



INFERENCE FOR ERGODIC MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS WITH POLYNOMIAL INTERACTIONS

Valentine Genon-Catalot, Catherine Larédo

► To cite this version:

Valentine Genon-Catalot, Catherine Larédo. INFERENCE FOR ERGODIC MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS WITH POLYNOMIAL INTERACTIONS. 2022. hal-03866218v1

HAL Id: hal-03866218

<https://hal.science/hal-03866218v1>

Preprint submitted on 24 Mar 2022 (v1), last revised 17 Feb 2023 (v3)

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.

INFERENCE FOR ERGODIC MCKEAN-VLASOV STOCHASTIC DIFFERENTIAL EQUATIONS WITH POLYNOMIAL INTERACTIONS.

V. GENON-CATALOT⁽¹⁾, C. LARÉDO²

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. March 24, 2022

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

AMS Classification. 60J60, 60J99, 62F12, 62M05

CONTENTS

1. Introduction	2
2. Probability preliminaries.	4
2.1. Assumptions	4
2.2. Stationary distributions	5
3. Parametric inference in centered stationary regime.	7
3.1. Theoretical likelihood inference.	8
3.2. Explicit estimators using empirical moments.	9
3.3. Another inference method.	11
3.4. Examples	11
4. Parametric inference in noncentered stationary regime.	14
5. Numerical considerations	15

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

⁽²⁾: MaIAGE, INRAE, Université Paris-Saclay et LPSM, Université Paris Cité, France
email: catherine.laredo@inrae.fr

5.1. Computation of the stationary distribution	15
5.2. Approximation of the stationary distribution	15
6. Concluding remarks	16
7. Proofs	17
8. Appendix	25
References	26

1. INTRODUCTION

McKean-Vlasov stochastic differential equations (SDE) are a class of processes which have aroused lots of recent contributions due to the wide field of potential applications. These SDEs with coefficients depending both on the state of the process and on its current distribution were first described by McKean (1966) to model plasma dynamics. They appear as limits of systems of interacting particles and arise in a wide variety of disciplines, where particles may represent atoms, cells, animals, neurons, people, rational agents, opinions, financial assets: see *e.g.* Benedetto *et al.* (1997) for the modeling of granular media, Baladron *et al.* (2012)) for neurosciences, Molginer and Edelstein-Keshet (1999), Carrillo *et al.* (2014) for population dynamics and ecology, Ball and Sirl (2020), Forien and Pardoux (2022) for epidemics dynamics, Giesecke *et al.* (2020) and references therein for finance.

A wide research field is devoted to develop probabilistic tools for the study of interacting particles and their limits (propagation of chaos) (see *e.g.* among many references Funaki (1984), Gärtner (1988), Méléard (1996), Benachour *et al.* (1998a, 1998b), Malrieu (2003), Cattiaux *et al.* (2008), Sznitman (1991) and Kolokoltsov (2010) for books). Herrmann *et al.* (2008) were concerned with small noise properties and large deviations results for these processes.

Therefore, the statistical inference of such models is an important issue. It started with Kasonga (1990) who studied parametric inference from the observation of systems of N interacting particles for a model with linear dependence on the parameters in the drift term. It was later extended by Bishwal (2011) and Chen (2021). Giesecke *et al.* (2020) studied inference based on the empirical distributions of the particle system. Della Maestra and Hoffmann (2021) were concerned with nonparametric inference for the drift term. Belomestny *et al.* (2021) studied the semiparametric estimation for a drift term containing both a parametric and a nonparametric part, Li *et al.* (2021) the identifiability of the interaction function in a model with only interaction between particles.

Another direction is to investigate the inference based on the limiting process of the interacting particle systems. This yields a McKean-Vlasov stochastic differential equation which describes the typical behavior of one isolated particle among others. It satisfies:

$$(1) \quad dX_t = b(t, X_t)dt + \sigma dW_t, \quad X_0 = \eta$$

where

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

$\Phi : \mathbb{R} \rightarrow \mathbb{R}$, (W_t) is a standard Brownian motion, η a random variable independent of (W_t) . The potential term V describes the geometry of the state 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 for this model has been already studied. Genon-Catalot and Larédo (2021a) 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 Genon-Catalot and Larédo (2021b), the parametric inference is studied from the continuous observation of a single path in the double asymptotic $\sigma \rightarrow 0$ and $T \rightarrow +\infty$. Sharrock *et al.* (2021) study *i.i.d.* observations of (1) and build an approximation of the likelihood to obtain offline and online estimations.

We consider here a specific family of one-dimensional McKean-Vlasov SDEs, with no potential term and with interaction term modeled by an odd increasing polynomial. We moreover restrict our study to McKean -Vlasov equations where existence, uniqueness of a solution, existence of a stationary measure and convergence towards this stationary measure are satisfied. Classical assumptions for existence and uniqueness of solutions to (1) are linear growth and Lipschitz continuity for the coefficients. Nevertheless, in order to prove results under mild growth conditions, several authors assume monotonicity and coercitivity assumptions instead (see *e.g.* Gärtner (1988), Benachour *et al.* (1998a), Malrieu (2003), Hermann *et al.* (2008), Cattiaux *et al.* (2008)). They generally assume that the interaction function Φ is increasing and odd. Moreover, considering odd increasing polynomials ensures existence and uniqueness of an invariant distribution (see *e.g.* Benachour *et al.* (1998a-b), Veretennikov (2006), Hermann and Tugaut (2010), Cattiaux *et al.* (2008), Eberle *et al.* (2019)).

In the present work, our aim is to estimate the parameters present in the interaction (or self-stabilizing) term from the continuous observation of one sample path on the time interval $[0, 2T]$ as T grows to infinity. The potential term is set to 0 and we assume that the process is in stationary regime.

Statistical inference for ergodic diffusion processes has a longstanding history. Among many references, we can quote the books by Kutoyants (2004), Iacus (2010), Kessler *et al.* (2012), Höpfner (2014). 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.* Bibby and Sørensen (1995), Kessler (1997), Hoffmann (1999), Dalalyan (2005), Comte *et al.* (2007), Comte and Genon-Catalot (2021); for multi-dimensional diffusions, *e.g.* Dalalyan and Reiss (2007), Nickl and Ray (2020). Ergodic diffusions with jumps are considered in Masuda (2007, 2019), Schmisser (2014), Amorino and Gloter (2020).

In contrast, to our knowledge, the inference for ergodic McKean-Vlasov SDEs has not been investigated. The statistical problem is very different from the case of usual stochastic differential equations. Let us consider the one-dimensional process defined by (1) with $V \equiv 0$, *i.e.*

$$(3) \quad dX_t = b(t, X_t)dt + \sigma dW_t, \quad \text{where}$$

$$(4) \quad b(t, x) = - \int_{\mathbb{R}} \Phi(x - y)u_t(dy) = -\Phi \star u_t(x), \quad u_t = \mathcal{L}(X_t),$$

$$(5) \quad \Phi(x) = \Phi(\mathbf{f}, x) = \sum_{j=0}^{k-1} f_{2j+1} x^{2j+1}, \quad f_1 > 0, \quad f_{2j+1} \geq 0, \quad j = 1, \dots, k-1,$$

(W_t) is a standard Brownian motion, σ is known and $\mathbf{f} = (f_1, \dots, f_{k-1})$ is an unknown parameter. A solution of (3) is a couple $((X_t, u_t), t \geq 0)$ composed with a process (X_t) and a family of distribution (u_t) satisfying (3)-(4). When defined, (X_t) is a time-*inhomogeneous* Markov process. Under appropriate conditions, the process (X_t) admits a stationary distribution. We assume here that (X_t) is in stationary regime.

As Φ is odd, whatever the initial distribution, the process (X_t) solving (3) has a constant expectation m (see Section 2). Contrary to classical stochastic differential equations, stationary distributions of model (3) are not uniquely determined except if the expectation of (X_t) is specified.

This is due to the fact that (X_t) and $(X_t - \mathbb{E}X_t)$ follow the same equation (3). If $\mathbb{E}(X_t) = 0$, (3) admits a unique invariant distribution with symmetric density $u(\mathbf{f}, x)$. If $\mathbb{E}(X_t) = m$, (3) admits a unique invariant distribution with density $u(\mathbf{f}, x - m)$. Therefore, we first study the estimation of \mathbf{f} when (X_t) is in centered stationary regime ($X_0 \sim u(\mathbf{f}, x)dx$). Then, we study the joint estimation of (m, \mathbf{f}) when the process is in non centered stationary regime ($X_0 \sim u(\mathbf{f}, x - m)dx$). When (X_t) is in stationary regime, $\mathcal{L}(X_t)$ does no longer depend on t and is equal to the stationary distribution $u(\mathbf{f}, x)dx$ in centered stationary regime and to $u(\mathbf{f}, x - m)dx$ in stationary regime with expectation m . Because of the specific form of the interaction function Φ (polynomial), the convolution product $\Phi \star u(\mathbf{f}, \cdot)$ (resp. $\Phi \star u(\mathbf{f}, \cdot - m)$) is explicitly given as a function of \mathbf{f} and the moments of the invariant distribution. 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.

