Semiparametric estimation of McKean-Vlasov SDEs
Denis Belomestny\textsuperscript{a,c,}\textsuperscript{*}, Vytautė Pilipauskaitė\textsuperscript{b,†} and Mark Podolskij\textsuperscript{b,‡}

\textsuperscript{a}Faculty of Mathematics, University of Duisburg-Essen, E-mail: \textsuperscript{*}denis.belomestny@uni-due.de
\textsuperscript{b}Department of Mathematics, University of Luxembourg, E-mail: \textsuperscript{†}vytaute.pilipauskaite@uni.lu; \textsuperscript{‡}mark.podolskij@uni.lu
\textsuperscript{c}National University Higher School of Economics, Moscow

Abstract. In this paper we study the problem of semiparametric estimation for a class of McKean-Vlasov stochastic differential equations. Our aim is to estimate the drift coefficient of a MV-SDE based on observations of the corresponding particle system. We propose a semiparametric estimation procedure and derive the rates of convergence for the resulting estimator. We further prove that the obtained rates are essentially optimal in the minimax sense.

MSC2020 subject classifications: Primary 62G20, 62M05; secondary 60G07, 60H10
Keywords: deconvolution, McKean-Vlasov SDEs, mean field models, multi-agent learning, minimax bounds, semiparametric estimation

1. Introduction

In the past fifty years diffusion processes found numerous applications in natural and social sciences, and a variety of statistical methods have been investigated in the setting of SDEs during the last few decades. Maximum likelihood estimation and Bayesian approach are the most well established parametric methods in the literature; we refer to the monograph [19]. When the likelihood function is not available in a closed form, quasi likelihood methods provide an alternative approach to parameter estimation, see [6] and references therein. The most recent contributions to nonparametric inference for diffusions are [26, 29]. These belong to the most successful tools when analysing estimation problems in different observation schemes of a diffusion.

Many diffusion models in natural and applied sciences can be viewed as continuous-time processes with complex and nonlinear probabilistic structure. For example, in statistical mechanics nonlinear diffusion models and particle systems have long and successful history. In general, nonlinear Markov processes are stochastic processes whose transition functions may depend not only on the current state of the process but also on the current distribution of the process. These processes were introduced by McKean [23] to model plasma dynamics. Later nonlinear Markov processes were studied by a number of authors; we mention here the books of Kolokoltsov [18] and Sznitman [30]. These processes arise naturally in the study of the limit behavior of a large number of weakly interacting particles and have a wide range of applications, including financial mathematics, population dynamics, multi-agent reinforcement learning and neuroscience (see, e.g., [10] and the references therein). This is in turn related to many complex phenomena arising in engineering and socioeconomic settings, among others, for instance, dynamic economic models involving competing agents [20], biological models on animal competition and conflicts [24], wireless power control [16], road traffic engineering [15], and shared data buffer modelling [1]. The particle systems are also used as an approximation of continuous models (as in e.g. vortex simulation for Euler equation, see [22]). In this context the mean field theory has been employed to bridge the interaction of particles at the microscopic scale and the mesoscopic features of the system. From a probabilistic point of view propagation of chaos, fluctuation analysis, and large deviations have been investigated for a variety of mean field models and nonlinear diffusions.
In recent years, one witnessed a growing interest in statistical problems for high dimensional diffusions in general and McKean-Vlasov (MV) SDEs in particular. Statistical inference for high dimensional Ornstein-Uhlenbeck models have been investigated in [7, 11]. The article [17] has developed maximum likelihood estimation for a system of interacting diffusion processes. The paper [4] discusses properties of the approximate maximum likelihood estimator in the discrete observation case and those of the sieve estimator in the continuous observation case, where the parameter to be estimated is a function of time. In the case of nonlinear parametrisation [28] has extended the results in [17] and also developed an online-estimation procedure. A parametric problem of estimating the coefficients of a MV-SDE under a small noise assumption has been studied in [12, 13, 27]. The article [14] treats the parameter estimation problem for a system of interacting jump-diffusion processes, widely applicable in financial mathematics. For more references we refer to [8, 28]. We note that in many known models for MV-SDEs (corresponding systems of interacting particles) the drift is given by the convolution of current distribution (respectively, current empirical distribution) of the process and polynomials or series of trigonometric functions, see e.g. a model in [14, Example 2.5] for systemic risk in a large system of interacting agents, Kuramoto-Shinomoto-Sakaguchi model in [2, 10, Equation 5.214], granular media equation with physical interpretation in [3, 5, 21], mean field control problems [9]. Our current work is mostly related to a recent paper [8], where based on observation of a trajectory of an interacting particle system over a fixed time horizon, the authors study nonparametric estimation of the solution (density) of the associated nonlinear Fokker-Planck equation, together with the drift function. The underlying statistical problem turns out to be rather challenging and [8] contains only partial results. In particular, the problem of estimating a distribution dependent drift function of a MV-SDE from the observations of the corresponding particle system at time $T > 0$ has not been yet studied in the literature.

In this paper we consider the case where the drift has a semiparametric form consisting of a polynomial part, a trigonometric part and a nonparametric interaction function convolved with an unknown marginal distribution of the underlying MV-SDE. The goal of this research is twofold: first to propose a kernel type estimator for the drift function based on the empirical characteristic function of the particles; and, second, to study its properties. Our estimation strategy uses the fact that the empirical measure of the particle system converges to the law of the corresponding MV-SDE for any fixed time as the number of particles diverges to infinity (a propagation of chaos result), while the latter converges to the invariant measure when time diverges to infinity. In the second step we employ the functional form of the invariant measure (cf. [5, 21]) to recover the unknown interaction function via deconvolution and Fourier inversion. This procedure requires a double asymptotic regime, where the number of particles $N \to \infty$ as well as the terminal time $T \to \infty$; we emphasise that only observations of the particle system at the terminal time $T$ are required. We derive upper bounds on $L^2$ risk of the proposed estimator and show that these bounds are essentially optimal in minimax sense for a properly chosen functional class of drift functions. In particular, we show that the convergence rates of our estimator are logarithmic under a polynomial tail behaviour of the non-parametric part of the interaction function. The slow rate of convergence is not surprising as our statistical problem can be compared with a classical deconvolution problem with super smooth noise. To the best of our knowledge, this is the first work containing minimax optimal procedure for semiparametric estimation of the coefficients of MV-SDEs from discrete observations of the corresponding particle system and hence fills an important gap in the current literature on statistical inference for MV-SDEs.

The structure of the paper is as follows. In Section 2 we introduce the main setup and recall some basic facts about MV-SDEs and related particle systems. In Section 3 we formulate our main statistical problem and describe the estimation procedure. Section 4 is devoted to the convergence analysis of the proposed algorithm. In particular, we derive upper bounds on $L^2$ risk of the suggested drift estimator. In Section 5 we complete our theoretical analysis by providing lower bounds for the nonparametric part of the model that essentially match upper bounds obtained in Section 4. Conclusions and outlook are presented in Section 6. All proofs are collected in Section 7.

2. The particle system model and propagation of chaos

Throughout the paper we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, on which all stochastic processes are defined. We focus on an $N$-dimensional system of stochastic differential equations given by

$$X_{t}^{i,N} = X_{0}^{i,N} + B_{t}^{i} - \frac{1}{2N} \sum_{j=1}^{N} \int_{0}^{t} \varphi'(X_{s}^{i,N} - X_{s}^{j,N}) \, ds, \quad 1 \leq i \leq N, \quad t \geq 0,$$

where $B_{t} = (B_{t}^{i})_{t \geq 0}$, $1 \leq i \leq N$, are independent standard Brownian motions and $X_{0}^{i}$, $1 \leq i \leq N$, are i.i.d. random variables with common distribution $\mu_{0}(dx)$, independent of $B_{t}^{i}$, $1 \leq i \leq N$, and

$$\int_{\mathbb{R}} x \mu_{0}(dx) = 0, \quad \int_{\mathbb{R}} |x|^{k} \mu_{0}(dx) < \infty \quad \forall k \geq 1.$$
Here the interaction function $\varphi'$ denotes the derivative of the function $\varphi \in C^2(\mathbb{R})$, which satisfies the following assumption:

(A) The function $\varphi$ is even (i.e., $\varphi(x) = \varphi(-x)$ for all $x \in \mathbb{R}$), strictly convex:

\begin{equation}
\varphi''(x) \geq \lambda > 0, \quad \forall x \in \mathbb{R},
\end{equation}

and locally Lipschitz with polynomial growth, that is

\[ |\varphi'(x) - \varphi'(y)| \leq |x - y|P(x) + P(y), \quad \forall x, y \in \mathbb{R}, \]

for a polynomial $P$.

