ESAIM: PS 26 (2022) 126–151
https://doi.org/10.1051/ps/2022001

ESAIM: Probability and Statistics
www.esaim-ps.org

OPTIMAL CONVERGENCE RATES FOR THE INVARIANT DENSITY ESTIMATION OF JUMP-DIFFUSION PROCESSES

Chiara Amorino[1,*,**] and Eulalia Nualart[2,***]

Abstract. We aim at estimating the invariant density associated to a stochastic differential equation with jumps in low dimension, which is for $d = 1$ and $d = 2$. We consider a class of fully non-linear jump diffusion processes whose invariant density belongs to some Hölder space. Firstly, in dimension one, we show that the kernel density estimator achieves the convergence rate $\frac{1}{T}$, which is the optimal rate in the absence of jumps. This improves the convergence rate obtained in Amorino and Gloter [*J. Stat. Plann. Inference* **213** (2021) 106–129], which depends on the Blumenthal-Getoor index for $d = 1$ and is equal to $\frac{\log T}{T}$ for $d = 2$. Secondly, when the jump and diffusion coefficients are constant and the jumps are finite, we show that is not possible to find an estimator with faster rates of estimation. Indeed, we get some lower bounds with the same rates $\{\frac{1}{T}, \frac{\log T}{T}\}$ in the mono and bi-dimensional cases, respectively. Finally, we obtain the asymptotic normality of the estimator in the one-dimensional case for the fully non-linear process.

Mathematics Subject Classification. 62G07, 62G20, 60J74.

Received January 15, 2021. Accepted January 17, 2022.

1. Introduction

Solutions to Lévy-driven stochastic differential equations have recently attracted a lot of attention in the literature due to its many applications in various areas such as finance, physics, and neuroscience. Indeed, it includes some important examples from finance such as the well-known Kou model in [31], the Barndorff-Nielsen-Shephard model [7], and the Merton model [37] to name just a few. An important example of application of jump-processes in neuroscience is the stochastic Morris-Lecar neuron model presented in [24]. As a consequence, statistical inference for jump processes has recently become an active domain of research.

** CA gratefully acknowledges financial support of ERC Consolidator Grant 815703 "STAMFORD: Statistical Methods for High Dimensional Diffusions".
*** EN acknowledges support from the Spanish MINECO grant PGC2018-101643-B-I00 and Ayudas Fundacion BBVA a Equipos de Investigación Científica 2017.

Keywords and phrases: Minimax risk, convergence rate, non-parametric statistics, ergodic diffusion with jumps, Lévy driven SDE, invariant density estimation.

[1] Université du Luxembourg, 4364 Esch-Sur-Alzette, Luxembourg.
[2] Universitat Pompeu Fabra and Barcelona School of Economics, Department of Economics and Business, Ramón Trias Fargas 25-27, 08005 Barcelona, Spain.

* Corresponding author: chiara.amorino@uni.lu

We consider the process $X = (X_t)_{t\geq 0}$ solution to the following stochastic differential equation with jumps:

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t a(X_s)\mathrm{d}B_s + \int_0^t \int_{\mathbb{R}_0^d} \gamma(X_{s^-})z(\nu(\mathrm{d}s,\mathrm{d}z) - F(z)\mathrm{d}z\mathrm{d}s), \tag{1.1}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion and ν is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated to a Lévy process $(L_t)_{t\geq 0}$ with Lévy density function F. We focus on the estimation of the invariant density μ associated to the jump-process solution to (1.1) in low dimension, which is for $d = 1$ and $d = 2$. In particular, assuming that a continuous record of $(X_t)_{t\in[0,T]}$ is available, our goal is to propose a non-parametric kernel estimator for the estimation of the stationary measure and to discuss its convergence rate for large T.

The same framework has been considered in some recent papers such as [1, 22] (Sect. 5.2), and [2]. In the first paper, it is shown that the kernel estimator achieves the following convergence rates for the pointwise estimation of the invariant density: $\frac{\log T}{T}$ for $d = 2$ and $\frac{(\log T)^{(2-\frac{(1+\alpha)}{2})\vee 1}}{T}$ for $d = 1$ (where α is the Blumenthal-Getoor index). We recall that, in the absence of jumps, the optimal convergence rate in the one-dimensional case is $\frac{1}{T}$, while the one found in [1] depends on the jumps and belongs to the interval $(\frac{\log T}{T}, \frac{(\log T)^{\frac{3}{2}}}{T})$.

In this paper, we wonder if such a deterioration on the rate is because of the presence of jumps or the used approach. Indeed, our purpose is to look for a new approach to recover a better convergence rate in the one-dimensional case (hopefully the same as in the continuous case) and to discuss the optimality of such a rate. This new approach will also lead to the asymptotic normality of the proposed estimator. After that, we will discuss the optimality of the convergence rate in the bi-dimensional case. This will close the circle of the analysis of the convergence rates for the estimation of the invariant density of jump-diffusions, as the convergence rates and their optimality in the case $d \geq 3$ have already been treated in detail in [2].

Beyond these works, to our best knowledge, the literature concerning non-parametric estimation of diffusion processes with jumps is not wide. One of the few examples is given by Funke and Schmisser: in [27] they investigate the non parametric adaptive estimation of the drift of an integrated jump diffusion process, while in [40], Schmisser deals with the non-parametric adaptive estimation of the coefficients of a jumps diffusion process. To name other examples, in [23] the authors estimate in a non-parametric way the drift of a diffusion with jumps driven by a Hawkes process, while in [3] the volatility and the jump coefficients are considered.

