Inference for ergodic McKean-Vlasov stochastic differential equations with polynomial interactions

Valentine Genon-Catalot, Catherine Laredo

Submitted to the Annales de l'Institut Henri Poincaré - Probabilités et Statistiques

Inference for ergodic McKean-Vlasov stochastic differential equations with polynomial interactions

Valentine Genon-Catalot[a], Catherine Larédo[b]

[a]MAP5, UMR 8145 CNRS, Université Paris Cité, France. E-mail: valentine.genon-catalot@mi-parisdescartes.fr
[b]MaIAGE, INRAE, Université Paris-Saclay et LPSM, Université Paris Cité, France. E-mail: catherine.laredo@inrae.fr

Abstract. We consider a specific family of one-dimensional McKean-Vlasov stochastic differential equations with no potential term and with interaction term modeled by an odd increasing polynomial. We assume that the observed process is in stationary regime and that the sample path is continuously observed on a time interval $[0, 2T]$. Due to the McKean-Vlasov structure, the drift function depends on the unknown marginal law of the process in addition to the unknown parameters present in the interaction function. This is why the exact likelihood function does not lead to computable estimators. We overcome this difficulty by a two-step approach leading to an approximate likelihood function. We then derive explicit estimators of the coefficients of the interaction term and prove their consistency and asymptotic normality with rate $\sqrt{T}$ as T grows to infinity. Examples illustrating the theory are proposed.

MSC2020 subject classifications: Primary 60J60, 60J99, 62F12, 62M05
Keywords: McKean-Vlasov stochastic differential equation, Continuous observations, Parametric inference, Invariant distribution, Asymptotic properties of estimators, Approximate likelihood, Long time asymptotics

1. Introduction

We consider the parametric inference for ergodic McKean-Vlasov stochastic differential equations (SDE). These SDEs with coefficients depending both on the state of the process and on its current distribution were first described by McKean [41] to model plasma dynamics. They appear when describing the limit behavior of a large population of interacting particles with an interaction function between the dynamical systems.
A wide field of research is devoted to developing probabilistic tools for the study of interacting particles and their limits (propagation of chaos) (see *e.g.* among many references [23], [45], [5], [6], [42], [12]; [51] and [37] for books). In [29], small noise properties and large deviations results for these processes are investigated.

The statistical inference for models of interacting particles has been little studied. But, these models were recently shown to describe observable dynamics in a wide variety of disciplines, where particles may represent atoms, cells, animals, neurons, people, rational agents, opinions, financial assets: see *e.g.* [7] for the modeling of granular media, [3] for neurosciences, [46], [11] for population dynamics and ecology, [4], [22] for epidemics dynamics, [28] and the references therein for finance. Therefore, the statistical inference for models of interacting particles has become an important issue. Two axes of research can be considered: inference based on the observation of the dynamics of the N interacting particles, inference based on the limiting process (*i.e.* McKean-Vlasov SDE), which describes the typical behaviour of one isolated particle among others. As far as the first approach is concerned, a first result was obtained by [34] who studied parametric inference for a model with linear dependence on the parameters in the drift term. It was later extended by [13], and by [10], [15] for a time-dependent model. Parametric inference based on martingale estimation functions was investigated in [48], a LAN property was recently proved for these systems ([20]); inference based on the empirical distributions of the particle system was studied in [28]. [19] are concerned with nonparametric inference for the drift term in a general model. Recently, [8] study the semi-parametric estimation for a drift term containing both a parametric and a nonparametric part, [2] the inference from discrete observations and [39] and [40] nonparametric inference from *i.i.d.* repetitions of interacting particle systems.
However, assuming that the whole N-particle system is observed might be too demanding and unrealistic. Hence, using

the convergence, as $N \to \infty$ of the N- particle system to McKean-Vlasov SDEs (*propagation of chaos*), makes worthy of interest the study of inference for these SDEs. This is the line of research developed here.

More precisely, we consider a McKean-Vlasov SDE having the specific and classical form

$$dX_t = \mu(t, X_t)dt + \sigma dW_t, \quad X_0 = \eta, \quad \text{with} \tag{1}$$

$$\mu(t,x) = V(x) - \int_{\mathbb{R}} \Phi(x-y)u(t,y)dy = -\Phi \star u(t,.)(x), \quad u(t,y)dy = \mathcal{L}(X_t), \tag{2}$$

where $\mathcal{L}(X_t)$ denotes the law of X_t with density $u(t,.)$, $V : \mathbb{R} \to \mathbb{R}$, $\Phi : \mathbb{R} \to \mathbb{R}$, (W_t) is a standard Brownian motion, η a random variable independent of (W_t). The potential term V describes the geometry of the space. The term Φ derives from the interaction between particles in the original system of particles. These equations differ from classical SDEs because of this interaction term which contains the current distribution of the state variable. Parametric inference studies for such models have started under different asymptotic frameworks. [26] consider the parametric inference for model (1) from a continuous observation on a fixed time interval $[0, T]$ of a single path and of n *i.i.d* paths in the asymptotic framework σ tends to 0. In [27], the parametric inference is studied from the continuous observation of a single path in the double asymptotic $\sigma \to 0$ and $T \to +\infty$. [50] study *i.i.d.* observations of (1) and build an approximation of the likelihood to obtain offline and online estimations.

In this paper, we are concerned with the parametric inference based on a continuous observation of a single path of (1)-(2) on a time interval $[0, 2T]$ with asymptotic properties as T tends to infinity. As studied in numerous papers (see *e.g.* [5], [12], [42], [39], [40]), we consider here models with no potential term ($V \equiv 0$) and odd interaction term which constitute an important class of McKean-Vlasov SDEs both for theoretical properties and for applications. With a nul potential term, considering an odd interaction term ensures the existence and uniqueness of solutions to (1) (see *e.g.* [24], [5], [42], [30], [12]) and existence of invariant distributions for the model (see *e.g.* [6], [52], [30], [12], [21]).

Statistical inference for ergodic diffusion processes has a longstanding history. Among many references, we can quote the books of [38], [33], [36], [32]. There are also lots of papers concerning parametric or nonparametric inference for ergodic diffusions based on continuous or discrete observations: for one dimensional diffusions, *e.g.* [9], [35], [31], [17], [16], [14]; for multi-dimensional diffusions, *e.g.* [18], [47]. Ergodic diffusions with jumps are considered in [43], [44], [49], [1].

To our knowledge, except in [50], the inference for McKean-Vlasov SDEs in stationary regime has not been investigated. The statistical problem is very different from the case of usual SDEs. For inference, the main difficulty lies in the presence of $\mathcal{L}(X_t)$ in the drift term. This is why, as in [8], we consider a specific family of interaction functions which are odd polynomials. More precisely, we consider the one-dimensional process defined by (1),(2) with $V \equiv 0$, Φ odd and increasing and $\mu(t,x) = \mu_{\mathbf{f}}(t,x)$ depending on an unknown parameter $\mathbf{f}$, *i.e.*

$$dX_t = \mu_{\mathbf{f}}(t, X_t)dt + \sigma dW_t, \quad \text{where} \tag{3}$$

$$\mu_{\mathbf{f}}(t,x) = -\int_{\mathbb{R}} \Phi(\mathbf{f}, x-y)u(t,\mathbf{f},y)dy = -\Phi(\mathbf{f},.) \star u(t,\mathbf{f},.)(x), \quad u(t,\mathbf{f},y)dy = \mathcal{L}(X_t), \tag{4}$$

(W_t) is a standard Brownian motion, σ is known and $\mathbf{f}$ is an unknown parameter. A solution of (3)-(4) is a couple $((X_t, u(t,\mathbf{f},.), t \geq 0)$ composed with a process (X_t) and a family of distributions $(u(t,\mathbf{f},x)dx)$ satisfying (3)-(4). When defined, (X_t) is a time-*inhomogeneous* Markov process which admits stationary distributions.

As Φ is odd, whatever the initial distribution, the process (X_t) solving (3)-(4) has a constant expectation m (see Section 2). Contrary to classical SDEs, stationary distributions of model (3)-(4) are not uniquely determined except if the expectation of (X_t) is specified. Under additional assumptions, stationary distributions for (3)-(4) exist and satisfy: If $\mathbb{E}(X_t) = 0$, (3)-(4) admits a unique invariant distribution with symmetric density $u_0(\mathbf{f}, x)$; if $\mathbb{E}(X_t) = m$, (3)-(4) admits a unique invariant distribution with density $u_m(\mathbf{f}, x) = u_0(\mathbf{f}, x - m)$.

When (X_t) is in stationary regime, $\mathcal{L}(X_t)$ does no longer depend on t and is equal to the stationary distribution $u_0(\mathbf{f}, x)dx$ in centered stationary regime and to $u_m(\mathbf{f}, x)dx = u_0(\mathbf{f}, x-m)dx$ in stationary regime with expectation m.

In this paper, we assume that the interaction Φ satisfies

$$\Phi(\mathbf{f}, x) = \sum_{j=0}^{k-1} f_{2j+1}x^{2j+1}, \; f_1 > 0, \; f_{2j+1} \geq 0, j = 1, \ldots, k-1, \quad \text{where } \mathbf{f} = (f_1, f_3, \ldots, f_{2k-1}). \tag{5}$$

Then, (X_t) satisfying (3)-(4)-(5) has stationary distributions. We first study the estimation of $\mathbf{f}$ when (X_t) is in centered stationary regime ($X_0 \sim u_0(\mathbf{f}, x)dx$). Then, we study the joint estimation of $(m, \mathbf{f})$ when the process is in non centered

stationary regime ($X_0 \sim u_m(\mathbf{f},x)dx = u_0(\mathbf{f},x-m)dx$). Because of the specific form of the interaction function Φ (polynomial), the convolution product $\Phi(\mathbf{f},.)\star u_0(\mathbf{f},.)$ (resp. $\Phi(\mathbf{f},.)\star u_m(\mathbf{f},.)$) is explicitly given as a function of $\mathbf{f}$ and the moments of the invariant distribution. This strongly simplifies the drift term. However, these moments have no explicit expression as functions of $\mathbf{f}$ and m. Therefore, the exact log-likelihood can be studied theoretically but does not lead to computable estimators.
Thus we first build estimators of the stationary distribution moments based on the sample path $(X_t, t\in[0,T])$. Then, to get an explicit contrast, we plug these moment estimators into the exact conditional log-likelihood of (3)-(4) given X_T, based on the sample $(X_t, t\in[T,2T])$. We prove that these estimators are consistent and asymptotically Gaussian with rate $\sqrt{T}$.

The paper is organized as follows. In Section 2, we detail the assumptions for existence and uniqueness of solutions and existence of invariant distributions. In particular, we describe these invariant distributions (Proposition 2.1). In Section 3 (resp. Section 4), we estimate $\mathbf{f}$ when the observed process is in centered stationary regime (resp. non centered stationary regime). We study the exact likelihood and prove that the maximum likelihood estimator is consistent and asymptotically Gaussian with rate $\sqrt{T}$ (Proposition 3.2) . However, this remains completely theoretical and the estimators are numerically intractable. Next, we study computable estimators of $\mathbf{f}$ for the centered process and for the non centered process. First, we rely on a two-step approach. We use the sample path on $[0,T]$ to estimate moments of the stationary distribution. Then, estimators are built using the resulting approximation of the likelihood on $[T,2T]$. The main results are stated in Theorem 3.1 and Theorem 4.1. Second, we also build empirical estimators based on some specific properties of model (3). Examples illustrating the methods are given. Section 5 contains concluding remarks. Proofs are gathered in Section 6. In the Appendix (Section 7), properties of the infinitesimal generator and a central limit theorem for ergodic diffusions are recalled.

2. Probability preliminaries for general interaction term.

In this section, we give sufficient conditions for existence and uniqueness of a solution to (3)-(4) and existence and uniqueness of a stationary distribution. We explain how the stationary distribution with specified expectation may be computed by an implicit fixed point equation. This is different from the case of classical SDEs. We describe the properties of (3)-(4) when the initial variable follows the stationary distribution.

2.1. Assumptions for a general interaction term Φ

The following assumptions may be found in Benachour *et al.* (1998a), Malrieu (2003) or Cattiaux *et al.* (2008).

- [H1] Φ is odd and increasing.
- [H2] Φ is locally Lipschitz with polynomial growth, *i.e.* there exist $c>0$, $\ell\in\mathbb{N}^*$ such that $\forall x,y\in\mathbb{R}$, $|\Phi(x)-\Phi(y)|\le c|x-y|(1+|x|^\ell+|y|^\ell)$.
- [H3] Φ is $C^1(\mathbb{R})$ and there exists a constant $\lambda>0$ such that $\forall x,\ \Phi'(x)\ge\lambda$.
- [H4] Φ' has ℓ polynomial growth: $\exists C>0, \ell\in\mathbb{N}^\star$, $\forall x\in\mathbb{R}$, $|\Phi'(x)|\le C(1+|x|^\ell)$.

Note that, under [H2], for all x, $|\Phi(x)|\le c|x|(1+|x|^\ell)$. Therefore, there exists $c_1>0$ such that $|\Phi(x)|\le c_1(1+|x|^{\ell+1})$ and this implies

$$\int|\Phi(x-y)|u(t,y)dy)\le c_1(1+2^\ell|x|^{\ell+1}+2^\ell\int|y|^{\ell+1}u(t,y)dy)).$$

So $\Phi\star u(t,.)$ is well defined as soon as $u(t,.)$ has a moment of order $\ell+1$.

Under [H1]-[H2], if $\mathbb{E}X_0^{2(\ell+1)^2}<+\infty$, equation (3) admits a unique strong solution. If $\mathbb{E}X_0^{2n}<+\infty$, then $\sup_{t\ge0}\mathbb{E}X_t^{2n}<+\infty$ (see Theorem 3.1 and Proposition 3.10 in Benachour *et al.*, 1998a, Theorem 2.13 in Hermann *et al.*, 2008).

Now the fact that there is no potential term and that Φ is odd has the following consequence.

Lemma 2.1. *Assume [H1]-[H3]. Let $(X_t,u(t,.))$ be a solution of (3)-(4). Then, for all t, $\mathbb{E}X_t=\mathbb{E}X_0$. Moreover, setting $Y_t=X_t-\mathbb{E}(X_0)$ and $v(t,.)=\mathcal{L}(Y_t)$, $(Y_t,v(t,.))$ is also a solution of (3)-(4).*

Proof By (3)-(4),

$$\mathbb{E}(X_t)=\mathbb{E}(X_0)-\int_0^t ds\,\mathbb{E}\int\Phi(X_s-y)u(s,y)dy).$$

Since Φ is odd, taking $(\overline{X}_t)$ an *i.i.d.* copy of (X_t), $\mathbb{E}\int \Phi(X_t - y)u(t,y)dy = \mathbb{E}\Phi(X_t - \overline{X}_t) = 0$. Thus,

$$\forall t, \quad \mathbb{E}(X_t) = \mathbb{E}(X_0). \tag{6}$$

This holds whatever the initial variable.
Now, considering $Y_t = X_t - \mathbb{E}(X_0)$ and let $v(t,y)dy$ be the distribution of $X_t - \mathbb{E}(X_0)$, we have

$$\begin{aligned} X_t - \mathbb{E}(X_0) &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - (y - \mathbb{E}(X_0)))u(s,y)dy + \sigma W_t \\ &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - z)u(s, z + \mathbb{E}(X_0))dz + \sigma W_t \\ &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - z)v(s,z)dz + \sigma W_t. \square \end{aligned}$$

This is why the specification of the process expectation is important especially for invariant distributions (see below).
Finally, let us state another useful property associated with this equation.