The particle system (2.1) has been originally studied in [3]. However, as pointed out in [21], the asymptotic properties of the system are rather ill-behaved in terms of uniformity and long term behaviour, and it is more appropriate to consider the projected particle system

\begin{equation}
Y_t^{i, N} := X_t^{i, N} - \frac{1}{N} \sum_{j=1}^N X_t^{j, N}, \quad 1 \leq i \leq N, \quad t \geq 0.
\end{equation}

The mean field equation, which determines the asymptotic behaviour of the $i$-th particle of the system at (2.3), is given by the 1-dimensional McKean-Vlasov equation

\begin{equation}
\dot{X}_t = \mathbb{X}_0 + B_t - \frac{1}{2} \int_0^t (\varphi' \ast \mu_s)(\mathbb{X}_s) \, ds, \quad t \geq 0,
\end{equation}

where $\mu_t(dx) = \mathbb{P}(X_t \in dx)$ with

\[ (\varphi' \ast \mu_t)(x) = \int_{\mathbb{R}} \varphi'(x - y)\mu_t(dy), \quad x \in \mathbb{R}, \quad t \geq 0, \]

the initial random variable $\mathbb{X}_0$ and standard Brownian motion $B$ being independent. Under Assumption (A) the existence and uniqueness of $Y^N$ and $\mathbb{X}$ have been established, see e.g. [3, Theorem 3.1 and Proposition 5.1], [5, discussion before Section 2.1 and Theorem 2.6]. Under Assumption (A) (in addition [5] assumes the polynomial growth of $\varphi''$ to weaken the assumption of the uniform convexity) the measure $\mu_t$ possesses a smooth Lebesgue density, which solves the partial differential equation

\[ \frac{\partial}{\partial t} \mu_t = \frac{1}{2} \frac{\partial^2}{\partial x^2} \mu_t + \frac{1}{2} \frac{\partial}{\partial x} ((\varphi' \ast \mu_t) \mu_t), \quad \mu_0(dx) = \mathbb{P}(\mathbb{X}_0 \in dx). \]

The stochastic differential equation (2.4) admits a unique invariant measure in the set of probability measures with all finite moments (see e.g. [3, Section 4], discussion before and remark after Theorem 4.1 in [5], also [21]). Its density is described by an integral equation of convolution type:

\begin{equation}
\pi(x) = Z_\pi^{-1} \exp(-\varphi \ast \pi(x)) \quad \text{with} \quad Z_\pi = \int_{\mathbb{R}} \exp(-\varphi \ast \pi(x)) \, dx.
\end{equation}

We note that $Z_\pi < \infty$ because the invariant measure admits all moments and $\varphi \geq \varphi(0)$ has a polynomial growth under Assumption (A), see also Lemma 3.1. Furthermore, the uniqueness of invariant measure implies its symmetry, since the functions $\pi(x)$ and $\pi(-x)$ both satisfy the integral equation in (2.5). In this article we will consider semiparametric estimation of the interaction function $\varphi'$ and the identity (2.5) will be key for our approach.

Next, we will state a propagation of chaos result and exponential convergence to equilibrium for the particle system (2.3). We recall that the Wasserstein $p$-distance between two probability measures $\mu_1, \mu_2$ on $\mathbb{R}$ is defined by

\[ W_p(\mu_1, \mu_2) := \left( \inf_{X_1 \sim \mu_1, X_2 \sim \mu_2} \mathbb{E}[|X_1 - X_2|^p] \right)^{1/p}, \]

where the infimum is taken over all couplings $(X_1, X_2)$ such that $X_i$ has the law $\mu_i$, $i = 1, 2$. The following theorem has been shown in [5, 21].
Theorem 2.1 (Theorems 5.1 and 6.2 in [21]). Let $X$ be i.i.d. copies of the process $X$ defined at (2.4) so that every $X^i$ is driven by the same Brownian motion as the $i$th particle of the system (2.1) and they are equal at time 0. Denote by

$$\Pi_{N,T} = N^{-1} \sum_{i=1}^{N} \delta_{Y_t^i,N}$$

the empirical distribution of the projected particle system $Y^i_{T,N}$, $1 \leq i \leq N$, and by $\Pi$ the law associated to the invariant density $\pi$. Under Assumption (A) there exist constants $C_1, C_2 > 0$ (independent of $N, T$) such that

$$\sup_{t \geq 0} \mathbb{E}[(Y_{t}^i, N - X_{t}^i)^2] \leq C_1 N^{-1}$$

and

$$\mathbb{E}[W_1^2(\Pi_{N,T}, \Pi)] \leq C_1 N^{-1} + C_2 \exp(-\lambda T) =: N_T^{-1}$$

where the constant $\lambda > 0$ has been introduced in (2.2).

The estimate (2.6) states that the invariant distribution $\Pi$ of the mean field equation (2.4) is well approximated by the empirical measure $\Pi_{N,T}$ and gives the error bound associated with this approximation. In the next section we will use this result in our estimation procedure.

3. Statistical problem and the estimation method

We assume that the data

$$Y^{1,N}_T, \ldots, Y^{N,N}_T$$

is observed and $N, T \to \infty$ (and, as a consequence, $N_T \to \infty$), and our goal is to estimate the interaction function $\phi'$ introduced in (2.1) in a semiparametric setting. We remark that the sampling scheme is rather natural as we observe the $N$-particle system only at the terminal time $T$ and don’t assume that the history of the particle system is known. According to the identity (2.5) and the statement (2.6), the considered data suffices to identify the interaction function $\phi'$ within the double asymptotic framework. Indeed, thanks to the uniform in time propagation of chaos, as $N \to \infty$, every $Y^i_{T,N}$ converges to $X_T$, where $X^1_T, \ldots, X^N_T$ are independent copies of $X_T$ with law $\mu_T$ in (2.4). The latter converges as $T \to \infty$ to the invariant law $\Pi$ of the process $X$.

Due to the complexity of the integral equation (2.5), which we will heavily rely on in our estimation procedure, we can not treat fully general interaction functions $\phi'$. Instead we consider a semiparametric model of the form

$$\phi(y) = \sum_{0 < j \leq J} a_j \varphi_j(y) + \beta(y), \quad y \in \mathbb{R},$$

where

$$\varphi_j(y) = y^{2j}, \quad 0 < j \leq J_1, \quad \varphi_j(y) = \cos(\theta_j y), \quad J_1 < j \leq J, \quad y \in \mathbb{R},$$

for some known distinct frequencies $\theta_{J_1+1} > 0, \ldots, \theta_J > 0$ and known positive integers $J_1 \leq J$. The parameters $a_1 > 0, a_2 \geq 0, \ldots, a_{J_1-1} \geq 0, a_{J_1} > 0, a_{J_1+1}, \ldots, a_J \in \mathbb{R}$ and the function $\beta$ are unknown. The nonparametric component $\beta \in C^2(\mathbb{R})$ is even, bounded and such that $\beta'$ is bounded, $||\beta'||_{L^1(\mathbb{R})} := \int_{\mathbb{R}} |\beta'(y)| dy < \infty$ (later we will also assume that $||\beta''||_{L^2(\mathbb{R})} := (\int_{\mathbb{R}} |\beta''(y)|^2 dy)^{1/2} < \infty$). The strict convexity condition (2.2) is induced by the assumption

$$2a_1 - \sum_{J_1 < j \leq J} \theta_j^2 |a_j| + \beta''(y) \geq \lambda > 0, \quad \forall y \in \mathbb{R}.$$

Our approach will be based upon the integral equation (2.5). Firstly, we give the representation of the invariant density $\pi$ in (2.5) in the setting (3.1). Since linear space of power (and trigonometric) functions is closed under convolution, the presence of the polynomial term translates from (3.1) to this representation and subsequently allows to bound $\pi$. It turns out that the invariant distribution has subgaussian tails and its density is infinitely smooth. Specifically, the technical
Lemma 7.1 gives upper bounds for $\pi$ and its derivatives which are required in the proofs, whereas the lower bound in (7.1) determines the threshold $\delta$ in the estimation procedure below. The presence of the trigonometric terms in (3.1) is for modelling purpose only and does not influence the estimation theory. It remains valid even in case $J_1 = J$, where trigonometric terms are excluded from the model. In the following, for any function $f \in L^1(\mathbb{R})$, we denote by $\mathcal{F}(f)$ the Fourier transform of $f$.

**Lemma 3.1.** For $\varphi$ given in (3.1), we have

$$(\varphi \ast \pi)(y) = \alpha_0 + \sum_{0 < j \leq J_1} \alpha_j \varphi_j(y) + (\beta \ast \pi)(y), \quad y \in \mathbb{R},$$

with $\alpha_0 = \sum_{0 < k \leq J_1} m_{2k} a_k$ and

$$(3.3) \quad \alpha_j = \sum_{j \leq k \leq J_1} \binom{2k}{2j} m_{2(k-j)} a_k, \quad 0 < j \leq J_1, \quad \alpha_j = a_j \mathcal{F}(\pi)(\theta_j), \quad J_1 < j \leq J,$$

where $m_k = \int_{\mathbb{R}} y^k \pi(y) dy$ denotes the $k$th moment of the invariant measure $\Pi$.

