the true value of the parameter,

$$K(\beta_0, \beta) = \int [\beta_1x + \beta_3(x^3 + 3x\gamma_2(\beta)) - (\beta_{10}x + \beta_{30}(x^3 + 3x\gamma_2(\beta_0)))]^2 \pi_{\beta_0}(x)dx.$$

Therefore, $K(\beta_0, \beta) = 0$ implies that $\beta_3 = \beta_{30}$ and $\beta_1 - \beta_{10} + 3\beta_{30}(\gamma_2(\beta) - \gamma_2(\beta_0)) = 0$. Thus, Assumption [S2] holds if

$$[C1] : \forall \beta_1 > 0, \beta_3 \geq 0, 1 + 3\beta_3 \frac{\partial \gamma_2}{\partial \beta_1}(\beta) \neq 0.$$

Due to the presence of $\gamma_2(\beta)$, $\Phi(\beta, \cdot) \star \pi_\beta(x)$ is not explicit so the MLE cannot be computed.

We can compute $I(\beta)$ (see (17)) and we get, for $z = (z_1 \ z_2)' \in \mathbb{R}^2$,

$$z'I(\beta)z = \int x^2 \left(z_1(1 + 3\beta_3 \frac{\partial \gamma_2}{\partial \beta_1}(\beta)) + z_2(3\gamma_2(\beta) + x^2 + 3\beta_3 \frac{\partial \gamma_2}{\partial \beta_3}(\beta)) \right)^2 \pi_\beta(x)dx.$$

Condition [C1] implies that $z'I(\beta)z \neq 0$. The Fisher information matrix is obtained by:

$$\begin{aligned} I_{11}(\beta) &= (1 + 3\beta_3 \frac{\partial \gamma_2}{\partial \beta_1}(\beta))^2 \gamma_2(\beta), \\ I_{12}(\beta) &= (1 + 3\beta_3 \frac{\partial \gamma_2}{\partial \beta_1}(\beta))(3(\beta_3 \frac{\partial \gamma_2}{\partial \beta_3}(\beta) + \gamma_2(\beta))\gamma_2(\beta) + \gamma_4(\beta)), \\ I_{22}(\beta) &= \gamma_6(\beta) + 6\gamma_2(\beta)\gamma_4(\beta) + 9\gamma_2(\beta)(\gamma_2(\beta) + \beta_3 \frac{\partial \gamma_2}{\partial \beta_3}(\beta))^2 + 6\beta_3\gamma_4(\beta) \frac{\partial \gamma_2}{\partial \beta_3}(\beta). \end{aligned}$$

Let us now study the Maximum Contrast method. We have $\tilde{K}(\beta_0, \beta) = \int [(\beta_1 - \beta_{10})x + (\beta_3 - \beta_{30})(x^3 + 3x\gamma_2(\beta_0))]^2 \pi_{\beta_0}(x)dx$. Therefore, Assumption [S4] is satisfied and β_1, β_3 are identifiable.

Note that, if β_1, β_3 are identifiable with the contrast method (*i.e.* can be consistently estimated), β_1, β_3 are identifiable with the likelihood method. Thus, [S3] holds and this shows that [C1] holds.

We have:

$$\frac{\partial \Phi(\beta, \cdot)}{\partial \beta_1} \star \pi_\beta(x) = x, \quad \frac{\partial \Phi(\beta, \cdot)}{\partial \beta_3} \star \pi_\beta(x) = x^3 + 3x^2\gamma_2(\beta).$$

We can compute $\mathcal{I}(\beta)$ (see (25)).

$$\mathcal{I}(\beta) = \begin{pmatrix} \gamma_2(\beta) & \gamma_4(\beta) + 3\gamma_2^2(\beta) \\ \gamma_4(\beta) + 3\gamma_2^2(\beta) & \gamma_6(\beta) + 6\gamma_2(\beta)\gamma_4(\beta) + 9\gamma_2^3(\beta) \end{pmatrix}$$

Thus, $z' \mathcal{I}(\beta) z = \int x^2 (z_1 + z_2(3\gamma_2(\beta) + x^2))^2 \pi_\beta(x) dx \Rightarrow$ Assumption [S5] is satisfied.

Let us study $\Lambda_T(\beta)$. Setting $\widehat{m}_i(T) = \int y^i \widehat{\pi}_T(y) dy$, we obtain that

$$\Phi(\beta, \cdot) \star \widehat{\pi}_T(x) = \beta_1 \widehat{\Gamma}_1(x) + \beta_3 \widehat{\Gamma}_2(x),$$

with $\widehat{\Gamma}_1(x) = x - \widehat{m}_1(T)$ and $\widehat{\Gamma}_2(x) = x^3 - 3x^2\widehat{m}_1(T) + 3x\widehat{m}_2(T) - \widehat{m}_3(T)$.

Therefore, the MCE $(\widehat{\beta}_1, \widehat{\beta}_3)$ satisfies

$$-\widehat{\Gamma}_T \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \int_T^{2T} \widehat{\Gamma}_1(X_s) dX_s \\ \int_T^{2T} \widehat{\Gamma}_2(X_s) dX_s \end{pmatrix} \text{ with } \widehat{\Gamma}_T = \begin{pmatrix} \int_T^{2T} \widehat{\Gamma}_1^2(X_s) ds & \int_T^{2T} \widehat{\Gamma}_1(X_s) \widehat{\Gamma}_2(X_s) ds \\ \int_T^{2T} \widehat{\Gamma}_1(X_s) \widehat{\Gamma}_2(X_s) ds & \int_T^{2T} \widehat{\Gamma}_2^2(X_s) ds \end{pmatrix}.$$

We can compute $\Sigma(\beta)$. First, let us obtain $\mathcal{H}(\beta)$. Since the odd moments of π_β are nul,

$$\begin{aligned} \Xi_1(\beta, y) &= \beta_1\gamma_2(\beta) + \beta_3\gamma_4(\beta) + 3\beta_3\gamma_2(\beta)y^2 \Rightarrow \Xi_{1,c}(\beta, y) = 3\beta_3\gamma_2(\beta)(y^2 - \gamma_2(\beta)) \\ \Xi_3(\theta, y) &= \beta_1(\gamma_4(\beta) + 3\gamma_2^2(\beta)) + \beta_3(\gamma_6(\beta) + 3\gamma_2(\beta)\gamma_4(\beta)) + 3\beta_3(\gamma_4(\beta) + 3\gamma_2^2(\beta))y^2 \\ &\Rightarrow \Xi_{3,c}(\beta, y) = 3\beta_3(\gamma_4(\beta) + 3\gamma_2^2(\beta))(y^2 - \gamma_2(\beta)). \end{aligned}$$

Let $g(\beta, y)$ denote the solution of $Lg(y) = -(y^2 - \gamma_2(\beta))$.

Then, $h_1(\beta, y) = 3\beta_3\gamma_2(\beta)g(\beta, y)$ and $h_3(\beta, y) = 3\beta_3(\gamma_4(\beta) + 3\gamma_2^2(\beta))g(\beta, y)$. Therefore,

$$\begin{aligned} \mathcal{H}(\beta) &= 9\beta_3^2 \int (g'(\beta, y))^2 \pi_\beta(y) dy \begin{pmatrix} \gamma_2^2(\beta) & \gamma_2(\beta)(\gamma_4(\beta) + 3\gamma_2^2(\beta)) \\ \gamma_2(\beta)(\gamma_4(\beta) + 3\gamma_2^2(\beta)) & (\gamma_4(\beta) + 3\gamma_2^2(\beta))^2 \end{pmatrix}, \\ \Sigma(\beta) &= \sigma^2(\mathcal{I}(\beta) + \mathcal{H}(\beta)). \end{aligned}$$

Therefore the asymptotic variance of the MCE is $\sigma^2(\mathcal{I}^{-1}(\beta) + \mathcal{I}^{-1}(\beta)\mathcal{H}(\beta)\mathcal{I}^{-1}(\beta))$.

This example is a special case of a previous work where estimation for ergodic McKean-Vlasov equations with polynomial interactions and no potential term was investigated ([25], Example 3 of Section 3.5.) The estimators, say $(\tilde{\beta}_1, \tilde{\beta}_3)$ are obtained by means of an explicit relation linking them to the empirical moments of (X_t) . Set $\hat{\gamma}_2(T) = \frac{1}{T} \int_0^T X_s^2 ds$, then

$$\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_3 \end{pmatrix} = - \begin{pmatrix} 1 & -3\hat{\gamma}_2(T) \\ 0 & 1 \end{pmatrix} \Psi_T^{-1} \begin{pmatrix} \int_T^{2T} X_s dX_s \\ \int_T^{2T} X_s^3 dX_s \end{pmatrix} \text{ where } \Psi_T = \begin{pmatrix} \int_T^{2T} X_s^2 ds & \int_T^{2T} X_s^4 ds \\ \int_T^{2T} X_s^4 ds & \int_T^{2T} X_s^6 ds \end{pmatrix}.$$

This estimator is consistent and the asymptotic variance of $\sqrt{T}(\tilde{\beta} - \beta)$ is $\sigma^2(\Sigma_1(\beta) + \Sigma_2(\beta))$ with $\Sigma_1(\beta) = \mathcal{I}(\beta)^{-1}$ and $\Sigma_2(\beta) = 9\beta_3^2 \int (g'(\beta, y))^2 \pi_\beta(y) dy \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Thus there is a slight loss of information when using the present nonparametric approach leading to $(\hat{\beta}_1, \hat{\beta}_3)$ instead of the approach of [25] leading to $(\tilde{\beta}_1, \tilde{\beta}_3)$.

Of course, the method of [25] only works when $b(\alpha, x) \equiv 0$ and $\Phi(\beta, x)$ is an odd polynomial. The present method works with a non nul potential term and a more general interaction term.