This is why we first 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 (3), 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 1). When the observed process is in stationary regime, it is identical to a classical ergodic diffusion process (Proposition 2). In Section 3, we estimate \mathbf{f} when the observed process is in centered stationary regime. We first study the exact likelihood and prove that the maximum likelihood estimator is consistent and asymptotically Gaussian with rate \sqrt{T} (Proposition 4). However, this remains completely theoretical and the estimators are numerically intractable. Next, we study computable estimators of \mathbf{f} for the centered process. Two kinds of estimators are proposed. First, estimators are built using an approximation of the likelihood on $[T, 2T]$ (Theorem 1). Second, we also build empirical estimators based on some specific properties of model (3). Examples are given. Next, we study the inference of (m, \mathbf{f}) when the process is no longer centered (Theorem 2). In Section 5, we recall some known results to compute or approximate the invariant distribution. Section 6 contains concluding remarks. Proofs are gathered in Section 7. In the Appendix (Section 8), a central limit theorem for ergodic diffusions is recalled.

2. PROBABILITY PRELIMINARIES.

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) when the initial variable follows the stationary distribution.

2.1. Assumptions. 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 $\forall x, y \in \mathbb{R}$, $|\Phi(x) - \Phi(y)| \leq c|x - y|(1 + |x|^\ell + |y|^\ell)$.
- [H3] Φ is C^1 and strictly convex on \mathbb{R}^+ : there exists a constant $\lambda > 0$ such that $\forall x \geq 0$, $\Phi'(x) \geq \lambda$.
- [H4] Φ' has ℓ polynomial growth: $\exists C > 0$, $\ell \in \mathbb{N}$, $\forall x \in \mathbb{R}$, $|\Phi'(x)| \leq C(1 + |x|^\ell)$.

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

$$\int |\Phi(x - y)|u_t(dy) \leq c_1(1 + 2^\ell|x|^{\ell+1} + 2^\ell \int |y|^{\ell+1}u_t(dy)).$$

So $\Phi \star u_t$ is well defined as soon as u_t has a $\ell + 1$ -th order moment.

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 $\sup_{t \geq 0} \mathbb{E}X_t^{2n} < +\infty$ (see Theorem 3.1 and Proposition 3.10 in Benachour *et al.*, 1998a, Theorem 2.13 in Hermann *et al.*, 2008).

Note that, by (3)-(4),

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

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

$$(6) \quad \mathbb{E}(X_t) = \mathbb{E}(X_0).$$

This holds whatever the initial variable.

Note also that $X_t - \mathbb{E}(X_0)$ is also solution of (3). Indeed,

$$\begin{aligned} X_t - \mathbb{E}(X_0) &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - (y - \mathbb{E}(X_0)))u_s(y)dy \\ &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - z)u_s(z + \mathbb{E}(X_0))dz \\ &= X_0 - \mathbb{E}(X_0) - \int_0^t ds \int \Phi(X_s - \mathbb{E}(X_0) - z)v_s(z)dz \end{aligned}$$

where v_s is the distribution of $X_s - \mathbb{E}(X_0)$.

This is why the specification of the expectation of the process is important especially for invariant distributions (see below).

Finally, let us state another useful property associated with this equation.

Lemma 1. *Assume [H1]-[H3]. Consider an even 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 $x \geq 0$, $\lambda x \leq \Phi(x) \leq \Phi \star u(x) \leq c_\ell x[1 + x^\ell + \int_0^\infty y^\ell u(y)dy]$ for some constant c_ℓ .

2.2. Stationary distributions.

2.2.1. Existence and uniqueness. By Lemma 2.2 in Hermann and Tugaut (2010), if there exists an invariant density whose $(8(\ell + 1)^2)$ -moment is finite, then it satisfies the implicit fixed point equation

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

where, by Lemma 1, $\nu(u)$ below is well defined and finite,

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

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,

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

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

$$X_t = X_0 - \int_0^t (X_s - \mathbb{E}X_s)ds + \sigma W_t = X_0 - \int_0^t (X_s - m)ds + \sigma W_t$$

where $m = \mathbb{E}X_0 = \mathbb{E}X_t$ for all t .

For this reason, many authors (Cattiaux *et al.*, 2008, Benachour *et al.*, 1998a, Malrieu, 2003) consider equation (3) assuming that $\mathbb{E}X_t = 0$ and prove the following result.

Proposition 1. (see Cattiaux *et al.*, 2008, Benachour *et al.*, 1998a or Malrieu 2003).

(i) Under [H1]-[H4], there exists a unique even density function $u(x)$ implicitly defined by

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

which satisfies (see [H3] for λ and Lemma 1):

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

(ii) Moreover, if $u(\cdot)$ is the density of X_0 and (X_t) is the unique solution of (3), then $u(\cdot)$ 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.

(iv) The convergence of $\mathcal{L}(X_t)$ to the stationary density u is exponential with respect to the Wasserstein distance \mathcal{W}_2 : $\mathcal{W}_2(\mathcal{L}(X_t), u) \leq Ce^{-\lambda t} \mathcal{W}_2(\mathcal{L}(X_0), u)$ (λ is defined in [H3]).

Consequently, the general equation (3) admits a unique invariant density $u_m(x)$ such that $\int yu_m(y)dy = m$. 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 $u_t(dy)$ of X_t defined in (3) satisfies

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

so that the following holds.

Proposition 2. Consider the stochastic differential equation

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

where $u_m(\cdot) = u(\cdot - m)$ and $u_0 = u$ 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 stochastic differential equation in stationary regime and is ergodic. Therefore, 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

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

The central limit theorem associated with this result is stated and detailed in the Appendix.

3. PARAMETRIC INFERENCE IN CENTERED STATIONARY REGIME.

From now on, we consider (X_t) defined by (3),(4),(5). Therefore, assumptions [H1]-[H4] are satisfied and equation (3) admits an invariant distribution which is unique when its expectation is specified. In this section, we assume that (X_t) is in centered stationary regime and consider the estimation of $\mathbf{f}' = (f_1, f_3, \dots, f_{2k-1})$, with $f_1 > 0, f_{2j+1} \geq 0, j = \dots, k-1$, which is an unknown parameter of $(0, +\infty) \times [0, +\infty)^{k-1}$, based on the continuous observation of $(X_t, t \leq 2T)$. According to Proposition 1, Equation (9) has a unique symmetric density solution $u(\mathbf{f}, \cdot)$ which depends on parameter \mathbf{f} . Note that $u(\mathbf{f}, \cdot)$ depends on \mathbf{f} and σ . As σ is known, in what follows, we omit the dependence w.r.t. σ in the notations. Therefore, we make here the assumptions:

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

Define, for $j \geq 0$,

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

Proposition 3. *Under [H5]-[H6], the drift $b(t, x) = b(x) = -\Phi(\mathbf{f}, \cdot) \star u(\mathbf{f}, x)$ is an odd polynomial such that*

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

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

$$(14) \quad \mathbf{b} = (b_{2i+1}, i = 0, \dots, k-1)' = M_k(\mathbf{f})\mathbf{f},$$

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

$$(15) \quad M_k(i, j) = 0 \text{ for } i > j, M_k(i, j) = - \binom{2j+1}{2(j-i)} \gamma_{2(j-i)}(\mathbf{f}) \text{ for } i \leq j.$$

Note that $\gamma_0(\mathbf{f}) = \gamma_0 = 1$ so that $M_k(i, i) = -1$.

The coefficients of $b(\mathbf{f}, x)$ are explicit functions of \mathbf{f} and of the moments of $u(\mathbf{f}, \cdot)$. Note that the matrix $M_k(\mathbf{f})$ depends on \mathbf{f} only through the moments $(\gamma_0, \gamma_2(\mathbf{f}), \dots, \gamma_{2(k-1)}(\mathbf{f}))$.

We define for $\mathbf{v} = (v_0, v_1, \dots, v_{k-1})'$ a vector of \mathbb{R}^k , $M_k^{(\mathbf{v})} = (M_k^{(\mathbf{v})}(i, j))_{0 \leq i, j \leq k-1}$ with

$$(16) \quad M_k^{(\mathbf{v})}(i, j) = 0 \text{ for } i > j, M_k^{(\mathbf{v})}(i, j) = - \binom{2j+1}{2(j-i)} v_{j-i} \text{ for } i \leq j.$$

Note that $M_k(\mathbf{f})$ defined in (15) satisfies $M_k(\mathbf{f}) = M_k^{(\gamma)}$ where $\gamma = (\gamma_0, \gamma_2(\mathbf{f}), \dots, \gamma_{2(k-1)}(\mathbf{f}))$.

Examples:. For $k = 1$, $b(\mathbf{f}, x) = -f_1x$, $M_1 = -[1]$.