**Proof.** Firstly, we note that for every $k \in \mathbb{N}$ the $k$-th moment of $\Pi$ exists. Indeed, it follows from weak convergence of $\mu_t$ to $\Pi$ (e.g., [21, Theorem 6.2]) and from uniform integrability of $|X|^k$ which in turn follows from moment bounds (e.g., [3, Proposition 3.10], [21, Lemma 5.2]), see also Lemma 7.1. We consider the sum of $a_k(\varphi_k \ast \pi)(x)$ over $0 < k \leq J_1$, where

$$(\varphi_k \ast \pi)(y) = \int (y - x)^{2k} \pi(x) dx = \sum_{j=0}^{k} \binom{2k}{2j} m_{2(k-j)} y^{2j}$$

since $\pi$ is symmetric. Then interchanging the order of summation, we get the formula for the coefficients $\alpha_j$, $0 \leq j \leq J_1$.

Similarly, for $J_1 < j \leq J$, the coefficients $\alpha_j$ are obtained through the identity

$$\int \cos(\theta_j(y - x)) \pi(x) dx = \cos(\theta_j y) \int \cos(\theta_j x) \pi(x) dx,$$

where we again have used the symmetry of $\pi$. This completes the proof of Lemma 3.1. \hfill \Box

We now proceed with the introduction of the estimation procedure, which consists of four steps:

(i) Estimate the derivative of the log-density

$$(3.4) \quad l(y) := (\log \pi)'(y) = \frac{\pi'(y)}{\pi(y)}, \quad y \in \mathbb{R},$$

via a kernel-type estimator $l_{N,T}$ based on the observed data $Y_{T}^{1, N}, \ldots, Y_{T}^{N, N}$.

(ii) Estimate the parameter $\alpha := (\alpha_1, \ldots, \alpha_J)^T \in \mathbb{R}^J$ using the minimum contrast method based on

$$l(y, \alpha) = - \sum_{j=1}^{J} \alpha_j \varphi_j(y),$$

which approximates $l(y) = l(y, \alpha) - (\beta' \ast \pi)(y)$ for large values of $y \in \mathbb{R}$.

(iii) Use the results of step (ii) to construct an estimator $\Psi_{N,T}$ of

$$\Psi(y) := -(\beta' \ast \pi)(y), \quad y \in \mathbb{R}.$$

(iv) Finally, apply the deconvolution

$$\mathcal{F}(\beta')(z) = -\frac{\mathcal{F}(\Psi)(z)}{\mathcal{F}(\pi)(z)} = -\frac{\mathcal{F}(\Psi)(z) \mathcal{F}(\pi)(z)}{|\mathcal{F}(\pi)(z)|^2}, \quad z \in \mathbb{R},$$

and Fourier inversion to obtain an estimator $\beta'_{N,T}$ of $\beta'$. 

Estimation of the function $l$. Following the latter we first introduce kernel estimators of $\pi$ and $\pi'$. For any function $f : \mathbb{R} \to \mathbb{R}$ and $u > 0$ we use the standard notation $f_u(x) := u^{-1} f(u^{-1} x)$. Let $K$ be a smooth kernel of order $m \geq 2$, that is
\[ \int_{\mathbb{R}} K(x) dx = 1, \quad \int_{\mathbb{R}} x^j K(x) dx = 0 \quad j = 1, \ldots, m - 1, \quad \int_{\mathbb{R}} x^m K(x) dx \neq 0. \]
Let $h_i = h_{N,T}^i$, $i = 0, 1$, be two bandwidth parameters vanishing as $N, T \to \infty$. We define
\[ \pi_{N,T}(y) := \frac{1}{N} \sum_{i=1}^{N} K_{h_0}(y - Y_{T}^{i,N}), \quad \pi'_{N,T}(y) := \frac{1}{Nh_1} \sum_{i=1}^{N} K'_{h_1}(y - Y_{T}^{i,N}), \quad y \in \mathbb{R}. \]

Next, we introduce a threshold $\delta = \delta_{N,T} \to 0$ as $N, T \to \infty$ and set
\[ l_{N,T}(y) := \pi'_{N,T}(y) \mathbb{I}_{\{\pi_{N,T}(y) > \delta\}}, \quad y \in \mathbb{R}, \]
which is an estimator of the function $l$ given at (3.4).

Estimation of the parametric part. Recall the identity $l(y) = l(y, \alpha) - (\beta' \ast \pi)(y)$. Using $(\beta' \ast \pi)(y) \to 0$ as $|y| \to \infty$ since $\beta' \in L^1(\mathbb{R})$, we will construct the minimum contrast estimator for $\alpha$. More specifically, we introduce an integrable weight function $w$ with support $[\epsilon, 1]$ ($\epsilon \in (0, 1)$) and a parameter $U = U_{N,T} \to \infty$ as $N, T \to \infty$. For $\alpha \in \mathbb{R}^J$, we define
\[ \alpha_{N,T} := \arg \min_{\alpha \in \mathbb{R}^J} \int (l_{N,T}(y) - l(y, \alpha))^2 w_U(y) dy. \]

We can use the relations (3.3) to estimate the coefficients $\alpha = (a_1, \ldots, a_J)^T$, based on the empirical moments $m_{2j;N,T}$, $1 \leq j \leq J_1$, and the empirical Fourier moments $F(\Pi_{N,T}(\theta_j))$, $J_1 < j \leq J$, of the particle system:
\[ m_{k;N,T} := \frac{1}{N} \sum_{i=1}^{N} (Y_{T}^{i,N})^k, \quad k \in \mathbb{N}, \quad F(\Pi_{N,T})(z) := \frac{1}{N} \sum_{i=1}^{N} \exp(izY_{T}^{i,N}), \quad z \in \mathbb{R}. \]

By solving the corresponding linear systems we so construct estimates $\alpha_{N,T}$ for the coefficients $\alpha$.

Estimation of the nonparametric part. Given the estimator $\alpha_{N,T}$ introduced in the previous step, we define
\[ \Psi_{N,T}(y) = (l_{N,T}(y) - l(y, \alpha_{N,T})) \mathbb{I}_{\{|y| \leq U\}}, \quad y \in \mathbb{R}, \]
which provides an estimator of the function $\Psi = -\beta' \ast \pi$. In the next step we choose another threshold $\omega = \omega_{N,T} \to 0$ and introduce
\[ F(\beta'_{N,T})(z) := -\frac{F(\Psi_{N,T})(z) F(\Pi_{N,T})(z)}{|F(\Pi_{N,T})(z)|^2} \mathbb{I}_{\{|F(\Pi_{N,T})(z)| > \omega\}}, \quad z \in \mathbb{R}. \]

Finally, we use the Fourier inversion to estimate the function $\beta'$:
\[ \beta'_{N,T}(y) := \frac{1}{2\pi} \int \exp(-izy) F(\beta'_{N,T})(z) dz, \quad y \in \mathbb{R}. \]

In the following we will derive asymptotic properties of all estimators introduced in this section.

4. The asymptotic theory

We start our asymptotic analysis with the estimator $l_{N,T}$. Fix some $m > 0$. In the following the bandwidth and threshold parameters are chosen as
\[ h_0 = N_T^{-\frac{1}{m+1}}, \quad h_1 = N_T^{-\frac{1}{m+1}}, \quad \delta = \delta_0 \exp(-\bar{\alpha}_1 U^{2J_1}), \]

(4.1)
where $N_T$ is the rate introduced in (2.6), $m$ is the order of the kernel $K$ and $\delta_0 := (2Z_{\pi})^{-1} \exp(-\alpha_0 - \sum_{i,j} |\alpha_j| - \| \beta \|_{\infty})$, $\alpha_1 := \sum_{0 < j \leq J} \alpha_j$. Here and in what follows, $\| f \|_{\infty} := \sup_{y \in \mathbb{R}} | f(y) |$ for $f: \mathbb{R} \to \mathbb{R}$. Furthermore, we write $a_n \lesssim b_n$ when there exists a constant $C > 0$, independent of $n$, such that $a_n \leq C b_n$. Our first result is the following proposition.

**Proposition 4.1.** Let $\delta, h_i, i = 0, 1$, be defined as in (4.1) and $U \geq 1$. Then

$$\sup_{|y| \leq U} \mathbb{E} \left[ |l_{N,T}(y) - l(y)|^2 \right] \lesssim \exp(\alpha_1 U^{2J_1}) \left( N_T^{-m/(2m+2)} + U^{2J_1-1} N_T^{-m/(2m+1)} \right).$$

Here $\lesssim$ stands for inequality up to a constant that depends on $m$ and the kernel $K$.

**Proof.** See Section 7.

**Remark 4.2.** It may seem tempting to let $m \to \infty$. However, the constants in the above inequality typically grow in $m$. This is related to the fact that although $\pi$ is infinitely smooth, it is not necessarily analytic.