On the other hand, the problem of invariant density estimation has been considered by many authors (see *e.g.* [5, 10, 20, 38, 44]) in several different frameworks: it is at the same time a long-standing problem and a highly active current topic of research. One of the reasons why the estimation of the invariant density has attracted the attention of many statisticians is the huge amount of numerical methods to which it is connected, the MCMC method above all. An approximation algorithm for the computation of the invariant density can be found for example in [33, 39]. Moreover, invariant distributions are essential for the analysis of the stability of stochastic differential systems (see *e.g.* [5, 28]).

In [5, 6, 11] some kernel estimators are used to estimate the marginal density of a continuous time process. When μ belongs to some Hölder class whose smoothness is β, they prove under some mixing conditions that their pointwise L^2 risk achieves the standard rate of convergence $T^{\frac{2\beta}{2\beta+1}}$ and the rates are minimax in their framework. Castellana and Leadbetter proved in [14] that, under condition **CL** below, the density can be estimated with the parametric rate $\frac{1}{T}$ by some non-parametric estimators (the kernel ones among them).

In order to introduce condition **CL** it is necessary to request that the process X belongs to a class of real processes with common marginal density μ with respect to the Lebesgue measure on $\mathbb{R}$ and such that the joint density of (X_s, X_t) exists for all $s \neq t$, it is measurable and satisfies $\mu_{(X_s,X_t)} = \mu_{(X_t,X_s)} = \mu_{(X_0,X_{t-s})}$ and it is denoted by $\mu_{|t-s|}$ for all $s, t \in \mathbb{R}$. We also denote by g_u the function $g_u(x, y) = \mu_u(x, y) - \mu(x)\mu(y)$. Then, condition **CL** writes as follows:

CL: $u \mapsto \|g_u\|_\infty$ is integrable on $(0, \infty)$ and $g_u(\cdot,\cdot)$ is continuous for each $u > 0$.

In our context, $g_u(x,y) = \mu(x)p_u(x,y) - \mu(x)\mu(y)$, where $p_u(x,y)$ is the transition density. More precisely, they shed light to the fact that local irregularities of the sample paths provide some additional information. Indeed, if the joint distribution of (X_0, X_t) is not too close to a singular distribution for $|t|$ small, then it is possible to achieve the superoptimal rate $\frac{1}{T}$ for the pointwise quadratic risk of the kernel estimator. Condition **CL** can be verified for ergodic continuous diffusion processes (see [45] for sufficient conditions). The paper of Castellana and Leadbetter led to a lot of works regarding the estimation of the common marginal distribution of a continuous time process. In [8–10, 16, 32] several related results and examples can be found.

An alternative to the kernel density estimator is given by the local time density estimator, which was proposed by Kutoyants in [21] in the case of diffusion processes and was extended by Bosq and Davydov in [12] to a more general context. The latest have proved that, under a condition which is mildly weaker than **CL**, the mean squared error of the local time estimator reaches the full rate $\frac{1}{T}$. Leblanc built in [34] a wavelet estimator of a density belonging to some general Besov space and proved that, if the process is geometrically strong mixing and a condition like **CL** is satisfied, then its L^p-integrated risk converges at rate $\frac{1}{T}$ as well. In [18] the authors built a projection estimator and showed that its L^2-integrated risk achieves the parametric rate $\frac{1}{T}$ under a condition named WCL, which is blandly different compared to **CL**.

WCL: There exists a positive integrable function k (defined on $\mathbb{R}$) such that

$$\sup_{y\in\mathbb{R}} \int_0^\infty |g_u(x,y)|\mathrm{d}u \leq k(x), \qquad \text{for all } x \in \mathbb{R}.$$

In this paper, we will show that our mono-dimensional jump-process satisfies a local irregularity condition **WCL1** and an asymptotic independence condition **WCL2** (see Prop. 3.2), two conditions in which the original condition **WCL** can be decomposed. In this way, it will be possible to show that the L^2 risk for the pointwise estimation of the invariant measure achieves the superoptimal rate $\frac{1}{T}$, using our kernel density estimator. Moreover, the same conditions will result in the asymptotic normality of the proposed estimator. Indeed, as we will see in the proof of Theorem 2.3, the main challenge in this part is to justify the use of dominated convergence theorem, which will ensured by conditions **WCL1** and **WCL2**. We will find in particular that, for any collection $(x_i)_{1\leq i\leq m}$ of real numbers, we have

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i) - \mu(x_i),\, 1 \leq i \leq m) \xrightarrow{\mathcal{D}} N^{(m)}(0, \Sigma^{(m)}) \text{ as } T \to \infty,$$

where $\hat{\mu}_{h,T}$ is the kernel density estimator and

$$\Sigma^{(m)} := (\sigma(x_i,x_j))_{1\leq i,j\leq m}, \qquad \sigma(x_i,x_j) := 2\int_0^\infty g_u(x_i,x_j)\mathrm{d}u.$$

We remark that the precise form of the equation above allows us to construct tests and confidence sets for the density.