Lemma 2.2. *Assume [H1]-[H3]. Consider a symmetric probability density u such that $\int_0^\infty y^{\ell+1}u(y)dy < +\infty$. Then, $\Phi \star u$ is well-defined and*

- *$\Phi \star u$ is odd.*
- *For all x, $\int_0^x \Phi \star u(y)dy \geq \lambda \frac{x^2}{2} + C$, for some constant C.*

2.2. *Stationary distributions*

2.2.1. *Existence and uniqueness*

By Lemma 2.2 in [30], if there exists an invariant density whose $(8(\ell+1)^2)$-moment is finite, then it satisfies the implicit fixed point equation

$$u(x) = \frac{\exp\left(-2\sigma^{-2}\int_0^x \Phi \star u(y)dy\right)}{\nu(u)} \tag{7}$$

where, by Lemma 2.2, $\nu(u)$ below is well defined and finite,

$$\nu(u) = \int_{\mathbb{R}} \exp\left(-2\sigma^{-2}\int_0^x \Phi \star u(y)dy\right)dx < +\infty. \tag{8}$$

Equation (7) does not possess a unique solution unless its expectation is specified. In other words, it has a unique solution with a given expectation.
As an example, consider the simple case $\Phi(x) = x$, then $\Phi \star u(x) = \int (x-y)u(y)dy = x - m$ with $m = \int yu(y)dy$. Thus,

$$u(x) = u_m(x) \propto \exp\left[-\frac{1}{\sigma^2}(x^2 - 2mx)\right] \propto \exp\left[-\frac{1}{\sigma^2}(x-m)^2\right].$$

Hence, the stationary distribution depends on the parameter m. This is consistent with the fact that equation (3) with $\Phi(x) = x$ writes

$$X_t = X_0 - \int_0^t (X_s - \mathbb{E}X_s)ds + \sigma W_t = X_0 - \int_0^t (X_s - m)ds + \sigma W_t, \quad \text{where, for all } t, m = \mathbb{E}X_0 = \mathbb{E}X_t.$$

For this reason, many authors ([12], [5], [42]) consider equations (3)-(4) under the assumption that $\mathbb{E}X_t = 0$ and prove the following result.

Proposition 2.1. *(see [12], [5], [42]).*
(i) Under [H1]-[H4], there exists a unique <u>symmetric</u> density function $u(x)$ implicitely defined by

$$u(x) = \frac{1}{\nu(u)} \exp\left(-2\sigma^{-2}\int_0^x \Phi \star u(y)dy\right) \tag{9}$$

which satisfies (see [H3] for λ *and Lemma 2.2):*

$$u(x) \leq \frac{1}{\nu(u)} \exp\left[-\sigma^{-2}\lambda x^2\right]. \tag{10}$$

(ii) If $u(.)$ *is the density of* X_0 *and* (X_t) *is the unique solution of* (3)-(4)*, then* $u(.)$ *is the density of* X_t*, for all* $t \geq 0$*.*
(iii) For any initial law satisfying the moment condition of order $8(\ell+1)^2$*,* $\mathcal{L}(X_t)$ *converges to the invariant symmetric law* u *as* t *tends to infinity.*

Consequently, equation (3)-(4) admits a unique invariant density $u_m(x)$ such that $\int y u_m(y)dy = m$. This density is equal to $u_m(x) = u(x-m)$ and is thus symmetric around m. By (10), u admits moments of any order.

2.2.2. Ergodicity
Let us now point out the following properties of the process (X_t) in stationary regime. The process defined in (3) is a time-inhomogeneous Markov process. However, when (X_t) is in stationary regime (with expectation m), (X_t) is identical to a time-homogeneous diffusion process.
Indeed, assume that the initial variable η has distribution $u_m(x)dx$ then, the density $u(t,y)dy$ of X_t defined in (3)-(4), satisfies

$$\forall t \geq 0,\ u(t,y) = u_m(y),$$

so that the following holds.

Proposition 2.2. *Consider the stochastic differential equation*

$$dY_t = b(Y_t)dt + \sigma dW_t, \quad b = -\Phi \star u_m, \tag{11}$$

where $u_m(.) = u_0(.-m)$ *and* u_0 *is the unique symmetric solution of* (9)*.*
Then (Y_t) *is a positive recurrent diffusion whose stationary distribution has density* $u_m(x)$*.*
If $Y_0 \sim u_m(x)dx$*, it is ergodic. Moreover,*
- If $Y_0 \neq X_0$*,* $(Y_t) \not\equiv (X_t)$*.*
- If $Y_0 = X_0 = \eta \sim u_m(x)dx$*, then* $X_t = Y_t$ *for all* $t \geq 0$*.*

Thus, when $X_0 \sim u_m(x)dx$, (X_t) is equal to the solution of a classical SDE in stationary regime and is ergodic. Let us stress the importance of this result. It allows to apply known results for classical ergodic SDEs. Consequently, we rely strongly on results stated in [25] which in particular sums up properties of the infinitesimal generator of ergodic diffusions. Thus, a law of large numbers holds for (X_t).
If f satisfies $\int |f(x)| u_m(x)dx < +\infty$, applying the ergodic theorem yields

$$\frac{1}{T}\int_0^T f(X_s)ds \to_{a.s.} \int f(x)u_m(x)dx. \tag{12}$$

The central limit theorem associated with this result is stated and detailed in the Appendix, together with important properties of the infinitesimal generator for stationary diffusions.

3. Parametric inference in centered stationary regime.

From now on, we consider that (X_t) is defined by (3)-(4)-(5) and recall that σ is known.
In this section, we assume that (X_t) is in centered stationary regime. Therefore, we make here the assumptions:

- [H5] $\Phi(x) = \Phi(\mathbf{f}, x) = \sum_{j=0}^{k-1} f_{2j+1}x^{2j+1},\ f_1 > 0,\ f_{2j+1} \geq 0, j = 0, \ldots, k-1.$
- [H6] $X_0 = \eta \sim u_0(\mathbf{f}, x)dx,$

where $u_0(\mathbf{f}, .)$ is the unique centered invariant density. Recall that $u_0(\mathbf{f}, .)$ is symmetric. For V a vector or M a matrix, denote by V' or M' the transposed vector or matrix.
We consider the estimation of the unknown parameter $\mathbf{f} = (f_1, f_3, \ldots, f_{2k-1})'$ based on the continuous observation of $(X_t, t \leq 2T)$.

We first describe some analytical properties of the drift term of (X_t), which leads to explicit expressions of $\mu_{\mathbf{f}}(t,x)$ in terms of moments of the stationary distribution (Section 3.1). Then, we study the theoretical likelihood (Section 3.2). Under [H5-[H6], Maximum Likelihood Estimators (MLE) exist but are intractable. Therefore, we develop in Section 3.3, a two-step approach, estimating first the empirical moments of $u_0(\mathbf{f},.)$ from the observation of (X_t) on $[0,T]$, and second building an approximate likelihood based on $(X_t), t\in[T,2T]$ which relies on these moments estimators. Finally, using another specific property of (X_t), we propose in Section 3.4 another inference method, which could serve to derive preliminary estimators. Examples are finally given in Section 3.5.

3.1. Analytical properties of the drift of (X_t)

Under[H5]-[H6] assumptions [H1]-[H4] are satisfied and Equation (3)-(4) admits an invariant distribution which is unique when its expectation is specified. Indeed, according to Proposition 2.1, Equation (9) has a unique symmetric density solution that we have denoted $u_0(\mathbf{f},.)$. Note that $u_0(\mathbf{f},.)$ depends on $\mathbf{f}$ and σ. As σ is known, in what follows, we omit the dependence w.r.t. σ in the notations. Define, for $j\geq 0$,

$$\gamma_{2j}(\mathbf{f})=\gamma_{2j}=\int_{\mathbb{R}} x^{2j}u_0(\mathbf{f},x)dx. \tag{13}$$

Set for $x\in\mathbb{R}$, define the vector

$$z_k(x)=(x,\ x^3\ \ldots, x^{2k-1})'. \tag{14}$$

Proposition 3.1. *Under [H5]-[H6], the drift $\mu_{\mathbf{f}}(t,x)=-\Phi(\mathbf{f},.)\star u_0(\mathbf{f},.)(x)$ is an odd polynomial such that*

$$\mu_{\mathbf{f}}(t,x)=b(\mathbf{f},x)=\sum_{i=0}^{k-1} b_{2i+1}(\mathbf{f})x^{2i+1}=\mathbf{b}(\mathbf{f})'z_k(x), \quad \textit{where} \tag{15}$$

$$\mathbf{b}(\mathbf{f})=(b_{2i+1}(\mathbf{f}), i=0,\ldots,k-1)' \tag{16}$$

$$b_{2i+1}(\mathbf{f})=-\sum_{j=i}^{k-1}\binom{2j+1}{2(j-i)}\gamma_{2(j-i)}(\mathbf{f})f_{2j+1},\ 0\leq i\leq k-1$$

where for $p\leq n$, $\binom{n}{p}$ is the binomial coefficient.

Thus,

$$\mathbf{b}(\mathbf{f})=M_k(\mathbf{f})\mathbf{f} \tag{17}$$

where $M_k(\mathbf{f})=(M_k(\mathbf{f},i,j)_{0\leq i,j\leq k-1})$ is the $k\times k$ upper triangular matrix given by

$$M_k(\mathbf{f},i,j)=0 \text{ for } i>j, M_k(\mathbf{f},i,j)=-\binom{2j+1}{2(j-i)}\gamma_{2(j-i)}(\mathbf{f}) \text{ for } i\leq j. \tag{18}$$

Note that $\gamma_0(\mathbf{f})=\gamma_0=1$ so that $M_k(\mathbf{f},i,i)=-1$ and that $M_k(\mathbf{f})$ depends on $\mathbf{f}$ only through the moments $(\gamma_0,\gamma_2(\mathbf{f}),\ldots,\gamma_{2(k-1)}(\mathbf{f}))$.
The coefficients of $b(\mathbf{f},x)$ are explicit functions of $\mathbf{f}$ and of the moments of $u_0(\mathbf{f},.)$. We define for $\mathbf{v}=(v_0,v_1,\ldots,v_{k-1})'$ a vector of $\mathbb{R}^k$, denote $M_k^{(\mathbf{v})}=(M_k^{(\mathbf{v})}(i,j)_{0\leq i,j\leq k-1}$ with

$$M_k^{(\mathbf{v})}(i,j)=0 \text{ for } i>j,\ \ M_k^{(\mathbf{v})}(i,j)=-\binom{2j+1}{2(j-i)}v_{j-i} \text{ for } i\leq j. \tag{19}$$

Note that $M_k(\mathbf{f})$ defined in (18) satisfies $M_k(\mathbf{f})=M_k^{(\boldsymbol{\gamma})}$ where $\boldsymbol{\gamma}=(\gamma_0,\gamma_2(\mathbf{f}),\ldots,\gamma_{2(k-1)}(\mathbf{f}))'$.

Proof of Proposition 3.1. First, since $u_0(\mathbf{f},.)$ is symmetric, odd moments of $u_0(\mathbf{f},.)$ are nul. Therefore,

$$\Phi(\mathbf{f},.)\star u_0(\mathbf{f},x)=\sum_{j=0}^{k-1} f_{2j+1}\int(x-y)^{2j+1}u_0(\mathbf{f},y)dy=\sum_{j=0}^{k-1} f_{2j+1}\sum_{\ell=0}^{j}\binom{2j+1}{2\ell}x^{2j+1-2\ell}\gamma_{2\ell}(\mathbf{f})$$

$$=\sum_{j=0}^{k-1} f_{2j+1}\sum_{i=0}^{j}\binom{2j+1}{2(j-i)}x^{2i+1}\gamma_{2(j-i)}(\mathbf{f})=\sum_{i=0}^{k-1}x^{2i+1}\sum_{j=i}^{k-1}\binom{2j+1}{2(j-i)}\gamma_{2(j-i)}(\mathbf{f})f_{2j+1}. \quad \square$$

Examples 3.1.

- *For* $k=1$, $b(\mathbf{f},x)=-f_1x$, $M_1=-[1]$.
- *For* $k=2$, $b(\mathbf{f},x)=-[(f_1+3\gamma_2(\mathbf{f})f_3)x+f_3x^3]$, $M_2(\mathbf{f})=-\begin{pmatrix}1 & 3\gamma_2(\mathbf{f})\\ 0 & 1\end{pmatrix}$.
- *For* $k=3$, $b(\mathbf{f},x)=-[(f_1+3\gamma_2(\mathbf{f})f_3+5\gamma_4(\mathbf{f})f_5)x+(f_3+10\gamma_2(\mathbf{f})f_5)x^3+x^5]$,

$$M_3(\mathbf{f})=-\begin{pmatrix}1 & \binom{3}{2}\gamma_2(\mathbf{f}) & \binom{5}{4}\gamma_4(\mathbf{f})\\ 0 & 1 & \binom{5}{2}\gamma_2(\mathbf{f})\\ 0 & 0 & 1\end{pmatrix}.$$

Remark 3.1. *Computation of the stationary distribution*
As a consequence of Proposition 3.1, the symmetric stationary distribution $u_0(\mathbf{f},.)$ *of* (3)-(4)-(5) *can be computed numerically (see e.g. [5]). Indeed, using* (13),

$$u_0(\mathbf{f},x)=\frac{1}{\nu(\mathbf{f},u_0)}\exp[-\frac{2}{\sigma^2}F(x)]dx,\qquad \nu(\mathbf{f},u_0)=\int\exp[-\frac{2}{\sigma^2}F(x)]dx$$

$$F(x)=\sum_{i=0}^{k-1}\frac{x^{2i+2}}{2i+2}\sum_{j=i}^{k-1}\binom{2j+1}{2(j-i)}\gamma_{2(j-i)}(\mathbf{f})f_{2j+1}=F_{\boldsymbol{\gamma}}(x).$$

Therefore, the stationary distribution only depends on its moments $\gamma_{2j}(\mathbf{f}); j=1,\dots,k-1$.
For each $\mathbf{f}$, *the vector* $\boldsymbol{\gamma}$ *is the unique solution of the system:*

$$\gamma_{2j}\int\exp[-\frac{2}{\sigma^2}F_{\boldsymbol{\gamma}}(x)]dx=\int x^{2j}\exp[-\frac{2}{\sigma^2}F_{\boldsymbol{\gamma}}(x)]dx,\ \ j=1,\dots,k-1,$$

and can therefore be numerically computed. Once $F_{\boldsymbol{\gamma}}$ *is obtained,* $u_0(\mathbf{f},.)$ *may also be numerically obtained.*

3.2. Theoretical likelihood inference

Let us introduce some notations for this section. Let $C([0,2T],\mathbb{R})$ denote the space of continuous functions defined on $[0,2T]$ and $\mathcal{C}_{2T}$ the associated Borel σ-algebra. The parameter set F is the subset of $\mathbb{R}^k$ defined by

(20) $$F=\{\mathbf{f}\in\mathbb{R}^k \text{ such that } \mathbf{f}'=(f_1,f_3,\dots,f_{2k-1}),\ f_1>0, f_{2j+1}\geq 0, j=0,\dots,k-1\}.$$

Let $\mathbf{f}_0$ denote the true value of the parameter and $\mathbb{P}_{\mathbf{f}}$ the distribution on $(C([0,2T],\mathcal{C}_{2T})$ of (X_t) defined by (3)-(4)-(5). Here, we look at maximum likelihood estimation based on $(X_t,t\in[0,T])$. The Girsanov formula holds and the conditional log-likelihood of $(X_t,t\in[0,T])$ given X_0 is given by, using (15):

(21) $$\ell_T(\mathbf{f})=\sigma^{-2}[\int_0^T b(\mathbf{f},X_s)dX_s-\frac{1}{2}\int_0^T b^2(\mathbf{f},X_s)ds].$$

Define the estimator

(22) $$\widehat{\widehat{\mathbf{f}}}_T=\arg\max_{\mathbf{f}\in\mathbb{R}^k}\ell_T(\mathbf{f}).$$

This estimator is purely theoretical as it is not given by explicit equations due to the presence of the moments of $u_0(\mathbf{f},.)$ in the drift $b(\mathbf{f},x)$ (see Proposition 3.1).