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

For $k = 3$, $b(\mathbf{f}, x) = -[(f_1 + 3\gamma_2 f_3 + 5\gamma_4 f_5)x + (f_3 + 10\gamma_2 f_5)x^3 + x^5]$,

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

Proof of Proposition 3. First, since $u(\mathbf{f}, \cdot)$ is symmetric, odd moments of $u(\mathbf{f}, \cdot)$ are nul. Therefore,

$$\begin{aligned} \Phi(\mathbf{f}, \cdot) \star u(\mathbf{f}, x) &= \sum_{j=0}^{k-1} f_{2j+1} \int (x-y)^{2j+1} u(\mathbf{f}, y) dy = \sum_{j=0}^{k-1} f_{2j+1} \sum_{\ell=0}^j \binom{2j+1}{2\ell} x^{2j+1-2\ell} \gamma_{2\ell}(\mathbf{f}) \\ &= \sum_{j=0}^{k-1} f_{2j+1} \sum_{i=0}^j \binom{2j+1}{2(j-i)} x^{2i+1} \gamma_{2(j-i)}(\mathbf{f}) = \sum_{i=0}^{k-1} x^{2i+1} \sum_{j=i}^{k-1} \binom{2j+1}{2(j-i)} \gamma_{2(j-i)}(\mathbf{f}) f_{2j+1}. \end{aligned}$$

□

3.1. Theoretical likelihood inference. 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:

$$(17) \quad \ell_T(\mathbf{f}) = \sigma^{-2} \left[\int_0^T b(\mathbf{f}, X_s) dX_s - \frac{1}{2} \int_0^T b^2(\mathbf{f}, X_s) ds \right].$$

Define the estimator

$$(18) \quad \widehat{\mathbf{f}}_T = \arg \max_{\mathbf{f} \in \mathbb{R}^k} \ell_T(\mathbf{f}).$$

This estimator is purely theoretical as it is not given by explicit equations due to the presence of the moments of $u(\mathbf{f}, \cdot)$ in the drift $b(\mathbf{f}, x)$ (see Proposition 3).

Proposition 4. Assume [H5]-[H6]. The following holds:

$$\frac{1}{T}(\ell_T(\mathbf{f}) - \ell_T(\mathbf{f}_0)) \rightarrow -\frac{1}{2\sigma^2} \int (b(\mathbf{f}, x) - b(\mathbf{f}_0, x))^2 u(\mathbf{f}_0, y) dy := -\frac{1}{2\sigma^2} K(\mathbf{f}_0, \mathbf{f}),$$

Assume moreover that the parameter set for \mathbf{f} is a compact subset of $(0, +\infty) \times [0, +\infty)^{k-1}$ and that $b(\mathbf{f}, x) = b(\mathbf{f}_0, x)$ for all x implies $\mathbf{f} = \mathbf{f}_0$. Then, the maximum likelihood estimator $\widehat{\mathbf{f}}_T$ is consistent.

The matrix $I_{\mathbf{f}} = (\int_{\mathbb{R}} [\frac{\partial b}{\partial f_i}(\mathbf{f}, x)] \frac{\partial b}{\partial f_{i'}}(\mathbf{f}, x) u(\mathbf{f}, x) dx)_{0 \leq i, i' \leq k}$ is invertible and

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

According to Proposition 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 $\widehat{\mathbf{f}}_T$ is numerically intractable (see e.g. Example 2).

3.2. 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 $u(\mathbf{f}, \cdot)$, 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} .

Using Proposition 3, let us consider the contrast function which is the log-likelihood given X_T of the process $(X_t, t \in [T, 2T])$ with polynomial drift $\sum_{i=0}^{k-1} b_{2i+1} x^{2i+1}$ and diffusion term σ :

$$(19) \quad U_T(\mathbf{b}) = \frac{1}{\sigma^2} \left(\int_T^{2T} \sum_{i=0}^{k-1} b_{2i+1} X_s^{2i+1} dX_s - \frac{1}{2} \int_T^{2T} \left[\sum_{i=0}^{k-1} b_{2i+1} X_s^{2i+1} \right]^2 ds \right).$$

The associated estimator is obtained maximizing U_T with respect to $\mathbf{b} = (b_{2i+1}, i = 0, \dots, k-1)'$. For this, define

$$(20) \quad z_k(x) = \begin{pmatrix} x \\ x^3 \\ \vdots \\ x^{2k-1} \end{pmatrix}, \quad Z_T = \int_T^{2T} z_k(X_s) dX_s.$$

Then, the estimator $\hat{\mathbf{b}}_T$ is defined by the equation

$$(21) \quad Z_T = \Psi_T \hat{\mathbf{b}}_T,$$

where

$$(22) \quad \Psi_T = \left(\int_T^{2T} z_k(X_s) [z_k(X_s)]' ds = \left(\int_T^{2T} X_s^{2i+2j+2} ds \right)_{0 \leq i, j \leq k-1} \right).$$

Let us define, using (13),

$$(23) \quad \Psi(\mathbf{f}) = (\gamma_{2(i+j+1)}(\mathbf{f}))_{0 \leq i, j \leq k-1}.$$

Proposition 5. Assume [H5]-[H6]. The matrix Ψ_T/T converges a.s. to $\Psi(\mathbf{f})$. The matrix $\Psi(\mathbf{f})$ is invertible, $\hat{\mathbf{b}}_T$ converges a.s. to \mathbf{b} and $\sqrt{T}(\hat{\mathbf{b}}_T - \mathbf{b})$ converges in distribution to the Gaussian law $\mathcal{N}_k(0, \sigma^2 \Psi(\mathbf{f})^{-1})$.

In a second step, we consider the consistent estimators of the moments of $u(\mathbf{f}, \cdot)$ built using the sample path $(X_t, t \in [0, T])$, defined by

$$(24) \quad \hat{\gamma}_{2j}(T) = \frac{1}{T} \int_0^T X_s^{2j} ds, j \geq 1 \quad (\text{we set } \hat{\gamma}_0(T) = \gamma_0 = 1).$$

By the ergodic theorem, the estimator $\hat{\gamma}_{2j}(T)$ converges almost surely to the moment $\gamma_{2j}(\mathbf{f})$ of $u(\mathbf{f}, \cdot)$. Moreover, let us define, $g'_0(x) = 0$ and, for $1 \leq \ell \leq k-1$,

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

Then, according to Genon-Catalot *et al.* (2000) (recalled in Section 8), the empirical moments of an ergodic diffusion (X_t) with invariant distribution $u(\mathbf{f}, \cdot)$ satisfy,

$$(26) \quad \sqrt{T}(\hat{\gamma}_{2\ell}(T) - \gamma_{2\ell}(\mathbf{f})) = \frac{\sigma}{\sqrt{T}} \int_0^T g'_\ell(\mathbf{f}, X_s) dW_s + o_P(1).$$

Hence, setting $V(\mathbf{f}) = (V_{l,m}(\mathbf{f}))_{0 \leq l, m \leq k-1}$ with $V_{l,m}(\mathbf{f}) = \int g'_l(\mathbf{f}, x) g'_m(\mathbf{f}, x) u(\mathbf{f}, x) dx$,

$$(27) \quad \sqrt{T} \begin{pmatrix} \hat{\gamma}_0(T) - \gamma_0 \\ \hat{\gamma}_2(T) - \gamma_2(\mathbf{f}) \\ \dots \\ \hat{\gamma}_{2(k-1)}(T) - \gamma_{2(k-1)}(\mathbf{f}) \end{pmatrix} \rightarrow \mathcal{N}_k(0, \sigma^2 V(\mathbf{f})).$$

Let us come back to the estimation of \mathbf{f} . Relation (14), $\mathbf{b} = M_k(\mathbf{f})\mathbf{f}$, of Proposition 3 suggests to consider the matrix \widehat{M}_k using (16) where the unknown moments of $u(\mathbf{f}, \cdot)$ are replaced by their consistent estimators built on the observation of (X_t) on $[0, T]$ given above:

$$(28) \quad \widehat{M}_k = M_k^{(\hat{\gamma}_0(T), \hat{\gamma}_2(T), \dots, \hat{\gamma}_{2(k-1)}(T))} = (\widehat{M}_k(i, j))_{0 \leq i, j \leq k-1}, \quad \text{with}$$

$$(29) \quad \widehat{M}_k(i, j) = - \binom{2j+1}{2(j-i)} \hat{\gamma}_{2(j-i)}(T) \mathbf{1}_{i \leq j}.$$

It follows from (26) that \widehat{M}_k converges *a.s.* to $M_k(\mathbf{f})$.

This justifies the definition of $\widehat{\mathbf{f}}_T$ by (see (21)):

$$(30) \quad \widehat{\mathbf{f}}_T = \widehat{M}_k^{-1} \widehat{\mathbf{b}}_T = \widehat{M}_k^{-1} \Psi_T^{-1} Z_T = (\Psi_T \widehat{M}_k)^{-1} Z_T.$$

Let us stress that as for the theoretical maximum likelihood estimator (18), this new estimator does not depend on σ .

Theorem 1. *Under the assumptions [H5]-[H6], the estimator $\widehat{\mathbf{f}}_T$ is consistent and satisfies*

$$(31) \quad \sqrt{T}(\widehat{\mathbf{f}}_T - \mathbf{f}) \rightarrow_{\mathcal{L}} \mathcal{N}_k(0, \sigma^2 \Sigma(\mathbf{f})) \quad \text{with } \Sigma(\mathbf{f}) = \Sigma_1(\mathbf{f}) + \Sigma_2(\mathbf{f})$$

$$(32) \quad \Sigma_1(\mathbf{f}) = M_k^{-1}(\mathbf{f}) \Psi^{-1}(\mathbf{f}) (M_k^{-1}(\mathbf{f}))'; \quad \Sigma_2(\mathbf{f}) = \int \beta(\mathbf{f}, x) \beta'(\mathbf{f}, x) u(\mathbf{f}, x) dx$$

with, using definitions (15), (16), and (23),

$$(33) \quad \beta(\mathbf{f}, x) = M_k(\mathbf{f})^{-1} M_k^{(\mathbf{v}(x))} \mathbf{f},$$

with $\mathbf{v}(x) = (0, g'_1(\mathbf{f}, x), \dots, g'_{k-1}(\mathbf{f}, x))$.