To be complete, we have seen that $\widehat{m}_1(T) = T^{-1} \int_0^T X_s ds$. For $i = 2, 3$, the moments $\widehat{m}_i(T)$ of $\widehat{\pi}_T$ satisfy, using that K is even and $\sigma_K^2 = \int z^2 K(z) dz < \infty$,

$\widehat{m}_2(T) = \frac{1}{T} \int_0^T ds \int_{\mathbb{R}} (X_s + hz)^2 K(z) dz = \frac{1}{T} \int_0^T X_s^2 ds + \sigma_K^2 h_T^2$,
 $\widehat{m}_3(T) = \frac{1}{T} \int_0^T ds \int (X_s + hz)^3 K(z) dz = \frac{1}{T} \int_0^T X_s^3 ds + 3\sigma_K^2 h_T^2 \widehat{m}_1(T)$.
 As $T \rightarrow \infty$, by the ergodic theorem, $\widehat{m}_1(T) \rightarrow \int y \pi_\beta(y) dy = 0$, $\widehat{m}_2(T) \rightarrow \int y^2 \pi_\beta(y) dy = \gamma_2(\beta)$
 and $\widehat{m}_3(T) \rightarrow \int y^3 \pi_\beta(y) dy = 0$. Therefore $\frac{1}{T}(\widehat{\Gamma}_T - \Psi_T) = o_P(1)$.

5. MAIN LEMMAS

Proposition 5, Proposition 6 and Theorem 3 rely on the following Lemmas. Lemma 2 is especially difficult to obtain and used for Theorem 3. These lemmas use Propositions 1 and 2.

Lemma 1. *Assume [H4] and that $\varphi, H, G : \mathbb{R} \rightarrow \mathbb{R}$ have polynomial growth, then, under \mathbb{P}_θ ,*

$$\begin{aligned}
 D_{11}(\varphi, H) &= \frac{1}{T} \int_T^{2T} \varphi(X_s) H \star (\widehat{\pi}_T - \mathbb{E}_\theta \widehat{\pi}_T)(X_s) ds = O_P\left(\frac{1}{\sqrt{T}}\right), \\
 D_{12}(\varphi, H) &= \frac{1}{T} \int_T^{2T} \varphi(X_s) H \star (\mathbb{E}_\theta \widehat{\pi}_T - \pi_\theta)(X_s) ds = O_P(h_T^2), \\
 D_{21}(H, G) &= \frac{1}{T} \int_T^{2T} H \star (\widehat{\pi}_T - \mathbb{E}_\theta \widehat{\pi}_T)(X_s) G \star (\widehat{\pi}_T - \mathbb{E}_\theta \widehat{\pi}_T)(X_s) ds = O_P\left(\frac{1}{T}\right), \\
 D_{22}(H, G) &= \frac{1}{T} \int_T^{2T} H \star (\mathbb{E}_\theta \widehat{\pi}_T - \pi_\theta)(X_s) G \star (\mathbb{E}_\theta \widehat{\pi}_T - \pi_\theta)(X_s) ds = O_P(h_T^4), \\
 D_{23}(H, G) &= \frac{1}{T} \int_T^{2T} H \star (\mathbb{E}_\theta \widehat{\pi}_T - \pi_\theta)(X_s) G \star (\widehat{\pi}_T - \mathbb{E}_\theta \widehat{\pi}_T)(X_s) ds = O_P\left(\frac{h_T^2}{\sqrt{T}}\right).
 \end{aligned}$$

Lemma 2. *Assume [H4] and that $\varphi, H : \mathbb{R} \rightarrow \mathbb{R}$ have polynomial growth. Then, under \mathbb{P}_θ ,*

$$\begin{aligned}
 \sqrt{T} D_{11}(\varphi, H) &= \frac{1}{T} \int_T^{2T} \varphi(X_s) H \star (\sqrt{T}(\widehat{\pi}_T - \mathbb{E}_\theta \widehat{\pi}_T))(X_s) ds \\
 &= \frac{\sigma}{\sqrt{T}} \int_0^T g'(X_s) dW_s + O_P(h_T \sqrt{T}) + O_P\left(\frac{1}{\sqrt{T}}\right).
 \end{aligned}$$

where g solution of $Lg = -\Psi_{\theta,c}$, $\Psi_\theta(y) = \int \varphi(x) H(x-y) \pi_\theta(x) dx$, $\Psi_{\theta,c}(x) = \Psi_\theta(x) - \int \Psi_\theta(y) \pi_\theta(y) dy$.

The reason why we separate Lemma 1 and Lemma 2 is that Lemma 1 is enough to study the second derivatives of $\Lambda_T(\theta)$ w.r.t. the parameters whereas Lemma 2 is required to study the first derivatives. We take $h_T = o(T^{-1/2})$ so that the middle term of $\sqrt{T} D_{11}(\varphi, H)$ tends to 0.

6. CONCLUDING REMARKS

In this paper, we study the estimation of $\theta = (\alpha, \beta)$ for the process (X_t) given by (1) when the process is in stationary regime. Thus the distribution of X_t is constant and equal to the invariant distribution π_θ . In such a case, (X_t) is equal to an ergodic diffusion. The exact MLE of θ obtained from the continuous observation of (X_t) on $[0, T]$ can be studied theoretically but does not lead to computable estimators since the drift term of (X_t) depends on π_θ and therefore is not explicitly known. To overcome this difficulty, we assume that (X_t) is observed on the time interval $[0, 2T]$ and we build an explicit contrast based on the conditional likelihood of $(X_t, t \in [T, 2T])$ given X_T , where π_θ in the drift is replaced by a nonparametric estimator $\widehat{\pi}_T$ computed from $(X_t, t \in [0, T])$. This leads us to study a weighted mean integrated risk for $\widehat{\pi}_T$ yielding a new result for this estimator. Then, we prove that the MCE is asymptotically Gaussian with rate \sqrt{T} with explicit asymptotic variance. Several classes of models are discussed.

In the continuation of this work, it is interesting to consider the same estimation problem for discrete observations or multidimensional McKean-Vlasov diffusions.

Here, our assumptions ensure uniqueness of an invariant distribution. But concrete models having more than one invariant distribution are proposed in [18] or [28] (see also [46]). The estimation problem in this case certainly raises new considerations, which are worthwhile.

7. APPENDIX

Based on the ergodic theorem (10), the central limit theorem for stochastic integrals (see *e.g.* [35], Chap.1) can be proved and states that, if $f \in \mathbb{L}^2(\pi_\theta(x)dx)$,

$$(32) \quad \frac{1}{\sqrt{T}} \int_0^T f(X_s) dW_s \rightarrow_{\mathcal{L}} \mathcal{N}(0, \int f^2(x) \pi_\theta(x) dx).$$

Consider now the central limit theorem associated with (10), *i.e.* the limiting distribution of $\sqrt{T}(\frac{1}{T} \int_0^T f(X_s) ds - \int f(x) \pi_\theta(x) dx)$. The results below can be found in [22].

Let L denote the infinitesimal generator of the SDE (9) defined in (11). The operator L acts on $\mathbb{L}^2(\pi_\theta(x)dx)$ and is defined on the domain \mathcal{D} (see (12)). Note that

$$(33) \quad Lg = \frac{\sigma^2}{2\pi_\theta} (g' \pi_\theta)'$$

Proposition 8. *Assume [H4]. Let $f \in \mathbb{L}^2(\pi_\theta(x)dx)$, set $f_c = f - \int_{\mathbb{R}} f(x) \pi_\theta(x) dx$ and denote by $\langle \cdot, \cdot \rangle_{\pi_\theta}$ the scalar product of $\mathbb{L}^2(\pi_\theta(x)dx)$. If $f_c \in \text{Range}(\mathcal{D})$, where $\text{Range}(\mathcal{D}) = L(\mathcal{D})$ is the image of \mathcal{D} by L , then, as T tends to infinity, under \mathbb{P}_θ ,*

$$(34) \quad \frac{1}{\sqrt{T}} \int_0^T f_c(X_s) ds \rightarrow_{\mathcal{L}} \mathcal{N}(0, \sigma^2(f_c))$$

where $\sigma^2(f_c) = -2\langle f_c, g \rangle_{\pi_\theta}$ and g is any element of \mathcal{D} satisfying $Lg = f_c$. Moreover,

$$\text{Var} \left(\frac{1}{\sqrt{T}} \int_0^T f_c(X_s) ds \right) \rightarrow \sigma^2(f_c).$$

The relation $-2\langle f_c, g \rangle_{\pi_\theta} = -2\langle Lg, g \rangle_{\pi_\theta} = \sigma^2 \int_{\mathbb{R}} (g'(x))^2 \pi_\theta(x) dx$ holds.

Proposition 8 is exactly Theorem 2.2 in [22]. It is a well known result on ergodic diffusions. In the proofs, result (34) will be clarified when used.

Corollary 2. *Let h_1, \dots, h_p be functions belonging to $\text{Range}(\mathcal{D})$ such that $\int h_j(x) \pi_\theta(x) dx = 0$, for $j = 1, \dots, p$. Define*

$$V_\theta(h_i, h_j) = 4\sigma^{-2} \int_{\mathbb{R}} \pi_\theta^{-1}(x) \left(\int_{-\infty}^x h_i(y) \pi_\theta(y) dy \int_{-\infty}^x h_j(y) \pi_\theta(y) dy \right) dx$$

so that $\sigma^2(h_i) = V_\theta(h_i, h_i)$. The vector $\frac{1}{\sqrt{T}} (\int_0^T h_i(X_s) ds, i = 1, \dots, p)' \rightarrow_{\mathcal{L}} \mathcal{N}_p(0, V(\theta))$ with $V(\theta) = (V_\theta(h_i, h_j), 1 \leq i, j \leq p)$.

Corollary 2 is a straightforward consequence of Proposition 8 using the Cramér-Wold device. The following proposition is little known, and its proof is given below.

Proposition 9. *Assume [H4]. Let $\gamma_\theta(x) = 2\sigma^{-1}(b(\alpha, x) + \Phi(\beta, \cdot) \star \pi_\theta(x))$. Then,*

$$(35) \quad \lim_{x \rightarrow +\infty} \gamma_\theta^{-1}(x) = 0, \quad \lim_{x \rightarrow -\infty} \gamma_\theta^{-1}(x) = 0.$$

This implies that $\text{Range}(\mathcal{D}) = \{h \in \mathbb{L}^2(\pi_\theta(x)dx), \int h(x) \pi_\theta(x) dx = 0\}$ so that the central limit theorem associated to (10) holds for all $f \in \mathbb{L}^2(\pi_\theta(x)dx)$.

Proof of Proposition 9. We rely on results stated in Proposition 2.2 in [22]. First note that $\forall g \in \mathcal{D}, \int Lg(x)\pi_\theta(x)dx = 0$ (see (12)). Thus $\text{Range}(\mathcal{D}) \subset \{h \in \mathbb{L}^2(\pi_\theta(x)dx), \int h(x)\pi_\theta(x)dx = 0\}$. But the other inclusion does not hold except if the process is ρ -mixing, that is if its generator has a spectral gap. In [22], it is proved that condition (35) is a necessary and sufficient condition for ρ -mixing (Proposition 2.8). However, this proposition is stated for SDEs with locally Lipschitz coefficients having linear growth, which were standard assumptions for classical SDEs.