Proposition 3.2. *Assume [H5]-[H6]. Then, under* $\mathbb{P}_{\mathbf{f}_0}$, *the following holds:*

$$\frac{1}{T}(\ell_T(\mathbf{f})-\ell_T(\mathbf{f}_0))\to-\frac{1}{2\sigma^2}\int(b(\mathbf{f},x)-b(\mathbf{f}_0,x))^2u_0(\mathbf{f}_0,x)dx:=-\frac{1}{2\sigma^2}K(\mathbf{f}_0,\mathbf{f})\quad a.s.$$

The identifiability assumption $\{K(\mathbf{f}_0,\mathbf{f})=0\Rightarrow\mathbf{f}=\mathbf{f}_0\}$ *is satisfied.*

If moreover the parameter set F *is compact, the maximum likelihood estimator* $\widehat{\widehat{\mathbf{f}}}_T$ *is consistent.*
The matrix $I(\mathbf{f}_0)=(\int_{\mathbb{R}}\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}_0,x)\frac{\partial b}{\partial f_{2j+1}}(\mathbf{f}_0,x)u_0(\mathbf{f}_0,x)dx)_{0\leq i,j\leq k-1}$ *is invertible and under* $\mathbb{P}_{\mathbf{f}_0}$,

$$\sqrt{T}(\widehat{\widehat{\mathbf{f}}}_T-\mathbf{f}_0)\to_{\mathcal{L}}\mathcal{N}\left(0,\sigma^2I(\mathbf{f}_0)^{-1}\right).$$

According to Proposition 2.2, when (X_t) is in stationary regime, $X_t \equiv Y_t$ defined in (11) which is a classical ergodic diffusion. Therefore, the proof is classical. The difficulty here lies in the fact that the drift has a complex dependence with respect to the unknown parameters so that the estimator $\widehat{\widehat{\mathbf{f}}}_T$ is numerically intractable (see *e.g.* Section 3.5, Example 2). In particular, the partial derivatives $\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},x)$, $\frac{\partial^2 b}{\partial f_{2i+1}\partial f_{2j+1}}(\mathbf{f},x)$ are non linear w.r.t. $\mathbf{f}$ as $b(\mathbf{f},x)$ includes in its definition $\mathbf{f}$ and the moments $\gamma_{2i}(\mathbf{f})$ of $u_0(\mathbf{f},x)$ which depend on $\mathbf{f}$. Thus the partial derivatives also depend on the partial derivatives of $\gamma_{2i}(\mathbf{f})$ w.r.t. $\mathbf{f}$.

3.3. *Explicit estimators using empirical moments.*

We assume here that the sample path (X_t) is continuously observed throughout the time interval $[0,2T]$. We use the first half of the sample path, $(X_t, t\in[0,T])$, to build empirical estimators of the moments of $u_0(\mathbf{f},.)$, and the second half, $(X_t, t\in[T,2T])$, to define a contrast in order to estimate the coefficients b_{2i+1} in the drift $b(\mathbf{f},x)$. Finally, we deduce estimators for the parameter $\mathbf{f}$.

3.3.1. *Estimation of the moments of the stationary distribution based on* $(X_t, t\in[0,T])$

Let us consider the empirical estimators of the moments of $u_0(\mathbf{f},.)$ built using the sample path $(X_t, t\in[0,T])$, defined by

$$\widehat{\gamma}_{2j}(T) = \frac{1}{T}\int_0^T X_s^{2j}ds, j\geq 1 \quad (\text{we set} \quad \widehat{\gamma}_0(T)=\gamma_0=1). \tag{23}$$

The following holds.

Proposition 3.3. *Assume [H5] and [H6]. As T tends to infinity, under $\mathbb{P}_{\mathbf{f}}$, for $\ell\geq 1$, $\widehat{\gamma}_{2\ell}(T)\to_{a.s.}\gamma_{2\ell}(\mathbf{f}) = \int x^{2\ell}u_0(\mathbf{f},x)dx$. Moreover,*

$$\sqrt{T}(\widehat{\gamma}_{2\ell}(T)-\gamma_{2\ell}(\mathbf{f})) = \frac{\sigma}{\sqrt{T}}\int_0^T g'_\ell(X_s-m)dW_s + o_P(1), \tag{24}$$

where for $1\leq \ell\leq k-1$,

$$g'_\ell(\mathbf{f},x) = -2\sigma^{-2}[u_0(\mathbf{f},x)]^{-1}\int_{-\infty}^x (y^{2\ell}-\gamma_{2\ell}(\mathbf{f}))u_0(\mathbf{f},y)dy \tag{25}$$

satisfies $\int (g'_\ell(x))^2 u_0(\mathbf{f},x)dx < +\infty$. Consequently, for all k, the vector $(\sqrt{T}(\widehat{\gamma}_{2\ell}(T)-\gamma_{2\ell}(\mathbf{f})), \ell=1,\dots,k-1)$ converges in distribution, under $\mathbb{P}_{\mathbf{f}}$, to
$\mathcal{N}_{k-1}(\mathbf{0},\sigma^2 V(\mathbf{f}))$ with $V(\mathbf{f}) = (V_{i,j}(\mathbf{f}))_{0\leq i,j\leq k-1}$ and $V_{i,j}(\mathbf{f}) = \int g'_i(\mathbf{f},x)g'_j(\mathbf{f},x)u_0(\mathbf{f},x)dx$.

3.3.2. *Estimation of* $\mathbf{b}(\mathbf{f}) = (b_{2i+1}(\mathbf{f}), i=0,\cdots,k-1)'$

The drift function $b(\mathbf{f},x)$ of (X_t) is as an odd polynomial of degree $2k-1$ w.r.t. x (see (15)):

$$dX_t = b(\mathbf{f},X_t)dt + \sigma dW_t, \qquad b(\mathbf{f},x) = \sum_{i=0}^{k-1} b_{2i+1}(\mathbf{f})x^{2i+1}, \quad X_0\sim u_0(\mathbf{f},x)dx.$$

The vector $\mathbf{b}(\mathbf{f}) = (b_{2i+1}(\mathbf{f}), i=0,\dots,k-1)'$ is given in Proposition 3.1. Consider the contrast function which is the log-likelihood given X_T of the process $(X_t, t\in[T,2T])$,

$$U_T(\mathbf{b}(\mathbf{f})) = \frac{1}{\sigma^2}\left(\int_T^{2T}[\sum_{i=0}^{k-1} b_{2i+1}(\mathbf{f})X_s^{2i+1}]dX_s - \frac{1}{2}\int_T^{2T}[\sum_{i=0}^{k-1} b_{2i+1}(\mathbf{f})X_s^{2i+1}]^2 ds\right). \tag{26}$$

We define the estimator $\widehat{\mathbf{b}(\mathbf{f})}_T$ of $\mathbf{b}(\mathbf{f})$ by maximizing U_T with respect to $\mathbf{b}(\mathbf{f}) = (b_{2i+1}(\mathbf{f}), i=0,\dots,k-1)'$.
For this, we set, using (14),

$$Z_T = \int_T^{2T} z_k(X_s)dX_s. \tag{27}$$

Then, $\widehat{\mathbf{b}(\mathbf{f})}_T$ satisfies

(28) $$Z_T = \Psi_T \widehat{\mathbf{b}(\mathbf{f})}_T,$$

where

(29) $$\Psi_T = \int_T^{2T} z_k(X_s)[z_k(X_s)]' ds = \left(\int_T^{2T} X_s^{2i+2j+2} ds\right)_{0\le i,j\le k-1}.$$

Let us define, using (13),

(30) $$\Psi(\mathbf{f}) = (\gamma_{2(i+j+1)}(\mathbf{f}))_{0\le i,j\le k-1}.$$

Proposition 3.4. *Assume [H5]-[H6]. Under $\mathbb{P}_{\mathbf{f}_0}$, the matrix Ψ_T/T converges a.s. to $\Psi(\mathbf{f}_0)$. The matrix $\Psi(\mathbf{f}_0)$ is invertible, $\widehat{\mathbf{b}(\mathbf{f})}_T = [\Psi_T]^{-1} Z_T$ converges a.s. to $\mathbf{b}(\mathbf{f}_0)$ and $\sqrt{T}(\widehat{\mathbf{b}(\mathbf{f})}_T - \mathbf{b}(\mathbf{f}_0))$ converges in distribution to the Gaussian law $\mathcal{N}_k(0, \sigma^2 \Psi(\mathbf{f}_0)^{-1})$.*

3.3.3. *Estimation of* $\mathbf{f}$

Let us come back to the estimation of $\mathbf{f}$. For this, we rely on relations (16)-(17) (see Proposition 3.1) which links $\mathbf{f}$ to $\mathbf{b}(\mathbf{f})$: $\mathbf{b}(\mathbf{f}) = M_k(\mathbf{f})\mathbf{f}$. It suggests to consider the matrix $\widehat{M}_k$ using (19) where the unknown moments of $u_0(\mathbf{f}, .)$ are replaced by their consistent estimators built on the observation of (X_t) on $[0, T]$ given above:

(31) $$\widehat{M}_k = M_k^{(\widehat{\gamma}_0(T), \widehat{\gamma}_2(T), \ldots, \widehat{\gamma}_{2(k-1)}(T))} = (\widehat{M}_k(i,j))_{0\le i,j\le k-1}, \quad \text{with}$$

(32) $$\widehat{M}_k(i,j) = -\binom{2j+1}{2(j-i)} \widehat{\gamma}_{2(j-i)}(T) \mathbf{1}_{i\le j}.$$

It follows from Proposition 3.3 that $\widehat{M}_k$ converges *a.s.* to $M_k(\mathbf{f})$ and that the whole vector $\sqrt{T}((\widehat{M}_k(i,j) - M_k(\mathbf{f}, i, j))_{0\le i,j\le k})$ is asymptotically Gaussian.
This justifies the definition of $\widehat{\mathbf{f}}_T$ by (see (28)):

(33) $$\widehat{\mathbf{f}}_T = (\widehat{M}_k)^{-1} \widehat{\mathbf{b}(\mathbf{f})}_T = (\widehat{M}_k)^{-1} \Psi_T^{-1} Z_T = (\Psi_T \widehat{M}_k)^{-1} Z_T.$$

Let us stress that, as for the theoretical maximum likelihood estimator (22), this new estimator does not depend on σ.

Theorem 3.1. *Under the assumptions [H5]-[H6], the estimator $\widehat{\mathbf{f}}_T$ is consistent and satisfies*

(34) $$\sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) \to_{\mathcal{L}} \mathcal{N}_k(0, \sigma^2 \Sigma(\mathbf{f})) \quad \textit{with } \Sigma(\mathbf{f}) = \Sigma_1(\mathbf{f}) + \Sigma_2(\mathbf{f})$$

(35) $$\Sigma_1(\mathbf{f}) = M_k^{-1}(\mathbf{f}) \Psi^{-1}(\mathbf{f}) (M_k^{-1}(\mathbf{f}))'; \quad \Sigma_2(\mathbf{f}) = \int \beta(\mathbf{f}, x) \beta'(\mathbf{f}, x) u_0(\mathbf{f}, x) dx$$

with, using definitions (18)*,* (19) *and* (30)*,*

(36) $$\beta(\mathbf{f}, x) = M_k(\mathbf{f})^{-1} M_k^{(\mathbf{v}(\mathbf{f},x))} \mathbf{f} \qquad \textit{with} \quad \mathbf{v}(\mathbf{f}, x) = \left(0, g_1'(\mathbf{f}, x), \ldots, g_{k-1}'(\mathbf{f}, x)\right).$$

By Proposition 3.4, $\widehat{\mathbf{b}(\mathbf{f})}_T$ converges *a.s.* to $\mathbf{b}(\mathbf{f})$ and, by Proposition 3.3, $\widehat{M}_k$ to $M_k(\mathbf{f})$ so that $\widehat{\mathbf{f}}_T$ is consistent. For the asymptotic normality, two terms appear. Heuristically, the first term $\Sigma_1(\mathbf{f})$ derives from the change of variable $\mathbf{b}(\mathbf{f}) \to \mathbf{f}$ and the second one $\Sigma_2(\mathbf{f})$ from the estimation of the moments of $u_0(\mathbf{f}, .)$ and the plug-in device in the estimation. The proof, detailed in the appendix, relies on the decomposition in the two main terms

$$\Psi(\mathbf{f}) M_k(\mathbf{f}) \sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}} \int_T^{2T} z_k(X_s) dW_s - \Psi(\mathbf{f}) \sqrt{T}(\widehat{M}_k - M_k(\mathbf{f}))\mathbf{f} + o_P(1).$$

According to Proposition 3.3, $\sqrt{T}(\widehat{M}_k(i,j) - M_k(\mathbf{f}, i, j)) = -\frac{\sigma}{\sqrt{T}} \binom{2j+1}{2(j-i)} \int_0^T g_{2(j-i)}'(\mathbf{f}, X_s) dW_s + o_P(1)$, the second term depends on $(X_t, t \le T)$ while the first term depends on $(X_t, T \le t \le 2T)$. These two terms are conditionally independent and lead to the two quantities appearing in $\Sigma(\mathbf{f})$.