We observe that the upper bound in Proposition 4.1 grows exponentially in $U$, which will strongly affect convergence rates for all parameters of the model. We will now find the explicit expression for the estimator of $\alpha$. For this purpose, in (3.5) we write $l(y, \alpha) = (y/U)^T \cdot \alpha^U$, where

$$l(y) = -(\varphi_1(y), \ldots, \varphi_{J_1}(y), \varphi_{J_1+1}(y), \ldots, \varphi_J(y))^T, \quad y \in \mathbb{R},$$

and

$$\alpha^U = (\alpha_1 U, \alpha_2 U^3, \ldots, \alpha_{J_1} U^{2J_1-1}, \alpha_{J_1+1}, \alpha_{J_1+2}, \ldots, \alpha_J)^T.$$  

Then the unknown $\alpha^U$ and analogously scaled estimator $\tilde{\alpha}_N^U$ satisfy the identities

$$Q \tilde{\alpha}_N^{U,T} = \int l_{N,T}(y) l(y/U) w_U(y) dy, \quad Q \alpha^U = \int (l(y) + (\beta' \ast \pi)(y)) l(y/U) w_U(y) dy,$$

where

$$Q := \int l(y) l(y)^T w(y) dy \in \mathbb{R}^{J \times J}.$$

Notice that the components of the function $l$ are linearly independent on any interval $[s, t]$, $s < t$, since $U \theta_{J_1+1}, \ldots, U \theta_J$ are all distinct. In this case the matrix $Q$ is positive definite and hence invertible.

In the next proposition we derive convergence rates for the estimates $\tilde{\alpha}_N^{U,T}$ and $\Psi_{N,T}$. By $\| \cdot \|_2$ we denote the Euclidean norm.

**Proposition 4.3.** Let $\delta, h_i, i = 0, 1$, be defined as in (4.1) and $U \geq 1$. Then

$$\mathbb{E} \left[ \| \tilde{\alpha}_N^{U,T} - \alpha^U \|_2^2 \right]^{1/2} \lesssim \exp(\alpha_1 U^{2J_1}) \left( N_T^{-m/(2m+2)} + U^{2J_1-1} N_T^{-m/(2m+1)} \right)$$

$$+ \frac{\exp(-\alpha_{J_1} (U/2)^{2J_1})}{(U/2)^{2J_1}} + U^{-1} \int_{|y| > U/2} |\beta'(y)| dy,$$

$$\mathbb{E} \left[ \left| \Psi_{N,T}(y) - \Psi(y) \right|^2 dy \right]^{1/2} \lesssim \exp(\alpha_1 U^{2J_1}) U^{1/2} \left( N_T^{-m/(2m+2)} + U^{2J_1-1} N_T^{-m/(2m+1)} \right)$$

$$+ \frac{\exp(-\alpha_{J_1} (U/2)^{2J_1})}{(U/2)^{2J_1-1}} + U^{-1/2} \int_{|y| > U/2} |\beta'(y)| dy + \left( \int_{|y| > U/2} |\beta'(y)|^2 dy \right)^{1/2}.$$  

**Proof.** See Section 7.
We remark that the convergence rates for the parametric part of the model are logarithmic, which is rather unusual for parametric estimation problems. Indeed, the logarithmic rate is obtained under a proper choice of $U$ when there exists $q > 0$ such that $0 < \lim \inf_{y \to \infty} F_{q,\beta'}(y) \leq \lim \sup_{y \to \infty} F_{q,\beta'}(y) < \infty$, where $F_{q,\beta'}(y) = y^q \int_y^\infty |\beta'(u)| du$, $y > 0$. We believe that the reason for such a slow convergence rate is a convolution type structure of the invariant density $\pi$.

**Remark 4.4.** The statement of Proposition 4.3 can be transferred from $\alpha$ to the original parameter $\beta$ via the identities (3.3). Indeed, we have $\alpha = B\alpha$ for some $J \times J$ matrix $B$ (given all $F(\pi)(\theta_j) \neq 0$) and construct the estimator $\alpha_{N,T} = B_{N,T}\alpha_{N,T}$ correspondingly. To get the rate of $\mathbb{E}[\|V(\alpha_{N,T} - \alpha)\|_2^2]$, where $V = \text{diag}(v_1, \ldots, v_J)$ with $v_j = U^{2j-1}$, $j \leq J_1$, and $v_j = 1$, $j > J_1$, we can use the decomposition $\alpha_{N,T} - \alpha = (B_{N,T} - B)\alpha_{N,T} + B(\alpha_{N,T} - \alpha)$. Since $B$ is upper triangular we have $V'BV^{-1}$ bounded, whereas Proposition 4.3 applies to $V(\alpha_{N,T} - \alpha) = \alpha_{U,N,T} - \alpha_U$. To deal with the first term in the decomposition, in particular, $B_{N,T} - B$, we can use Theorem 2.1 and the moment bounds from [5, 21], whence $\mathbb{E}[|m_{2j;N,T} - m_{2j}|^2] \lesssim N_T^{-1}$ since $X_2$ has all moments in our setting. At last, the convergence rates for the original parameters $\alpha$ stay logarithmic as those obtained for $\alpha$ under a proper choice of $U$ when the integrated tail of $\beta'$ has a power-like decay.

It is evident from Proposition 4.3 that the rate of convergence for $\Psi_{N,T}$ crucially depends on the tail behaviour of the function $\beta'$. The next corollary, which is an immediate consequence of the previous result, gives the precise bound.

**Corollary 4.5.** In the setting of Proposition 4.3 assume that there exists $p > 0$ such that $\lim \inf_{y \to \infty} \Phi_{p,\beta'}(y) > 0$, where

$$\Phi_{p,\beta'}(y) := y^{p-(1/2)} \int_y^\infty \frac{|\beta'(u)|}{y^p} \left(\int_y^\infty |\beta'(u)|^2 du\right)^{1/2}, \quad y > 0.$$  

Choose $U = (c \log N_T)^{1/(2J_1)}$ for some $0 < c < m/(2(m+2)\bar{\alpha}_1)$. Then

$$\mathbb{E} \left[ \int_R |\Psi_{N,T}(y) - \Psi(y)|^2 dy \right] \lesssim (\log N_T)^{-p/J_1} \Phi_{p,\beta'}((c \log N_T)^{1/(2J_1)} \epsilon/2).$$

Moreover, if $\lim \sup_{y \to \infty} \Phi_{p,\beta'}(y) < \infty$, then

$$\mathbb{E} \left[ \int_R |\Psi_{N,T}(y) - \Psi(y)|^2 dy \right] \lesssim (\log N_T)^{-p/J_1}.$$  

We see that the convergence rate in (4.3) depends on the nonparametric part of the model through the parameter $p$ and on the highest degree polynomial in the parametric part through $J_1$ (in contrast, the trigonometric part of the model does not influence the convergence rate). This phenomenon will translate to the estimation of $\beta'$.

The following proposition will play a crucial role for the analysis of the estimator $\beta'_{N,T}$.

**Proposition 4.6.** In the setting of Proposition 4.3 assume $\beta'' \in L^2(\mathbb{R})$. Then

$$\mathbb{E} \left[ \int_R |\beta'_{N,T}(y) - \beta'(y)|^2 dy \right] \lesssim \omega^{-2} \left( \mathbb{E} \left[ \int_R |\Psi_{N,T}(y) - \Psi(y)|^2 dy \right] + N_T^{-1} \right)$$

$$\quad + N_T^{-1} \int_{|\mathcal{F}(\pi)(z)| > 2\omega} |\mathcal{F}(\beta')(z)|^2 |\mathcal{F}(\pi)(z)|^{-2} z^2 dz$$

$$\quad + \int_{|\mathcal{F}(\pi)(z)| \leq 2\omega} |\mathcal{F}(\beta')(z)|^2 dz.$$  

**Proof.** See Section 7.

We observe that, up to the presence of the factor $\omega^{-2} \to \infty$ which can be chosen arbitrarily, the convergence rate for the nonparametric part $\beta'$ is transferred from Proposition 4.3. Indeed, the next result shows that the second and the third terms in (4.4) are negligible under appropriate technical assumption on $\beta'$, which depends on the unknown $\lambda$ in (3.2). More specifically, this assumption implies that $\mathcal{F}(\beta')$ has compact support, bounded in terms of $\lambda$. An application of Lemma 7.2 shows $\mathcal{F}(\pi)$ is bounded from below.