We have found the convergence rate $\left\{\frac{1}{T}, \frac{\log T}{T}\right\}$ for the risk associated to our kernel density estimator for the estimation of the invariant density for $d = 1$ and $d = 2$. Then, some questions naturally arise: are the convergence rates the best possible or is it possible to improve them by using other estimators? In order to answer, we consider a simpler model where both the volatility and the jump coefficient are constant and the intensity of the jumps is finite. Then, we look for a lower bound for the risk at a point $x \in \mathbb{R}^d$ defined as in equation (2.8) below. The first idea is to use the two hypothesis method (see Sect. 2.3 in [43]). To do that, the knowledge of the link between the drift b and the invariant density μ_b is essential. In absence of jumps such link is explicit, but in our context it is more challenging. As shown in [2, 19], it is possible to find the link knowing that the invariant measure has to satisfy $A^*\mu_b = 0$, where A^* is the adjoint of the generator of the considered diffusion. This method allows us to show that the superoptimal rate $\frac{1}{T}$ is the best possible for the estimation of the invariant density in $d = 1$, but it fails in the bi-dimensional case (see Rem. 1 below for details). Finally, we

use a finite number of hypotheses to prove a lower bound in the bi-dimensional case. This requires a detailed analysis of the Kullback divergence between the probability laws associated to the different hypotheses. Thanks to that, it is possible to recover the optimal rate $\frac{\log T}{T}$ in the two-dimensional case.

The paper is organised as follows. In Section 2 we give the assumptions on our model and we provide our main results. Section 3 is devoted to state and prove some preliminary results needed for the proofs of the main results. To conclude, in Section 4 we give the proof of Theorems 2.1, 2.3, 2.4, and 2.5, where our main results are gathered.

Throughout all the paper c and λ are constants that may change from line to line. Their dependence on T or other fixed constants will be implied from the statements.

2. Model assumption and main results

We consider the following stochastic differential equation with jumps

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t a(X_s)\mathrm{d}B_s + \int_0^t \int_{\mathbb{R}_0^d} \gamma(X_{s^-})z(\nu(\mathrm{d}s,\mathrm{d}z) - F(z)\mathrm{d}z\mathrm{d}s), \tag{2.1}$$

where $t \geq 0$, $d \in \{1,2\}$, $\mathbb{R}_0^d = \mathbb{R}^d \backslash \{0\}$, the initial condition X_0 is a $\mathbb{R}^d$-valued random variable, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable functions, $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, and ν is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated to a Lévy process $(L_t)_{t\geq 0}$ with Lévy density function F. All sources of randomness are mutually independent.

We consider the following assumptions on the coefficients and on the Lévy density F:

- **A1** The functions b, γ and aa^T are globally Lipschitz and bounded. Moreover, $\inf_{x\in\mathbb{R}^d} aa^T(x) \geq c\mathrm{Id}$, for some constant $c > 0$, where Id denotes the $d \times d$ identity matrix and $\inf_{x\in\mathbb{R}^d} \det(\gamma(x)) > 0$.
- **A2** $\langle x, b(x)\rangle \leq -c_1|x| + c_2$, for all $|x| \geq \rho$, for some $\rho, c_1, c_2 > 0$.
- **A3** $\mathrm{Supp}(F) = \mathbb{R}_0^d$ and for all $z \in \mathbb{R}_0^d$, $F(z) \leq \frac{c_3}{|z|^{d+\alpha}}$, for some $\alpha \in (0,2), c_3 > 0$.
- **A4** There exist $\epsilon_0 > 0$ and $c_4 > 0$ such that $\int_{\mathbb{R}_0^d} |z|^2 e^{\epsilon_0|z|} F(z)\mathrm{d}z \leq c_4$.
- **A5** If $\alpha = 1$, $\int_{r<|z|<R} zF(z)\mathrm{d}z = 0$, for any $0 < r < R < \infty$.

Assumption **A1** ensures that equation (2.1) admits a unique càdlàg adapted solution $X = (X_t)_{t\geq 0}$ satisfying the strong Markov property, see *e.g.* [4]. Moreover, it is shown in Lemma 2 [1] that if we further assume Assumptions **A2-A4**, then the process X is exponentially ergodic and exponentially β-mixing. Therefore the process is stationary and, in particular, it has a unique invariant distribution π, which we assume it has a density μ with respect to the Lebesgue measure. Finally, Assumption **A5** ensures the existence of the transition density of X denoted by $p_t(x,y)$ which satisfies the following upper bound (see [1], Lem. 1): for all $T \geq 0$, there exist $c > 0$ and $\lambda > 0$ such that for any $t \in [0,T]$ and $x, y \in \mathbb{R}^d$,

$$p_t(x,y) \leq c\left(t^{-d/2}e^{-\lambda\frac{|y-x|^2}{t}} + \frac{t}{(t^{1/2}+|y-x|)^{d+\alpha}}\right). \tag{2.2}$$