3.4. Another inference method.

We assume that the observation is $(X_t, t\in[0,T])$. This method is based on a special property of model (3)-(4)-(5). There is an explicit relation linking the vector $\mathbf{f}$ and the vector $(\gamma_{2i}(\mathbf{f}), i=0,\dots,k-1)'$. Indeed, writing the Ito formula yields

$$X_t^{2\ell}=X_0^{2\ell}+2\ell\int_0^t X_s^{2\ell-1}\left(-\sum_{j=0}^{k-1} f_{2j+1}(\sum_{m=0}^{j}\binom{2j+1}{2m}X_s^{2j+1-2m}\gamma_{2m}(\mathbf{f}))ds+\sigma dW_s\right)+\sigma^2\ell(2\ell-1)\int_0^t X_s^{2\ell-2}ds.$$

Taking expectations and using that the process is in centered stationary regime yields

$$\forall t\geq 0,\quad 0=-2\ell t\sum_{j=0}^{k-1}\left(\sum_{m=0}^{j}\binom{2j+1}{2m}\gamma_{2m}(\mathbf{f})\gamma_{2(j+\ell-m)}(\mathbf{f})\right)f_{2j+1}+\sigma^2\ell(2\ell-1)t\gamma_{2\ell-2}(\mathbf{f}).$$

We set:

$$B(\mathbf{f}):=(\sigma^2(2\ell-1)\gamma_{2\ell-2}(\mathbf{f}),\ell=1,\dots,k)',\text{ and} \tag{37}$$

$$\Gamma(\mathbf{f})=(\Gamma_{\ell j}(\mathbf{f}))_{1\leq\ell\leq k,0\leq j\leq k-1}\text{ with }\Gamma_{\ell,j}(\mathbf{f})=2\sum_{m=0}^{j}\binom{2j+1}{2m}\gamma_{2m}(\mathbf{f})\gamma_{2(j+\ell-m)}(\mathbf{f}). \tag{38}$$

Then, $B(\mathbf{f})=\Gamma(\mathbf{f})\mathbf{f}$. The matrix $\Gamma(\mathbf{f})$ is necessarily invertible.
Substituting in (37) each moment by its empirical estimator (23) yields the two estimators $\widetilde{B}_T,\widetilde{\Gamma}_T$ and the relation defining the moment estimator of $\mathbf{f}$:

$$\widetilde{\mathbf{f}}_T=(\widetilde{\Gamma}_T)^{-1}\widetilde{B}_T, \tag{39}$$

which is by construction consistent and asymptotically Gaussian. We only need to compute the asymptotic covariance matrix. Contrary to the previous estimators, a drawback of the estimator $\widetilde{\mathbf{f}}_T$ is that it explicitely depends on σ and thus requires its precise knowledge.

Proposition 3.5. *Assume [H5]-[H6]. The estimator defined by* (39) *is consistent and such that* $\sqrt{T}(\widetilde{\mathbf{f}}_T-\mathbf{f})$ *converges in distribution to* $\mathcal{N}(\mathbf{0},\Gamma^{-1}(\mathbf{f})K\Gamma^{-1}(\mathbf{f})')$ *where* K *is given in the proof (see* (65)*).*

3.5. Examples

We illustrate the previous theory on several examples.

Example 1: $\Phi(\mathbf{f},x)=fx, f>0$.
The centered stationary distribution is the Gaussian law $u_0(f,x)dx=\mathcal{N}(0,\sigma^2/2f)$. Equation (3) writes
$dX_t=-f\int(X_t-y)u(f,y)dydt+\sigma dW_t=-fX_tdt+\sigma dW_t$. The estimator $\widehat{f}_T$ is equal to the maximum likelihood estimator:

$$\widehat{f}_T=-\frac{\int_0^T X_sdX_s}{\int_0^T X_s^2ds}=\widehat{\widehat{\mathbf{f}}}_T.$$

As $T^{-1}\int_0^T X_s^2ds$ converges *a.s.* to $\sigma^2/2f$, we obtain the classical result that $\sqrt{T}(\widehat{f}_T-f)$ converges in distribution to $\mathcal{N}(0,2f)$. With the notations of Theorem 3.1, $\Sigma_1(f)=2f/\sigma^2,\Sigma_2(f)=0$.
The second method estimator, based on the relation $\gamma_2(f)=\sigma^2/2f$, is given by:

$$\widetilde{f}_T=\frac{\sigma^2T}{2\int_0^T X_s^2ds}.$$

The generator L of (X_t) is $Lg=\frac{\sigma^2}{2}g''-fxg'$. The equation $Lg_2(x)=(\sigma^2/2f)-x^2$ admits an explicit solution g_2 such that $g_2'(x)=x/f$. Thus,

$T^{-1/2}\int_0^T(X_s^2-(\sigma^2/2f))ds$ converges in distribution to $\mathcal{N}(0,\sigma^2 V)$ with $V=\int(x/f)^2u_0(f,x)dx=\dfrac{\sigma^2}{2f^3}$. This yields that $\sqrt{T}(\widetilde{f}_T-f)$ converges in distribution to $\mathcal{N}(0,2f)$. In this special example, $\widehat{f}_T$ and $\widetilde{f}_T$ have the same asymptotic distribution. Note that $\widehat{f}_T$ can be computed without knowing σ^2 which is preferable.

Example 2: $\Phi(f,x)=fx^3, f>0$.
The function $\Phi(f,x)=fx^3$ does not satisfy all our (sufficient) assumptions but the existence and uniqueness of an invariant density can be checked directly. The stationary density $u_0(f,.)$ is unique and defined by the implicit equation (9). As $u_0(f,.)$ is symmetric, $\int(x-y)^3u_0(f,y)dy=x^3+3x\gamma_2(f)$. Therefore, equation (3) starting with $X_0\sim u_0(f,x)dx$, writes:

(40)
$$dX_t=-f(X_t^3+3X_t\,\gamma_2(f))dt+\sigma dW_t,\; X_0\sim u_0(f,x)dx.$$

where

$$u_0(f,x)=\exp\left[-\sigma^{-2}f(\frac{x^4}{2}+3x^2\gamma_2(f))\right]/\nu(u_0(f,.))$$

and $\gamma_2(f)$ is implicitly given as the unique solution (see [5] and Remark 3.1) of

(41)
$$\int_{\mathbb{R}}x^2\exp\left[-\sigma^{-2}f(\frac{x^4}{2}+3x^2\gamma_2(f))\right]dx=\gamma_2(f)\int_{\mathbb{R}}\exp\left[-\sigma^{-2}f(\frac{x^4}{2}+3x^2\gamma_2(f))\right]dx.$$

Let us start with the exact maximum likelihood estimator. It is defined as the solution of $\ell_T'(\widehat{\widehat{\mathbf{f}}}_T))=0$, *i.e.*

$$\int_0^T[X_t^3+3X_t(\gamma_2(\widehat{\widehat{\mathbf{f}}}_T)-\widehat{\widehat{\mathbf{f}}}_T\gamma_2'(\widehat{\widehat{\mathbf{f}}}_T)]dX_t=-\int_0^T[2\widehat{\widehat{\mathbf{f}}}_T(X_t^3+3X_t(\gamma_2(\widehat{\widehat{\mathbf{f}}}_T))+(\widehat{\widehat{\mathbf{f}}}_T)^2 3X_t\gamma_2'(\widehat{\widehat{\mathbf{f}}}_T)dt.$$

Differentiating (41) w.r.t. f allows to obtain an expression of $\gamma_2'(f)$ as a function of $(f,\gamma_2(f),\gamma_4(f),\gamma_6(f)$. But this does not help in obtaining an explicit equation $\widehat{\widehat{\mathbf{f}}}_T$. This illustrates the fact that the exact MLE is intractable.

Let us now look at the maximum contrast estimator of f based on $(X_t,t\in[0,2T]$. We are not in the framework of Theorem 3.1 since $f_1=0$. But we can compute explicitly the estimator of f and get using (23)

$$\widehat{f}_T=-\frac{\int_T^{2T}(X_t^3+3X_t\widehat{\gamma}_2(T))dX_t}{\int_T^{2T}(X_t^3+3X_t\widehat{\gamma}_2(T))^2dt}:=-\frac{N_T}{D_T}.$$

Define the two quantities $a(f)$ and $c(f)$,

(42)
$$a(f)=\gamma_6(f)+9\gamma_2^3(f)+6\gamma_2(f)\gamma_4(f);\quad c(f)=3(\gamma_4(f)+3\gamma_2^2(f)).$$

As $T\to\infty$, $\frac{D_T}{T}\to\int_{\mathbb{R}}(x^3+3x\gamma_2(f))^2u_0(f,x)dx=a(f)$. We write

$$\sqrt{T}(\widehat{f}_T-f)=-\frac{\sigma T^{-1/2}}{D_T/T}\int_T^{2T}(X_t^3+3X_t\widehat{\gamma}_2(T))dW_t+f\sqrt{T}(\gamma_2(f)-\widehat{\gamma}_2(T))\frac{\int_T^{2T}3X_t(X_t^3+3X_t\widehat{\gamma}_2(T))dt/T}{D_T/T}.$$

We have that $\frac{1}{T}\int_T^{2T}3X_t(X_t^3+3X_t\widehat{\gamma}_2(T))dt\to c(f)$ and

$$\frac{1}{\sqrt{T}}\int_T^{2T}(X_t^3+3X_t\widehat{\gamma}_2(T))dW_t=\frac{1}{\sqrt{T}}\int_T^{2T}(X_t^3+3X_t\gamma_2(f))dW_t+o_P(1).$$

Thus,

$$\sqrt{T}(\widehat{f}_T-f)=-\frac{\sigma}{a(f)}\frac{1}{\sqrt{T}}\int_T^{2T}(X_t^3+3X_t\gamma_2(f))dW_t+\sqrt{T}(\gamma_2(f)-\widehat{\gamma}_2(T))\frac{c(f)}{a(f)}f+o_P(1).$$

Now, using (25),

$$\sqrt{T}(\widehat{f}_T-f)=\frac{\sigma\sqrt{2}}{\sqrt{2T}}\int_0^{2T}\left(\frac{1}{a(f)}(X_t^3+3X_t\gamma_2(f))1_{[T,2T]}(s)+f\frac{c(f)}{a(f)}g_1'(X_s)1_{[0,T]}(s)\right)dW_s+o_P(1).$$

Using the notations of Theorem 3.1, $\Sigma_1(f) = \dfrac{1}{a(f)}, \quad \Sigma_2(f) = f^2 \dfrac{c^2(f)}{a^2(f)} \displaystyle\int (g_1'(x))^2 u_0(f,x)dx,$

$$\sqrt{T}(\widehat{f}_T - f) \to_{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(\Sigma_1(f) + \Sigma_2(f))\right).$$

Let us now look at the second method for f. The Ito formula yields:

$$\mathbb{E}X_t^2 = \mathbb{E}X_0^2 - 2f\int_0^t [\mathbb{E}(X_s(X_s^3 + 3X_s\gamma_2(f))]ds + \sigma^2 t.$$

By the strict stationarity, we get: $f(2\gamma_4(f) + 6\gamma_2^2(f)) = \sigma^2$. Thus, we can define an estimator of f by

$$\widetilde{f}_T = \frac{\sigma^2}{2\widehat{\gamma}_4(T) + 6(\widehat{\gamma}_2(T))^2}.$$

The Delta method yields $\sqrt{T}(\widetilde{f}_T - f) = -\dfrac{2}{\sigma^2} f^2[6\gamma_2(f)(\widehat{\gamma}_2(T) - \gamma_2(f)) + (\widehat{\gamma}_4(T) - \gamma_4(f))] + o_P(1)$. Using that $\sqrt{T}(\widehat{\gamma}_2(T) - \gamma_2(f), \widehat{\gamma}_4(T) - \gamma_4(f))$ is asymptotically Gaussian with covariance given in Proposition 3.3, we get that $\sqrt{T}(\widetilde{f}_T - f) \to \mathcal{N}(0, \widetilde{V}(f))$ with $\widetilde{V}(f) = 4\sigma^{-2}f^4 \displaystyle\int (6\gamma_2(f)g_1'(x) + g_2'(x))^2 u_0(f,x)dx$.

Example 3: $\Phi(\mathbf{f}, x) = f_1 x + f_3 x^3$, $f_1 > 0, f_3 \geq 0$ $(k = 2)$.
We have that $b(\mathbf{f}, x) = -(f_1 x + f_3(x^3 + 3x\gamma_2(\mathbf{f}))) = -((f_1 + 3\gamma_2(\mathbf{f})f_3)x + f_3 x^3)$, and

$$\begin{pmatrix} b_1(\mathbf{f}) \\ b_3(\mathbf{f}) \end{pmatrix} = M_2(\mathbf{f}) \begin{pmatrix} f_1 \\ f_3 \end{pmatrix}, \quad M_2(\mathbf{f}) = -\begin{pmatrix} 1 & 3\gamma_2(\mathbf{f}) \\ 0 & 1 \end{pmatrix}.$$

This yields

$$\begin{pmatrix} \widehat{f}_{T,1} \\ \widehat{f}_{T,3} \end{pmatrix} = -\begin{pmatrix} 1 & -3\widehat{\gamma}_2(T) \\ 0 & 1 \end{pmatrix} \Psi_T^{-1} \begin{pmatrix} \int_T^{2T} X_s dX_s \\ \int_T^{2T} X_s^3 dX_s \end{pmatrix} \text{ where } \Psi_T = \begin{pmatrix} \int_T^{2T} X_s^2 ds & \int_T^{2T} X_s^4 ds \\ \int_T^{2T} X_s^4 ds & \int_T^{2T} X_s^6 ds \end{pmatrix}.$$

According to Theorem 3.1, the asymptotic variance of $\sqrt{T}(\widehat{f}_T - f)$ is $\sigma^2(\Sigma_1(\mathbf{f}) + \Sigma_2(\mathbf{f}))$ where

$$\Sigma_1(\mathbf{f}) = M_2^{-1}(\mathbf{f})\Psi^{-1}(\mathbf{f})(M_2^{-1}(\mathbf{f}))' \quad \text{and} \quad \Sigma_2(\mathbf{f}) = \int \beta(\mathbf{f}, x)\beta'(\mathbf{f}, x)u_0(\mathbf{f}, x)dx,$$

with

$$\Psi(\mathbf{f}) = \begin{pmatrix} \gamma_2(\mathbf{f}) & \gamma_4(\mathbf{f}) \\ \gamma_4(\mathbf{f}) & \gamma_6(\mathbf{f}) \end{pmatrix}, \quad \beta(x) = \beta(\mathbf{f}, x) = M_2(\mathbf{f})^{-1} M_2(0, g_1'(x))\mathbf{f}.$$

Therefore we get that $\Sigma_1(\mathbf{f}) = (\gamma_2(\mathbf{f})\gamma_6(\mathbf{f}) - \gamma_4^2(\mathbf{f}))^{-1} \begin{pmatrix} \gamma_6 + 6\gamma_2\gamma_4 + 9\gamma_2^3 & -\gamma_4 - 3\gamma_2^2 \\ -\gamma_4 - 3\gamma_2^2 & \gamma_2 \end{pmatrix}$.
Now, for $\mathbf{v}(x) = (0, g_1'(\mathbf{f}, x))$, $M_2^{(\mathbf{v}(x))} = -\begin{pmatrix} 0 & 3g_1'(\mathbf{f}, x) \\ 0 & 0 \end{pmatrix}$, so that
$\beta(\mathbf{f}, x) = \begin{pmatrix} 3f_3 g_1'(\mathbf{f}, x) \\ 0 \end{pmatrix}$ and $\Sigma_2(\mathbf{f}) = 9f_3^2 \begin{pmatrix} \int (g_1'(\mathbf{f}, x))^2 u_0(\mathbf{f}, x)dx & 0 \\ 0 & 0 \end{pmatrix}$.