Here, this assumption is not satisfied by (Y_t) defined in (9). Nevertheless, this assumption is not mandatory, the only requirement being that the SDE admits a unique strong solution and an invariant distribution. (see the proof in the Appendix p.1074-1077). Therefore, we can apply these results here and check that γ_θ satisfies (35) to get Proposition 9.

Set, for the proof, $b(\alpha, x) = b(x), \Phi(\beta, x) = \Phi(x), \pi_\theta(x) = \pi(x), K(\alpha) = K, \lambda(\beta) = \lambda$ and $\gamma(x) = 2\sigma^{-1}(b(x) + \Phi \star \pi(x))$.

We have $\gamma(x) = 2\sigma^{-1}(V'(x) + W' \star \pi)$. As $V''(x) \geq K$, for $x \geq 0, V'(x) \geq Kx + V'(0)$. For $x \leq 0, V'(x) \leq Kx + V'(0)$. As $W''(x) \geq \lambda$ and W' is odd, we have for $x \geq 0, W'(x) \geq \lambda x$. For $x \leq 0, W'(x) \leq \lambda x$. Thus, for $x \geq 0$,

$$(36) \quad V'(x) + W' \star \pi(x) \geq (K + \lambda)x + V'(0) - \lambda \int y\pi(y)dy,$$

which implies $\lim_{x \rightarrow +\infty} \gamma(x) = +\infty$ as $K + \lambda > 0$. Analogously, for $x \leq 0$,

$$(37) \quad V'(x) + W' \star \pi(x) \leq (K + \lambda)x + V'(0) - \lambda \int y\pi(y)dy,$$

which implies $\lim_{x \rightarrow -\infty} \gamma(x) = -\infty$. Hence, $\lim_{x \rightarrow -\infty} \gamma^{-1}(x) = 0$. \square

Remark 1. Using (33) and (12), we can solve $Lg = f_c = f - \int_{\mathbb{R}} f(y)\pi_\theta(y)dy$ and obtain

$$g'(x) = \frac{2}{\sigma^2\pi_\theta(x)} \int_{-\infty}^x f_c(y)\pi_\theta(y)dy = -\frac{2}{\sigma^2\pi_\theta(x)} \int_x^{+\infty} f_c(y)\pi_\theta(y)dy.$$

The above relation holds since $\int_{-\infty}^{+\infty} f_c(x)\pi_\theta(x)dx = 0$. By Proposition 9, Condition (35) holds, so that, for all $f \in \mathbb{L}^2(\pi_\theta(x)dx)$, the integral

$$\sigma^2(f_c) = \sigma^2 \int_{\mathbb{R}} (g'(x))^2 \pi_\theta(x)dx = 4\sigma^{-2} \int_{\mathbb{R}} \pi_\theta^{-1}(x) \left(\int_{-\infty}^x f_c(y)\pi_\theta(y)dy \right)^2 dx < \infty.$$

We can choose $g(u) = \int_0^u g'(v)dv$. Note that $\sigma^2(f_c) < \infty$ is not obvious as $\int \pi_\theta^{-1}(x)dx = +\infty$.

8. PROOFS OF SECTION 2.

Proof of Proposition 1. Under the assumptions of the proposition, (1) admits a unique invariant distribution having a finite second order moment. For the proof, set $V(\alpha, \cdot) = V, W(\beta, \cdot) = W, \pi_\theta = \pi, K(\alpha) = K, \lambda(\beta) = \lambda$.

By [H1], for all $x \in \mathbb{R}, V(x) \geq K\frac{x^2}{2} + V'(0)x$. And, as W is even, for all $x \in \mathbb{R}, W(x) \geq \lambda\frac{x^2}{2}$. Therefore,

$$\begin{aligned} W \star \pi(x) &= \int_{\mathbb{R}} W(x-y)\pi(y)dy \geq \int_{\mathbb{R}} \lambda \frac{(x-y)^2}{2} \pi(y)dy \\ &= \lambda \frac{x^2}{2} - \lambda x \int_{\mathbb{R}} y\pi(y)dy + \frac{\lambda}{2} \int_{\mathbb{R}} y^2 \pi(y)dy. \end{aligned} \quad \text{Hence, the result.} \square$$

9. PROOFS OF SECTION 3

We first study the properties of the nonparametric estimator of π_θ .

9.1. Proofs of Section 3.2. We study the properties of the nonparametric estimator of π_θ under \mathbb{P}_θ .

Proof of Theorem 1. (1) The beginning of the proof is classical (see Proposition 3.3 in [13] or [51]). Set $\pi_\theta = \pi$, $\mathbb{E}_\theta = \mathbb{E}$. We have:

$$\mathbb{E}\widehat{\pi}_T(x) = \int_{\mathbb{R}} K_{h_T}(y-x)\pi(y)dy = K_{h_T} \star \pi(x) = \int K(v)\pi(x+vh_T)dv.$$

By the Taylor formula, $\pi(x+vh_T) - \pi(x) = vh_T \int_0^1 \pi'(x+\tau vh_T)d\tau$. As $\int vK(v)dv = 0$,

$$\mathbb{E}\widehat{\pi}_T(x) - \pi(x) = \int vh_T K(v)F(x,v)dv, \quad F(x,v) = \int_0^1 [\pi'(x+\tau vh_T) - \pi'(x)]d\tau.$$

We apply the generalized Minkowski inequality which states that, for all r ,

$$\left[\int_X \left(\int_V F(x,v)d\nu(v) \right)^r d\mu(x) \right]^{\frac{1}{r}} \leq \int_V \left(\int_X F^r(x,v)d\mu(x) \right)^{\frac{1}{r}} d\nu(v)$$

with $r = 2$, $d\mu(x) = (1+|x|^p)dx$, $d\nu(v) = |v|K(v)dv$. This yields

$$(38) \quad \int_{\mathbb{R}} (1+|x|^p)(\mathbb{E}\widehat{\pi}_T(x) - \pi(x))^2 dx \leq h_T^2 \left(\int |v|K(v)dv \left[\int F^2(x,v)(1+|x|^p)dx \right]^{1/2} \right)^2.$$

Next, we apply a second time the generalized Minkowski inequality with $r = 2$, $d\mu(x) = (1+|x|^p)dx$, $d\nu(\tau) = 1_{[0,1]}(\tau)d\tau$. This yields

$$\left[\int F^2(x,v)(1+|x|^p)dx \right]^{1/2} \leq \int_0^1 d\tau \left(\int (1+|x|^p)[\pi'(x+\tau vh_T) - \pi'(x)]^2 dx \right)^{1/2}$$

Now, we study, for $t \in \mathbb{R}$,

$$\left(\int (1+|x|^p)[\pi'(x+t) - \pi'(x)]^2 dx \right)^{1/2}$$

Using the Taylor formula and the Cauchy-Schwarz inequality yields

$[\pi'(x+t) - \pi'(x)]^2 \leq t^2 \int_0^1 [\pi''(x+ut)]^2 du$ where

$$\pi''(x) = \pi(x)h(x), \quad h(x) = -\frac{2}{\sigma^2}[b'(x) + \Phi' \star \pi(x)] + \frac{4}{\sigma^4}[b(x) + \Phi \star \pi(x)]^2$$

By [S1], we check that $|h(x)| \lesssim (1+|x|^{2\ell})$. Thus,

$$[\pi'(x+t) - \pi'(x)]^2 \lesssim t^2 \int_0^1 \pi^2(x+ut)(1+x^{2\ell} + (ut)^{2\ell})du.$$

This implies,

$$\begin{aligned} \int (1+|x|^p)[\pi'(x+t) - \pi'(x)]^2 dx &\lesssim t^2 \int (1+|x|^p) \int_0^1 \pi^2(x+ut)(1+x^{2\ell} + (ut)^{2\ell})du \\ &\lesssim t^2 \int_0^1 du \int (1+|x|^p)\pi^2(x+ut)(1+x^{2\ell} + (ut)^{2\ell})dx \\ &\lesssim t^2 \int_0^1 du \int (1+|y-ut|^p)\pi^2(y)(1+(y-ut)^{2\ell} + t^{2\ell})dy. \end{aligned}$$

Now, $|y - ut|^p \lesssim (|y|^p + |t|^p)$ and $(y - ut)^{2\ell} \lesssim (y^{2\ell} + t^{2\ell})$. Thus,

$$\int (1 + |x|^p)[\pi'(x+t) - \pi'(x)]^2 dx \lesssim t^2 \int (1 + (|y|^p + |t|^p))(1 + y^{2\ell} + t^{2\ell})\pi^2(y)dy = t^2 f(t),$$

where f is a positive continuous function on \mathbb{R} . Next, using (38)

$$\int_{\mathbb{R}} (1 + |x|^p)(\mathbb{E}\widehat{\pi}_T(x) - \pi(x))^2 dx \lesssim h_T^2 \left[\int dv |v|K(v) \int_0^1 d\tau |\tau v h_T| \sqrt{f(\tau v h_T)} d\tau \right]^2.$$

We note that $0 < h_T < 1$, K has compact support $[-A, A]$ and finally obtain

$$\int_{\mathbb{R}} (1 + |x|^p)(\mathbb{E}\widehat{\pi}_T(x) - \pi(x))^2 dx \lesssim h_T^4 \left(\int v^2 K(v) dv \right)^2 \sup_{t \in [-A, A]} f(t).$$

(2) Next, we study the mean integrated variance term. The proof uses definitions (12), (33), Proposition 1, Proposition 9 and Remark 1. Let $F_\pi(x) = \int_{-\infty}^x \pi(v)dv$ and set $h_T = h$, $K_{h_T} = K_h$ (see (13)). Recall that K is even.