By Proposition 5, $\widehat{\mathbf{b}}_T$ converges *a.s.* to \mathbf{b} and, by (26), \widehat{M}_k to M_k 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 $\mathbf{b} \rightarrow \mathbf{f}$ and the second one $\Sigma_2(\mathbf{f})$ from the estimation of the moments of $u(\mathbf{f}, \cdot)$ and the plug-in device in the estimation. The proof, detailed in the appendix, relies on the decomposition in the two main terms

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

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

3.3. 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

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

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:

$$(34) \quad B(\mathbf{f}) := (\sigma^2(2\ell-1) \gamma_{2\ell-2}(\mathbf{f}), \ell = 1, \dots, k)', \text{ and}$$

$$(35) \quad \Gamma(\mathbf{f}) = (\Gamma_{\ell j}(\mathbf{f}))_{1 \leq \ell \leq k, 0 \leq j \leq k-1} \text{ with } \Gamma_{\ell, j}(\mathbf{f}) = 2 \sum_{m=0}^j \binom{2j+1}{2m} \gamma_{2m}(\mathbf{f}) \gamma_{2(j+\ell-m)}(\mathbf{f})$$

Then, $B(\mathbf{f}) = \Gamma(\mathbf{f})\mathbf{f}$. The matrix $\Gamma(\mathbf{f})$ is necessarily invertible.

Substituting in (34) each moment by its empirical estimator (24) yields the two estimators $\tilde{B}_T, \tilde{\Gamma}_T$ and the relation defining the moment estimator of \mathbf{f} :

$$(36) \quad \tilde{\mathbf{f}}_T = (\tilde{\Gamma}_T)^{-1} \tilde{B}_T$$

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

Proposition 6. *Assume [H5]-[H6]. The estimator defined by (36) is consistent and such that $\sqrt{T}(\tilde{\mathbf{f}}_T - \mathbf{f})$ 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 (60)).*

3.4. 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 $u(f, x)dx = \mathcal{N}(0, \sigma^2/2f)$. Equation (3) writes $dX_t = -f \int (X_t - y)u(f, y)dydt + \sigma dW_t = -fX_tdt + \sigma dW_t$. The estimator \hat{f}_T is equal to the maximum likelihood estimator:

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

As $T^{-1} \int_0^T X_s^2 ds$ converges *a.s.* to $\sigma^2/2f$, we obtain the classical result that $\sqrt{T}(\hat{f}_T - f)$ converges in distribution to $\mathcal{N}(0, 2f)$. With the notations of Theorem 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:

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

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 $V = \int (x/f)^2 u(f, x) dx = \sigma^4/2f^3$. This yields that $\sqrt{T}(\tilde{f}_T - f)$ converges in distribution to $\mathcal{N}(0, 2f)$. In this special example, \hat{f}_T and \tilde{f}_T have the same asymptotic distribution. Note that \hat{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 distribution $u(f, \cdot)$ is unique and defined by the implicit equation (9). As $u(f, \cdot)$ is symmetric, $\int (x - y)^3 u(f, y) dy = x^3 + 3x\gamma_2(f)$. Therefore, equation (3) starting with $X_0 \sim u(f, x)dx$, writes:

$$(37) \quad dX_t = -f(X_t^3 + 3X_t \gamma_2(f))dt + \sigma dW_t, \quad X_0 \sim u(f, x)dx.$$

where

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

and $\gamma_2(f)$ is implicitly given as the unique solution (see Benachour *et al.* (1998a)) of

$$(38) \quad \int_{\mathbb{R}} x^2 \exp \left[-\sigma^{-2} f \left(\frac{x^4}{2} + 3x^2 \gamma_2(f) \right) \right] dx = \gamma_2(f) \int_{\mathbb{R}} \exp \left[-\sigma^{-2} f \left(\frac{x^4}{2} + 3x^2 \gamma_2(f) \right) \right] dx.$$

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

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

Differentiating this identity w.r.t. f , equation (38) 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 $\hat{\mathbf{f}}_T$. 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 1 since $f_1 = 0$. But we can compute explicitly the estimator of f and get using (24)

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

Define

$$(39) \quad a(f) = \gamma_6(f) + 9\gamma_2^3(f) + 6\gamma_2(f)\gamma_4(f); \quad c(f) = 3(\gamma_4(f) + 3\gamma_2^2(f)).$$

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

$$\begin{aligned} \sqrt{T}(\hat{f}_T - f) &= - \frac{\sigma}{D_T/T} T^{-1/2} \int_T^{2T} (X_t^3 + 3X_t \hat{\gamma}_2(T)) dW_t \\ &\quad + f \sqrt{T}(\gamma_2(f) - \hat{\gamma}_2(T)) \frac{\int_T^{2T} 3X_t(X_t^3 + 3X_t \hat{\gamma}_2(T)) dt/T}{D_T/T} \end{aligned}$$

We have that $\frac{1}{T} \int_T^{2T} 3X_t(X_t^3 + 3X_t\hat{\gamma}_2(T))dt \rightarrow c(f)$ and,
 $\frac{1}{\sqrt{T}} \int_T^{2T} (X_t^3 + 3X_t\hat{\gamma}_2(T))dW_t = \frac{1}{\sqrt{T}} \int_T^{2T} (X_t^3 + 3X_t\gamma_2(f))dW_t + o_P(1)$. Thus,

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

Now, using (25) and (26)

$$\begin{aligned}\sqrt{T}(\hat{f}_T - f) &= \frac{\sigma\sqrt{2}}{\sqrt{2T}} \int_0^{2T} \left(\frac{1}{a(f)} (X_t^3 + 3X_t\gamma_2(f))1_{[T,2T]}(s) + f \frac{c(f)}{a(f)} g'_1(X_s)1_{[0,T]}(s) \right) dW_s \\ &\quad + o_P(1).\end{aligned}$$

Using the notations of Theorem 1, $\Sigma_1(f) = \frac{1}{a(f)}$, $\Sigma_2(f) = f^2 \frac{c^2(f)}{a^2(f)} \int (g'_1(x))^2 u(f, x) dx$,

$$\sqrt{T}(\hat{f}_T - f) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \sigma^2(\Sigma_1(f) + \Sigma_2(f))).$$

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

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

We obtain the limiting distribution of $\sqrt{T}(\tilde{f}_T - f)$ by the delta method using that $\sqrt{T}(\hat{\gamma}_2(T) - \gamma_2(f), \hat{\gamma}_4(T) - \gamma_4(f))$ is asymptotically Gaussian with covariance given in (27).

• **Example 3:** $\Phi(\mathbf{f}, x) = f_1x + f_3x^3$, $f_1 > 0$, $f_3 \geq 0$ ($k = 2$).

We have that $b(\mathbf{f}, x) = -(f_1x + f_3(x^3 + 3x\gamma_2(\mathbf{f}))) = -((f_1 + 3\gamma_2(\mathbf{f})f_3)x + f_3x^3)$, and

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

This yields

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

According to Theorem 1, the asymptotic variance of $\sqrt{T}(\hat{f}_T - f)$ is $\sigma^2(\Sigma_1(\mathbf{f}) + \Sigma_2(\mathbf{f}))$ where $\Sigma_1(\mathbf{f}) = M_2^{-1}(\mathbf{f})\Psi^{-1}(\mathbf{f})(M_2^{-1}(\mathbf{f}))'$ and $\Sigma_2(\mathbf{f}) = \int \beta(\mathbf{f}, x)\beta'(\mathbf{f}, x)u(\mathbf{f}, x)dx$, 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 $\mathbf{v}(x) = (0, g'_1(x))$, $M_2^{(\mathbf{v}(x))} = - \begin{pmatrix} 0 & 3g'_1(x) \\ 0 & 0 \end{pmatrix}$, so that

$$\beta(\mathbf{f}, x) = \begin{pmatrix} 3f_3g'_1(x) \\ 0 \end{pmatrix}, \text{ and } \Sigma_2(\mathbf{f}) = 9f_3^2 \begin{pmatrix} \int (g'_1(x))^2 u(\mathbf{f}, x) dx & 0 \\ 0 & 0 \end{pmatrix}.$$

4. PARAMETRIC INFERENCE IN NONCENTERED STATIONARY REGIME.

In this section, we assume that the process (3) is in noncentered stationary regime. Thus, we assume

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

i.e. we observe the process such that:

$$(40) \quad dX_t = - \int \Phi \star \mathcal{L}(X_t)(X_t - y)dy \, dt + \sigma dW_t, \quad X_0 = \eta \sim u_m(x)dx,$$

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

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

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

$$(41) \quad b(\mathbf{f}, m, x) = - \int \Phi(x - m - (y - m))u_m(y)dy = -\Phi \star u(x - m) = b(\mathbf{f}, x - m).$$