4. Parametric inference in non centered stationary regime.

In this section, we assume that the process (3)-(4)-(5) is in non centered stationary regime. Thus, we assume [H5] and

- [H7] $X_0 \sim u_m(\mathbf{f}, .) = u_0(\mathbf{f}, . - m)$,

i.e. we observe the process such that:

$$dX_t = -\int \Phi \star \mathcal{L}(X_t)(X_t - y)dy\, dt + \sigma dW_t, \quad X_0 = \eta \sim u_m(\mathbf{f}, x)dx, \tag{43}$$

where $u_m(\mathbf{f},.) = u_0(\mathbf{f},. - m)$ and $u_0(\mathbf{f},.)$ is the unique symmetric solution of (9). Hence, for all $t \geq 0$, $\mathcal{L}(X_t) = u_m(\mathbf{f},x)dx$, so that

$$dX_t = b(\mathbf{f},m,X_t)\,dt + \sigma dW_t, \quad X_0 = \eta \sim u_m(\mathbf{f},x)dx, \quad b(\mathbf{f},m,x) = -\Phi \star u_m(\mathbf{f},.)(x)$$

In this case, $\mathbb{E}(X_t) = m$ for all t and m must be estimated in addition to $\mathbf{f}$. Now, we have, using (16),

$$b(\mathbf{f},m,x) = -\int \Phi(x-m-(y-m))u_m(\mathbf{f},y)dy = -\Phi \star u_0(\mathbf{f},.)(x-m) = b(\mathbf{f},x-m). \tag{44}$$

Hence, the drift term $b(\mathbf{f},m,x)$ satisfies, using Proposition 3.1 and (16)

$$b(\mathbf{f},m,x) = b(\mathbf{f},x-m) = \sum_{i=0}^{k-1} b_{2i+1}(\mathbf{f})(x-m)^{2i+1} \tag{45}$$

4.1. Estimation of the moments of u_m

Define the centered moments of (X_t), for $j \geq 0$

$$\gamma_{2\ell}(\mathbf{f},m) = \int_{\mathbb{R}} (x-m)^{2\ell} u_m(\mathbf{f},x)dx. \tag{46}$$

Since $u_m(\mathbf{f},x) = u_0(\mathbf{f},x-m)$, we have, using (13), $\gamma_{2\ell}(\mathbf{f},m) = \gamma_{2\ell}(\mathbf{f})$.
As m is unknown, we also need an estimator for it. We define

$$\widehat{m} = T^{-1}\int_0^T X_s ds \quad \text{and } \widetilde{\gamma}_{2\ell}(T) = T^{-1}\int_0^T (X_s - \widehat{m})^{2\ell} ds, \;\; \ell \geq 0. \tag{47}$$

The following holds.

Proposition 4.1. *Assume [H5]-[H7]. As $T \to \infty$, $\widehat{m} \to_{a.s.} m$ and for $\ell \geq 1$, $\widetilde{\gamma}_{2\ell}(T) \to_{a.s.} \gamma_{2\ell}(\mathbf{f}) = \int x^{2\ell} u_0(\mathbf{f},x)dx$. Moreover,*

$$\sqrt{T}(\widehat{m} - m) = \frac{\sigma}{\sqrt{T}}\int_0^T h_0'(X_s - m)dW_s + o_P(1), \tag{48}$$

$$\sqrt{T}(\widetilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}(\mathbf{f})) = \frac{\sigma}{\sqrt{T}}\int_0^T h_\ell'(X_s - m)dW_s + o_P(1), \tag{49}$$

where (see (25))

$$h_0'(x) = -\frac{2}{\sigma^2 u_0(\mathbf{f},x)}\int_{-\infty}^x v\, u_0(\mathbf{f},v)dv, \quad h_\ell'(x) = g_\ell'(x) = -\frac{2}{\sigma^2 u_0(\mathbf{f},x)}\int_{-\infty}^x (v^{2\ell} - \gamma_{2\ell}(\mathbf{f}))u_0(\mathbf{f},v)dv$$

satisfy for all ℓ, $\int (h_\ell'(x))^2 u_0(\mathbf{f},x)dx < +\infty$. Consequently, for all k, the vector $(\sqrt{T}(\widehat{m}-m), \sqrt{T}(\widetilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}(\mathbf{f})), \ell = 1,\ldots,k-1)$ converges in distribution to $\mathcal{N}_k(\mathbf{0}, \sigma^2(\int h_i'(v)h_j'(v)u_0(\mathbf{f},v)dv))_{0\leq i,j\leq k})$

4.2. Explicit estimators of $(m,\mathbf{f})$

We consider the contrast, using (47) ,

$$\Lambda_T(\mathbf{b}(\mathbf{f})) = \int_T^{2T} b(\mathbf{f}, X_s - \widehat{m})dX_s - \frac{1}{2}\int_T^{2T} b^2(\mathbf{f}, X_s - \widehat{m})ds. \tag{50}$$

As previously, we proceed in two steps. First we define $\widetilde{\mathbf{b}(\mathbf{f})}_T$ the estimator of $\mathbf{b}(\mathbf{f})$ by

$$\widetilde{\Psi}_T \widetilde{\mathbf{b}(\mathbf{f})}_T = \widetilde{Z}_T \tag{51}$$

with (see (14) and (33))

(52) $$\widetilde{Z}_T = \int_T^{2T} z_k(X_s - \widehat{m})dX_s,\ \widetilde{\Psi}_T = (\int_T^{2T} (X_s - \widehat{m})^{2i+2j+2}ds)_{0\le i,j\le k-1}.$$

Now, define the estimator $\widetilde{\mathbf{f}}_T$ by

(53) $$\widetilde{\mathbf{b}(\mathbf{f})}_T = \widetilde{M}_k \widetilde{\mathbf{f}}_T, \quad \widetilde{M}_k = M_k^{(\widetilde{\gamma}_{2\ell}(T),\ell=0,\dots,k-1)}.$$

Theorem 4.1. *Assume [H5] and [H7].*
• *The estimator $\widetilde{\mathbf{f}}_T$ is consistent and $\sqrt{T}(\widetilde{\mathbf{f}}_T - \mathbf{f})$ converges in distribution to $\mathcal{N}_k(\mathbf{0}, \sigma^2\Sigma(\mathbf{f}))$ with $\Sigma(\mathbf{f})$ defined in (34).*
• *The joint asymptotic distribution of $(\widehat{m}, \widetilde{\mathbf{f}}_T)$ is as follows*

$$\sqrt{T}\begin{pmatrix}\widehat{m} - m\\ \widetilde{\mathbf{f}}_T - \mathbf{f}\end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_{1+k}\left(\mathbf{0}, \sigma^2\begin{pmatrix}\int [h_0'(x)]^2 u_0(\mathbf{f},x)dx & \int h_0'(x)[\beta(x)]' u_0(\mathbf{f},x)dx\\ \int h_0'(x)\beta(x)u_0(\mathbf{f},x)dx & \Sigma(\mathbf{f})\end{pmatrix}\right),$$

where h_0' is defined in Proposition 4.1, $\beta = \beta(\mathbf{f}, .)$ is defined in (36) and $\Sigma(\mathbf{f})$ is defined in (34).

It is interesting to note that $\widetilde{\mathbf{f}}_T$ and $\widehat{\mathbf{f}}_T$ have the same asymptotic distribution.

Example 1. Consider the Ornstein-Uhlenbeck process in non centered stationary regime: $dX_t = -f(X_t - m)dt + \sigma dW_t$ with stationary distribution equal to $u_m(f,x)dx = \mathcal{N}(m, \sigma^2/2f)$. The MLE based on $(X_t, t \in [0,T])$ can be computed in this model:

$$\widehat{\widehat{\mathbf{f}}}_T = -\frac{\int_0^T (X_t - \widehat{\widehat{m}}_T)dX_t}{\int_0^T (X_t - \widehat{\widehat{m}}_T)^2 dt}, \quad \widehat{\widehat{m}}_T = \frac{X_T - X_0}{T\widehat{\widehat{\mathbf{f}}}_T} + \frac{1}{T}\int_0^T X_s ds, \quad I(f) = \begin{pmatrix} f^2 & 0\\ 0 & \sigma^2/(2f)\end{pmatrix}.$$

The asymptotic distribution of $\sqrt{T}(\widehat{\widehat{m}}_T - m, \widehat{\widehat{\mathbf{f}}}_T - f)$ is the Gaussian law $\mathcal{N}_2(0, \sigma^2 I^{-1}(f))$. The maximum contrast estimator is given by:

$$\widetilde{f}_T = -\frac{\int_T^{2T}(X_t - \widehat{m})dX_t}{\int_T^{2T}(X_t - \widehat{m})^2 dt}, \quad \widehat{m} = \frac{1}{T}\int_0^T X_s ds.$$

We have $M_1 = -1$, $\Psi(f) = \sigma^2/2f$, $\Sigma_1(f) = 2f/\sigma^2$, $\Sigma_2(f) = 0$. The contrast estimator has the same asymptotic distribution as the exact MLE.

5. Concluding remarks

In this paper, we study the estimation of an unknown parameter $\mathbf{f} = (f_{2j+1}, j = 0,\dots,k-1)'$ in the interaction term $\Phi(\mathbf{f}, x)$ from the continuous observation of the McKean-Vlasov process

(54) $$dX_t = -\Phi(\mathbf{f}, .) \star \mathcal{L}(X_t)(X_t)\, dt + \sigma dW_t$$

with $\Phi(\mathbf{f}, x) = \sum_{j=0}^{k-1} f_{2j+1}x^{2j+1},\ f_1 > 0, f_{2j+1} \ge 0, j = 1,\dots,k-1$, throughout the time interval $[0, 2T]$. Here $\mathcal{L}(X_t)$ represents the law of X_t. The interaction term $\Phi(\mathbf{f}, x)$ is an odd increasing polynomial with known degree $2k-1$ so that $\Phi(\mathbf{f}, .) \star \mathcal{L}(X_t)$ only depends on $\mathbf{f}$ and the moments of $\mathcal{L}(X_t)$. Contrary to SDEs, stationary distributions of model (54) are uniquely determined only if the expectation of (X_t) is specified. We assume here that (X_t) is in stationary regime with given expectation. Hence its moments do not depend on t. The exact log-likelihood can be studied theoretically (Proposition 3.2) but does not lead to computable estimators. This is why we use a two-step procedure. First we build estimators of the stationary distribution moments based on the sample path $(X_t, t \in [0,T])$. Then, to build an explicit contrast, we plug these moment estimators into the exact conditional log-likelihood given X_T of (54), based on the sample path $(X_t, t \in [T, 2T])$. We prove that these estimators are consistent and asymptotically Gaussian with rate $\sqrt{T}$.

In here, we assume that the degree of the interaction function Φ is known. When it is unknown, the question of estimating this degree is of interest but beyond the scope of this paper.

Extension of this approach to multidimensional models McKean-Vlasov SDEs is possible. Indeed, following [29], we may consider an interaction of the form $\Phi(x) = \frac{x}{\|x\|}\varphi(\|x\|)$ with $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\varphi(r) = \sum_{j=0}^{k-1} f_{2j+1}r^{2j+1}$ with

$f_1 > 0, f_{2j+1} \geq 0, j = 1, \dots k-1$. In this case, $\Phi = \nabla W$, with $W(x) = \frac{1}{2}\sum_{j=0}^{k-1} \frac{1}{2j+1} f_{2j+1} \|x\|^{2j}$. The process admits a unique stationary distribution u_m with given expectation m (see *e.g.* [42]) and $\Phi \star u_m$ only depends on moments of u_m.

Extensions of this work naturally comprise the introduction of an additional potential term $V(\alpha, x)$ in the drift of equation (54) and a more general form of the interaction $\Phi(\beta, .)$ with unknown parameters α, β.

In practice, only discretizations of the sample path are generally available. This study could be extended to take into account discrete observations.

6. Proofs

Proof of Lemma 2.2. Since u admits a $\ell + 1$-th order moment, $\Phi \star u$ is well-defined. As Φ is odd and u is symmetric, we have:

$$\Phi \star u(-x) = \int \Phi(-x-y)u(y)dy = -\int \Phi(x+y)u(y)dy$$
$$= -\int \Phi(x-y)u(-y)dy = -\int \Phi(x-y)u(y)dy = -\Phi \star u(x).$$

Let $W(x) = \int_0^x \Phi(y)dy$. Then, $W \star u(x) = \int W(x-y)u(y)dy$ satisfies $(W \star u)'(x) = \Phi \star u(x)$ so that $W \star u(x) = \int_0^x \Phi \star u(y)dy + W \star u(0)$.