We start by a property of the invariant distribution. For all (x, z) ,

$$\frac{\pi(z)}{\pi(x)} = \exp \left\{ -\frac{2}{\sigma^2} [V(z) - V(x) + W \star \pi(z) - W \star \pi(x)] \right\}.$$

Using (36), we get, for $0 \leq x \leq z$,

$$V(z) - V(x) + W \star \pi(z) - W \star \pi(x) \geq \int_x^z [(K + \lambda)t + c] dt = \frac{(K + \lambda)}{2} (z^2 - x^2) + c(z - x),$$

where $K = K(\alpha)$, $\lambda = \lambda(\beta)$, $c = \frac{\partial V}{\partial x}(0) - \lambda \int y \pi(y) dy$. Therefore,

$$(39) \quad \text{for } 0 \leq x \leq z, \quad \frac{\pi(z)}{\pi(x)} \leq \exp [C(z - x)], \quad C = -\frac{2}{\sigma^2} c.$$

Analogously, using (37),

$$(40) \quad \text{for } z \leq x \leq 0, \quad \frac{\pi(z)}{\pi(x)} \leq \exp [C(z - x)], \quad C = -\frac{2}{\sigma^2} c.$$

Now, we have:

$$\widehat{\pi}_T(x) - \mathbb{E}\widehat{\pi}_T(x) = \frac{1}{T} \int_0^T (K_h(X_t - x) - \mathbb{E}K_h(X_t - x)) dt.$$

For each x , we look for $f_x \in \mathcal{D}$ such that $Lf_x(y) = -[K_h(y - x) - \int_{\mathbb{R}} K_h(z - x)\pi(z)dz]$. Using (33) and Remark 1, we take $f_x(u) = \int_0^u g_x(v)dv$ where

$$\begin{aligned} g_x(u) &= \frac{2}{\sigma^2 \pi(u)} \int_{-\infty}^u \left(K_h(y - x) - \int_{\mathbb{R}} K_h(z - x)\pi(z)dz \right) \pi(y) dy \\ &= -\frac{2}{\sigma^2 \pi(u)} \int_u^{+\infty} \left(K_h(y - x) - \int_{\mathbb{R}} K_h(z - x)\pi(z)dz \right) \pi(y) dy \end{aligned}$$

Equivalently,

$$\begin{aligned} g_x(u) &= -\frac{2}{\sigma^2 \pi(u)} \left[\int_{-\infty}^u K_h(y - x)\pi(y)dy - F_\pi(u) \int_{\mathbb{R}} K_h(z - x)\pi(z)dz \right] \\ &= -\frac{2}{\sigma^2 \pi(u)} \left[\int_u^{+\infty} K_h(y - x)\pi(y)dy - (1 - F_\pi(u)) \int_{\mathbb{R}} K_h(z - x)\pi(z)dz \right]. \end{aligned}$$

By the Ito formula,

$$f_x(X_T) - f_x(X_0) = \sigma \int_0^T g_x(X_t) dW_t + \int_0^T Lf_x(X_t) dt.$$

Therefore,

$$\widehat{\pi}_T(x) - \mathbb{E}\widehat{\pi}_T(x) = -\frac{\sigma}{T} \int_0^T g_x(X_t) dW_t + \frac{1}{T} [f_x(X_T) - f_x(X_0)].$$

This implies

$$\mathbb{E}(\widehat{\pi}_T(x) - \mathbb{E}\widehat{\pi}_T(x))^2 \leq 2\frac{\sigma^2}{T} \mathbb{E}g_x^2(X_0) + \frac{4}{T^2} \mathbb{E}f_x^2(X_0).$$

Therefore, it remains to prove that

$$(41) \quad I = \int_{\mathbb{R}} (1 + |x|^p) \mathbb{E}g_x^2(X_0) dx < +\infty, \quad J = \int_{\mathbb{R}} (1 + |x|^p) \mathbb{E}f_x^2(X_0) dx < +\infty.$$

Consider first I :

$$I = \int_{\mathbb{R}} (1 + |x|^p) \int_{\mathbb{R}} g_x^2(u) \pi(u) du dx = \frac{4}{\sigma^4} (I_+ + I_-)$$

with

$$\begin{aligned} I_+ &= \int_{u>0} \frac{1}{\pi(u)} du \int_{\mathbb{R}} (1 + |x|^p) \left[\int_u^{+\infty} \left(K_h(y-x) - \int_{\mathbb{R}} K_h(z-x) \pi(z) dz \right) \pi(y) dy \right]^2 dx \\ &\leq 2(I_+^{(1)} + I_+^{(2)}), \end{aligned}$$

where

$$\begin{aligned} I_+^{(1)} &= \int_{u>0} \frac{1}{\pi(u)} du \int_{\mathbb{R}} (1 + |x|^p) \left(\int_u^{+\infty} K_h(y-x) \pi(y) dy \right)^2 dx \\ I_+^{(2)} &= \int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)} \int_{\mathbb{R}} (1 + |x|^p) \left[\int_{\mathbb{R}} K_h(z-x) \pi(z) dz \right]^2 dx \end{aligned}$$

and

$$\begin{aligned} I_- &= \int_{u<0} \frac{1}{\pi(u)} du \int_{\mathbb{R}} ((1 + |x|^p) \left[\int_{-\infty}^u \left(K_h(y-x) - \int_{\mathbb{R}} K_h(z-x) \pi(z) dz \right) \pi(y) dy \right]^2 dx \\ &\leq 2(I_-^{(1)} + I_-^{(2)}), \end{aligned}$$

with

$$\begin{aligned} I_-^{(1)} &= \int_{u<0} \frac{1}{\pi(u)} du \int_{\mathbb{R}} (1 + |x|^p) \left(\int_{-\infty}^u K_h(y-x) \pi(y) dy \right)^2 dx \\ I_-^{(2)} &= \int_{u<0} du \frac{(F_\pi(u))^2}{\pi(u)} \int_{\mathbb{R}} (1 + |x|^p) \left[\int_{\mathbb{R}} K_h(z-x) \pi(z) dz \right]^2 dx. \end{aligned}$$

Consider first $I_+^{(1)}$. By the change of variables $v = (y-x)/h$, we obtain

$$\begin{aligned} I_+^{(1)} &= \int_{u>0} \frac{1}{\pi(u)} du \int_{\mathbb{R}} (1 + |x|^p) \left(\int_{(x+hv)\geq u} K(v) \pi(x+hv) dv \right)^2 dx \\ &= \int_{u>0, v, v' \in [-A, A]} \frac{1}{\pi(u)} K(v) K(v') du dv dv' \int_{(x+hv)\geq u, (x+hv')\geq u} \pi(x+hv) \pi(x+hv') (1 + |x|^p) dx. \end{aligned}$$

By (39), for $0 \leq u \leq x + hv$, $\frac{\pi(x+uv)}{\pi(u)} \leq \exp C(x + hv - u)$ and for $0 \leq u \leq x + hv'$, $\frac{\pi(x+uv')}{\pi(u)} \leq \exp C(x + hv' - u)$. This implies, for $0 < h < 1, v, v' \in [-A, A]$,

$$\begin{aligned} \pi(x + hv)\pi(x + hv') &\leq \pi^{3/2}(u)\pi^{1/2}(x + hv') \exp[(3C/2)(x - u) + Chv + (C/2)hv'] \\ &\leq \pi^{3/2}(u)\pi^{1/2}(x + hv') \exp[(3C/2)(x - u)] \exp(3A|C|/2). \end{aligned}$$

Thus,

$$I_+^{(1)} \leq \int_{u>0, -A \leq v, v' \leq A} e^{(-2Cu)} \frac{\pi^{3/2}(u)}{\pi(u)} K(v)K(v') dudvdv' \int_{\mathbb{R}} e^{[(3C/2)x]} \pi^{1/2}(x + hv')(1 + |x|^p) dx.$$

Now, since $v' \in [-A, A]$,

$$\begin{aligned} &\int_{\mathbb{R}} \exp[(3C/2)x] \pi^{1/2}(x + hv')(1 + |x|^p) dx \lesssim \\ &\int_{\mathbb{R}} \exp[(3|C|/2)(|z| + A)] \pi^{1/2}(z)(1 + (|z| + A)^p) dz := B < +\infty. \end{aligned}$$

Thus,

$$\begin{aligned} I_+^{(1)} &\lesssim B \int_{u>0, v, v' \in [-A, A]} \exp(-2Cu) \pi^{1/2}(u) K(v)K(v') dudvdv' \\ &= B \int_{u>0} \exp(-2Cu) \pi^{1/2}(u) du < +\infty. \end{aligned}$$

Now, we look at the other term $I_+^{(2)}$. It holds that

$$\begin{aligned} I_+^{(2)} &= \int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)} \int_{v, v' \in [-A, A]} K(v)K(v') \int_{\mathbb{R}} (1 + |x|^p) [\pi(x + hv)\pi(x + hv')] dx dv dv' \\ &\lesssim \int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)} \int_{\mathbb{R}} (1 + (|z| + A)^p) \pi(z) dz \lesssim \int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)}, \end{aligned}$$

as π is bounded.

Now,

$$\int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)} = 2 \int_0^{+\infty} \frac{du}{\pi(u)} \int_{u < v < v'} \pi(v)\pi(v') dv dv'$$

For $0 \leq u \leq v \leq v'$, using (39),

$$\pi(v)\pi(v') = [\pi(v)\pi(v')]^{1/4} [\pi(v)\pi(v')]^{3/4} \leq [\pi(v)\pi(v')]^{1/4} [\pi(u)]^{3/2} \exp(C(v + v' - 2u)).$$

Therefore,

$$\int_{u>0} du \frac{(1 - F_\pi(u))^2}{\pi(u)} \leq 2 \int_0^{+\infty} \pi^{1/2}(u) \exp(-2Cu) du \left(\int_0^{+\infty} \exp(Cv) \pi^{1/4}(v) dv \right)^2 < +\infty.$$

The term I_- can be treated analogously using (40). Hence the first integral I of (41) is finite. We turn to the second one.

$$J = \int (1 + |x|^p) \pi(u) \left(\int_0^u g_x(v) dv \right)^2 dx du \leq J_+ + J_-, \quad \text{with}$$

$$J_+ = \int (1 + |x|^p) \int_{u>0} u \pi(u) \int_0^u g_x^2(v) dv dudx, \quad J_- = \int (1 + |x|^p) \int_{u<0} \pi(u) |u| \int_u^0 g_x^2(v) dv dudx.$$

We only treat J_+ as J_- is analogous. We have

$$g_x(v) = -\frac{2}{\sigma^2\pi(v)} \left[\int_{x+hz \geq v} K(z)\pi(zh+x)dz - (1-F_\pi(v)) \int K(z)\pi(x+zh)dz \right],$$

so that

$$g_x^2(v) \leq \frac{4}{\sigma^4} \frac{1}{\pi^2(v)} \left[\left(\int_{x+hz \geq v} K(z)\pi(zh+x)dz \right)^2 + (1-F_\pi(v))^2 \left(\int K(z)\pi(x+zh)dz \right)^2 \right].$$