**Proposition 4.7.** In the setting of Proposition 4.6 assume that $\beta'$ is an entire function of the first order and type less than $\vartheta > 0$, i.e.

$$|\beta'(z)| \leq A \exp(\vartheta |z|), \quad z \in \mathbb{C},$$  

then
with $A > 0$ and $\theta \leq \lambda^{1/2}$. Let $\liminf_{y \to \infty} \Phi_{p,\beta}(y) > 0$ for $p > 0$. Choose $U = (C \log N_T)^{1/(2J)}$ for some $0 < C < m/(2(m+2)\bar{\alpha}_1)$. Then

$$\mathbb{E} \left[ \int_{\mathbb{R}} |\beta_{N,T}(y) - \beta'(y)|^2 dy \right] \leq \omega^{-2}(\log N_T)^{-p/J} \Phi^2_{p,\beta}((C \log N_T)^{1/(2J)} + 2).$$

Moreover, if $\limsup_{y \to \infty} \Phi_{p,\beta}(y) > 0$, then

$$\mathbb{E} \left[ \int_{\mathbb{R}} |\beta_{N,T}(y) - \beta'(y)|^2 dy \right] \leq \omega^{-2}(\log N_T)^{-p/J}.$$

**Proof.** See Section 7. □

**Example 4.8.** As an example of functions satisfying the conditions of Proposition 4.7 we may consider

$$\beta_1(y) = \frac{1 - \cos(by)}{y^2}, \quad b > 0, \quad \beta_2(y) = \frac{\sin^2k(y)}{y^{2k}}, \quad k \in \mathbb{N}.$$  

One can easily check that both function fulfill the necessary assumptions with $p = 3/2$ in case of $\beta_1$ and $p = 2k - 1/2$ in case of $\beta_2$. □

**Remark 4.9.** We remark that the bandwidth parameters $h_1$ and $h_2$ (as well as the thresholds $\delta$ and $U$) solely depend on the rate $N_T$, which however cannot be computed explicitly as it involves an unknown parameter $\lambda$. In practice, assuming that $\lambda > 0$ is not too small, the rate $N^{-1}$ will dominate $\exp(-\lambda T)$ and one would simply use $N_T = N$. □

### 5. Lower bound for the nonparametric component

In this section we will derive the lower bound for the estimation of the nonparametric component $\beta'$. We consider the simple model with i.i.d. observations $Y_1, \ldots, Y_N$ having the following density:

$$\pi_\beta(y) = Z_{\pi_\beta}^{-1} \exp \left( - \left( \sum_{j=1}^{J} a_j \phi_j + \beta \right) * \pi_\beta \right)(y), \quad y \in \mathbb{R},$$

where

$$Z_{\pi_\beta} = \int_{\mathbb{R}} \exp \left( - \left( \sum_{j=1}^{J} a_j \phi_j + \beta \right) * \pi_\beta \right)(y) dy$$

and $\phi_j(y) = y^{2j}, \quad y \in \mathbb{R}, \quad 1 \leq j \leq J$, for given $J \geq 1$. We assume that the constants $a_1 > 0, a_2 \geq 0, \ldots, a_{J-1} \geq 0, a_J > 0$ are known, the function $\beta(y)$ is even and such that $\phi''(y) \geq \lambda, \quad y \in \mathbb{R}$, with known $\lambda > 0$. We remark that the nonparametric estimation problem is similar in spirit to the classical deconvolution problem, but we can not apply the same techniques to derive the lower bound.

We consider the following class of functions

$$\mathcal{F}_{p,C,C_0,a,\lambda} := \left\{ f \in C^2_b(\mathbb{R}) : \|f\|_{\infty} \leq C_0, \quad \|f'\|_{\infty} \leq C_1, \quad \inf_{y \in \mathbb{R}} f'''(y) \geq -C_2, \quad \limsup_{y \to \infty} y^{2p} \int_{y}^{\infty} |f'(u)|^2 du \leq C \right\},$$

where $C^2_b(\mathbb{R})$ denotes the space of twice continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that $f, f', f''$ are bounded. Moreover, $C, C_0 > 0$ and

$$C_1 := \lambda^{1/2} \left( 1 - \sum_{1 < j \leq J} 2j c_j a_j / \lambda^j \right) > 0, \quad C_2 := 2a_1 - \lambda > 0,$$

where $c_j^2 := 2(2(j-1))!/(2j-1)!$. $1 < j \leq J$. We remark that the condition $C_1 > 0$ holds whenever $\lambda > 0$ is sufficiently large. We also note that assumption (A) is automatically satisfied when $\beta \in \mathcal{F}_{p,C,C_0,a,\lambda}$. Indeed, $\mathcal{F}_{p,C,C_0,a,\lambda}$ is chosen in accordance with assumptions of our estimation procedure; in particular, it reflects the tail behaviour of the function $\beta'$. The main result of this section is the minimax bound over the functional class $\mathcal{F}_{p,C,C_0,a,\lambda}$.  

Theorem 5.1. Denote the law of $Y_1$ by $\Pi \beta$, and consider the functional class $F_{P,C,X_0,a,\lambda}$ for $p > 1/2$. Then there exists a constant $c_0 > 0$ (depending on $p$, $C$, $C_0$, $a$, $\lambda$) such that
\[
\inf_{\beta_N} \sup_{\beta \in F_{P,C,X_0,a,\lambda}} \Pi_{\beta}^\otimes N \left( \| \beta_N' - \beta' \|_{L^2(\mathbb{R})}^2 > c_0 (\log N)^{-p/J} \right) > 0.
\]

Proof. See Section 7. □

We remark that our estimator $\beta_N, T$ matches the lower bound up to the factor $\omega^-2$, which can be chosen to diverge to $\infty$ at an arbitrary slow rate.

Remark 5.2. The result of Theorem 5.1 can be compared to a classical deconvolution problem. Consider a model
\[
Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, N,
\]
where $(X_i)_{i \geq 1}$ and $(\varepsilon_i)_{i \geq 1}$ are mutually independent i.i.d. sequences. Assume that $X_1$ (resp. $\varepsilon_1$) has a Lebesgue density $f$ (resp. $g$), and we are in the super smooth setting, that is,
\[
F(g)(z) \sim \exp(-\text{const} \cdot |z|^{2J}), \quad |z| \to \infty.
\]
When the density $f$ satisfies the condition $\int_{\mathbb{R}} |F(f)(z)|^2 |z|^{2p} dz \leq C$, $p > 1/2$, it is well known that the minimax rate for the estimation of the density $f$ becomes $(\log N)^{-p/J}$ (see e.g. Theorem 2.14(b) in [25]). While in the classical deconvolution problem the assumptions are imposed in the Fourier domain, we have comparable assumptions on the functions themselves. Notice that due to the structure of the model $\pi$ plays the role of the noise density and the condition on $F(g)$ can be compared to the decay of $\pi_{\beta}$, which is determined by the leading polynomial of degree $J$. On the other hand, the integral condition on $F(f)$ is related to the corresponding tail condition on $\beta'$.

6. Conclusions and outlook

In this work we study the problem of estimating the drift of a MV-SDE based on observations of the corresponding particle system. We propose a kernel-type estimator and provide theoretical analysis of its convergence. In particular, for the nonparametric part of the model, we derive minimax convergence rates and show rate-optimality of our proposed estimator. As a promising future research direction, one can consider the case of continuous time observations and high-dimensional MV-SDEs with general form of the drift function. In another direction the problem of estimating diffusion coefficient remains completely open.

7. Proofs

7.1. Preliminary results

Lemma 7.1. Set $\tilde{\varphi}_1(y) := \sum_{0 \leq j \leq 4} \alpha_j \varphi_j(y)$. Then $\pi \in C^\infty(\mathbb{R})$ and for every $n$,
\[
|\pi^{(n)}(y)| \lesssim (1 + |\varphi_1'(y)|)^n \exp(-\varphi_1'(y)), \quad y \in \mathbb{R}.
\]

Proof. We decompose $\tilde{\varphi} = \varphi \ast \pi$ as $\tilde{\varphi} = \alpha_0 + \tilde{\varphi}_1 + \tilde{\varphi}_2 + \beta \ast \pi$. Here $\tilde{\varphi}_2 := \sum_{1 \leq j \leq 4} \alpha_j \varphi_j$ is bounded, and $\|\beta \ast \pi\|_\infty \leq \|\beta\|_\infty \|\pi\|_{L^1(\mathbb{R})} < \infty$. Hence, we obtain
\[
\pi(y) = Z \pi^{-1} \exp(-\tilde{\varphi}(y)) \lesssim \exp(-\tilde{\varphi}_1(y)), \quad y \in \mathbb{R}.
\]

We now consider its derivative. Since $\beta$ has a bounded derivative, so does $\beta \ast \pi$. We obtain $\pi' = -\tilde{\varphi}' \pi$, where $\tilde{\varphi}' = \tilde{\varphi}_1' + \tilde{\varphi}_2' + \beta' \ast \pi$ satisfies
\[
|\tilde{\varphi}'(y)| \lesssim 1 + |\varphi_1'(y)|, \quad y \in \mathbb{R}.
\]
That is, statement of the proposition holds for $n = 1$. For $n \geq 1$ it follows by induction using
\[
\pi^{(n+1)} = (\pi^n)' = - \sum_{k=0}^{n} \binom{n}{k} \tilde{\varphi}^{(k+1)} \pi^{(n-k)}.
\]
with \( \varphi^{(k+1)} = \varphi_1^{(k+1)} + \varphi_2^{(k+1)} + (\beta' \ast \pi)^{(k)} \), where \((\beta' \ast \pi)^{(k)} = \beta' \ast \pi^{(k)}\) is bounded when \(\|\pi^{(k)}\|_\infty < \infty, \|\beta'\|_{L^1(\mathbb{R})} < \infty\).