Now, let $x \geq 0$, we have $W''(x) = \Phi'(x) \geq \lambda$, thus $W'(x) \geq \lambda x + W'(0) = \lambda x$. This implies, $W(x) \geq \frac{\lambda x^2}{2} + W(0)$ and thus, as $\int y u(y)dy = 0$,

$$W \star u(x) \geq \int [\frac{\lambda(x-y)^2}{2} + W(0)]u(y)dy = \frac{\lambda x^2}{2} + \frac{\lambda}{2}\int y^2 u(y)dy + W(0).$$

Consequently,

$$\int_0^x \Phi \star u(y)dy \geq \frac{\lambda x^2}{2} + \frac{\lambda}{2}\int y^2 u(y)dy + W(0) - W \star u(0) = \frac{\lambda x^2}{2} + C.$$

As $\int_0^x \Phi \star u(y)dy$ is symmetric, the result holds for all x. □

Proof of Proposition 2.2. The result for (Y_t) is standard. By computing the scale and the speed density, we obtain that (Y_t) is positive recurrent and admits u_m as invariant density (see *e.g.* [38]). When $Y_0 = \eta \sim u_m$, by the uniqueness of solution, we obtain that $Y_t \equiv X_t$ for all $t \geq 0$. □

Proof of Proposition 3.2. Recall that $X_t \equiv Y_t$ (see Proposition 2.2). We have, applying the ergodic theorem, as $u_0(\mathbf{f}_0, .)$ has moments of any order by (10),

$$\frac{1}{T}(\ell_T(\mathbf{f}) - \ell_T(\mathbf{f}_0)) \to_{\mathbb{P}_{\mathbf{f}_0}} -\frac{1}{2}\int (b(\mathbf{f}, x) - b(\mathbf{f}_0, x))^2 u_0(\mathbf{f}_0, x)dx = -\frac{1}{2\sigma^2}K(\mathbf{f}_0, \mathbf{f}) \quad \text{a.s..}$$

Now, $K(\mathbf{f}_0, \mathbf{f}) = 0$ is equivalent to "for all x, $b(\mathbf{f}, x) = b(\mathbf{f}_0, x)$", as $u(\mathbf{f}_0, .)$ is positive and continuous on $\mathbb{R}$. This in turn implies that $u(\mathbf{f}, .) \equiv u(\mathbf{f}_0, .)$, as two diffusions with the same drift and diffusion coefficients have the same invariant density, and $M_k(\mathbf{f})\mathbf{f} = M_k(\mathbf{f}_0)\mathbf{f}_0$ (see Proposition 3.1). As $u(\mathbf{f}, .) \equiv u(\mathbf{f}_0, .)$, their moments are identical, *i.e.* $\gamma_{2\ell}(\mathbf{f}) = \gamma_{2\ell}(\mathbf{f}_0)$ for all ℓ. Thus, $M_k(\mathbf{f}) = M_k(\mathbf{f}_0)$. As $M_k(\mathbf{f}_0)$ is invertible, we conclude $\mathbf{f} = \mathbf{f}_0$.

Now, the proof of consistency of the maximum likelihood estimator standardly follows.

Next,

$$\sigma^2 \frac{\partial \ell_T}{\partial f_{2i+1}}(\mathbf{f}) = \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s)dX_s - \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s)b(\mathbf{f}, X_s)ds$$
$$= \sigma \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s)dW_s$$
$$\sigma^2 \frac{\partial^2 \ell_T}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f}) = \int_0^T \frac{\partial^2 b}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f}, X_s)(dX_s - b(\mathbf{f}, X_s)ds)$$

$$-\int_0^T (\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},X_s)\frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f},X_s))ds$$
$$=\sigma\int_0^T \frac{\partial^2 b}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f},X_s)dW_s - \int_0^T (\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},X_s)\frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f},X_s))ds.$$

The functions $x\to\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},x), x\to\frac{\partial^2 b}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f},x)$ are polynomial and thus integrable with respect to $u_0(\mathbf{f},.)$. Under $\mathbb{P}_{\mathbf{f}}$, by the ergodic theorem and the central limit theorem for stochastic integrals, for all i,i',

$$\frac{\sigma^2}{\sqrt{T}}(\frac{\partial \ell_T}{\partial f_{2i+1}}(\mathbf{f}), i=0,\dots k-1)' \to_{\mathcal{L}} \mathcal{N}(0,\sigma^2 I(\mathbf{f})),$$
$$(\frac{1}{T}\int_0^T \frac{\partial^2 b}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f},X_s)dW_s)_{i,i'} \ \to\ 0, \qquad (\frac{\sigma^2}{T}\frac{\partial^2 \ell_T}{\partial f_{2i+1}\partial f_{2i'+1}}(\mathbf{f}))_{i,i'} \to -I(\mathbf{f}),$$

where $I(\mathbf{f}) = (\int [\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},x)]\frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f},x)]u_0(\mathbf{f},x)dx)_{0\le i,i'\le k-1}$.
For any vector $a=(a_1\dots a_k)'$, $a'I(\mathbf{f})a = \int[\sum_{i=0}^{k-1} a_i\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f},x)]^2 u_0(\mathbf{f},x)dx > 0$ as the function under the integral is a polynomial and $u_0(\mathbf{f},x)$ is positive for all x. By standard methods, we can prove that the maximum likelihood $\widehat{\widehat{\mathbf{f}}}_T$ associated with (21) satisfies that, under $\mathbb{P}_{\mathbf{f}_0}$,

$$\sqrt{T}(\widehat{\widehat{\mathbf{f}}}_T - \mathbf{f}_0) \to_{\mathcal{L}} \mathcal{N}(0,\sigma^2 I^{-1}(\mathbf{f}_0)).$$

□

Proof of Proposition 3.3. We rely on results of [25] recalled in Section 7 (Proposition 7.1). These results concern the infinitesimal generator L of $\mathcal{D}$ of (11) here for $m=0$. The definitions of $L,\mathcal{D}$ are given in (74)-(75)) (see Appendix). For g an element of $\mathcal{D}$, we have by the Ito formula,

$$-\int_0^T Lg(X_s)ds = g(X_T) - g(X_0) + \sigma\int_0^T g'(X_s)dW_s,$$

where $Lg(x) = b(\mathbf{f},x)g'(x) + \frac{\sigma^2}{2}g''(x)$. By Proposition 7.1, the range of $\mathcal{D}$ is exactly the set of functions $h\in\mathbb{L}^2(u_0(\mathbf{f},x)dx)$ such that $\int h(x)u_0(\mathbf{f},x)dx = 0$. Therefore, let $g_\ell(\mathbf{f},.)$ be any element of $\mathcal{D}$ such that $Lg_\ell(\mathbf{f},x) = -(x^{2\ell}-\gamma_{2\ell}(\mathbf{f}))$. We have (see (78))

$$g'_\ell(\mathbf{f},x) = g'_\ell(x) = -2\sigma^{-2}u_0^{-1}(\mathbf{f},x)\int_{-\infty}^x (y^{2\ell}-\gamma_{2\ell})u_0(\mathbf{f},y)dy. \tag{55}$$

As, by the definition of $\mathcal{D}$, g_ℓ belongs to $\mathbb{L}^2(u_0(x)dx)$, and (X_t) is stationary with marginal distribution $u_0(x)dx$,

$$\mathbb{E}(g_\ell(X_T)-g_\ell(X_0))^2 \le 2\int g_\ell^2(x)u_0(\mathbf{f},x)dx. \tag{56}$$

Thus,

$$\sqrt{T}(\widehat{\gamma}_{2\ell}(T)-\gamma_{2\ell}) = -\frac{1}{\sqrt{T}}\int_0^T Lg_\ell(X_s)ds = \frac{\sigma}{\sqrt{T}}\int_0^T g'_\ell(X_s)dW_s + \frac{1}{\sqrt{T}}(g_\ell(X_T)-g_\ell(X_0))$$
$$= \frac{\sigma}{\sqrt{T}}\int_0^{2T}\mathbf{1}_{[0,T]}(s)g'_\ell(X_s)dW_s + o_P(1), \tag{57}$$

by (56). Proposition 7.1, for all ℓ, $\int(g'_\ell(x))^2u_0(,x)dx < +\infty$ so that $V(\mathbf{f})$ is well defined. So the vector $T^{1/2}(\widehat{\gamma}_{2\ell}(T)-\gamma_{2\ell})'_{\ell=1,\dots,k}$ converges in distribution to $\mathcal{N}_k(0,\sigma^2V(\mathbf{f}))$ with $V(\mathbf{f}) = (V_{i,j}(\mathbf{f})_{0\le i,j\le k-1}$ and $V_{i,j}(\mathbf{f}) = \int g'_i(x)g'_j(x)u_0(\mathbf{f},x)dx$.

Proof of Proposition 3.4. This result is classical. By the ergodic theorem applied to $(X_t \equiv Y_t)$, we have that Ψ_T/T converges *a.s.* to $\Psi(\mathbf{f})$. For any vector $a' = (a_0, \dots, a_{k-1})$,

$$a'\Psi(\mathbf{f})a = \int_{\mathbb{R}} (\sum_{\ell=0}^{k-1} a_\ell x^{2\ell+1})^2 u_0(\mathbf{f}, x)dx > 0,$$

as the integrand is a polynomial and u is $\mathbb{R}$-supported. Thus, $\Psi(\mathbf{f})$ is positive definite.
We write:

$$\int_T^{2T} X_s^{2i+1} dX_s = \int_T^{2T} X_s^{2i+1} \sum_{j=0}^{k-1} b_{2j+1}(\mathbf{f}) X_s^{2j+1} ds + \sigma \int_T^{2T} X_s^{2i+1} dW_s.$$

Thus, for large enough T, (see (14))

$$(\frac{1}{T}\Psi_T)^{-1}\frac{1}{T}Z_T = \widehat{\mathbf{b}(\mathbf{f})}_T = \mathbf{b}(\mathbf{f}) + (\frac{1}{T}\Psi_T)^{-1}\frac{\sigma}{T}\int_T^{2T} z_k(X_s)dW_s$$

As Ψ_T/T converges *a.s.* to $\Psi(\mathbf{f})$, the vector of stochastic integrals $\frac{\sigma}{T}\int_T^{2T} z_k(X_s)dW_s$ converges *a.s.* to 0. Moreover, $\frac{\sigma}{\sqrt{T}}\int_T^{2T} z_k(X_s)dW_s$ converges in distribution to $\mathcal{N}_k(0, \sigma^2\Psi(\mathbf{f}))$. Consequently, $\widehat{\mathbf{b}(\mathbf{f})}_T$ converges to $\mathbf{b}(\mathbf{f})$ and $\sqrt{T}(\widehat{\mathbf{b}(\mathbf{f})}_T - \mathbf{b})$ converges in distribution to the Gaussian law $\mathcal{N}_k(0, \sigma^2\Psi^{-1}(\mathbf{f}))$. □

Proof of Theorem 3.1. For the proof, we set $\Psi(\mathbf{f}) = \Psi$, $M_k(\mathbf{f}) = M_k$, $u_0(\mathbf{f}, .) = u_0(.)$, $\gamma_{2\ell}(\mathbf{f}) = \gamma_{2\ell}$, $V = V(\mathbf{f})$ and $\widehat{\mathbf{b}}_T = \widehat{\mathbf{b}(\mathbf{f})}_T$. The relation $\widehat{\mathbf{b}}_T = \widehat{M}_k\widehat{\mathbf{f}}_T$ implies $Z_T = \Psi_T\widehat{M}_k\widehat{\mathbf{f}}_T$ (see (28)). We have

$$Z_T = \Psi_T\mathbf{b} + \sigma\int_T^{2T} z_k(X_s)dW_s,$$

$$\Psi_T\widehat{M}_k\widehat{\mathbf{f}}_T = \Psi_T M_k\mathbf{f} + \Psi_T M_k(\widehat{\mathbf{f}}_T - \mathbf{f}) + \Psi_T(\widehat{M}_k - M_k)\mathbf{f} + \Psi_T(\widehat{M}_k - M_k)(\widehat{\mathbf{f}}_T - \mathbf{f}).$$

Therefore, noting that $\Psi_T M_k\mathbf{f} = \Psi_T\mathbf{b}$, we obtain:

$$\frac{1}{T}\Psi_T M_k\sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}}\int_T^{2T} z(X_s)dW_s - \Psi\sqrt{T}(\widehat{M}_k - M_k)\mathbf{f} - R_T \tag{58}$$

$$= \Psi M_k\sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) + S_T, \text{ with} \tag{59}$$

$$R_T = (\frac{1}{T}\Psi_T - \Psi)\sqrt{T}(\widehat{M}_k - M_k)\mathbf{f} + \frac{1}{T}\Psi_T\sqrt{T}(\widehat{M}_k - M_k)(\widehat{\mathbf{f}}_T - \mathbf{f}), \tag{60}$$

$$S_T = \sqrt{T}(\frac{1}{T}\Psi_T - \Psi)M_k(\widehat{\mathbf{f}}_T - \mathbf{f}). \tag{61}$$

Finally,

$$\Psi M_k\sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}}\int_T^{2T} z(X_s)dW_s - \Psi\sqrt{T}(\widehat{M}_k - M_k)\mathbf{f} - R_T - S_T.$$

It is the sum of two main terms and two remainders. The second term $\sqrt{T}(\widehat{M}_k - M_k)$ depends on the observation $(X_t, t \in [0, T])$ while the first one depends on the sample path $(X_t, t \in [T, 2T])$.
To study $\sqrt{T}(\widehat{M}_k - M_k)$, we use the fact that the vector of centered and normalized moments $T^{1/2}((\widehat{\gamma}_{2\ell}(T) - \gamma_{2\ell})'_{\ell=1,\dots,k}$ converges in distribution to $\mathcal{N}_k(0, \sigma^2 V)$ with $V = (V_{ij})_{0\leq i,j\leq k-1}$ and $V_{ij} = \int g_i'(x)g_j'(x)u(x)dx$ (Proposition 3.3).
Consequently, as we have $\sqrt{T}(\widehat{M}_k - M_k) = O_P(1)$, $(\Psi_T/T) - \Psi = o_P(1)$, $\widehat{\mathbf{f}} - \mathbf{f} = o_P(1)$, we concude that the remainder term $R_T = o_P(1)$.
We can treat analogously each term of $\sqrt{T}((\Psi_T/T) - \Psi)$ and prove that $\sqrt{T}((\Psi_T/T) - \Psi) = O_P(1)$. Consequently, $S_T = o_P(1)$.