We assume that the process is observed continuously $X = (X_t)_{t\in[0,T]}$ in a time interval $[0,T]$ such that T tends to ∞. In the paper [1] cited above, the nonparametric estimation of μ is studied *via* the kernel estimator which is defined as follows. We assume that μ belongs to the Hölder space $\mathcal{H}_d(\beta,\mathcal{L})$ where $\beta = (\beta_1,\ldots,\beta_d)$, $\beta_i \geq 1$ and $\mathcal{L} = (\mathcal{L}_1,\ldots,\mathcal{L}_d)$, $\mathcal{L}_i > 0$, which means that for all $i \in \{1,\ldots,d\}$, $k = 0,1,\ldots,\lfloor\beta_i\rfloor$ and $t \in \mathbb{R}$,

$$\left\|D_i^{(k)}\mu\right\|_\infty \leq \mathcal{L} \quad \text{and} \quad \left\|D_i^{(\lfloor\beta_i\rfloor)}\mu(.+te_i) - D_i^{(\lfloor\beta_i\rfloor)}\mu(.)\right\|_\infty \leq \mathcal{L}_i|t|^{\beta_i-\lfloor\beta_i\rfloor},$$

where $D_i^{(k)}$ denotes the kth order partial derivative of μ w.r.t the ith component, $\lfloor \beta_i \rfloor$ is the integer part of β_i, and $e_1, \dots, e_d$ is the canonical basis of $\mathbb{R}^d$. That is, all the partial derivatives of μ up to order $\lfloor \beta \rfloor$ are bounded and the $\lfloor \beta \rfloor$th partial derivative is Hölder continuous of order $\beta - \lfloor \beta \rfloor$ in any direction. We recall that it is natural in our context to assume that the invariant density belongs to a Hölder class as above. In fact, the proof of the bias bound (2.5) stated below gives a direct application of this assumption, see the proof of Proposition 2 in [1]. Other examples of nonparametric estimation over Hölder classes can be found in [29, 30, 35, 42].

We set

$$\hat{\mu}_{h,T}(x) = \frac{1}{T\prod_{i=1}^d h_i} \int_0^T \prod_{i=1}^d K\left(\frac{x_i - X_t^i}{h_i}\right) \mathrm{d}t =: \frac{1}{T}\int_0^T \mathbb{K}_h(x - X_t)\mathrm{d}t,$$

where $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $h = (h_1, \dots, h_d)$ is a bandwidth and $K : \mathbb{R} \to \mathbb{R}$ is a kernel function satisfying

$$\int_{\mathbb{R}} K(x)\mathrm{d}x = 1, \quad \|K\|_\infty < \infty, \quad \mathrm{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^i \mathrm{d}x = 0,$$

for all $i \in \{0, \dots, M\}$ with $M \geq \max_i \beta_i$.

We first consider equation (2.1) with $d = 1$ and show that the kernel estimator reaches the optimal rate T^{-1}, as it is for the stochastic differential equation (2.1) without jumps. For this, we need the following additional assumption on a.

A6 If $d = 1$, $a^2 \in C_b^2(\mathbb{R})$, that is, a^2 is twice continuously differentiable with bounded first and second derivatives.

Assumption **A6** is needed in order to show the results gathered in Theorems 2.1 and 2.3, while for the other results only assumptions **A1**–**A5** will be required.

Theorem 2.1. *Let X be the solution to (2.1) on $[0, T]$ with $d = 1$. Suppose that Assumptions* **A1**–**A6** *hold and $\mu \in \mathcal{H}_1(\beta, \mathcal{L})$, with $\beta \geq 1$. Then there exists $c > 0$ such that for all $T > 0$, $h \leq 1$, and $x \in \mathbb{R}$,*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq ce^{\epsilon|x|}(h^{2\beta} + \frac{1}{T}), \tag{2.3}$$

where $0 < \epsilon \leq \min(\frac{\epsilon_0}{\|\gamma\|_\infty}, \epsilon_0)$, with $\epsilon_0 > 0$ as in Assumption **A4** *In particular, choosing $h(T) = \frac{1}{\sqrt{T}}$, we conclude that for $T \geq 1$,*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq \frac{ce^{\epsilon|x|}}{T}. \tag{2.4}$$

We observe that both the bandwidth and the upper bound do not depend on the unknown smoothness of the invariant density β, so there is no need to propose a data driven bandwidth adaptive selection procedure as in the case $d > 2$ (see [1]).

Theorem 2.1 improves the upper bound obtained in [1] which was of the form $\frac{(\log T)^{(2-\frac{1+\alpha}{2})\vee 1}}{T}$. The price to pay is that the constant in the upper bound depends on x (see Rem. 2.2). However, we are able to find a convergence rate which is optimal, as we will see in Theorem 2.4. As in [1], we will use the bias-variance

decomposition (see [17], Prop. 1)

$$\begin{aligned}\mathbb{E}[|\hat{\mu}_{h,T}(x)-\mu(x)|^2] &\le |\mathbb{E}[\hat{\mu}_{h,T}(x)]-\mu(x)|^2+\mathbb{E}[|\hat{\mu}_{h,T}(x)-\mathbb{E}[\hat{\mu}_{h,T}(x)]|^2]\\ &\le c\left(h^{2\beta}+T^{-2}\mathrm{Var}\left(\int_0^T \mathbb{K}(x-X_t)\mathrm{d}t\right)\right),\end{aligned} \tag{2.5}$$

for some constant $c>0$. For the proof of the bias bound $ch^{2\beta}$ in the same setting of this paper see the proof of Proposition 2 in [1].