**Lemma 7.2. Moments of the density \(\pi\) in (2.5) satisfy**

\[
m_{2k} \leq (2k - 1)!/\lambda^k, \quad k \in \mathbb{N}.
\]

**Proof.** Set

\[
I_k := \int_{\mathbb{R}} y^{2k-1} (\varphi' \ast \pi)(y) \pi(y) dy.
\]

Since \(\varphi(y)\) is even, we have \(\varphi'(y) = -\varphi'(-y)\), implying

\[
2I_k = \int \int (y^{2k-1} - x^{2k-1}) \varphi'(y - x) \pi(x) \pi(y) dx dy.
\]

By the mean value theorem we conclude that

\[
\varphi'(y - x) = \varphi'(y - x) - \varphi'(0) = (y - x) \varphi''(z)
\]

for some \(z \in (x, y)\), and by the convexity assumption,

\[
2I_k \geq \lambda \int \int (y^{2k-1} - x^{2k-1}) (y - x) \pi(x) \pi(y) dx dy = 2\lambda m_{2k}.
\]

On the other hand,

\[
I_k = - \int y^{2k-1} d\pi(y) = (2k - 1) \int y^{2k-1} \pi(y) dy = (2k - 1) m_{2(k-1)}.
\]

We conclude that \(m_{2k} \leq m_{2(k-1)} (2k - 1)/\lambda\). Induction provides the desired result. \(\square\)

### 7.2. Proof of Proposition 4.1

Using

\[
|l_{N,T}(y) - l(y)| \leq \left| \frac{\pi'_{N,T}(y)}{\pi_{N,T}(y)} - \frac{\pi'(y)}{\pi_{N,T}(y)} \right| + \left| \frac{\pi'(y)}{\pi_{N,T}(y)} - \frac{\pi'(y)}{\pi(y)} \right|
\]

on \(\pi_{N,T}(y) > \delta\), we get

\[
|l_{N,T}(y) - l(y)| \leq \delta^{-1}|\pi'_{N,T}(y) - \pi'(y)| + \delta^{-1}|l(y)||\pi(y) - \pi_{N,T}(y)| + |l(y)||1_{\{\pi_{N,T}(y) \leq \delta\}}|
\]

Here using \(|l(y)| \leq 2 \sum_{0 < j \leq J_1} j \alpha_j |y|^{2j-1} + \sum_{j_1 < j \leq J} \theta_j |\alpha_j| + \|\beta\|_{\infty}\) we get

\[
|l(y)| \lesssim U^{2J_1-1}
\]

for all \(|y| \leq U\). Since \(|(\varphi \ast \pi)(y)| \leq \sum_{0 < j \leq J_1} \alpha_j y^{2j} + \sum_{j_1 < j \leq J} |\alpha_j| + \|\beta\|_{\infty}\), the chosen \(\delta\) satisfies

\[
2\delta \leq Z_-^{-1} \exp(-|\varphi \ast \pi(y)|) = \pi(y)
\]

for all \(|y| \leq U\). Hence, it follows that for all \(|y| \leq U\),

\[
\mathbb{P}(\pi_{N,T}(y) \leq \delta) = \mathbb{P}(\pi(y) - \pi_{N,T}(y) \geq \pi(y) - \delta) \leq \mathbb{P}(\|\pi - \pi_{N,T}\|_{\infty} \geq \delta) \leq \delta^{-2} \mathbb{E}[\|\pi - \pi_{N,T}\|_{\infty}^2].
\]

Now, with \(\pi_{N,T}(y) = (K_{h_0} \ast \Pi_{N,T})(y)\), we have

\[
|(K_{h_0} \ast (\Pi_{N,T} - \Pi))(y)| \leq \text{Lip}(K_{h_0}(y - \cdot)) W_1(\Pi_{N,T}, \Pi) = h_0^{-1} \text{Lip}(K) W_1(\Pi_{N,T}, \Pi),
\]
where \( \mathbb{E}[W_T^2(\Pi_{N,T}, \Pi)] \leq N_T^{-1} \) due to (2.6). Furthermore, we have

\[
(K_{h_0} \ast \Pi)(y) - \pi(y) = \int K(x)(\pi(y + xh_0) - \pi(y))dx,
\]

where \( \pi \in C^\infty(\mathbb{R}) \) satisfies \(|\pi^{(n)}(y)| \lesssim (1 + |y|^{2J-1})\exp(-\alpha_J y^{2J_1})\), \( y \in \mathbb{R}, \ n \in \mathbb{N} \), by Lemma 7.1. Using a Taylor expansion of \( \pi \) and that the kernel \( K \) is of order \( m \), we obtain

\[
(K_{h_0} \ast \Pi)(y) - \pi(y) = \int K(x)R_{m-1}(y + xh_0)dx,
\]

where

\[
|R_{m-1}(y + xh_0)| \lesssim |x|^m h_0^m
\]

and so

\[
|(K_{h_0} \ast \Pi)(y) - \pi(y)| \lesssim h_0^m
\]

uniformly in \( y \in \mathbb{R} \). Similarly,

\[
|h_1^{-1}(K_{h_1} \ast \Pi)(y) - \pi'(y)| \lesssim h_1^m,
\]

because

\[
h_1^{-1}(K'_{h_1} \ast \Pi)(y) = ((K_{h_1})' \ast \Pi)(y) = \int K(x)\pi'(y + xh_1)dx.
\]

We thus deduce that

\[
\mathbb{E}[\|\pi_N - \pi\|^2_\infty] \lesssim N_T^{-1} h_0^{-2} + h_0^{2m}, \quad \mathbb{E}[\|\pi_N' - \pi'\|^2_\infty] \lesssim N_T^{-1} h_1^{-4} + h_1^{2m},
\]

where our choice of \( h_0, h_1 \) yields the optimal rate in upper bounds:

\[
\mathbb{E}[\|\pi_N - \pi\|^2_\infty] \lesssim N_T^{-m}, \quad \mathbb{E}[\|\pi_N' - \pi'\|^2_\infty] \lesssim N_T^{-m/2}.
\]

7.3. Proof of Proposition 4.3

Recall the definition of the matrix \( Q \in \mathbb{R}^{J \times J} \) introduced at (4.2). Since \( Q \) is invertible, we deduce that

\[
\mathbb{E} \left[ \|\alpha_{N,T} - \alpha\|_U^2 \right] \leq \mathbb{E} \left[ \|l_{N,T}(y) - l(y) - (\beta' \ast \pi)(y)\|^2 \right]^{1/2} \mathbb{E} [\|\bar{w}\|_{L^1(\mathbb{R})}] 
\]

where \( \bar{w}(y) := \|Q^{-1}l(y)\|_2 |w(y)|, \ y \in \mathbb{R} \). Moreover, \( \|\bar{w}\|_{L^1(\mathbb{R})} < \infty \) and \( \|\bar{w}\|_\infty < \infty \) are uniformly bounded in \( U \). Hence,

\[
\mathbb{E} \left[ \|\alpha_{N,T} - \alpha\|_U^2 \right] \leq \sup_{|y| \leq U} \mathbb{E} \left[ \|l_{N,T}(y) - l(y)\|^2 \right]^{1/2} + U^{-1} \int_{|y| \leq U} (|\beta' \ast \pi|(y))dy.
\]

As for the last term, we have

\[
\int_{|y| \leq U} (|\beta' \ast \pi|(y))dy \leq \int_{|y| \leq U} \int |\beta'(y - x)| \pi(x)dx dy
\]

\[
= \left( \int_{-\infty}^{U/2} + \int_{U/2}^{\infty} \right) \left( \int_{U-x}^{\infty} |\beta'(y)|dy \right) \pi(x) dx \leq \int_{U/2}^{\infty} |\beta'(y)|dy \leq \|\beta'\|_{L^1(\mathbb{R})} \int_{U/2}^{\infty} \pi(x) dx.
\]

Finally, note that

\[
\int_{u}^{\infty} \pi(x) dx \lesssim \int_{u}^{\infty} \pi_1(x) dx,
\]
where \( \pi_1(x) = \exp(-\alpha_J x^{2J_1}) \) satisfies \( \pi'_1(x) = -2J_1 \alpha_J x^{2J_1-1} \pi_1(x) \) and so,
\[
\int_u^\infty \pi_1(x) dx \leq \frac{1}{u^{2J_1-1}} \int_u^\infty x^{2J_1-1} \pi_1(x) dx = \frac{\pi_1(u)}{2J_1 \alpha_J u^{2J_1-1}}.
\]