Therefore, from (58) and (59),

$$\sqrt{T}(\widehat{\mathbf{f}}_T-\mathbf{f})=\frac{\sigma}{\sqrt{T}}\int_T^{2T}M_k^{-1}\Psi^{-1}z_k(X_s)dW_s-M_k^{-1}\sqrt{T}(\widehat{M}_k-M_k)\mathbf{f}+o_P(1)$$

$$(62)\qquad =\frac{\sigma}{\sqrt{T}}\int_0^{2T}\left(\mathbf{1}_{[T,2T]}(s)\alpha(\mathbf{f},X_s)-\mathbf{1}_{[0,T]}(s)\beta(\mathbf{f},X_s)\right)dW_s+o_P(1),$$

where

$$(63)\qquad \alpha_\ell(\mathbf{f},x)=\sum_{u=0}^{k-1}[M_k^{-1}]_{\ell u}\sum_{j=0}^{k-1}[\Psi^{-1}]_{uj}x^{2j+1}=[M_k^{-1}\Psi^{-1}z_k(x)]_\ell\,,$$

$$(64)\qquad \beta_\ell(\mathbf{f},x)=\sum_{j=0}^{k-1}[M_k^{-1}]_{\ell j}\sum_{v=j}^{k-1}\binom{2v+1}{v-j}g'_{v-j}(x)f_{2v+1}\mathbf{1}_{j\le v}=[M_k^{-1}M_k^{(g'(x))}\mathbf{f}]_\ell,\text{ and}$$

$M_k^{(g'(x))}$ is the matrix $M_k^{(v)}$ for $v=(g'_0(x),\dots,g'_{k-1}(x))'$ (with $g'_0=0$). Finally, applying the ergodic theorem for stochastic integrals yields

$$\sqrt{T}(\widehat{\mathbf{f}}-\mathbf{f})\to_{\mathcal{L}}\mathcal{N}_k(0,\sigma^2\Sigma(\mathbf{f}))\quad\text{where}\quad \Sigma(\mathbf{f})=\Sigma_1(\mathbf{f})+\Sigma_2(\mathbf{f})$$

with

$$\Sigma_1(\mathbf{f})=\int\alpha(\mathbf{f},x)[\alpha(\mathbf{f},x)]'u_0(\mathbf{f},x)dx=M_k^{-1}(\mathbf{f})\Psi^{-1}(\mathbf{f})(M_k^{-1}(\mathbf{f}))'$$

and $\Sigma_2(\mathbf{f})=\int\beta(\mathbf{f},x)[\beta(\mathbf{f},x)]'u_0(\mathbf{f},x)dx$.
□

Proof of Proposition 3.5. Define $D_k=diag((2\ell-1),\ell=1,\dots,k)$ the diagonal matrix with diagonal element $\ell(2\ell-1)$. We have $B(\mathbf{f})=\sigma^2D_k(1\ \gamma_2(\mathbf{f})\ \dots\gamma_{2(k-1)}(\mathbf{f}))'$. The vector $\sqrt{T}(1\ \widehat{\gamma}_2(T)\ \dots\widehat{\gamma}_{2(k-1)}(T))'-(1\ \gamma_2(\mathbf{f})\ \dots\gamma_{2(k-1)}(\mathbf{f}))'$ converges in distribution to $\mathcal{N}_k(\mathbf{0},\sigma^2V(\mathbf{f}))$ where $V(\mathbf{f})$ is defined in Proposition 3.3. Consequently, the vector $\sqrt{T}(\widetilde{B}_T-B(\mathbf{f}))$ converges in distribution to $\mathcal{N}(\mathbf{0},K)$ with

$$(65)\qquad K=\sigma^4D_kV(\mathbf{f})D_k.$$

The result of Proposition 3.5 follows.
□

Proof of Proposition 4.1. This proposition is analogous to Proposition 3.3.
For simplicity, set $u_0(\mathbf{f},.)=u$, $u_m(\mathbf{f},.)=u_m$, $\gamma_{2\ell}(\mathbf{f})=\gamma_{2\ell}$. By the ergodic theorem $\widehat{m}=T^{-1}\int_0^TX_sds\to\int xu_m(x)dx=m$. For $\widetilde{\gamma}_{2\ell}(T)$, we write:

$$\widetilde{\gamma}_{2\ell}(T)=\frac{1}{T}\int_0^T(X_s-m+m-\widehat{m})^{2\ell}ds=\frac{1}{T}\int_0^T(X_s-m)^{2\ell}ds$$
$$+(m-\widehat{m})\sum_{k=1}^{2\ell}\binom{2\ell}{k}\frac{1}{T}\int_0^T(X_s-m)^{2\ell-k}ds(m-\widehat{m})^{k-1}.$$

The first term of the sum converges to $\int(x-m)^{2\ell}u_m(x)dx=\int(x-m)^{2\ell}u(x-m)dx=\int x^{2\ell}u(x)dx=\gamma_{2\ell}$. The second term tends to 0 as $\widehat{m}-m$ tends to 0 and is multiplied by a factor tending to a limit.
We rely again on the results of Proposition 7.1. The infinitesimal generator L_m of (43) is given by

$$L_mg(x)=\frac{\sigma^2}{2u(x-m)}(g'u(.-m))'(x).$$

Thus, the equality

$$g(X_T)-g(X_0)=\int_0^TL_mg(X_s)ds+\sigma\int_0^Tg'(X_s)dW_s$$

implies, as (X_t) is stationary,

$$-\frac{1}{\sqrt{T}}\int_0^T L_m g(X_s)ds = \frac{\sigma}{\sqrt{T}}\int_0^T g'(X_s)dW_s + o_P(1).$$

For f a u_m square integrable function, the solution of $L_m g = -(f - \int f(x)u_m(x)dx)$ is

$$g'(x) = -\frac{2}{\sigma^2 u(x-m)}\int_{-\infty}^{x-m} u(v)\left(f(m+v) - \int f(m+y)u(y)dy\right)dv.$$

For $f(x) = x$, we get $g'(x) = -\frac{2}{\sigma^2 u(x-m)}\int_{-\infty}^{x-m} vu(v)dv := h_0'(x-m)$. Thus,

$$\sqrt{T}(\widehat{m} - m) = \frac{\sigma}{\sqrt{T}}\int_0^T h_0'(X_s - m)dW_s + o_P(1). \tag{66}$$

For $f(x) = (x-m)^{2\ell}$, we get $g'(x) = -\frac{2}{\sigma^2 u(x-m)}\int_{-\infty}^{x-m}(v^{2\ell} - \gamma_{2\ell})u(v)dv = g_\ell'(x-m) := h_\ell'(x-m)$ (see (55)). Therefore,

$$\frac{1}{\sqrt{T}}\int_0^T \left((X_s - m)^{2\ell} - \gamma_{2\ell}\right) = \frac{\sigma}{\sqrt{T}}\int_0^T h_\ell'(X_s - m)dW_s + o_P(1), \tag{67}$$

where, by Proposition 7.1, all the integrals $\int (h_\ell'(x))^2 u_0(x)dx = \int h_\ell'(x-m))^2 u(x-m)dx < +\infty$. Now, splitting $X_s - \widehat{m} = X_s - m + m - \widehat{m}$ yields,

$$\sqrt{T}(\widetilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}) = \frac{1}{\sqrt{T}}\int_0^T \left((X_s - m)^{2\ell} - \gamma_{2\ell}\right) + \sqrt{T}(m - \widehat{m})\binom{2\ell}{1}\frac{1}{T}\int_0^T (X_s - m)^{2\ell-1}ds$$
$$+ \sqrt{T}(m - \widehat{m})^2\sum_{k=2}^{2\ell}\binom{2\ell}{k}(m - \widehat{m})^{k-2}\frac{1}{T}\int_0^T (X_s - m)^{2\ell-k}ds.$$

For the second term, note that $\frac{1}{T}\int_0^T (X_s - m)^{2\ell-1}ds \to \int (x-m)^{2\ell-1}u_m(x)dx = \int x^{2\ell-1}u_0(x)dx = 0$ as u_0 is symmetric and $2\ell - 1$ is odd. Therefore, the second term is $o_P(1)$ as well as the third term. Therefore, we have obtained (48) and (49). The convergence in distribution result follows. □

Proof of Theorem 4.1 . Here again, we set $\Psi(\mathbf{f}) = \Psi$, $M_k(\mathbf{f}) = M_k$. We proceed as in Proposition 3.4 and Theorem 3.1. First, we prove, using (52),

$$\frac{\widetilde{\Psi}_T}{T} \to_{a.s.} \Psi \quad \text{and} \quad \sqrt{T}(\frac{\widetilde{\Psi}_T}{T} - \Psi) = O_P(1) \quad \text{and} \tag{68}$$

$$V_T = \sqrt{T}\left(\frac{\widetilde{Z}_T}{T} - \frac{\widetilde{\Psi}_T}{T}\mathbf{b}\right) \to_{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^2\Psi). \tag{69}$$

From these two results, as $\widetilde{\mathbf{b}(\mathbf{f})}_T := \widetilde{\mathbf{b}}_T = (\frac{\widetilde{\Psi}_T}{T})^{-1}\frac{\widetilde{Z}_T}{T}$ (see (51)), we deduce:

$$\sqrt{T}(\widetilde{\mathbf{b}}_T - \mathbf{b}) = (\frac{\widetilde{\Psi}_T}{T})^{-1}V_T \to_{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^2\Psi^{-1}). \tag{70}$$

Proof of (68): We have:

$$[\frac{\widetilde{\Psi}_T}{T}]_{ij} = \frac{1}{T}\int_T^{2T}(X_s - m + m - \widehat{m})^{2(i+j)+2}ds = \frac{1}{T}\int_T^{2T}(X_s - m)^{2(i+j)+2}ds$$
$$+(m - \widehat{m})\sum_{r=1}^{2(i+j)+2}\binom{2(i+j)+2}{r}(m - \widehat{m})^{r-1}\frac{1}{T}\int_T^{2T}(X_s - m)^{2(i+j)+2-r}ds,$$

where $\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+2-r}ds \to \int (x-m)^{2(i+j)+2-r}u_m(x)dx = \int x^{2(i+j)+2-r}u_0(x)dx$ and $m-\widehat{m}=o_P(1)$. Thus,

$$[\frac{\widetilde{\Psi}_T}{T}]_{ij} = \frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+2}ds + o_P(1) \to_{a.s.} \int (x-m)^{2(i+j)+2}u_m(x)dx = \gamma_{2(i+j)+2} = \Psi_{ij}.$$

Next,

$$\begin{aligned}
&\sqrt{T}([\frac{\widetilde{\Psi}_T}{T}]_{ij} - \gamma_{2(i+j)+2}) = \sqrt{T}(\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+2}ds - \gamma_{2(i+j)+2})\\
&+\sqrt{T}(m-\widehat{m})\binom{2(i+j)+2}{1}\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+1}ds\\
&+\sqrt{T}(m-\widehat{m})^2\sum_{r=2}^{2(i+j)+2}\binom{2(i+j)+2}{r}(m-\widehat{m})^{r-1}\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+2-r}ds,
\end{aligned}$$

As $2(i+j)+1$ is odd, $\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+1}ds \to_{a.s.} \int (x-m)^{2(i+j)+1}u_m(x)dx = 0$. Thus, the second term above is $\sqrt{T}(m-\widehat{m})\times o_{a.s.}(1) = O_P(1)$. The third term is $\frac{1}{\sqrt{T}}T(m-\widehat{m})^2\times O_P(1) = o_P(1)$.
For the first term, we prove as in Proposition 4.1 that

$$\begin{aligned}
\sqrt{T}(\frac{1}{T}\int_T^{2T}(X_s-m)^{2(i+j)+2}ds - \gamma_{2(i+j)+2}) &= \frac{\sigma}{\sqrt{T}}\int_T^{2T}h'_{i+j+1}(X_s-m)dW_s + o_P(1)\\
&\to_{\mathcal{L}} \mathcal{N}(0,\sigma^2\int [h'_{i+j+1}(x)]^2u(x)dx).
\end{aligned}$$

The proof of (68) is complete.

Proof of (69): We write:

$$\begin{aligned}
\frac{1}{T}\widetilde{Z}_{T,i} &= \frac{1}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+1}dX_s = \frac{1}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+1}b(X_s-\widehat{m},\mathbf{f})ds + T_{2,i} + T_{3,i}\\
&= \sum_{j=0}^{k-1}b_{2j+1}\frac{1}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+2j+1}ds + T_{2,i} + T_{3,i} = [\frac{\widetilde{\psi}_T}{T}\mathbf{b}]_i + T_{2,i} + T_{3,i},
\end{aligned}$$

where

$$T_{2,i} = \frac{1}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+1}(b(\mathbf{f},X_s-m) - b(\mathbf{f},X_s-\widehat{m}))ds, \tag{71}$$

$$T_{3,i} = \frac{\sigma}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+1}dW_s. \tag{72}$$

We have

$$T_{2,i} = (\widehat{m}-m)\sum_{j=0}^{k-1}b_{2j+1}\sum_{\ell=0}^{2j}T_{2,i,j,\ell}, \quad T_{2,i,j,\ell} = \frac{1}{T}\int_T^{2T}(X_s-\widehat{m})^{2i+1+\ell}(X_s-m)^{2j-\ell}ds.$$

Now,

$$\begin{aligned}
T_{2,i,j,\ell} &= \frac{1}{T}\int_T^{2T}(X_s-m+m-\widehat{m})^{2i+1+\ell}(X_s-m)^{2j-\ell}ds\\
&= \frac{1}{T}\int_T^{2T}\sum_{r=0}^{2i+1+\ell}\binom{2i+1+\ell}{r}(m-\widehat{m})^r(X_s-m)^{2i+1+2j-r}ds\\
&= \frac{1}{T}\int_T^{2T}(X_s-m)^{2i+1+2j}ds
\end{aligned}$$

$$+ (m - \widehat{m}) \sum_{r=1}^{2i+1+\ell} \binom{2i+1+\ell}{r} (m-\widehat{m})^{r-1} \frac{1}{T}\int_T^{2T} (X_s - m)^{2i+1+2j-r} ds$$
$$= o_P(1).$$

Indeed, $\frac{1}{T}\int_T^{2T}(X_s-m)^{2i+1+2j}ds \to 0$ since $2i+2j+1$ is odd. And the second term tends to 0. Thus, for $i = 0,1,\dots,k-1$

$$\sqrt{T}\,T_{2,i} = \sqrt{T}(\widehat{m} - m) \times o_P(1) = o_P(1). \tag{73}$$

Now, $\sqrt{T}\,T_3$ is a martingale such that $< \sqrt{T}\,T_3 >_T = \sigma^2 \widetilde{\Psi}_T/T \to \sigma^2\Psi$. Therefore, $\sqrt{T}\,T_3$ converges in distribution to $\mathcal{N}(\mathbf{0}, \sigma^2\Psi)$.
Finally, we have obtained $\sqrt{T}\left(\frac{\widetilde{Z}_T}{T} - \frac{\widetilde{\Psi}_T}{T}\mathbf{b}\right) = \sqrt{T}\,T_3 + o_P(1)$. The proof of (69) is achieved and (70) follows.