Then in [1] bounds on the transition semigroup and on the transition density (see (2.2) above) give an upper bound for the variance depending on the bandwidth. Here, we use a similar approach as in [14, 18] to obtain a bandwidth-free rate for the variance of smoothing density estimators (which include the kernel estimator). For Markov diffusions, the sufficient conditions can be decomposed into a local irregularity condition **WCL1** plus an asymptotic independence condition **WCL2**. There exist two positive integrable functions k_1 and k_2 (defined on $\mathbb{R}$) and $u_0>0$ such that

$$\textbf{WCL1:}\ \sup_{y\in\mathbb{R}}\int_0^{u_0}|g_u(x,y)|\,\mathrm{d}u<k_1(x),\qquad \text{for all } x\in\mathbb{R},$$

$$\textbf{WCL2:}\ \sup_{y\in\mathbb{R}}\int_{u_0}^{\infty}|g_u(x,y)|\,\mathrm{d}u<k_2(x),\qquad \text{for all } x\in\mathbb{R}$$

where $g_u(x,y):=\mu(x)p_u(x,y)-\mu(x)\mu(y)$. In order to show these conditions, some further bounds on the transition density $p_t(x,y)$ involving partial derivatives are needed (see Lem. 3.1 below), for which the additional condition **A6** is required.

Remark 2.2. The term $e^{\epsilon|x|}$ that appears in the bounds (2.3) and (2.4) comes from the fact that we are able to show condition **WCL2** with $k_2(x)=\mu(x)(1+f^*(x))$, where f^* is the Lyapunov function constructed in [1], defined as a C^∞ approximation of $e^{\epsilon|x|}$ (see the proof of Prop. 3.2). We know that $\int_{\mathbb{R}}\mu(x)f^*(x)\mathrm{d}x<\infty$, as shown in [36], but this is not sufficient as it was in [18] in order to bound the variance term in (2.5) since here we are dealing with the kernel estimator. In order to remove the term $e^{\epsilon|x|}$ an additional assumption would be needed that ensures that $\sup_{x\in\mathbb{R}}\mu(x)f^*(x)<\infty$.

As shown in [13], conditions **WLC1** and **WLC2** are also useful to show the asymptotic normality of the kernel density estimator, as proved in the next theorem.

Theorem 2.3. *Let X be the solution to (2.1) on $[0,T]$ with $d=1$. Suppose that Assumptions* **A1**-**A6** *hold and $\mu\in\mathcal{H}_1(\beta,\mathcal{L})$, with $\beta\ge 1$. Consider the bandwidth $h(T)=(\frac{1}{T})^{\frac{1}{2}-\epsilon}$, where $\epsilon\in(0,\frac{1}{2})$. Then, for any collection $(x_i)_{1\le i\le m}$ of distinct real numbers*

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i)-\mathbb{E}[\hat{\mu}_{h,T}(x_i)],\,1\le i\le m)\xrightarrow{\mathcal{D}}N^{(m)}(0,\Sigma^{(m)})\ \text{as } T\to\infty, \tag{2.6}$$

where

$$\Sigma^{(m)}:=(\sigma(x_i,x_j))_{1\le i,j\le m},\qquad \sigma(x_i,x_j):=2\int_0^\infty g_u(x_i,x_j)\mathrm{d}u.$$

Observe that using the choice of $h(T)=(\frac{1}{T})^{\frac{1}{2}-\epsilon}$, with $\epsilon>0$ in the bias bound (2.5), we get that for any $x\in\mathbb{R}$ and $T\ge 1$,

$$\sqrt{T}|\mathbb{E}[\hat{\mu}_{h,T}(x)]-\mu(x)|\le cT^{-\frac{1}{2}(\beta-1-2\beta\epsilon)}.$$

Therefore, choosing $\beta > 1$ and $\epsilon < \frac{\beta-1}{2\beta}$ and applying Theorem 2.3, we conclude that as $T \to \infty$

$$\sqrt{T}(\hat{\mu}_{h,T}(x_i) - \mu(x_i),\, 1 \le i \le m) \xrightarrow{\mathcal{D}} N^{(m)}(0, \Sigma^{(m)}).$$

We are also interested in obtaining lower bounds in dimension $d \in \{1, 2\}$. For the computations of the lower bounds we consider the particular case of equation (2.1) given by

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + aB_t + \int_0^t \int_{\mathbb{R}_0^d} \gamma z(\nu(\mathrm{d}s, \mathrm{d}z) - F(z)\mathrm{d}z\mathrm{d}s), \tag{2.7}$$

where a and γ are $d \times d$ constant matrices and the rest of terms are as in equation (2.1).

We next introduce the following set of drift functions of equation (2.7). We say that a bounded and Lipschitz function $b : \mathbb{R}^d \to \mathbb{R}^d$ belongs to $\Sigma(\beta, \mathcal{L})$ if the unique invariant density μ_b of the solution $X = (X_t)_{t\ge 0}$ to (2.7) belongs to $\mathcal{H}_d(\beta, 2\mathcal{L})$ for some $\beta, \mathcal{L} \in \mathbb{R}^d$, $\beta_i \ge 1$, $\mathcal{L}_i > 0$. A detailed description of the set $\Sigma(\beta, \mathcal{L})$ will be given in Section 4.3, where two explicit examples of drift coefficients b_0 and b_1 belonging to $\Sigma(\beta, \mathcal{L})$ will be introduced.

We denote by $\mathbb{P}_b^{(T)}$ and $\mathbb{E}_b^{(T)}$ the law and expectation of the solution $(X_t)_{t\in[0,T]}$. We define the minimax risk at a point $x \in \mathbb{R}^d$ by

$$\mathcal{R}_T^x(\beta, \mathcal{L}) := \inf_{\tilde{\mu}_T} \mathcal{R}(\tilde{\mu}_T(x)) := \inf_{\tilde{\mu}_T} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b^{(T)}[(\tilde{\mu}_T(x) - \mu_b(x))^2], \tag{2.8}$$

where the infimum is taken on all possible estimators of the invariant density.