Now we consider the bound for
\[
\Psi_{N,T}(y) = (l_{N,T}(y) - l(y, \alpha_{N,T})) \mathbb{I}_{|y| \leq \epsilon U}, \quad y \in \mathbb{R},
\]
with
\[
\Psi(y) = l(y) - l(y, \alpha) = -(\beta' \ast \pi)(y), \quad y \in \mathbb{R}.
\]
We have
\[
\int_{|y| \leq \epsilon U} \mathbb{E}[|\Psi_{N,T}(y) - \Psi(y)|^2] dy \leq \epsilon U \sup_{|y| \leq \epsilon U} \mathbb{E}[|\Psi_{N,T}(y) - \Psi(y)|^2],
\]
where
\[
\sup_{|y| \leq \epsilon U} \mathbb{E}[|\Psi_{N,T}(y) - \Psi(y)|^2] \lesssim \sup_{|y| \leq \epsilon U} \mathbb{E}[|l_{N,T}(y) - l(y)|^2] + \mathbb{E}[\|\alpha_{N,T}^l - \alpha^U\|_2^2].
\]
Finally, we deduce
\[
\left( \int_{|y| > \epsilon U} |\Psi(y)|^2 dy \right)^\frac{1}{2} \leq \left( \int_{|y| > \frac{\epsilon U}{2}} |\beta'(y-x)|^2 dy \right)^\frac{1}{2} \pi(x) dx
\]
\[
\leq \left( \int_{|y| > \frac{\epsilon U}{2}} |\beta'(y)|^2 dy \right)^\frac{1}{2} + \|\beta'\|_{L^2(\mathbb{R})} \int_{|x| > \frac{\epsilon U}{2}} \pi(x) dx,
\]
which completes the proof of Proposition 4.3.

7.4. Proof of Proposition 4.6

Define a function \( \beta_{N,T}^+ \) via the formula
\[
\mathcal{F}(\beta_{N,T}^+)(z) = \mathcal{F}(\beta')(z) \mathbb{I}_{|\mathcal{F}(\Pi_{N,T})(z)| > \omega}.
\]
Write \( \mathcal{F}(\Pi) = \mathcal{F}(\pi) \). Use the identity
\[
\mathcal{F}(\beta_{N,T}^+ - \beta_{N,T}^-)(z) = (-\mathcal{F}(\Psi_{N,T} - \Psi)(z) + \mathcal{F}(\beta')(z) \mathcal{F}(\Pi - \Pi_{N,T})(z)) \frac{\mathcal{F}(\Pi_{N,T})(z)}{|\mathcal{F}(\Pi_{N,T})(z)|^2} \mathbb{I}_{|\mathcal{F}(\Pi_{N,T})(z)| > \omega},
\]
where by the Kantorovich–Rubinstein dual formulation, we have
\[
|\mathcal{F}(\Pi - \Pi_{N,T})(z)| \leq |z| W_1(\Pi_{N,T}, \Pi).
\]
As a result,
\[
\mathbb{E} \left[ \int_{\mathbb{R}} |\beta_{N,T}^+(y) - \beta_{N,T}^-(y)|^2 dy \right]^{1/2} \leq \omega^{-1} \mathbb{E} \left[ \int_{\mathbb{R}} |\Psi_{N,T}(y) - \Psi(y)|^2 dy \right]^{1/2}
\]
\[
+ \omega^{-1} \left( \mathbb{E}[W_1^2(\Pi_{N,T}, \Pi)] \int_{\mathbb{R}} |\beta''(y)|^2 dy \right)^{1/2}.
\]
Furthermore, it holds that
\[
\mathbb{E} \left[ \int_{\mathbb{R}} |\beta_{N,T}^+(y) - \beta'(y)|^2 dy \right] = \frac{1}{2\pi} \left( \int_{|\mathcal{F}(\pi)(z)| > 2\omega} + \int_{|\mathcal{F}(\pi)(z)| \leq 2\omega} \right) |\mathcal{F}(\beta')(z)|^2 \mathbb{P}(|\mathcal{F}(\Pi_{N,T})(z)| \leq \omega) dz.
\]
If $\omega < |\mathcal{F}(\pi)(z)|/2$, then we have by the Markov and Jensen’s inequalities
\[
\mathbb{P}(|\mathcal{F}(\Pi_{N,T})(z)| \leq \omega) \leq \mathbb{P}(|\mathcal{F}(\Pi_{N,T} - \Pi)(z)| \geq |\mathcal{F}(\pi)(z)| - \omega) \leq \frac{\mathbb{E}||| \mathcal{F}(\Pi_{N,T} - \Pi)(z) ||^2}{(|\mathcal{F}(\pi)(z)| - \omega)^2}.
\]
Consequently, we obtain
\[
\mathbb{E} \left[ \int R |\beta_{N,T}^*(y) - \beta^*(y)|^2dy \right] \leq \frac{1}{2\pi} \left( \int |\mathcal{F}(\pi(z)| > 2\omega |\mathcal{F}(\beta^*)(z)|^2 \mathbb{E}W_2^2(\Pi_{N,T}, \Pi)|z|^2 |\mathcal{F}(\pi)(z)|^2/4 dz + \int |\mathcal{F}(\beta^*)(z)|^2dz, \right)
\]
which completes the proof of Proposition 4.6.

7.5. Proof of Proposition 4.7

We use Proposition 4.3. It suffices to show that on the r.h.s. of (4.4) the first term dominates. For this purpose, we decompose each of the last two terms into two integrals over $|z| \leq \vartheta$ and $|z| > \vartheta$, respectively. For all $|z| > \vartheta$, it holds $|\mathcal{F}(\pi)(z)| \geq 1/2$. Indeed, since $\pi$ is an even density, we have $\mathcal{F}(\pi)(z) = 1 + \int (\exp(izy) - 1 - izy)\pi(y)dy$. Using $|\exp(iz) - 1| \leq |z|^2/2, x \in \mathbb{R}$, we get $|\mathcal{F}(\pi)(z)| \geq 1 - m_2|z|^2/2$. By Lemma 7.2, we have $m_2 \leq 1/\lambda \leq 1/\vartheta^2$. On the other hand, the Paley–Wiener theorem implies that $\mathcal{F}(\beta^*)(z)$ vanishes for $|z| > \vartheta$. Finally, it remains
\[
\mathbb{E} \left[ \int R |\beta_{N,T}^*(y) - \beta^*(y)|^2dy \right] \lesssim \omega^{-2} \left( \mathbb{E} \left[ |\Psi_{N,T}(y) - \Psi(y)|^2dy \right] + 1/N_T \right) + \vartheta^2/N_T,
\]
where Corollary 4.5 provides an upper bound on the dominating term.

7.6. Proof of Theorem 5.1

We will use the classical two hypotheses approach presented in the monograph [31]. More specifically, we will find functions $\beta_0, \beta_1 \in \mathcal{F}_{p,c,c_0,a_0}$ such that
\[
||\beta_0 - \beta_1||_{L^2(\mathbb{R})} = \text{const} \cdot (\log N)^{-p/2} \quad \text{and} \quad K(\Pi_{\beta_1}, \Pi_{\beta_0}) \leq 1,
\]
where $K(\Pi_{\beta_1}, \Pi_{\beta_0})$ denotes the Kullback-Leibler divergence.

We start with some preliminary estimates. Let $\beta \in \mathcal{F}_{p,c,c_0,a_0}$. Due to the inequality $(\log \pi_\beta)''(y) \leq -2a - (\beta'' + \pi_\beta)(y) \leq -\lambda$ the probability measure $\Pi_\beta$ satisfies the logarithmic Sobolev inequality with constant $2/\lambda$:
\[
\text{Ent}_{\Pi_\beta}(f^2) \leq \frac{2}{\lambda} \int |f(y)|^2\Pi_\beta(dy),
\]
where
\[
\text{Ent}_{\Pi_\beta}(f^2) := \int f^2(y) \log f^2(y)\Pi_\beta(dy) - \int f^2(y)\Pi_\beta(dy) \log \left( \int f^2(y)\Pi_\beta(dy) \right),
\]
for every smooth function $f : \mathbb{R} \to \mathbb{R}$ with $\int R |f'(y)|^2\Pi_\beta(dy) < \infty$. Hence, for any $\beta_0, \beta_1 \in \mathcal{F}_{p,c,c_0,a_0}$, we can bound the Kullback-Leibler divergence as
\[
K(\Pi_{\beta_1}, \Pi_{\beta_0}) := \int \pi_{\beta_1}(y) \log \frac{\pi_{\beta_1}(y)}{\pi_{\beta_0}(y)} dy \leq \frac{1}{2\lambda} \int \pi_{\beta_1}(y)|g(y)|^2 dy
\]
with a function
\[
g = \frac{\pi'_{\beta_1}}{\pi_{\beta_1}} - \frac{\pi'_{\beta_0}}{\pi_{\beta_0}} = \left( \sum_{1 \leq j \leq J} a_j \varphi_j + \beta_0 \right)' \pi_{\beta_0} - \left( \sum_{1 \leq j \leq J} a_j \varphi_j + \beta_1 \right)' \pi_{\beta_1}.
\]
We further decompose it as $g = \sum_{1 \leq j \leq J} a_j g_j + g_1 + g_0$ with
\[
g_j := \varphi_j' \pi_{\beta_0} - \pi_{\beta_1}, \quad g_1 := \beta_1' \pi_{\beta_0} - \pi_{\beta_1}, \quad g_0 := (\beta_0' - \beta_1') \pi_{\beta_0}.
\]
In general, it seems hard to assess the functions \( g_j \) for \( j \geq 1 \) directly (in contrast to \( g_0 \)). Instead, we will show that

\[
\frac{1}{2\lambda} \int_R \pi_{\beta_1}(y) g(y) dy - g_0(y) \leq \frac{\gamma^2}{\lambda} K(\Pi_{\beta_1}, \Pi_{\beta_0})
\]

for some \( \gamma \in (0, 1) \), and as a consequence of the inequality (7.2) we deduce that

\[
K(\Pi_{\beta_1}, \Pi_{\beta_0}) \leq \int_R \pi_{\beta_1}(y) g_0(y) dy.
\]

The latter bound will be estimated directly for a proper choice of functions \( \beta_0, \beta_1 \).