Now, we can complete the proof of Theorem 4.1. On one hand, we have the relation (see (72)-(73)):

$$\sqrt{T}\frac{\widetilde{Z}_T}{T} = \frac{\widetilde{\Psi}_T}{T}\mathbf{b} + \sqrt{T}\,T_3 + o_P(1).$$

On the other hand, we have:

$$\widetilde{\Psi}_T \widetilde{M}_k \widetilde{\mathbf{f}}_T = \widetilde{\Psi}_T M_k \mathbf{f} + \widetilde{\Psi}_T M_k(\widetilde{\mathbf{f}}_T - \mathbf{f}) + \widetilde{\Psi}_T(\widetilde{M}_k - M_k)\mathbf{f} + \widetilde{\Psi}_T(\widetilde{M}_k - M_k)(\widetilde{\mathbf{f}}_T - \mathbf{f}).$$

Note that $\mathbf{b} = M_k\mathbf{f}$ and $\widetilde{Z}_T = \widetilde{\Psi}_T\widetilde{\mathbf{b}}_T = \widetilde{\Psi}_T\widetilde{M}_k\widetilde{\mathbf{f}}_T$. Therefore, we obtain the relation:

$$\frac{\widetilde{\Psi}_T}{T} M_k \sqrt{T}(\widetilde{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}}\left(\int_T^{2T}(X_s - \widehat{m})^{2i+1}dW_s\right)_{i=0,\dots,k-1} + o_P(1)$$
$$- \frac{\widetilde{\Psi}_T}{T}\sqrt{T}(\widetilde{M}_k - M_k)\mathbf{f} - \frac{\widetilde{\Psi}_T}{T}\sqrt{T}(\widetilde{M}_k - M_k)(\widetilde{\mathbf{f}}_T - \mathbf{f}).$$

This yields:

$$\Psi M_k \sqrt{T}(\widetilde{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}}\int_T^{2T} z(X_s - m)dW_s - \Psi\sqrt{T}(\widetilde{M}_k - M_k)\mathbf{f}$$
$$- \widetilde{R}_T - \widetilde{S}_T + K_T + o_P(1),$$

where

$$\widetilde{R}_T = (\frac{\widetilde{\Psi}_T}{T} - \Psi)\sqrt{T}(\widetilde{M}_k - M_k)\mathbf{f} + \frac{\widetilde{\Psi}_T}{T}\sqrt{T}(\widetilde{M}_k - M_k)(\widetilde{\mathbf{f}}_T - \mathbf{f}),$$
$$\widetilde{S}_T = \sqrt{T}(\frac{\widetilde{\Psi}_T}{T} - \Psi)M_k(\widetilde{\mathbf{f}}_T - \mathbf{f}),$$
$$K_T = \frac{\sigma}{\sqrt{T}}\int_T^{2T}[z(X_s - \widehat{m}) - z(X_s - m)]dW_s.$$

As previously, we prove that $\widetilde{R}_T = o_P(1)$, $\widetilde{S}_T = o_P(1)$ using Proposition 4.1. We have to look at K_T. We have:

$$K_{T,i} = \sigma(m - \widehat{m})\frac{1}{\sqrt{T}}\int_T^{2T}\sum_{\ell=0}^{2i}(X_s - m + m - \widehat{m})^{2i-\ell}(X_s - m)^{\ell}dW_s$$
$$= \sigma(m - \widehat{m})\sum_{\ell=0}^{2i}\sum_{j=0}^{2i-\ell}(m - \widehat{m})^j\binom{2i-\ell}{j}\frac{1}{\sqrt{T}}\int_T^{2T}(X_s - m)^{2i-j}dW_s.$$

Each term $\frac{1}{\sqrt{T}}\int_T^{2T}(X_s-m)^{2i-j}dW_s$ converges in distribution while $m-\widehat{m}$ tends to 0. So $K_{T,i}=o_P(1)$ for $i=0,\ldots,k-1$. Now, the term $\sqrt{T}(\widetilde{M}_k-M_k)\mathbf{f}$ can be treated as previously in Theorem 33 and we can write:

$$\begin{aligned}\sqrt{T}(\widetilde{\mathbf{f}}_T-\mathbf{f}) &= \frac{\sigma}{\sqrt{T}}(\Psi M_k)^{-1}\int_T^{2T} z_k(X_s-m)dW_s-\sqrt{T}M_k^{-1}(\widetilde{M}_k-M_k)\mathbf{f}+o_P(1)\\ &= \frac{\sigma}{\sqrt{T}}\int_0^{2T}\left((\Psi M_k)^{-1}z_k(X_s-m)\mathbf{1}_{[T,2T]}(s)+\mathbf{1}_{[0,T]}(s)\beta(X_s-m)\right)dW_s+o_P(1)\\ &= \frac{\sigma}{\sqrt{T}}\int_0^{2T}\left(\mathbf{1}_{[T,2T]}(s)\alpha(X_s-m)-\mathbf{1}_{[0,T]}(s)\beta(X_s-m)\right)dW_s+o_P(1),\end{aligned}$$

with $\alpha(x)=\alpha(\mathbf{f},x),\beta(x)=\beta(\mathbf{f},x)$ defined in (63) and (64). Therefore, $\sqrt{T}(\widetilde{\mathbf{f}}_T-\mathbf{f})$ converges in distribution to $\mathcal{N}(0,\sigma^2\Sigma(\mathbf{f}))$ with $\Sigma(\mathbf{f})$ defined in (34).

The result concerning the joint distribution follows from (48) and (58).□

7. Appendix

We now state the central limit theorem associated with (12), describe the properties of the infinitesimal generator of (11) and the conditions for ρ-mixing (see *e.g.* [25] and references therein).
Let (Y_t) be the solution of (11) and denote by L the infinitesimal generator of the SDE (11),

$$Lg=(\sigma^2/2)g''-\Phi\star u_m(.)g'=\frac{\sigma^2}{2u_m(w)}\left(g'u_m\right)'. \tag{74}$$

The operator L acting on $\mathbb{L}^2(u_m(x)dx)$ has domain $\mathcal{D}$ given by

$$\mathcal{D}=\{g\in\mathbb{L}^2(u_m(x)dx), g' \text{ absolutely continuous}, Lg\in\mathbb{L}^2(u_m(x)dx),\lim_{|x|\to\infty}g'(x)u_m(x)=0\}. \tag{75}$$

For all $g\in\mathcal{D}$, $\int Lg(x)u_m(x)dx=0$.

Proposition 7.1. *Let $f\in\mathbb{L}^2(u_m(x)dx)$, set $f_c=f-\int_{\mathbb{R}}f(x)u_m(x)dx$ and denote by $\langle .,.\rangle_{u_m}$ the scalar product of $\mathbb{L}^2(u_m(x)dx)$.*

1. *If $f_c\in Range(\mathcal{D})$, then, as T tends to infinity, the solution (Y_t) of (11) satisfies*

$$\frac{1}{\sqrt{T}}\int_0^T f_c(Y_s)ds\to_{\mathcal{L}}\mathcal{N}(0,\sigma^2(f_c)) \tag{76}$$

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

$$Var\left(\frac{1}{\sqrt{T}}\int_0^T f_c(Y_s)ds\right)\to\sigma^2(f_c). \tag{77}$$

The following relation holds:

$$\sigma^2(f_c)=-2\langle f_c,g\rangle_{u_m}=-2\langle Lg,g\rangle_{u_m}=\sigma^2\int_{\mathbb{R}}(g'(x))^2u_m(x)dx<+\infty$$

2. *In model (11), $Range(\mathcal{D})=\{h\in\mathbb{L}^2(u_m(x)dx),\int h(x)u_m(x)dx=0\}$. Therefore, (76)-(77) hold for all $f\in\mathbb{L}^2(u_m(x)dx)$.*

Proposition 7.1 requires some comments. Its first part ((76)-(77)) is classical. However, the last part, *i.e.* that (76)-(77) hold for all $f\in\mathbb{L}^2(u_m(x)dx)$, is less known and not obvious. This ensures that, for all $h\in\mathbb{L}^2(u_m(x)dx)$, such that $\int h(x)u_m(x)dx=0$, there exists $g\in\mathcal{D}$ such that $Lg=h$. In particular, this holds true for $h=f_c$ and in Theorem 3.1, for $f_c(x)=x^{2\ell}-\gamma_{2\ell}$.
This is obtained as follows. First, $Range\,\mathcal{D}=\{h\in\mathbb{L}^2(u_m(x)dx),\int h(x)u_m(x)dx=0\}$ if and only if L has a spectral gap which holds true if the process is ρ-mixing. In [25], a necessary and sufficient condition for ρ-mixing is proved for one-dimensional ergodic diffusions. In Proposition 7.1, we check that this condition holds for (11).

Using (74), equation $Lg = f_c = f - \int_{\mathbb{R}} f(y)u_m(y)dy$ can be solved. Only g' is needed for $\sigma^2(f_c)$. Using (74), as $\int_{-\infty}^{+\infty} f_c(y)u_m(y)dy = 0$, we have

$$g'_{f_c}(x) = g'(x) = 2\sigma^{-2}u_m^{-1}(x)\int_{-\infty}^{x} f_c(y)u_m(y)dy = -2\sigma^{-2}u_m^{-1}(x)\int_{x}^{+\infty} f_c(y)u_m(y)dy. \tag{78}$$

By Proposition 7.1, the integral

$$\sigma^2(f_c) = \sigma^2 \int_{\mathbb{R}} (g'(x))^2 u_m(x)dx = 4\sigma^{-2}\int_{\mathbb{R}} u_m^{-1}(x)\left(\int_{-\infty}^{x} f_c(y)u_m(y)dy\right)^2 dx \tag{79}$$

is finite for all $f \in \mathbb{L}^2(u_m(x)dx)$.
Note that the fact that (79) is finite is not obvious as $\int u_m^{-1}(x)dx = +\infty$. However, as $\int_{-\infty}^{+\infty} f_c(y)u_m(y)dy = 0$, the convergence of (79) is possible but the exact proof is not immediate.

Corollary 7.1. *Let $h_1, \ldots, h_p$ be functions belonging to Range$(\mathcal{D})$ and such that $\int h_j(x)u_m(x)dx = 0$, for $j = 1, \ldots, p$. Define*

$$V(h_i, h_j) = \sigma^2 \int_{\mathbb{R}} g'_{h_i}(x)g'_{h_j}(x)u_m(x)dx$$

so that $\sigma^2(h_i) = V(h_i, h_i)$. The vector $\frac{1}{\sqrt{T}}(\int_0^T h_i(Y_s)ds, i = 1, \ldots, p)' \to_{\mathcal{L}} \mathcal{N}_p(0, V)$ with $V = (V(h_i, h_j), 1 \le i, j \le p)$.

Proof of Proposition 7.1.

1. The result is given in Theorem 2.2 in [25].
2. We always have that Range$(\mathcal{D}) \subset \{h \in \mathbb{L}^2(u_m(x)dx), \int h(x)u_m(x)dx = 0\}$.
   This inclusion is an equality if and only if the process is ρ-mixing. Let $\gamma(x) = -2\sigma^{-1}(\Phi \star u_m)'(x)$. We can check that

$$\lim_{x \to +\infty} \gamma^{-1}(x) = 0, \quad \lim_{x \to -\infty} \gamma^{-1}(x) = 0. \tag{80}$$

   Thus, by Proposition 2.8 of the latter paper, as the limits above exist and are finite, (X_t) is ρ-mixing. The ρ-mixing property is equivalent to the fact that 0 is a simple eigenvalue and an isolated point of the spectrum of L. This implies that Range$(\mathcal{D}) = \{h \in \mathbb{L}^2(u_m(x)dx), \int h(x)u_m(x)dx = 0\}$. Therefore, (76)-(77) hold for all $f \in \mathbb{L}^2(u_m(x)dx)$.

□

Proof of Corollary 7.1. The proof follows by application of the Cramér-Wold device. □

References

[1] Amorino, C. and Gloter, A. (2020). Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scandinavian Journal of Statistics*, **47** (2), 279-246.

[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] 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.

[6] Benachour, S., Roynette, B. and Vallois, P. (1998b). Nonlinear self-stabilizing processes - II Convergence to invariant probability. *Stoch. Proc. Appl.* **75**, 203-224.

[7] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31** (5), 615-641.

[8] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2021). Semiparametric estimation of McKean-Vlasov stochastic differential equations. Preprint Arkiv.

[9] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion *Bernoulli* **1**, 17-39.

[10] Bishwal, J.P.N., (2011). Estimation in Interacting Diffusions: Continuous and Discrete Sampling. *Applied Mathematics*, **2**, 1154-1158.

[11] 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.
[12] 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.
[13] Chen, X. (2021). Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electron. Commun. Probab.* **26**, 1-13.
[14] Comte, F. and Genon-Catalot, V. (2021). Drift estimation on non compact support for diffusion models. *Stochastic Processes and their Applications* **134**, 174-207.
[15] Comte, F. and Genon-Catalot, V. (2022). Nonparametric adaptive estimation for interacting particle systems. *Preprint hal-03696877*.
[16] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**, 514-543.
[17] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *The Annals of Statistics*, **33**, 2507-2528.
[18] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probability Theory and Related Fields*, **137**, 25-47.
[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. *Preprint arXiv: 2205.05932*.
[21] Eberle, A., Guillin, A. and Zimmer, R. (2019). Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes, *Transactions of the American Mathematical Society* **371**, 7135-7173.
[22] Forien, R. and Pardoux, E. (2022). Household epidemic models and McKean-Vlasov Poisson driven SDEs. *Preprint arXiv:1907.03001, 2019*. *Annals of Applied Probability* to appear.
[23] Funaki, T. (1984). A certain class of diffusions processes associated with nonlinear parabolic equations, Z. Wahrscheinlichleitstheor. Verwandte Geb. **67**, 331-348.
[24] Gärtner, J. (1988). On the McKean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, **137**, 197-248.
[25] Genon-Catalot, V., Jeantheau, T. and. Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051-1079.
[26] 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.
[27] 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.
[28] Giesecke, K., Schwenkler, G. and Sirignano, J.A. (2020). Inference for large financial systems. *Mathematical Finance*, **30**, 3-46.
[29] Herrmann, S., Imkeller, P. and Peithmann, D. (2008). Large deviations and a Kramers'type law for self-stabilizing diffusions. *The Annals of Applied Probability*, **18**, 1379-1423.
[30] Herrmann, S. and Tugaut, J. (2010). Non uniqueness of stationary measures for self-stabilizing diffusions. *Stoch. Proc. Appl.* **120**, 1215-1246.
[31] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, 135-163.
[32] Höpfner, R. (2014). *Asymptotic Statistics with a View to Stochastic Processes.* Walter de Gruyter, Berlin/Boston.
[33] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples*. Springer.
[34] Kasonga, R.A. (1990). Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, **50**, 865-875.
[35] Kessler, M. (1997) . Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
[36] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations*. CRC press. Taylor & Francis Group. Boca Raton.
[37] Kolokoltsov, V.N. (2010). *Non linear Markov processes and kinetic equations* **182**. Cambridge University press.
[38] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes*. Springer, London.
[39] Li, Z., Lu, F., Maggioni, M. Tang, S. and Zhang, C. (2021). On the identifiability of interaction functions in systems of interacting particles. *Stochastic Processes and their Applications* **132**, 135-163.
[40] 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.
[41] 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.
[42] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of applied Probability* **13**, 540-560.
[43] Masuda, H. (2007). Ergodicity and exponential beta-mixing for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, **117**, 35-56.
[44] Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochastic Processes and their Applications*, **129**, 1013-1059.
[45] 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.
[46] Molginer, A. and Edelstein-Keshet, L. (1999). A non-local model for a swarm. *Journal of Mathematical Biology*, **38**, 534-570.
[47] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Annals of Statistics*, **48**, 1383-1408.
[48] Pavliotis, G.A. and Zanoni, A. (2022). Eigenfunction martingale estimators for interacting particle systems and their mean field limit. *Siam. J. Appl. Dyn. Systems*. to appear.
[49] Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications*, **124**, 883-914.
[50] Sharrock, L., Kantas, N., Parpas, P. and Pavliotis, G.A. (2021). Parameter Estimation for the McKean-Vlasov Stochastic Differential Equation. *Preprint Arkiv*.
[51] 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.

[52] Veretennikov, A. Yu., (2006). On ergodic measures for McKean-Vlasov stochastic equations. In *Monte Carlo and. Quasi-Monte Carlo Methods 2004*, 471-486. Springer.