Thus, $J_+ \leq \frac{4}{\sigma^4}(J_+^{(1)} + J_+^{(2)})$ where

$$J_+^{(1)} = \int (1+|x|^p) \int_{u>0} \pi(u)u \int_0^u \frac{1}{\pi^2(v)} \left[\left(\int_{x+hz \geq v} K(z)\pi(zh+x)dz \right)^2 \right] dudvdx$$

$$J_+^{(2)} = \int (1+|x|^p) \int_{u>0} \pi(u)u \int_0^u \frac{1}{\pi^2(v)} (1-F_\pi(v))^2 \left(\int K(z)\pi(x+zh)dz \right)^2 dudvdx.$$

We look at $J_+^{(1)}$.

$$J_+^{(1)} = \int (1+|x|^p) \int_{u>0} \pi(u)u \int_0^u \frac{1}{\pi^2(v)} \int_{x+zh \geq v, x+z'h \geq v} K(z)K(z')\pi(zh+x)\pi(z'h+x)dzdz'dudvdx.$$

For $0 < v < u$ and $v \leq x+zh, v \leq x+z'h$, we write

$$\frac{\pi(u)\pi(x+zh)\pi(x+z'h)}{\pi^2(v)} = \pi^{1/3}(u) \left[\frac{\pi(u)}{\pi(v)} \right]^{2/3} \left[\frac{\pi(x+zh)}{\pi(v)} \right]^{2/3} \left[\frac{\pi(x+z'h)}{\pi(v)} \right]^{2/3} [\pi(x+zh)\pi(x+z'h)]^{1/3}$$

where

$$\left[\frac{\pi(u)}{\pi(v)} \right]^{2/3} \left[\frac{\pi(x+zh)}{\pi(v)} \right]^{2/3} \left[\frac{\pi(x+z'h)}{\pi(v)} \right]^{2/3} \leq \exp(2/3)C(u-v) \exp((2/3)C(x+zh-v)) \exp((2/3)C(x+z'h-v)).$$

Therefore, as π is bounded,

$$J_+^{(1)} \lesssim \int_{u>0} u\pi^{1/3}(u) \int_0^u \exp(-2Cv)dv$$

$$\left[\int K(z)K(z') \int_{\mathbb{R}} [\pi(x+zh)]^{1/3} (1+|x|^p) \exp[(2/3)C(x+(z+z')h)]dx \right] dzdz'du$$

$$\lesssim \int_{u>0} u\pi^{1/3}(u) \int_0^u \exp(-2Cv)dvdu \times \int_{\mathbb{R}} \pi^{1/3}(y)(1+|y|+A)^p \exp[(2/3)Cy]dy < +\infty.$$

The fact that $J_+^{(2)} < +\infty$ is simpler and omitted.

Therefore, we conclude that (41) holds which implies the result. \square

Proof of Theorem 2. Set $\pi_\theta = \pi$, $\mathbb{E}_\theta = \mathbb{E}$ again. We study

$$(42) \quad I_T = \int dy \sqrt{T} (\widehat{\pi}_T(y) - \mathbb{E}\widehat{\pi}_T(y)) \psi(y) dy,$$

with, for some non negative c , $|\psi(x)| + |\psi'(x)| \leq c(1+|x|^c)$. Set

$$F_h(\xi, y) = K_h(\xi - y) - K_h \star \pi(y) = K_h(\xi - y) - \int K(v)\pi(y+vh)dv.$$

Thus, $I_T = \frac{1}{\sqrt{T}} \int_0^T S_{h_T}(X_s) ds$ with

$$\begin{aligned} S_h(\xi) &= \int dy \psi(y) F_h(\xi, y) = \int K(\tau) \psi(\xi + \tau h) d\tau - \int dy \psi(y) \int K(v) \pi(y + vh) dv \\ (43) \quad &= S(\xi) + \int K(\tau) [\psi(\xi + \tau h) - \psi(\xi)] d\tau - \int dy \psi(y) \int K(v) [\pi(y + vh) - \pi(y)] dv \end{aligned}$$

and $S(\xi) = \psi(\xi) - \int \psi(y) \pi(y) dy$. As ψ is C^1 and $0 < h < 1$,

$$|\psi(\xi + h\tau) - \psi(\xi)| = |h\tau \int_0^1 \psi'(\xi + hu\tau) du| \lesssim |h\tau| (1 + |\xi|^c + |\tau|^c).$$

Therefore, as K is compactly supported,

$$\left| \int K(\tau) [\psi(\xi + \tau h) - \psi(\xi)] d\tau \right| \lesssim |h| \left((1 + |\xi|^c) \int |K(\tau)| d\tau + \int |K(\tau)| |\tau|^c d\tau \right) \lesssim |h| (1 + |\xi|^c).$$

For the third term of (43), we can write:

$$\begin{aligned} \int dy \psi(y) \int K(v) [\pi(y + vh) - \pi(y)] dv &= \int dy \psi(y) \int K(v) hv \int_0^1 \pi'(y + h\tau v) d\tau dv \\ &= \int_0^1 d\tau \int hv K(v) dv \int dz \psi(z - h\tau v) \pi'(z). \end{aligned}$$

By our assumptions, $|\pi'(z)| \lesssim \pi(z)(1 + |z|^c)$ and for $v \in [-A, A]$, $|\psi(z - h\tau v)| \lesssim (1 + |z|^c + A^c)$. This yields $\int dy \psi(y) \int K(v) [\pi(y + vh) - \pi(y)] dv \lesssim |h|$. Therefore,

$$|S_h(\xi) - S(\xi)| \lesssim |h| (1 + |\xi|^c).$$

This yields:

$$\mathbb{E}_\theta \left(\frac{1}{\sqrt{T}} \int_0^T [S_{h_T}(X_s) - S(X_s)] ds \right)^2 \leq T \int [S_{h_T}(x) - S(x)]^2 \pi(x) dx = Th_T^2 O(1),$$

which implies

$$I_T = \frac{1}{\sqrt{T}} \int_0^T S(X_s) ds + O(\sqrt{Th_T}).$$

By Propositions 8 and 9, we can find $g \in \mathbb{L}^2(\pi)$ such that $Lg = -S$ with $g(x) = \int_0^x g'(u) du$, g' given by (21) and

$$\frac{1}{\sqrt{T}} \int_0^T S(X_s) ds = \frac{\sigma}{\sqrt{T}} \int_0^T g'(X_s) dW_s + \frac{1}{\sqrt{T}} (g(X_0) - g(X_T))$$

where $\frac{1}{\sqrt{T}} (g(X_0) - g(X_T)) = O_P(\frac{1}{\sqrt{T}})$ as (X_t) is stationary and $g \in \mathbb{L}^2(\pi)$.

Therefore, (42) can be written as

$$I_T = \frac{\sigma}{\sqrt{T}} \int_0^T g'(X_s) dW_s + O_P(\frac{1}{\sqrt{T}}) + O_P(\sqrt{Th_T}).$$

The proof of Theorem 2 is complete. \square

Proof of Corollary 1. We write

$$\sqrt{T}(\hat{\pi}_T(y) - \pi(y)) = \sqrt{T}(\hat{\pi}_T(y) - \mathbb{E}\hat{\pi}_T(y)) + \sqrt{T}(\mathbb{E}\hat{\pi}_T(y) - \pi(y)).$$

Next, choosing $p - 2c > 1$,

$$\begin{aligned} \left(\int \psi(y) \sqrt{T} (\mathbb{E} \hat{\pi}_T(y) - \pi(y)) dy \right)^2 &\leq \int \frac{\psi^2(y)}{1 + |y|^p} dy \times \int dy (\mathbb{E} (\sqrt{T} (\hat{\pi}_T(y) - \pi(y))))^2 (1 + |y|^p) dy \\ &\lesssim Th_T^4. \end{aligned}$$

Thus,

$$\int \psi(y) \sqrt{T} (\hat{\pi}_T(y) - \pi(y)) dy = \int \psi(y) \sqrt{T} (\hat{\pi}_T(y) - \mathbb{E} \hat{\pi}_T(y)) dy + O(\sqrt{T} h_T^2).$$

Thanks to Theorem 2 and the central limit theorem (32), the proof of Corollary 1 is complete. \square

9.2. Proofs of Section 3.3. Assumptions [H2]-[H3] ensure that all the functions used in the proofs below satisfy the assumptions of Lemma 1 and Lemma 2. As for the likelihood, we have to study $\Lambda_T(\theta) - \Lambda_T(\theta_0)$.

Proof of Proposition 5. We can write

$$\begin{aligned} \sigma^2 (\Lambda_T(\theta) - \Lambda_T(\theta_0)) &= -\frac{1}{2} \int_T^{2T} (\hat{S}(\theta, X_s) - \hat{S}(\theta_0, X_s))^2 ds \\ &\quad + \int_T^{2T} (\hat{S}(\theta, X_s) - \hat{S}(\theta_0, X_s)) \Phi(\beta_0, \cdot) \star (\hat{\pi}_T - \pi_{\theta_0})(X_s) ds \\ &\quad + \sigma \int_T^{2T} (\hat{S}(\theta, X_s) - \hat{S}(\theta_0, X_s)) dW_s = A_T + B_T + C_T. \end{aligned}$$

Under \mathbb{P}_{θ_0} , by Lemma 1,

$$\frac{1}{T} \int_T^{2T} (\Phi(\beta, \cdot) - \Phi(\beta_0, \cdot)) \star (\hat{\pi}_T - \pi_{\theta_0})(X_s) ds = O_P\left(\frac{1}{\sqrt{T}}\right) + O_P(h_T^2).$$

Therefore, $\frac{1}{T} A_T \rightarrow -\frac{1}{2} \int [-b(\alpha, x) + b(\alpha_0, x) - (\Phi(\beta, \cdot) - \Phi(\beta_0, \cdot)) \star \pi_{\theta_0}(x)]^2 \pi_{\theta_0}(x) dx$.

For the second term, $\frac{1}{T} B_T = O_P\left(\frac{1}{\sqrt{T}}\right) + O_P(h_T^2) \rightarrow 0$ by Lemma 1.

For C_T , we have that $\frac{1}{T} < C_T \leq \frac{1}{T} A_T$, so that, by the central limit theorem for stochastic integrals, $\frac{1}{\sqrt{T}} C_T$ converges in distribution. This implies $\frac{1}{T} C_T \rightarrow 0$.