As m is unknown, we set $\hat{m} = T^{-1} \int_0^T X_s ds$ and consider the contrast

$$(42) \quad \Lambda_T(\mathbf{f}) = \int_T^{2T} b(\mathbf{f}, X_s - \hat{m})dX_s - \frac{1}{2} \int_T^{2T} b^2(\mathbf{f}, X_s - \hat{m})ds.$$

As previously, we proceed in two steps. First we define $\tilde{\mathbf{b}}_T$ by

$$(43) \quad \tilde{\Psi}_T \tilde{\mathbf{b}}_T = \tilde{Z}_T$$

with (see (30))

$$(44) \quad \tilde{Z}_T = \int_T^{2T} z(X_s - \hat{m})dX_s, \quad \tilde{\Psi}_T = \left(\int_T^{2T} (X_s - \hat{m})^{2i+2j+2} ds \right)_{0 \leq i, j \leq k-1}.$$

Now, we set $\tilde{\gamma}_{2\ell}(T) = T^{-1} \int_0^T (X_s - \hat{m})^{2\ell} ds$, $\ell \geq 0$ and define the estimator $\tilde{\mathbf{f}}_T$ by

$$(45) \quad \tilde{\mathbf{b}}_T = \tilde{M}_k \tilde{\mathbf{f}}_T, \quad \tilde{M}_k = M_k^{(\tilde{\gamma}_{2\ell}(T), \ell=0, \dots, k-1)}.$$

Proposition 7. Assume [H5] and [H7]. As T tends to infinity, $\hat{m} \rightarrow_{a.s.} m$ and for $\ell \geq 1$, $\tilde{\gamma}_{2\ell}(T) \rightarrow_{a.s.} \gamma_{2\ell}(\mathbf{f}) = \int x^{2\ell} u(\mathbf{f}, x)dx$. Moreover,

$$(46) \quad \sqrt{T}(\hat{m} - m) = \frac{\sigma}{\sqrt{T}} \int_0^T h'_0(X_s - m)dW_s + o_P(1),$$

$$(47) \quad \sqrt{T}(\tilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}(\mathbf{f})) = \frac{\sigma}{\sqrt{T}} \int_0^T h'_\ell(X_s - m)dW_s + o_P(1),$$

where $h'_0(x) = -\frac{2}{\sigma^2 u(\mathbf{f}, x)} \int_{-\infty}^x v u(\mathbf{f}, v)dv$ and $h'_\ell(x) = g'_\ell(x) = -\frac{2}{\sigma^2 u(\mathbf{f}, x)} \int_{-\infty}^x (v^{2\ell} - \gamma_{2\ell}(\mathbf{f}))u(\mathbf{f}, v)dv$ (see (25)).

Consequently, for all k , the vector $(\sqrt{T}(\hat{m} - m), \sqrt{T}(\tilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}(\mathbf{f})), \ell = 1, \dots, k)$ converges in distribution to $\mathcal{N}_{k+1}(\mathbf{0}, \sigma^2(\int h'_i(v)h'_j(v)u(\mathbf{f}, v)dv))_{0 \leq i, j \leq k}$

Theorem 2. Assume [H5] and [H7].

- The estimator $\tilde{\mathbf{f}}_T$ is consistent and $\sqrt{T}(\tilde{\mathbf{f}}_T - \mathbf{f})$ converges in distribution to $\mathcal{N}_k(\mathbf{0}, \sigma^2 \Sigma(\mathbf{f}))$ where $\Sigma(\mathbf{f})$ is defined in (31).

- The joint asymptotic distribution of $(\hat{m}, \tilde{\mathbf{f}}_T)$ is as follows

$$(\sqrt{T}(\hat{m} - m), \sqrt{T}(\tilde{\mathbf{f}}_T - \mathbf{f})) \rightarrow_{\mathcal{L}} \mathcal{N}_{1+k} \left(\mathbf{0}, \sigma^2 \begin{pmatrix} \int h'_0(x)^2 u(x)dx & \int h'_0(x)[\beta(x)]'u(x)dx \\ \int h'_0(x)\beta(x)u(x)dx & \Sigma(\mathbf{f}) \end{pmatrix} \right),$$

where h'_0 is defined in Proposition 7, $\beta = \beta(\mathbf{f}, \cdot)$ is defined in (33) and $\Sigma(\mathbf{f})$ is defined in (31).

Note that $\tilde{\mathbf{f}}_T$ and $\hat{\mathbf{f}}_T$ 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 $u_m(x)dx = \mathcal{N}(m, \sigma^2/2f)$. The MLE based on $(X_t, t \in [0, T])$ can be computed in this model:

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

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

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

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. NUMERICAL CONSIDERATIONS

5.1. Computation of the stationary distribution. The symmetric stationary distribution u of (3) can be computed numerically (see *e.g.* Benachour *et al.*, 1998a). Indeed,

$$u(x) = \frac{1}{\nu(u)} \exp\left[-\frac{2}{\sigma^2} F(x)\right] dx, \quad \nu(u) = \int \exp\left[-\frac{2}{\sigma^2} F(x)\right] dx$$

$$F(x) = \sum_{\ell=0}^{k-1} \frac{x^{2\ell+2}}{2\ell+2} \sum_{i=\ell}^{k-1} f_{2i+1} \binom{2i+1}{2(i-\ell)} \gamma_{2(i-\ell)} = F_\gamma(x), \quad \gamma_{2\ell} = \int x^{2\ell} u(x) dx,$$

and $\gamma = (\gamma_{2j}, j = 1, \dots, k-1)'$. For each \mathbf{f} , the vector γ is the unique solution of the system:

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

and can therefore be numerically computed. Once F_γ is obtained, u may also be numerically obtained.

5.2. Approximation of the stationary distribution. Here is another way of approximating the stationary distribution u (see Malrieu (2003)). Consider the system of N interacting particles given by:

$$(48) \quad d\xi_t^{i,N} = -\frac{1}{N} \sum_{j=1}^N \Phi(\xi_t^{i,N} - \xi_t^{j,N}) dt + \sigma dB_t^i, \quad \xi_0^{i,N} = \xi_0^i, \quad i = 1, \dots, N,$$

where $(B^i, i = 1, \dots, N)$ are independent Brownian motions, $(\xi_0^i, i = 1, \dots, N)$ are *i.i.d.* random variables independent of $(B^i, i = 1, \dots, N)$ and Φ satisfies [H1]-[H4]. Set

$$X_t^{i,N} = \xi_t^{i,N} - \frac{1}{N} \sum_{j=1}^N \xi_t^{j,N}, \quad X_t^i = X_0^i - \int_0^t \int \Phi(X_s^i - y) u_s(dy) dt + \sigma B_t^i, \quad X_0^i = \xi_0^i - \mathbb{E} \xi_0^i$$

where u_s is the distribution of X_s^i . Then, for any fixed p , the vector of process $((X_t^{i,N}), i = 1, \dots, p)$ converges to the vector of process $((X_t^i), i = 1, \dots, p)$ which are *i.i.d.* with distribution

equal to the distribution of (3).

Moreover, the N -dimensional process $(X_t^{i,N}, i = 1 \dots, N)$ admits an invariant distribution ν_N whose first marginal ν_N^1 converges to the invariant distribution u of (X_t^1) (in the \mathcal{W}_2 Wasserstein distance).

The distribution ν_N is given by the density:

$$\nu_N(x_1, \dots, x_N) = \frac{1}{Z_N} \mathbf{1}_{\mathcal{M}}(x_1, \dots, x_N) \exp \left[-\frac{1}{N\sigma^2} \sum_{i,j=1}^N W(x_i - x_j) \right], \quad W(x) = \int_0^x \Phi(y) dy,$$

where $\mathcal{M} = \{(x_1, \dots, x_N), \sum_{i=1}^N x_i = 0\}$ and Z_N is the norming constant. Thus, for large N , ν_N^1 approximates $u(x)dx$.

6. 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

$$(49) \quad dX_t = -\Phi(\mathbf{f}, \cdot) \star \mathcal{L}(X_t)(X_t) dt + \sigma dW_t$$

with $\Phi(\mathbf{f}, x) = \sum_{j=0}^{k-1} f_{2j+1} x^{2j+1}$, $f_1 > 0, f_{2j+1} \geq 0, j = 1, \dots, k-1$, 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 $2k-1$. Under appropriate conditions, the process (X_t) admits a stationary distribution. We assume here that (X_t) is in stationary regime. The statistical problem is completely different from the case of usual stochastic differential equations. It follows from equation (49) that, whatever the initial distribution, the process (X_t) has a constant expectation m . Contrary to classical stochastic differential equations, stationary distributions of model (49) are uniquely determined only if the expectation of (X_t) is specified. Equation (49) is not precise enough for this. If $\mathbb{E}(X_t) = 0$, (49) admits a unique invariant distribution with symmetric density $u(\mathbf{f}, x)$ recalled in Proposition 1. If $\mathbb{E}(X_t) = m$, (49) admits a unique invariant distribution with density $u(\mathbf{f}, x - m)$. Therefore, we first study the estimation of \mathbf{f} when (X_t) is in centered stationary regime ($X_0 \sim u(\mathbf{f}, x)dx$). Then, we study the joint estimation of (m, \mathbf{f}) when the process is in non centered stationary regime ($X_0 \sim u(\mathbf{f}, x - m)dx$).