The following lower bounds hold true.

Theorem 2.4. *Let X be the solution to (2.7) on $[0, T]$ with $d = 1$. Suppose that Assumptions* **A1**–**A5** *hold, that $\int_{\mathbb{R}} F(z)\mathrm{d}z < \infty$ and that $\mu_b \in \mathcal{H}_1(\beta, \mathcal{L})$, with $\beta \ge 1$. Then, there exists $T_0 > 0$ and $c > 0$ such that, for all $T \ge T_0$,*

$$\inf_{x\in\mathbb{R}} \mathcal{R}_T^x(\beta, \mathcal{L}) \ge \frac{c}{T}.$$

Theorem 2.5. *Let X be the solution to (2.7) on $[0, T]$ with $d = 2$. Suppose that Assumptions* **A1**–**A5** *hold, that $\int_{\mathbb{R}^2} F(z)\mathrm{d}z < \infty$ and that $\mu_b \in \mathcal{H}_2(\beta, \mathcal{L})$, with $\beta_i \ge 1$ for $i = 1, 2$. Assume that for all $i \in \{1, 2\}$ and $j \ne i$,*

$$|(aa^T)_{ij}(aa^T)_{jj}^{-1}| \le \frac{1}{2}. \tag{2.9}$$

Then, there exists $T_0 > 0$ and $c > 0$ such that, for $T \ge T_0$,

$$\inf_{\tilde{\mu}_T} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b^{(T)} \left[\sup_{x\in\mathbb{R}^2} (\tilde{\mu}_T(x) - \mu_b(x))^2 \right] \ge c\frac{\log T}{T}.$$

Recall for these two theorems, a and γ are $d \times d$ constant matrices. In this case, when $d = 1$, Assumption **A1** is equivalent to say that $a \ne 0$ and $\gamma > 0$, while when $d = 2$, it is equivalent to say that $\det(a) \ne 0$ and $\det(\gamma) > 0$. Moreover, hypotheses **A3**–**A5** imply that the unique solution to equation (2.7) admits a unique invariant measure π_b, which we assume has a density μ_b with respect to the Lebesgue measure, as before.

Comparing these lower bounds with the upper bound of Theorem 2.1 for the case $d = 1$ and Proposition 4 in [1] for the two-dimensional case, we conclude that the convergence rate $\{\frac{1}{T}, \frac{\log T}{T}\}$ are the best possible for the estimation of the invariant density in dimension $d \in \{1, 2\}$.

The proof of Theorem 2.4 follows along the same lines as that of Theorem 2 in [2], where a lower bound for the estimation of the invariant density for the solution to (2.7) for $d \geq 3$ is obtained. The proof is based on the two hypotheses method, explained for example in Section 2.3 of [43]. However, this method does not work for the two-dimensional case as explained in Remark 4.2 below. Instead, we use the Kullback's version of the finite number of hypotheses method as stated in Lemma C.1 of [41], see Lemma 3.4 below. Observe that this method gives a slightly weaker lower bound as we get a $\sup_x$ inside the expectation, while the method in [2] provides an $\inf_x$ outside the expectation.

3. Preliminary results

The proof of Theorems 2.1 and 2.3 will use the following bounds on the transition density.

Lemma 3.1. *Let X be the solution to (2.1) on $[0,T]$ with $d=1$. Suppose that Assumptions* **A1**–**A6** *hold. Then, there exist jointly continuous processes Z, A and B on $\mathbb{R}_+ \times \mathbb{R}^2$ such that for all $t \geq 0$ and $x, y \in \mathbb{R}$,*

$$p_t(x,y) = Z_t(x,y) + A_t(x,y) + B_t(x,y) \tag{3.1}$$

satisfying that for all $T>0$, there exist $c>0$ and $\lambda > 0$ such that for any $x, y \in \mathbb{R}$ and $t \in [0,T]$

$$\left| \frac{\partial^2}{\partial y^2} Z_t(x,y) \right| \leq c\, t^{-3/2} e^{-\lambda \frac{|y-x|^2}{t}}, \tag{3.2}$$

$$|A_t(x,y)| \leq c\,(t^{3/2}(|y-x| + \sqrt{t})^{-1-\alpha} + e^{-\lambda \frac{|y-z|^2}{t}}), \tag{3.3}$$

and

$$|B_t(x,y)| \leq c\,(1 + t^{2-\alpha/2})(|y-x| + \sqrt{t})^{-1-\alpha}. \tag{3.4}$$

Proof. By Duhamel's formula (1.12) of [15], the transition density of the solution to (2.1) satisfies that for all $t \geq 0$ and $x, y \in \mathbb{R}$,

$$p_t(x,y) = Z_t(x,y) + A_t(x,y) + B_t(x,y)$$

where $Z_t(x,y)$ is the transition density of the solution to (2.1) with $b = \gamma = 0$, and A_t and B_t are defined as follows

$$A_t(x,y) := \int_0^t \int_{\mathbb{R}} p_r(x,z)\, b(z) \frac{\partial}{\partial z} Z_{t-r}(z,y)\, \mathrm{d}z\, \mathrm{d}r,$$

and

$$B_t(x,y) := \int_0^t \int_{\mathbb{R}} p_r(x,z) \int_{\mathbb{R}} \big(Z_{t-r}(z+\xi, y) - Z_{t-r}(z,y) \\ - \mathbf{1}_{|\xi| \leq 1}\, \xi\, \frac{\partial}{\partial z} Z_{t-r}(z,y) \big) \frac{k(z,\xi)}{|\xi|^{1+\alpha}}\, \mathrm{d}\xi\, \mathrm{d}z\, \mathrm{d}r,$$

where $k(z,\xi) = \frac{1}{\gamma(z)} |\xi|^{1+\alpha} F(\frac{\xi}{\gamma(z)})$. This shows the decomposition formula (3.1).