Joining these results yields that, under \mathbb{P}_{θ_0} , using (14), $\frac{1}{T} (\Lambda_T(\theta) - \Lambda_T(\theta_0)) \rightarrow -\frac{1}{2\sigma^2} \tilde{K}(\theta_0, \theta)$. \square

Proof of Proposition 6. To obtain the limiting distribution of the normalized MCE under \mathbb{P}_θ , we need study the derivatives of the contrast with respect to the parameters $\alpha = (\alpha_j, j = 1, \dots, d), \beta = (\beta_k, k = 1, \dots, d')$. Using (3) yields

$$\begin{aligned} \sigma^2 \frac{\partial \Lambda_T}{\partial \alpha_j}(\theta) &= - \int_T^{2T} \left[\frac{\partial b}{\partial \alpha_j}(\alpha, X_t) \right] dX_t + \int_T^{2T} \hat{S}(\theta, X_t) \frac{\partial b}{\partial \alpha_j}(\alpha, X_t) dt \\ \sigma^2 \frac{\partial \Lambda_T}{\partial \beta_k}(\theta) &= - \int_T^{2T} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \hat{\pi}_T(X_t) dX_t + \int_T^{2T} \hat{S}(\theta, X_t) \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \hat{\pi}_T(X_t) dt \end{aligned}$$

$$\begin{aligned}
\sigma^2 \frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \alpha_{j'}}(\theta) &= - \int_T^{2T} \frac{\partial^2 b}{\partial \alpha_j \partial \alpha_{j'}}(\alpha, X_t) dX_t \\
&+ \int_T^{2T} \widehat{S}(\theta, X_t) \frac{\partial^2 b}{\partial \alpha_j \partial \alpha_{j'}}(\alpha, X_t) dt - \int_T^{2T} \frac{\partial b}{\partial \alpha_j}(\alpha, X_t) \frac{\partial b}{\partial \alpha_{j'}}(\alpha, X_t) dt \\
\sigma^2 \frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \beta_k}(\theta) &= - \int_T^{2T} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \widehat{\pi}_T(X_t) \frac{\partial b}{\partial \alpha_j}(\alpha, X_t) dt \\
\sigma^2 \frac{\partial^2 \Lambda_T}{\partial \beta_k \partial \beta_{k'}}(\theta) &= - \int_T^{2T} \frac{\partial^2 \Phi}{\partial \beta_k \partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dX_t + \int_T^{2T} \widehat{S}(\theta, X_t) \frac{\partial^2 \Phi}{\partial \beta_k \partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dt \\
&- \int_T^{2T} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \widehat{\pi}_T(X_t) \frac{\partial \Phi}{\partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dt
\end{aligned}$$

Therefore, using that $\widehat{S}(\theta, x) - S(\theta, x) = -\Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(x)$ (see (24)),

$$\begin{aligned}
\sigma^2 \frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \alpha_{j'}}(\theta) &= - \int_T^{2T} \frac{\partial^2 b}{\partial \alpha_j \partial \alpha_{j'}}(\alpha, X_t) \Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_t) dt - \sigma \int_T^{2T} \frac{\partial^2 b}{\partial \alpha_j \partial \alpha_{j'}}(\alpha, X_t) dW_t \\
&- \int_T^{2T} \frac{\partial b}{\partial \alpha_j}(\alpha, X_t) \frac{\partial b}{\partial \alpha_{j'}}(\alpha, X_t) dt = A_T + B_T + C_T
\end{aligned}$$

Applying Lemma 1 yields that $\frac{1}{T} A_T$ is $o_P(1)$. By the central limit theorem for stochastic integrals, $\frac{1}{\sqrt{T}} B_T$ converges in distribution so that $\frac{1}{T} B_T$ is $o_P(1)$.

The last term satisfies $\frac{1}{T} C_T \rightarrow - \int \frac{\partial b}{\partial \alpha_j}(\alpha, x) \frac{\partial b}{\partial \alpha_{j'}}(\alpha, x) \pi_\theta(x) dx$ a.s..

Lemma 1 and the ergodic theorem yield that $\frac{\sigma^2}{T} \frac{\partial^2 \Lambda_T}{\partial \alpha_j \partial \beta_k}(\theta) \rightarrow_{a.s.} - \int \frac{\partial b}{\partial \alpha_j}(\alpha, x) \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(x) \pi_\theta(x) dx$.

Finally,

$$\begin{aligned}
\sigma^2 \frac{\partial^2 \Lambda_T}{\partial \beta_k \partial \beta_{k'}}(\theta) &= \int_T^{2T} \frac{\partial^2 \Phi}{\partial \beta_k \partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) \Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(x) dt + \sigma \int_T^{2T} \frac{\partial^2 \Phi}{\partial \beta_k \partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dW_t \\
&- \int_T^{2T} \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \widehat{\pi}_T(X_t) \frac{\partial \Phi}{\partial \beta_{k'}}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dt
\end{aligned}$$

Similarly, the first two terms of the equation above are $o_P(T)$ and

$$\frac{\sigma^2}{T} \frac{\partial^2 \Lambda_T}{\partial \beta_k \partial \beta_{k'}}(\theta) \rightarrow - \int \frac{\partial \Phi}{\partial \beta_k}(\beta, \cdot) \star \pi_\theta(x) \frac{\partial \Phi}{\partial \beta_{k'}}(\beta, \cdot) \star \pi_\theta(x) \pi_\theta(x) dx. \quad \square$$

Proof of Theorem 3. To simplify notations, we do the proof for $d = d' = 1$. We have:

$$\frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \alpha}(\theta) = - \frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) dW_t + \frac{1}{T} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) [\Phi(\beta, \cdot) \star (\sqrt{T}(\pi_\theta - \widehat{\pi}_T))](X_t) dt$$

Using the notations of Lemmas 1 and 2, we can write:

$$- \frac{1}{T} \int_T^{2T} \frac{\partial b}{\partial \alpha}(\alpha, X_t) [\Phi(\beta, \cdot) \star (\sqrt{T}(\pi_\theta - \widehat{\pi}_T))](X_t) dt = \sqrt{T} D_{11}(\varphi_0, H_0) + \sqrt{T} D_{12}(\varphi_0, H_0)$$

with

$$\varphi_0 = \frac{\partial b}{\partial \alpha}(\alpha, \cdot), H_0 = \Phi(\beta, \cdot).$$

Thus

$$\sqrt{T} D_{11}(\varphi_0, H_0) + \sqrt{T} D_{12}(\varphi_0, H_0) = \frac{\sigma}{\sqrt{T}} \int_0^T g'(\theta, X_s) dW_s + O_P(h_T \sqrt{T}) + O_P\left(\frac{1}{\sqrt{T}}\right) + O_P(\sqrt{T} h_T^2)$$

where $g(\theta, \cdot)$ is given by $Lg(\theta, \cdot)(y) = -S(y)$ with

$$S(y) = \Psi(\theta, y) - \int \Psi(\theta, z)\pi_\theta(z)dz, \quad \Psi(\theta, y) = \int \frac{\partial b}{\partial \alpha}(\alpha, x)\Phi(\beta, x - y)\pi_\theta(x)dx$$

and L is the infinitesimal generator given by (11)-(33). Therefore,

$$\frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \alpha}(\theta) = -\frac{\sigma}{\sqrt{T}} \int_0^{2T} \left(\mathbf{1}_{[T, 2T]}(t) \frac{\partial b}{\partial \alpha}(\alpha, X_t) + \mathbf{1}_{[0, T]}(t) g'(\theta, X_t) \right) dW_t + o_P(1).$$

Next,

$$\frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \beta}(\theta) = -\frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \widehat{\pi}_T(X_t) dW_t + \int_T^{2T} (\widehat{S}(\theta, X_s) - S(\theta, X_s)) \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \widehat{\pi}_T(X_s)$$

Therefore, using (2),(24)

$$\begin{aligned} \frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \beta}(\theta) &= -\frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(X_t) dW_t \\ &\quad - \frac{1}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(X_s) \times \Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_s) ds \\ &\quad - \frac{1}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_s) \Phi(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_s) ds \\ &\quad - \frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_t) dW_t. \end{aligned}$$

The third integral is in the form of $D(H, G) = \frac{1}{\sqrt{T}} \int_T^{2T} H \star (\widehat{\pi}_T - \pi_\theta)(X_t) G \star (\widehat{\pi}_T - \pi_\theta)(X_t) dt$ which is $O_P(1/\sqrt{T})$ by Lemma 1.

We have $\frac{1}{T} \int_T^{2T} [\frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_t)]^2 dt = o_P(1)$ by Lemma 1, so that,

$$\frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star (\widehat{\pi}_T - \pi_\theta)(X_t) dW_t = o_P(1).$$

The second term is

$$\begin{aligned} B_{2,T} &= -\frac{1}{T} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(X_s) \times \Phi(\beta, \cdot) \star (\sqrt{T}(\widehat{\pi}_T - \pi_\theta))(X_s) ds \\ &= \sqrt{T} \int (\widehat{\pi}_T(y) - \pi_\theta(y)) \frac{1}{T} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(X_s) \times \Phi(\beta, X_s - y) ds. \end{aligned}$$

Setting $\Xi(\theta, y) = \int \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(x) \times \Phi(\beta, x - y)\pi_\theta(x)dx$, we have, applying Theorem 2

$$B_{2,T} = -\frac{\sigma}{\sqrt{T}} \int_0^T h'(X_s) dW_s + o_P(1),$$

with h satisfying $Lh = -(\Xi(\theta, \cdot) - \int \Xi(\theta, y)\pi_\theta(y)dy)$. Therefore,

$$\frac{\sigma^2}{\sqrt{T}} \frac{\partial \Lambda_T}{\partial \beta}(\theta) = -\frac{\sigma}{\sqrt{T}} \int_T^{2T} \frac{\partial \Phi}{\partial \beta}(\beta, \cdot) \star \pi_\theta(X_t) dW_t - \frac{\sigma}{\sqrt{T}} \int_0^T h'(X_s) dW_s + o_P(1).$$

The convergence in distribution follows from the central limit theorem (32). \square

10. PROOFS OF SECTION 5.

Proof of Lemma 1.