We proceed with showing (7.4). We will find a constant \( \gamma_j > 0 \) such that

\[
\frac{1}{2\lambda} \int_R \pi_{\beta_1}(y) g_j(y) dy \leq \gamma_j^2 K(\Pi_{\beta_1}, \Pi_{\beta_0}), \quad 1 \leq j \leq J.
\]

Since \( \| g_1 \|_\infty \leq \| f_1' \|_\infty \| \pi_{\beta_0} - \pi_{\beta_1} \|_L^1(\mathbb{R}) \) and \( \| \pi_{\beta_0} - \pi_{\beta_1} \|_L^1(\mathbb{R}) \leq \sqrt{2K(\Pi_{\beta_1}, \Pi_{\beta_0})} \), we have

\[
\gamma_j^2 = \frac{\| f_1' \|_\infty^2}{\lambda}.
\]

For \( 1 < j \leq J \), we have

\[
|g_j(y)|^2 \leq (2j)^2 \int_R (y - x)^{2(2j - 1)} \left( \sqrt{\pi_{\beta_1}(x)} + \sqrt{\pi_{\beta_0}(x)} \right)^2 dx \cdot \| \pi_{\beta_1} - \pi_{\beta_0} \|_{L^2(\mathbb{R})}^2
\]

where \( \| \pi_{\beta_1} - \pi_{\beta_0} \|_{L^2(\mathbb{R})}^2 \leq K(\Pi_{\beta_1}, \Pi_{\beta_0}) \) and \( \| \pi_{\beta_1} + \pi_{\beta_0} \|_2^2 \leq 2(\pi_{\beta_1} + \pi_{\beta_0}) \). Furthermore,

\[
\int_R \pi_{\beta_1}(y) \int_R (y - x)^{2k} \pi_{\beta_1}(x) dx dy = \sum_{j=0}^k \binom{2k}{2j} m_j^{\pi_{\beta_1}} m_{2(k-j)}^{\pi_{\beta_1}} \leq \frac{C_k}{\lambda^k},
\]

by Lemma 7.2 with \( m_j^{\pi_{\beta_1}} = \int_R y^j \pi_{\beta_1}(y) dy, \ i = 0, 1, \) and

\[
C_k = \sum_{j=0}^k \binom{2k}{2j} (2j - 1)!(2(k - j) - 1)! = \frac{(2k)!}{k!2^k} \sum_{j=0}^k \binom{k}{j} = \frac{(2k)!}{k!}.
\]

Hence, (7.5) holds true with

\[
\gamma_j^2 = \frac{2(2j)^2 C_{2j-1}}{\lambda^{2j}}, \quad 1 < j \leq J.
\]

We conclude that

\[
(K(\Pi_{\beta_1}, \Pi_{\beta_0}))^{1/2} \leq \left( \sum_{1 < j \leq J} a_j \gamma_j + \gamma_1 \right) (K(\Pi_{\beta_1}, \Pi_{\beta_0}))^{1/2} + \left( \frac{1}{2\lambda} \int_R |g_0(y)|^2 \pi_{\beta_1}(y) dy \right)^{1/2}.
\]

Consequently, we deduce the inequality

\[
K(\Pi_{\beta_1}, \Pi_{\beta_0}) \leq \frac{1}{2\lambda(1 - \sum_{1 < j \leq J} a_j \gamma_j - \gamma_1)^2} \int_R |g_0(y)|^2 \pi_{\beta_1}(y) dy.
\]

Due to (7.6) we only need to handle the last term in (7.3). For this purpose we use following construction: We introduce the constants \( \rho > 0, \ M > 0 \) (\( \rho \to 0 \) and \( M \to \infty \) to be chosen later) and a function \( \phi \in C^2(\mathbb{R}) \) with

\[
\text{supp}(\phi) = [-2, -1] \cup [1, 2],
\]

and set

\[
\beta_0(y) = f_0(y), \quad \beta_1(y) = f_0(y) + \rho M \phi(y/M).
\]
On the other hand, we get $f_0 \in F_{p,C,C_0/2,(a_1/2,\ldots,a_J/2)},\lambda/2 \subset F_{p,C,C_0,a,\lambda}$. To ensure that $\beta_1 \in F_{p,C,C_0,a,\lambda}$ we assume

$$\rho M \|\phi\|_{\infty} \leq C_0/2, \quad \rho \|\phi'\|_{\infty} \leq C_1 - 1/2^{1/2}, \quad (\rho/M)\|\phi''\|_{\infty} \leq C_2/2,$$

(7.7)

$$\rho^2 M^{2p+1} \int_{\mathbb{R}} y^{2p}|\phi'(y)|^2 dy \leq C/4.$$  

In particular, condition (7.7) allows us to choose and later use $\rho = cM^{-p-1/2}$ for some $c > 0$. Furthermore, the condition $p > 1/2$ is required to ensure that $\rho M = O(1)$. We obviously have that

$$\|\beta'_0 - \beta'_1\|_{L^2(\mathbb{R})} = \rho^2 M \|\phi'\|_{L^2(\mathbb{R})}.$$  

Next, we will bound the right hand side of (7.6), where recall

$$g_0(y) = ((\beta'_1 - \beta'_0) \ast \pi_{\beta_0})(y).$$  

For this purpose we note that $g_0(y) = -g_0(y)$ and then decompose $\int_{\mathbb{R}} = 2(\int_{0}^{k} + \int_{\infty}^{\infty})$, where the threshold $k < M$ will be chosen later. Since $Z_{\pi_{\beta_1}} \geq \int_{\mathbb{R}} \exp(-\sum_{j=0}^{\infty} \alpha_j y^{2j} - \|\beta_1\|_{\infty}) dy$ and $\|g_0\|_{\infty} \leq \|\beta'_0 - \beta'_1\|_{\infty}$ we deduce that

$$\int_{k}^{\infty} \|g_0(y)|^2 \pi_{\beta_1}(y) dy \lesssim \rho^2 \int_{k}^{\infty} \exp(-a_J y^{2j}) dy \lesssim \rho^2 \frac{\exp(-a_J k^{2j})}{k^{2j-1}}.$$  

On the other hand, we get

$$\int_{0}^{k} \|g_0(y)|^2 \pi_{\beta_1}(y) dy \lesssim \rho^2 k \sup_{y \in [0,k]} |(\phi'(-/M) \ast \pi_{\beta_0})(y)|^2$$  

and

$$|(\phi'(-/M) \ast \pi_{\beta_0})(y)| \lesssim \int_{-\infty}^{k-M} \pi_{\beta_0}(y) dy \lesssim \frac{\exp(-a_J (M-k)^{2j})}{(M-k)^{2j-1}}.$$  

Consequently, choosing $k = M/2$ we deduce from (7.6) that

$$K(\Pi_{\beta_1}, \Pi_{\beta_0}) \lesssim \exp(-a_J (M/2)^{2j})$$

(recall that $\rho = O(M^{-p-1/2})$). Now, choosing $M = 2((\log N)/a_J)^{1/(2J)}$ we finally obtain that

$$K(\Pi_{\beta_1}, \Pi_{\beta_0}) = NK(\Pi_{\beta_1}, \Pi_{\beta_0}) \lesssim 1.$$  

On the other hand, since $\rho = cM^{-p-1/2}$ we deduce that

$$\|\beta'_0 - \beta'_1\|_{L^2(\mathbb{R})} = \text{const} \cdot (\log N)^{-p/J},$$  

which by [31, (2.9) and Theorem 2.2(iii)] completes the proof of our Theorem 5.1.

Acknowledgments

The authors would like to thank two anonymous referees for their useful comments.

Funding

The authors gratefully acknowledge financial support of ERC Consolidator Grant 815703 “STAMFORD: Statistical Methods for High Dimensional Diffusions”.

Here $f_0 \in F_{p,C,C_0/2,(a_1/2,\ldots,a_J/2)},\lambda/2 \subset F_{p,C,C_0,a,\lambda}$. To ensure that $\beta_1 \in F_{p,C,C_0,a,\lambda}$ we assume
References