By (6.1) in Theorem 7 of [26], using the fact that a^2 is bounded together with **A6**, we have that for all $T > 0$, there exist $c, \lambda > 0$ such that for all $x, y \in \mathbb{R}$ and $t \in [0, T]$

$$\left| \frac{\partial^2}{\partial y^2} Z_t(x, y) \right| \leq c\, t^{-3/2} e^{-\lambda \frac{|y-x|^2}{t}},$$

(which proves (3.2)) and

$$\left| \frac{\partial}{\partial x} Z_t(x, y) \right| \leq c\, t^{-1} e^{-\lambda \frac{|y-x|^2}{t}}. \tag{3.5}$$

In particular, using (3.5) and the fact that b is bounded, we get that

$$|A_t(x, y)| \leq c \int_0^t \int_{\mathbb{R}} p_r(x, z)(t - r)^{-1} e^{-\lambda \frac{|y-z|^2}{t-r}} \, \mathrm{d}z \, \mathrm{d}r.$$

Moreover, using (2.2) together with (2.6) and (2.8) of [15] with $\gamma_1 = -1$ and $\gamma_2 = 2$, and $\gamma_1 = 0$ and $\gamma_2 = -1$, respectively, we conclude that (3.3) holds true.

On the other hand, appealing to Corollary 2.4(i) of [15], from hypotheses **A1**, **A3** and **A5**, we get that for all $T > 0$, there exists $c > 0$ such that for all $x, y \in \mathbb{R}$ and $t \in [0, T]$,

$$|B_t(x, y)| \leq c \int_0^t \int_{\mathbb{R}} p_r(x, z)(|y - z| + \sqrt{t - r})^{-1-\alpha} \, \mathrm{d}z \, \mathrm{d}r.$$

Finally, using again (2.2) together with (2.5) and (2.6) of [15] with $\gamma_1 = 0$ and $\gamma_2 = 2$, and $\gamma_1 = 0$ and $\gamma_2 = 0$, respectively, we obtain (3.4).

The proof of the Lemma is completed. □

The key point of the proof of Theorem 2.1 consists in showing that conditions **WCL1** and **WCL2** hold true, which is proved in the next proposition.

Proposition 3.2. *Let X be the solution to (2.1) on $[0, T]$ with $d = 1$. Suppose that Assumptions* **A1-A6** *hold. Then, conditions* **WCL1** *and* **WCL2** *are satisfied.*

Proof. We start considering **WCL1**. The density estimate (2.2) yields

$$p_t(x, y) \leq c t^{-\frac{1}{2}} + \tilde{c} t^{\frac{1-\alpha}{2}} \leq \bar{c} t^{-\frac{1}{2}} \qquad 0 < t \leq 2, \tag{3.6}$$

which combined with $\sup_{y \in \mathbb{R}} \mu(y) < \infty$ gives **WCL1** with $k_1(x) = \mu(x)$ and $u_0 = 2$. In order to show **WCL2**, we set $\varphi(\xi) := \mathbb{E}[\exp(i\xi X_t)]$ and $\varphi_x(\xi, t) := \mathbb{E}[\exp(i\xi X_t) | X_0 = x]$ and we claim that there exists $\hat{c} > 0$ such that for all $\xi \in \mathbb{R}$,

$$|\varphi(\xi)| \leq \hat{c}(1 + |\xi|)^{-2}. \tag{3.7}$$

Moreover, there exists $\tilde{c} > 0$, such that for all $t \geq 2$, $x \in \mathbb{R}$, and $\xi \in \mathbb{R}$,

$$|\varphi_x(\xi, t)| \leq \tilde{c}(1 + |\xi|)^{-2}. \tag{3.8}$$

Recall from Lemma 2 in [1] and its proof that the process X is exponentially β-mixing and there exists $\rho > 0$ such that for all $x \in \mathbb{R}$ and $t > 0$,

$$\|P_t(x,\cdot) - \mu(\cdot)\|_{TV} \leq (1 + f^*(x))e^{-\rho t}, \tag{3.9}$$

where $(P_t)_{t\in\mathbb{R}}$ is the transition semigroup of our process X, $\|\cdot\|_{TV}$ is the total variation norm and $f^*(x)$ is a Lyapounov function. Specifically, $f^*(x)$ is defined as $e^{\epsilon|x|}$ for $|x| \geq 1$, with $\epsilon \leq \min(\frac{\epsilon_0}{\|\gamma\|_\infty}, \epsilon_0)$ ($\epsilon_0 > 0$ as in Asm. **A4**). In order to avoid any regularity problem in 0, f^* is introduced as piecewise function. For $|x| < 1$ it is defined as a C^∞ approximation of $e^{\epsilon|x|}$, such that f^* is C^∞ on $\mathbb{R}$.