When (X_t) is in stationary regime, $\mathcal{L}(X_t)$ does no longer depend on t and is equal to the stationary distribution $u(\mathbf{f}, x)dx$ in centered stationary regime and to $u(\mathbf{f}, x - m)dx$ in stationary regime with expectation m . Due to the polynomial form of the interaction function Φ , the convolution product $\Phi \star u(\mathbf{f}, \cdot)$ (resp. $\Phi \star u(\mathbf{f}, \cdot - m)$) is explicitly given as a function of \mathbf{f} and the moments of the invariant distribution. These moments have no explicit expression as functions of \mathbf{f} and m . Therefore, the exact log-likelihood can be studied theoretically (Proposition 4) but does not lead to computable estimators.

This is why we first 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 (49), based on the sample $(X_t, t \in [T, 2T])$. We prove that these estimators are consistent and asymptotically Gaussian with rate \sqrt{T} .

Extensions of this work naturally comprise the introduction of an additional potential term $V(\alpha, x)$ in the drift of equation (49).

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

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.

7. PROOFS

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

$$\begin{aligned}\Phi \star u(-x) &= \int \Phi(-x - y)u(y)dy = - \int \Phi(x + y)u(y)dy \\ &= - \int \Phi(x - y)u(-y)dy = - \int \Phi(x - y)u(y)dy = -\Phi \star u(x).\end{aligned}$$

Let $x \geq 0$, as Φ is convex on \mathbb{R}^+ by [H3] and u is even,

$$\begin{aligned}\Phi \star u(x) &= \Phi(x) + \int (\Phi(x - y) - \Phi(x))u(y)dy \\ &= \Phi(x) + \frac{1}{2} \int (\Phi(x - y) + \Phi(x + y) - 2\Phi(x))u(y)dy \\ &\geq \Phi(x).\end{aligned}$$

Using that Φ is odd, we have:

$$\Phi \star u(x) = \Phi(x) + \frac{1}{2} \int (\Phi(x + y) - \Phi(y - x) - 2\Phi(x))u(y)dy = \frac{1}{2} \int (\Phi(x + y) - \Phi(y - x))u(y)dy.$$

By [H2], we deduce:

$$\Phi \star u(x) \leq cx \int (2 + |x + y|^\ell + |x - y|^\ell)dy \leq c_\ell x(1 + x^\ell + \int_0^\infty y^\ell u(y)dy).$$

□

Proof of Proposition 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.* Kutoyants, 2004). When $Y_0 = \xi_0 = \eta \sim u_m$, by the uniqueness of solution, we obtain that $Y_t \equiv \xi_t$ for all $t \geq 0$. □

Proof of Proposition 4. Recall that $X_t \equiv Y_t$ (see Proposition 2). We have, applying the ergodic theorem, as u has moments of any order by (10),

$$\frac{1}{T}(\ell_T(\mathbf{f}) - \ell_T(\mathbf{f}_0)) \rightarrow_{\mathbb{P}_{\mathbf{f}_0}} -\frac{1}{2} \int (b(\mathbf{f}, x) - b(\mathbf{f}_0, x))^2 u_{\mathbf{f}_0}(y)dy = -\frac{1}{2\sigma^2} K(\mathbf{f}_0, \mathbf{f}).$$

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 $u(\mathbf{f}, \cdot) \equiv u(\mathbf{f}_0, \cdot)$ and $M_k(\mathbf{f})\mathbf{f} = M_k(\mathbf{f}_0)\mathbf{f}_0$ (see Proposition 3). 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,

$$\begin{aligned}
\sigma^2 \frac{\partial \ell_T}{\partial f_{2i+1}}(\mathbf{f}) &= \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s) dX_s - \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s) b(\mathbf{f}, X_s) ds \\
&= \sigma \int_0^T \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s) dW_s \\
\sigma^2 \frac{\partial^2 \ell_T}{\partial f_{2i+1} \partial f_{2i'+1}}(\mathbf{f}) &= \int_0^T \frac{\partial^2 b}{\partial f_{2i+1} \partial f_{2i'+1}}(\mathbf{f}, X_s) (dX_s - b(\mathbf{f}, X_s) ds) \\
&\quad - \int_0^T \left(\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s) \frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f}, X_s) \right) ds \\
&= \sigma \int_0^T \frac{\partial^2 b}{\partial f_{2i+1} \partial f_{2i'+1}}(\mathbf{f}, X_s) dW_s - \int_0^T \left(\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, X_s) \frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f}, X_s) \right) ds.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\sigma^2}{\sqrt{T}} \left(\frac{\partial \ell_T}{\partial f_{2i+1}}(\mathbf{f}), i = 0, \dots, k-1 \right)' &\rightarrow_{\mathcal{L}} \mathcal{N}(0, \sigma^2 I_{\mathbf{f}}), \\
\left(\frac{1}{T} \int_0^T \frac{\partial^2 b}{\partial f_{2i+1} \partial f_{2i'+1}}(\mathbf{f}, X_s) dW_s \right)_{i, i'} &\rightarrow 0, \quad \left(\frac{\sigma^2}{T} \frac{\partial^2 \ell_T}{\partial f_{2i+1} \partial f_{2i'+1}}(\mathbf{f}) \right)_{i, i'} \rightarrow -I_{\mathbf{f}},
\end{aligned}$$

where $I_{\mathbf{f}} = (\int_{\mathbb{R}} [\frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, x)] \frac{\partial b}{\partial f_{2i'+1}}(\mathbf{f}, x) u(\mathbf{f}, x) dx)$.

For any vector $a = (a_1 \dots a_k)'$, $a' I_{\mathbf{f}} a = \int [\sum_{i=0}^{k-1} a_i \frac{\partial b}{\partial f_{2i+1}}(\mathbf{f}, x)]^2 u(\mathbf{f}, x) dx > 0$ as the function under the integral is a polynomial and $u(\mathbf{f}, x)$ is positive for all x . By standard method, we can prove that the maximum likelihood $\hat{\mathbf{f}}_T$ associated with (17) satisfies

$$\sqrt{T}(\hat{\mathbf{f}}_T - \mathbf{f}) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \sigma^2 I_{\mathbf{f}}^{-1}).$$

□

Proof of Proposition 5. 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 $a' = (a_0, \dots, a_{k-1})$,

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

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

We write:

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

Thus, for large enough T ,

$$\left(\frac{1}{T} \Psi_T \right)^{-1} \frac{1}{T} Z_T = \hat{\mathbf{b}}_T = \mathbf{b} + \left(\frac{1}{T} \Psi_T \right)^{-1} \frac{\sigma}{T} \begin{pmatrix} \vdots \\ \int_T^{2T} X_s^{2i+1} dW_s \\ \vdots \end{pmatrix}_{0 \leq i \leq k-1}.$$

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

Proof of Theorem 1. For the proof, we set $\Psi(\mathbf{f}) = \Psi$, $M_k(\mathbf{f}) = M_k$, $u(\mathbf{f}, \cdot) = u(\cdot)$ and $\gamma_{2\ell}(\mathbf{f}) = \gamma_2$. 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 (21)). We have

$$\begin{aligned} Z_T &= \Psi_T \mathbf{b} + \sigma \int_T^{2T} z(X_s) dW_s \\ \Psi_T \widehat{M}_k \widehat{\mathbf{f}}_T &= \Psi_T M_k \mathbf{f} + \Psi_T M_k (\widehat{\mathbf{f}}_T - \mathbf{f}) + \Psi_T (\widehat{M}_k - M_k) \mathbf{f} + \Psi_T (\widehat{M}_k - M_k) (\widehat{\mathbf{f}}_T - \mathbf{f}). \end{aligned}$$

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

$$(50) \quad \frac{1}{T} \Psi_T M_k \sqrt{T} (\widehat{\mathbf{f}}_T - \mathbf{f}) = \frac{\sigma}{\sqrt{T}} \int_T^{2T} z(X_s) dW_s - \Psi \sqrt{T} (\widehat{M}_k - M_k) \mathbf{f} - R_T$$

$$(51) \quad = \Psi M_k \sqrt{T} (\widehat{\mathbf{f}}_T - \mathbf{f}) + S_T$$

with

$$(52) \quad R_T = \left(\frac{1}{T} \Psi_T - \Psi \right) \sqrt{T} (\widehat{M}_k - M_k) \mathbf{f} + \frac{1}{T} \Psi_T \sqrt{T} (\widehat{M}_k - M_k) (\widehat{\mathbf{f}}_T - \mathbf{f})$$

$$(53) \quad S_T = \sqrt{T} \left(\frac{1}{T} \Psi_T - \Psi \right) M_k (\widehat{\mathbf{f}}_T - \mathbf{f}).$$

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 have to precise the limiting distribution of the vector of centered and normalized moments $T^{1/2}((\widehat{\gamma}_{2\ell}(T) - \gamma_{2\ell})'_{\ell=1, \dots, k})$. For g an element of \mathcal{D} (see (69)-(70)), we have

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

where $Lg(x) = b(\mathbf{f}, x)g'(x) + \frac{\sigma^2}{2}g''(x)$. Let g_ℓ be any element of \mathcal{D} such that $Lg_\ell(x) = -(x^{2\ell} - \gamma_{2\ell})$. We have (see (73))

$$(54) \quad g'_\ell(f, x) = g'_\ell(x) = -2\sigma^{-2}u^{-1}(x) \int_{-\infty}^x (y^{2\ell} - \gamma_{2\ell})u(y)dy.$$

As g_ℓ belongs to $\mathbb{L}^2(u(x)dx)$, and (X_t) is stationary with marginal distribution $u(x)dx$,

$$(55) \quad \mathbb{E}_\beta(g_\ell(X_T) - g_\ell(X_0))^2 \leq 2 \int g_\ell^2(x)u(x)dx.$$

Thus,

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

by (55). So the vector $T^{1/2}(\widehat{\gamma}_{2\ell}(T) - \gamma_{2\ell})'_{\ell=1,\dots,k}$ converges in distribution to $\mathcal{N}_k(0, \sigma^2 V)$ with $V = (V_{ij})_{0 \leq i,j \leq k-1}$ and $V_{ij} = \int g'_i(x) g'_j(x) u(x) dx$.