◇ We have:

$$\begin{aligned} D_{11}(\varphi, H) &= \frac{1}{T} \int_T^{2T} \varphi(X_s) \left[\int H(X_s - y) (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)) dy \right] ds \\ &= \int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)) \left[\frac{1}{T} \int_T^{2T} \varphi(X_s) H(X_s - y) ds \right] dy \end{aligned}$$

By assumption, there exists a constant $c > 0$ such that, for all x , $|\varphi(x)| + |H(x)| \leq c(1 + |x|^c)$. Therefore, $|\varphi(x)H(x - y)| \lesssim (1 + |x|^{2c})(1 + |y|^{2c})$. Choosing $p - 4c > 1$, this implies

$$|D_{11}(\varphi, H)| \lesssim \left[\int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y))^2 (1 + |y|^p) dy \int \frac{(1 + |y^{2c})^2}{1 + |y|^p} dy \right]^{1/2} \frac{1}{T} \int_T^{2T} (1 + |X_s|^{2c}) ds$$

Now, $\frac{1}{T} \int_T^{2T} (1 + |X_s|^{2c}) ds = O_P(1)$ and by the concavity of $x \rightarrow \sqrt{x}$ and Proposition 2, we obtain

$$\begin{aligned} \mathbb{E}_\theta \left[\int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y))^2 (1 + |y|^p) dy \right]^{1/2} &\leq \\ \left[\mathbb{E}_\theta \int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y))^2 (1 + |y|^p) dy \right]^{1/2} &= O\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

◇ The proof for $D_{12}(\varphi, H)$ is identical except that we replace $\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)$ by $\mathbb{E}_\theta \hat{\pi}_T(y) - \pi_\theta(y)$ and use Proposition 1 instead of Proposition 2.

◇ We have

$$\begin{aligned} D_{21}(H, G) &= \frac{1}{T} \int_T^{2T} \left[\int \int H(X_s - y) (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)) G(X_s - z) (\hat{\pi}_T(z) - \mathbb{E}_\theta \hat{\pi}_T(z)) dy dz \right] ds \\ &= \int \int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)) (\hat{\pi}_T(z) - \mathbb{E}_\theta \hat{\pi}_T(z)) \left[\frac{1}{T} \int_T^{2T} H(X_s - y) G(X_s - z) ds \right] dy dz \end{aligned}$$

Using that $H^2(x - y) \lesssim (1 + |x|^{2c})(1 + y^{2c})$ and $G^2(x - z) \lesssim (1 + |x|^{2c})(1 + z^{2c})$ yields

$$\left| \frac{1}{T} \int_T^{2T} H(X_s - y) G(X_s - z) ds \right| \lesssim \frac{1}{T} \int_T^{2T} (1 + X_s^{2c}) ds [(1 + y^{2c})(1 + z^{2c})]^{1/2}.$$

Thus, choosing p such that $p - 2c > 1$ and applying Proposition 2 yields

$$\begin{aligned} |D_{21}(H, G)| &\lesssim \left[\int |\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y)| (1 + y^{2c})^{1/2} \right]^2 \frac{1}{T} \int_T^{2T} (1 + X_s^{2c}) ds \\ &\lesssim \int (\hat{\pi}_T(y) - \mathbb{E}_\theta \hat{\pi}_T(y))^2 (1 + |y|^p) dy \int \frac{1 + y^{2c}}{1 + |y|^p} dy \frac{1}{T} \int_T^{2T} (1 + X_s^{2c}) ds \\ &= O_P\left(\frac{1}{T}\right). \end{aligned}$$

◇ For $D_{22}(H, G)$, we proceed analogously applying Proposition 1.

◇ For $D_{23}(H, G)$, we use the Cauchy Schwarz inequality and the above to conclude.

□

Proof of Lemma 2. By assumption, there exists a constant $c > 0$ such that , for all x , $|\varphi(x)| + |H(x)| \leq c(1 + |x|^c)$. We can write

$$\sqrt{T}(D_{11}(\varphi, H)) = I_T + J_T,$$

where

$$I_T = \int dy \sqrt{T} (\widehat{\pi}_T(y) - \mathbb{E}_\theta \widehat{\pi}_T(y)) \psi_\theta(y) dy, \quad \psi_\theta(y) = \int \varphi(x) H(x-y) \pi_\theta(x) dx,$$

$$J_T = \int dy \sqrt{T} (\widehat{\pi}_T(y) - \mathbb{E}_\theta \widehat{\pi}_T(y)) \Delta_T(X_s, y)$$

$$\text{where} \quad \Delta_T(X_s, y) = \frac{1}{T} \int_T^{2T} \varphi(X_s) H(X_s - y) ds - \psi_\theta(y).$$

The term I_T has been studied in Theorem 2 and satisfies

$$I_T = \frac{\sigma}{\sqrt{T}} \int_0^T g'(X_s) dW_s + O_P\left(\frac{1}{\sqrt{T}}\right) + O_P(\sqrt{T} h_T)$$

with g given in Lemma 2. We now prove that $J_T = O(1/\sqrt{T})$.

We have

$$J_T^2 \leq \int (1 + |y|^p) T (\widehat{\pi}_T(y) - \mathbb{E}_\theta \widehat{\pi}_T(y))^2 dy \times \int \Delta_T^2(X_s, y) \frac{dy}{(1 + |y|^p)}.$$

We have already seen that the first factor above is $O_P(1)$. It remains to check that the second one tends to 0 for well chosen p .

Set

$$L(\xi, y) = \varphi(\xi) H(\xi - y) - \psi_\theta(y).$$

Define G_y such that $LG_y = -L(\xi, y)$, *i.e.* $G_y(x) = \int_0^x G'_y(\xi) d\xi$ with

$$G'_y(\xi) = \frac{-2}{\sigma^2 \pi_\theta(\xi)} \int_{-\infty}^{\xi} L(z, y) \pi_\theta(z) dz = \frac{2}{\sigma^2 \pi_\theta(\xi)} \int_{\xi}^{+\infty} L(z, y) \pi_\theta(z) dz.$$

We have

$$\mathbb{E} \Delta_T^2(X_s, y) \leq \frac{2\sigma^2}{T} \int_{\mathbb{R}} (G'_y(x))^2 \pi_\theta(x) dx + \frac{4}{T^2} \int_{\mathbb{R}} G_y^2(x) \pi_\theta(x) dx.$$

Let us set

$$\mathcal{I} = \int_{\mathbb{R}} \frac{1}{1 + |y|^p} \left[\int_{\mathbb{R}} (G'_y(x))^2 \pi_\theta(x) dx \right] dy, \quad \mathcal{J} = \int_{\mathbb{R}} \frac{1}{1 + |y|^p} \left[\int_{\mathbb{R}} G_y^2(x) \pi_\theta(x) dx \right] dy.$$

We have $\mathcal{I} = \mathcal{I}^+ + \mathcal{I}^-$ where

$$\begin{aligned} \mathcal{I}^+ &= \frac{4}{\sigma^4} \int_{\mathbb{R}} \frac{1}{1 + |y|^p} \left[\int_0^{+\infty} \frac{1}{\pi_\theta(x)} \left(\int_x^{+\infty} L(z, y) \pi_\theta(z) dz \right)^2 dx \right] dy \\ \mathcal{I}^- &= \frac{4}{\sigma^4} \int_{\mathbb{R}} \frac{1}{1 + |y|^p} \left[\int_{-\infty}^0 \frac{1}{\pi_\theta(x)} \left(\int_{-\infty}^x L(z, y) \pi_\theta(z) dz \right)^2 dx \right] dy \end{aligned}$$

We only treat \mathcal{I}^+ as the other one is analogous. By the assumptions of polynomial growth, we have $|L(z, y)| \lesssim 1 + |y|^{2c} + |z|^{2c}$ and if $z < z'$, $|L(z, y)L(z', y)| \lesssim 1 + |y|^{4c} + |z'|^{4c}$. Therefore, we

can write:

$$\begin{aligned} \mathcal{I}^+ &= \frac{4}{\sigma^4} \int_{y \in \mathbb{R}, x > 0, z > x, z' > x} \frac{1}{1 + |y|^p} \frac{1}{\pi_\theta(x)} L(z, y) L(z', y) \pi_\theta(z) \pi_\theta(z') dy dx dz dz' \\ &= \frac{8}{\sigma^4} \int_{0 < z < z'} \pi_\theta(z) \pi_\theta(z') \int_0^z \frac{1}{\pi_\theta(x)} dx \left(\int_{\mathbb{R}} \frac{dy}{1 + |y|^p} L(z, y) L(z', y) \right) dz dz'. \end{aligned}$$

Choosing $p - 4c > 1$, this yields, using (39),

$$\begin{aligned} |\mathcal{I}^+| &\lesssim \int_{0 < z < z'} \pi_\theta(z) \pi_\theta(z') \int_0^z \frac{1}{\pi_\theta(x)} dx (1 + |z'|^{4c}) dz dz' \\ &\lesssim \int_{z > 0} \pi_\theta(z) [\pi_\theta(z)]^{1/2} \int_0^z \frac{1}{\pi_\theta(x)} dx \int_z^{+\infty} \left(\frac{\pi_\theta(z')}{\pi_\theta(z)} \right)^{1/2} (1 + |z'|^{4c}) [\pi_\theta(z')]^{1/2} dz' dz \\ &\lesssim \int_{z > 0} \pi_\theta(z) [\pi_\theta(z)]^{1/2} \int_0^z \frac{1}{\pi_\theta(x)} dx e^{-(C/2)z} \int_0^{+\infty} e^{(C/2)z'} (1 + |z'|^{4c}) [\pi_\theta(z')]^{1/2} dz' dz \\ &\lesssim \int_{z > 0} [\pi_\theta(z)]^{1/2} \int_0^z \frac{\pi_\theta(z)}{\pi_\theta(x)} dx e^{-(C/2)z} dz = \int_{z > 0} [\pi_\theta(z)]^{1/2} \int_0^z e^{C(z-x)} dx e^{-(C/2)z} dz \\ &\lesssim \int_{z > 0} [\pi_\theta(z)]^{1/2} \left(|e^{(C/2)z} - e^{-(C/2)z}| \right) dz < +\infty. \end{aligned}$$

Now, we look at $\mathcal{J} = \mathcal{J}^+ + \mathcal{J}^-$ with

$$\begin{aligned} \mathcal{J}^+ &= \int \frac{1}{1 + |y|^p} \left[\int_0^{+\infty} \left(\int_0^x G'_y(v) dv \right)^2 \pi_\theta(x) dx \right] dy, \\ \mathcal{J}^- &= \int \frac{1}{1 + |y|^p} \left[\int_{-\infty}^0 \left(\int_x^0 G'_y(v) dv \right)^2 \pi_\theta(x) dx \right] dy. \end{aligned}$$