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 conclude 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 (50) and (51),

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

where

$$(58) \quad \alpha_\ell(\mathbf{f}, x) = \sum_{u=0}^{k-1} [M_k^{-1}]_{\ell u} \sum_{j=0}^{k-1} [\Psi^{-1}]_{uj} x^{2j+1} = [M_k^{-1} \Psi^{-1} \begin{pmatrix} x \\ x^3 \\ \vdots \\ x^{2k-1} \end{pmatrix}]_\ell$$

$$(59) \quad \beta_\ell(\mathbf{f}, x) = \sum_{j=0}^{k-1} [M_k^{-1}]_{\ell j} \sum_{v=j}^{k-1} \binom{2v+1}{v-j} g'_{v-j}(x) f_{2v+1} \mathbf{1}_{j \leq v} = [M_k^{-1} M_k^{g'} \mathbf{f}]_\ell,$$

where $M_k^{g'}$ is the matrix M_k where all γ_{2j} are replaced by $g'_j(x)$ (with $g'_0 = 0$). Finally, we apply the ergodic theorem for stochastic integrals to obtain:

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

with

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

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

□

Proof of Proposition 6. Define $D_k = \text{diag}((2\ell - 1), \ell = 1, \dots, k)$ the diagonal matrix with diagonal element $\ell(2\ell - 1)$. We have $B(\mathbf{f}) = \sigma^2 D_k (1 \ \gamma_2(\mathbf{f}) \ \dots \ \gamma_{2(k-1)}(\mathbf{f}))'$. 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 $\mathcal{N}_k(\mathbf{0}, \sigma^2 V(\mathbf{f}))$ where $V(\mathbf{f})$ is defined in (27). Consequently, the vector $\sqrt{T}(\widetilde{B}_T - B(\mathbf{f}))$ converges in distribution to $\mathcal{N}(\mathbf{0}, K)$ with

$$(60) \quad K = \sigma^4 D_k V(\mathbf{f}) D_k.$$

The result of Proposition 6 follows.

□

Proof of Proposition 7. By the ergodic theorem $\hat{m} = T^{-1} \int_0^T X_s ds \rightarrow \int x u_m(x) dx = m$. For $\tilde{\gamma}_{2\ell}(T)$, we write:

$$\begin{aligned} \tilde{\gamma}_{2\ell}(T) &= \frac{1}{T} \int_0^T (X_s - m + m - \hat{m})^{2\ell} ds = \frac{1}{T} \int_0^T (X_s - m)^{2\ell} ds \\ &\quad + (m - \hat{m}) \sum_{k=1}^{2\ell} \binom{2\ell}{k} \frac{1}{T} \int_0^T (X_s - m)^{2\ell-k} ds (m - \hat{m})^{k-1}. \end{aligned}$$

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 $\hat{m} - m$ tends to 0 and is multiplied by a factor tending to a limit.

The infinitesimal generator $L - m$ of (40) is given by

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

Thus, the equality

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

implies, as (X_t) is stationary,

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

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 $g'(x) = -\frac{2}{\sigma^2 u(x - m)} \int_{-\infty}^{x-m} v u(v) dv := h'_0(x - m)$. Thus,

$$(61) \quad \sqrt{T}(\hat{m} - m) = \frac{\sigma}{\sqrt{T}} \int_0^T h'_0(X_s - m) dW_s + o_P(1).$$

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 (54)).

Therefore,

$$(62) \quad \frac{1}{\sqrt{T}} \int_0^T ((X_s - m)^{2\ell} - \gamma_{2\ell}) = \frac{\sigma}{\sqrt{T}} \int_0^T h'_\ell(X_s - m) dW_s + o_P(1).$$

Now, splitting $X_s - \hat{m} = X_s - m + m - \hat{m}$ yields,

$$\begin{aligned} \sqrt{T}(\tilde{\gamma}_{2\ell}(T) - \gamma_{2\ell}) &= \frac{1}{\sqrt{T}} \int_0^T ((X_s - m)^{2\ell} - \gamma_{2\ell}) + \sqrt{T}(m - \hat{m}) \binom{2\ell}{1} \frac{1}{T} \int_0^T (X_s - m)^{2\ell-1} ds \\ &\quad + \sqrt{T}(m - \hat{m})^2 \sum_{k=2}^{2\ell} \binom{2\ell}{k} (m - \hat{m})^{k-2} \frac{1}{T} \int_0^T (X_s - m)^{2\ell-k} ds \end{aligned}$$

For the second term, note that $\frac{1}{T} \int_0^T (X_s - m)^{2\ell-1} ds \rightarrow \int (x - m)^{2\ell-1} u_m(x) dx = \int x^{2\ell-1} u(x) dx = 0$ as u 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 (46) and (47). The convergence in distribution result follows. \square

Proof of Theorem 2 . Here again, we set $\Psi(\mathbf{f}) = \Psi$, $M_k(\mathbf{f}) = M_k$. We proceed as in Proposition 5 and Theorem 1. First, we prove, using (44),

$$(63) \quad \frac{\tilde{\Psi}_T}{T} \rightarrow_{a.s.} \Psi \quad \text{and} \quad \sqrt{T}(\frac{\tilde{\Psi}_T}{T} - \Psi) = O_P(1) \quad \text{and}$$

$$(64) \quad V_T = \sqrt{T} \left(\frac{\tilde{Z}_T}{T} - \frac{\tilde{\Psi}_T}{T} \mathbf{b} \right) \rightarrow_{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^2 \Psi).$$

From these two results, as $\tilde{\mathbf{b}}_T = (\frac{\tilde{\Psi}_T}{T})^{-1} \frac{\tilde{Z}_T}{T}$ (see (43)), we deduce:

$$(65) \quad \sqrt{T}(\tilde{\mathbf{b}}_T - \mathbf{b}) = (\frac{\tilde{\Psi}_T}{T})^{-1} V_T \rightarrow_{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^2 \Psi^{-1}).$$

Proof of (63): We have:

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

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

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

Next,

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

As $2(i+j)+1$ is odd, $\frac{1}{T} \int_T^{2T} (X_s - m)^{2(i+j)+1} ds \rightarrow_{a.s.} \int (x - m)^{2(i+j)+1} u_m(x) dx = 0$. Thus, the second term above is $\sqrt{T}(m - \hat{m}) \times o_{a.s.}(1) = o_P(1)$. The third term is $\frac{1}{\sqrt{T}} T(m - \hat{m})^2 \times o_P(1) = o_P(1)$.

For the first term, we prove as in Proposition 7 that

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

The proof of (63) is complete.

Proof of (64): We write:

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

where

$$(66) \quad T_{2,i} = \frac{1}{T} \int_T^{2T} (X_s - \hat{m})^{2i+1} (b(\mathbf{f}, X_s - m) - b(\mathbf{f}, X_s - \hat{m})) ds$$

$$(67) \quad T_{3,i} = \frac{\sigma}{T} \int_T^{2T} (X_s - \hat{m})^{2i+1} dW_s$$

We have:

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

Now,

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

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

$$(68) \quad \sqrt{T} T_{2,i} = \sqrt{T} (\hat{m} - m) \times o_P(1) = o_P(1).$$

Now, $\sqrt{T} T_3$ is a martingale such that $\langle \sqrt{T} T_3 \rangle_T = \sigma^2 \tilde{\Psi}_T / T \rightarrow \sigma^2 \Psi$. Therefore, $\sqrt{T} T_3$ converges in distribution to $\mathcal{N}(\mathbf{0}, \sigma^2 \Psi)$.

Finally, we have obtained $\sqrt{T} \left(\frac{\tilde{Z}_T}{T} - \frac{\tilde{\psi}_T}{T} \mathbf{b} \right) = \sqrt{T} T_3 + o_P(1)$. The proof of (64) is achieved and (65) follows.

Now, we can complete the proof of Theorem 2. On one hand, we have the relation (see (67)-(68)):

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

On the other hand, we have:

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

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

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

This yields:

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

where

$$\begin{aligned} \tilde{R}_T &= \left(\frac{\tilde{\Psi}_T}{T} - \Psi \right) \sqrt{T} (\tilde{M}_k - M_k) \mathbf{f} + \frac{\tilde{\Psi}_T}{T} \sqrt{T} (\tilde{M}_k - M_k) (\tilde{\mathbf{f}}_T - \mathbf{f}), \\ \tilde{S}_T &= \sqrt{T} \left(\frac{\tilde{\Psi}_T}{T} - \Psi \right) M_k (\tilde{\mathbf{f}}_T - \mathbf{f}), \\ K_T &= \frac{\sigma}{\sqrt{T}} \int_T^{2T} [z(X_s - \hat{m}) - z(X_s - m)] dW_s. \end{aligned}$$

As previously, we prove that $\tilde{R}_T = o_P(1)$, $\tilde{S}_T = o_P(1)$ using Proposition 7. We have to look at K_T . We have:

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

Each term $\frac{1}{\sqrt{T}} \int_T^{2T} (X_s - m)^{2i-j} dW_s$ converges in distribution while $m - \hat{m}$ tends to 0. So $K_{T,i} = o_P(1)$ for $i = 0, \dots, k-1$. Now, the term $\sqrt{T} (\tilde{M}_k - M_k) \mathbf{f}$ can be treated as previously in Theorem 30 and we can write:

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

with $\alpha(x) = \alpha(\mathbf{f}, x)$, $\beta(x) = \beta(\mathbf{f}, x)$ defined in (58) and (59). Therefore, $\sqrt{T} (\tilde{\mathbf{f}}_T - \mathbf{f})$ converges in distribution to $\mathcal{N}(0, \sigma^2 \Sigma(\mathbf{f}))$ with $\Sigma(\mathbf{f})$ defined in (31).

The result concerning the joint distribution follows from (46) and (50). \square

8. APPENDIX

We now state the central limit theorem associated with (12) (see *e.g.* Genon-Catalot *et. al.*, 2000 and references therein). Let L denote the infinitesimal generator of the SDE (11),

$$(69) \quad Lg = (\sigma^2/2)g'' - \Phi \star u_m(\cdot)g' = \frac{\sigma^2}{2u_m} (g'u_m)'.$$

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

$$(70) \quad \mathcal{D} = \{g \in \mathbb{L}^2(u_m(x)dx), g' \text{ absolutely continuous}, Lg \in \mathbb{L}^2(u_m(x)dx), \lim_{|x| \rightarrow \infty} g'(x)/s(x) = 0\}.$$

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

Proposition 8. *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 $\langle \cdot, \cdot \rangle_{u_m}$ the scalar product of $\mathbb{L}^2(u_m(x)dx)$. If $f_c \in \text{Range}(\mathcal{D})$, then, as T tends to infinity,*

$$(71) \quad \frac{1}{\sqrt{T}} \int_0^T f_c(Y_s)ds \rightarrow_{\mathcal{L}} \mathcal{N}(0, \sigma^2(f_c))$$

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,

$$(72) \quad \text{Var} \left(\frac{1}{\sqrt{T}} \int_0^T f_c(Y_s)ds \right) \rightarrow \sigma^2(f_c).$$

The following relation holds:

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

Moreover, in model (11), $\text{Range}(\mathcal{D}) = \{h \in \mathbb{L}^2(u_m(x)dx), \int h(x)u_m(x)dx = 0\}$. Therefore, (71)-(72) hold for all $f \in \mathbb{L}^2(u_m(x)dx)$.

Proposition 8 requires some comments. Its first part ((71)-(72)) is classical. However, the last part, *i.e.* that (71)-(72) hold for all $f \in \mathbb{L}^2(u_m(x)dx)$, is less known and not obvious.

Using (69), 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 (69), as $\int_{-\infty}^{+\infty} f_c(y)u_m(y)dy = 0$, we have

$$(73) \quad g'_{f_c}(x) = g'(x) = 2\sigma^{-2}u_m^{-1}(x) \int_{-\infty}^x f_c(y)u_m(y)dy = -2\sigma^{-2}u_m^{-1}(x) \int_x^{+\infty} f_c(y)u_m(y)dy.$$

By Proposition 8, the integral

$$(74) \quad \sigma^2(f_c) = \sigma^2 \int_{\mathbb{R}} (g'(x))^2 u_m(x)dx = 4\sigma^{-2} \int_{\mathbb{R}} u_m^{-1}(x) \left(\int_{-\infty}^x f_c(y)u_m(y)dy \right)^2 dx$$

is finite for all $f \in \mathbb{L}^2(u_m(x)dx)$.

Note that the fact that (74) 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 (74) is possible but the exact proof is not immediate.

Corollary 1. *Let h_1, \dots, h_p be functions belonging to $\text{Range}(\mathcal{D})$ and such that $\int h_j(x)u_m(x)dx = 0$, for $j = 1, \dots, p$. 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 $\frac{1}{\sqrt{T}}(\int_0^T h_i(Y_s)ds, i = 1, \dots, p)' \rightarrow_{\mathcal{L}} \mathcal{N}_p(0, V)$ with $V = (V(h_i, h_j), 1 \leq i, j \leq p)$.

Proof of Proposition 8. The result is given in Theorem 2.2 in Genon-Catalot *et al.*, 2000. We know that $\text{Range}(\mathcal{D}) \subset \{h \in \mathbb{L}^2(u_m(x)dx), \int h(x)u_m(x)dx = 0\}$.

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

$$(75) \quad \lim_{x \rightarrow +\infty} \gamma^{-1}(x) = 0, \quad \lim_{x \rightarrow -\infty} \gamma^{-1}(x) = 0.$$

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 $\text{Range}(\mathcal{D}) = \{h \in \mathbb{L}^2(u_m(x)dx), \int h(x)u_m(x)dx = 0\}$. Therefore, (71)-(72) hold for all $f \in \mathbb{L}^2(u_m(x)dx)$. \square

Proof of Corollary 1. The proof follows by application of the Cramér-Wold device. \square

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] 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.
- [3] 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.
- [4] 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.
- [5] Benachour, S., Roynette, B. and Vallois, P. (1998b). Nonlinear self-stabilizing processes - II Convergence to invariant probability. *Stoch. Proc. Appl.* **75**, 203-224.
- [6] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31** (5), 615-641.
- [7] Belomestny, D., Pilipauskaitė, V. and Podolskij, M. (2021). Semiparametric estimation of McKean-Vlasov stochastic differential equations. Preprint Arkiv.
- [8] Bibby, B.M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion *Bernoulli* **1**, 17-39.
- [9] Bishwal, J.P.N., (2011). Estimation in Interacting Diffusions: Continuous and Discrete Sampling. *Applied Mathematics*, **2**, 1154-1158.
- [10] 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.
- [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. and Genon-Catalot, V. (2021). Drift estimation on non compact support for diffusion models. *Stochastic Processes and their Applications* **134**, 174-207.
- [14] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**, 514-543.
- [15] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *The Annals of Statistics*, **33**, 2507-2528.
- [16] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probability Theory and Related Fields*, **137**, 25-47.

- [17] Della Maestra, L. and Hoffmann, M. (2021). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. *Probability Theory and Related Fields*, <https://doi.org/10.1007/s00440-021-01044-6>.
- [18] 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.
- [19] 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.
- [20] Funaki, T. (1984). A certain class of diffusions processes associated with nonlinear parabolic equations, *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **67**, 331-348.
- [21] Gärtner, J. (1988). On the McKean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, **137**, 197-248.
- [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] Giesecke, K., Schwenkler, G. and Sirignano, J.A. (2020). Inference for large financial systems. *Mathematical Finance*, **30**, 3-46.
- [26] 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.
- [27] Herrmann, S. and Tugaut, J. (2010). Non uniqueness of stationary measures for self-stabilizing diffusions. *Stoch. Proc. Appl.* **120**, 1215-1246.
- [28] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, 135-163.
- [29] Höpfner, R. (2014). *Asymptotic Statistics with a View to Stochastic Processes*. Walter de Gruyter, Berlin/Boston.
- [30] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples*. Springer.
- [31] Kasonga, R.A. (1990). Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, **50**, 865-875.
- [32] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
- [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] 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.
- [37] McKean, H.P., Jr (1966). A class of Markov processes associated with nonlinear parabolic equation. *Proceedings NatL. Acad. Sci. U.S.A.* **56**, 1907-1911.
- [38] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of applied Probability* **13**, 540-560.
- [39] Masuda, H. (2007). Ergodicity and exponential beta-mixing for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, **117**, 35-56.
- [40] Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochastic Processes and their Applications*, **129**, 1013-1059.
- [41] 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.
- [42] Molginer, A. and Edelstein-Keshet, L. (1999). A non-local model for a swarm. *Journal of Mathematical Biology*, **38**, 534-570.
- [43] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Annals of Statistics*, **48**, 1383-1408.
- [44] Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications*, **124**, 883-914.

- [45] Sharrock, L., Kantas, N., Parpas, P. and Pavliotis, G.A. (2021). Parameter Estimation for the McKean-Vlasov Stochastic Differential Equation. *Preprint Arkiv*.
- [46] 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.
- [47] Veretennikov, A. Yu., (2006). On ergodic measures for McKean-Vlasov stochastic equations. In *Monte Carlo and. Quasi-Monte Carlo Methods 2004*, 471-486. Springer.