We have, for $p - 4c > 1$,

$$\begin{aligned} \mathcal{J}^+ &\leq \int \frac{1}{1 + |y|^p} \left[\int_0^{+\infty} x \int_0^x [G'_y(v)]^2 dv \pi_\theta(x) dx \right] dy \\ &\lesssim \int \frac{1}{1 + |y|^p} \left(\int_{x > 0, 0 < v < x, v < z < z'} dx dv dz dz' |L(z, y) L(z', y)| \frac{\pi_\theta(x) \pi_\theta(z) \pi_\theta(z')}{\pi_\theta^2(v)} \right) dy \\ &\lesssim \int_{x > 0, 0 < v < x, v < z < z'} dx dv dz dz' (1 + |z'|^{4c}) \frac{\pi_\theta(x) \pi_\theta(z) \pi_\theta(z')}{\pi_\theta^2(v)}. \end{aligned}$$

We can write, for $0 < v < x, v < z < z'$, using (39),

$$\begin{aligned} \frac{\pi_\theta(x) \pi_\theta(z) \pi_\theta(z')}{\pi_\theta^2(v)} &= [\pi_\theta(x)]^{1/3} \left[\frac{\pi_\theta(x)}{\pi_\theta(v)} \right]^{2/3} \left[\frac{\pi_\theta(z)}{\pi_\theta(v)} \right]^{2/3} \left[\frac{\pi_\theta(z')}{\pi_\theta(v)} \right]^{2/3} [\pi_\theta(z) \pi_\theta(z')]^{1/3} \\ &\lesssim [\pi_\theta(x)]^{1/3} [\pi_\theta(z) \pi_\theta(z')]^{1/3} e^{(2/3)C(x-v)} e^{(2/3)C(z-v)} e^{(2/3)C(z'-v)}. \end{aligned}$$

Thus,

$$\begin{aligned} \mathcal{J}^+ &\lesssim \int_{x > 0, 0 < v < x, v < z} dx dv dz \left(x [\pi_\theta(x)]^{1/3} e^{(2/3)Cz} e^{-Cv} [\pi_\theta(z)]^{1/3} \int_z^{+\infty} [\pi_\theta(z')]^{1/3} (1 + |z'|^{4c}) e^{(2/3)Cz'} dz' \right) \\ &\lesssim \int_0^{+\infty} x \left(\int_0^x e^{-Cv} dv \right) [\pi_\theta(x)]^{1/3} dx \int_0^{+\infty} [\pi_\theta(z)]^{1/3} e^{(2/3)Cz} dz < +\infty. \end{aligned}$$

We can proceed analogously for \mathcal{J}^- .

Finally, we find that $J_T = O_P((1/\sqrt{T}))$. \square

REFERENCES

- [1] Amorino, C. and Gloter, A. (2020). Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. *Journal of Statistical Planning and Inference*, **213**, 106-129.
- [2] Amorino, C., Heidari, A., Pilipauskaitė, V. and Podolskij, M. (2022). Parameter estimation of discretely observed interacting particle systems. *Preprint, ArXiv.2208.11965*.
- [3] Baladron, J., Fasoli, D., Faugeras, O. and Touboul, J. (2012). Mean field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *The Journal of Mathematical Neuroscience*, **2(1):10**, 1-50.
- [4] Ball, F. and Sirl, D. (2020) Stochastic *SIR* in Structured Populations. *Stochastic Epidemic Models with Inference*, **Part II**, 123-240. Britton, T. and Pardoux, E., Editors. Lecture Notes in Mathematics, **2255**, Mathematical Biosciences Subseries, Springer.
- [5] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2021). Semiparametric estimation of McKean-Vlasov SDEs. *Preprint arXiv:2107.00539*. To appear in *Ann. Inst. Henri Poincaré Probab. Stat.*
- [6] Benachour, S., Roynette, B. and Vallois, P. (1998a). Nonlinear self-stabilizing processes - I Existence, invariant probability, propagation of chaos. *Stoch. Proc. Appl.* **75**, 173-201.
- [7] Benachour, S., Roynette, B. and Vallois, P. (1998b). Nonlinear self-stabilizing processes - II Convergence to invariant probability. *Stoch. Proc. Appl.* **75**, 203-224.
- [8] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31** (5), 615-641.
- [9] Carrillo, L.A., Choi, Y.-P., Hauray, M. (2014). The derivation of swarming models: Mean-field limit and Wasserstein distances. In: Muntean A., Toschi F. (eds) *Collective Dynamics from Bacteria to Crowds. CISM International Centre for Mechanical Sciences* **553**. Springer, Vienna.
- [10] Castellana J.V. and Leadbetter, M.R. (1986). On smoothed probability density estimation for stationary processes. *Stoch. Proc. and Appl.* **21**, 179-193.
- [11] Cattiaux, P., Guillin, A. and Malrieu, F. (2008). Probabilistic approach for granular media equations in the non-uniformly convex case, *Probab. Theory Relat. Fields* **140**, 19-40.
- [12] Chen, X. (2021). Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electron. Commun. Probab.* **26**, 1-13.
- [13] Comte, F. (2017). *Estimation nonparamétrique*. 2nd Edition. Spartacus IDH, Paris.
- [14] Comte, F. and Genon-Catalot, V. (2023). Nonparametric adaptive estimation for interacting particle systems. *Preprint hal-03696877*. To appear in *Scand. J. Statist.*
- [15] Comte, F. and Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. *Stoch. Proc. and Appl.* **115**, 797-826.
- [16] Dalalyan, A. and Reiss, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Relat. Fields* **134** 248-282.
- [17] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probability Theory and Related Fields*, **137**, 25-47.
- [18] Dawson, D.A. (1983). Critical Dynamics and Fluctuations for a Mean-Field Model of Cooperative Behavior. *Journal of Statistical Physics* **31** 29-85.
- [19] Della Maestra, L. and Hoffmann, M. (2022). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probability Theory and Related Fields*, **182**, 551-613.
- [20] Della Maestra, L. and Hoffmann, M. (2022). The LAN property for McKean-Vlasov models in a mean-field regime. *Stochastic Processes and their Applications*, **155**, 109-146.
- [21] Forien, R. and Pardoux, E. (2022). Household epidemic models and McKean-Vlasov Poisson driven SDEs. *Annals of Applied Probability* **32**, 1210-1233.
- [22] Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051-1079.
- [23] Genon-Catalot, V. and Larédo, C. (2021a). Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. *Stoch. Proc. Appl.* **142**, 513-548.
- [24] Genon-Catalot, V. and Larédo, C. (2021b). Parametric inference for small variance and long time horizon McKean-Vlasov diffusion models. *Electronic Journal of Statistics* **15**, 5811-5854.
- [25] Genon-Catalot, V. and Larédo, C. (2023). Inference for ergodic McKean-Vlasov stochastic differential equations with polynomial interactions. *Preprint hal 03618498, v3*.
- [26] Giesecke, K., Schwenkler, G. and Sirignano, J.A. (2020). Inference for large financial systems. *Mathematical Finance*, **30**, 3-46.

- [27] Herrmann, S., Inkeller, P. and Peithmann, D. (2008). Large deviations and a Kramers' type law for self-stabilizing diffusions. *The Annals of Applied Probability*, **18**, 1379-1423.
- [28] Herrmann, S. and Tugaut, J. (2010). Non uniqueness of stationary measures for self-stabilizing diffusions. *Stoch. Proc. Appl.* **120**, 1215-1246.
- [29] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, 135-163.
- [30] Höpfner, R. (2014). *Asymptotic Statistics with a View to Stochastic Processes*. Walter de Gruyter, Berlin/Boston.
- [31] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples*. Springer.
- [32] Kasonga, R.A. (1990). Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, **50**, 865-875.
- [33] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations*. CRC press. Taylor & Francis Group. Boca Raton.
- [34] Kolokoltsov, V.N. (2010). *Non linear Markov processes and kinetic equations* **182**. Cambridge University press.
- [35] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes*. Springer, London.
- [36] Leblanc, F. (1997). Density estimation for a class of continuous time processes. *Mathematical Methods of Statistics* **6**, 171-199.
- [37] Lu, F, Maggioni, M. and Tang, S. (2022). Learning Interaction Kernels in Stochastic Systems of Interacting Particles from Multiple Trajectories. *Foundations of Computational Mathematics* **22**, 1013-1067.
- [38] Mac Kean, H.P., Jr (1966). A class of Markov processes associated with nonlinear parabolic equation. *Proceedings NatL. Acad. Sci. U.S.A.* **56**, 1907- 1911.
- [39] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of applied Probability* **13**, 540-560.
- [40] Marie, N. and Rosier, A. (2023). Nadaraya-Watson Estimator for I.I.D. Paths of Diffusion Processes. *Scandinavian Journal of Statistics* to appear, *preprint hal-0322653v3*.
- [41] Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochastic Processes and their Applications*, **129**, 1013-1059.
- [42] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations*, Lecture Notes in Mathematics, **1627**, 42-95, Springer.
- [43] Molginer, A. and Edelstein-Keshet, L. (1999). A non-local model for a swarm. *J. Math. Biol.* **38**, 534-570.
- [44] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Annals of Statistics*, **48**, 1383-1408.
- [45] Pavliotis, G.A. and Zanoni, A. (2022a). Eigenfunction martingale estimators for interacting particle systems and their mean field limit. *Siam. J. Appl. Dyn. Systems.* to appear.
- [46] Pavliotis, G.A. and Zanoni, A. (2022b). A method of moments for interacting particle systems and their mean-field limit. *Preprint arXiv:2212.00403v1*.
- [47] Rogers, L.C.G. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales, Volume 2 Itô Calculus*. University of Bath Publisher. Cambridge University Press.
- [48] Sharrock, L., Kantas, N., Parpas, P. and Pavliotis, G.A. (2021). Parameter Estimation for the McKean-Vlasov Stochastic Differential Equation. *Preprint Arkiv*.
- [49] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46**, 3451 - 3480.
- [50] Sznitman, A.-S. (1991). Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX-1989. Lecture Notes in Math.* **1464**, 165-251. Springer, Berlin.
- [51] Tsybakov, A. (2009). *Introduction to nonparametric estimation*, Springer series in statistics. Springer, New York