We now prove that inequalities (3.7), (3.8) and (3.9) imply **WCL2**. Using the inverse Fourier transform, we have

$$2\pi(p_t(x,y) - \mu(y)) = \int_{\mathbb{R}} \exp(-i\xi y)(\varphi_x(\xi,t) - \varphi(t))\mathrm{d}\xi.$$

Then, using (3.7) and (3.8) we get, for $t \geq 2$,

$$2\pi|p_t(x,y) - \mu(y)| \leq 2(\tilde{c} + \hat{c})^{\frac{p-1}{p}} (\sup_{\xi\in\mathbb{R}} |\varphi_x(\xi,t) - \varphi(\xi)|)^{\frac{1}{p}} \int_{\mathbb{R}^+} (1+\xi)^{-2\frac{p-1}{p}} \mathrm{d}\xi,$$

where we have used that $1 = \frac{1}{p} + \frac{p-1}{p}$. We can choose $p > 2$, so that $2\frac{p-1}{p} > 1$. We get that there exists a finite constant c such that, for all $t \geq 2$ and $x, y \in \mathbb{R}$,

$$|g_t(x,y)| = \mu(x)|p_t(x,y) - \mu(y)| \leq c\mu(x)(\sup_{\xi\in\mathbb{R}} |\varphi_x(\xi,t) - \varphi(\xi)|)^{\frac{1}{p}},$$

where we observe that the right hand side is independent of y. By using the fact that

$$\sup_{\lambda\in\mathbb{R}} |\varphi_x(\lambda,t) - \varphi(\lambda)| \leq \|P_t(x,\cdot) - \mu(\cdot)\|_{TV}$$

together with (3.9) we obtain that there exist $c > 0$ and $\rho > 0$ such that for all $x, y \in \mathbb{R}$ and $t \geq 2$,

$$|g_t(x,y)| \leq c\mu(x)(1 + f^*(x))e^{-\rho t},$$

as f^* is positive, and so

$$\sup_{y\in\mathbb{R}} \int_2^\infty |g_t(x,y)|\,\mathrm{d}t \leq c\mu(x)(1 + f^*(x)) \int_2^\infty e^{-\rho t}\,\mathrm{d}t,$$

which implies **WCL2** with $k_2(x) = c\mu(x)(1 + f^*(x))$.

We are left to show (3.7) and (3.8). We start showing (3.8). Using (3.1) and integrating by parts yields

$$\begin{aligned}
|\varphi_x(\xi,t)| &= \left| \int_{\mathbb{R}} \exp(i\xi y) p_t(x,y) \mathrm{d}y \right| \\
&= \left| \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(i\xi y) p_{t-1}(x,z) p_1(z,y) \mathrm{d}y \, \mathrm{d}z \right| \\
&= \left| \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(i\xi y) p_{t-1}(x,z) \left(Z_1(z,y) + A_1(z,y) + B_1(z,y) \right) \mathrm{d}y \, \mathrm{d}z \right| \\
&\leq |\xi|^{-2} \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) \left| \frac{\partial^2}{\partial y^2} Z_1(z,y) \right| \mathrm{d}y \, \mathrm{d}z \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) |A_1(z,y)| \mathrm{d}y \, \mathrm{d}z + \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) |B_1(z,y)| \mathrm{d}y \, \mathrm{d}z \\
&=: |\xi|^{-2} (I_1 + I_2 + I_3).
\end{aligned}$$

Appealing to (3.2), we obtain that

$$I_1 \leq c \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) e^{-\lambda |y-z|^2} \mathrm{d}y \, \mathrm{d}z = c,$$

where c is independent of t and x as $\int_{\mathbb{R}} p_{t-1}(x,z) \mathrm{d}z = 1$. Using (3.3), we get that

$$I_2 \leq c \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) (|y-z|+1)^{-1-\alpha} + e^{-\lambda |y-z|^2}) \mathrm{d}y \, \mathrm{d}z = c,$$

as the $\mathrm{d}y$ integral is finite since $\alpha \in (0,2)$. Similarly, by (3.4),

$$I_3 \leq c \int_{\mathbb{R}} \int_{\mathbb{R}} p_{t-1}(x,z) (|y-x|+1)^{-1-\alpha} \mathrm{d}y \, \mathrm{d}z \leq c.$$

Thus, we have proved that $|\varphi_x(\xi,t)| \leq c|\xi|^{-2}$. Since $|\varphi_x(\xi,t)| \leq 1$, this implies (3.8). Similarly,

$$\begin{aligned}
|\varphi(\xi)| &= \left| \int_{\mathbb{R}} \exp(i\xi y) \mu(y) \mathrm{d}y \right| \\
&\leq \left| \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(i\xi y) \mu(z) Z_1(z,y) \mathrm{d}y \, \mathrm{d}z \right| \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}} \mu(z) |A_1(z,y)| \mathrm{d}y \, \mathrm{d}z + \int_{\mathbb{R}} \int_{\mathbb{R}} \mu(z) |B_1(z,y)| \mathrm{d}y \, \mathrm{d}z \\
&\leq |\xi|^{-2} \int_{\mathbb{R}} \int_{\mathbb{R}} \mu(z) \left| \frac{\partial^2}{\partial y^2} Z_1(z,y) \right| \mathrm{d}y \, \mathrm{d}z \\
&\quad + \int_{\mathbb{R}} \int_{\mathbb{R}} \mu(z) |A_1(z,y)| \mathrm{d}y \, \mathrm{d}z + \int_{\mathbb{R}} \int_{\mathbb{R}} \mu(z) |B_1(z,y)| \mathrm{d}y \, \mathrm{d}z \\
&\leq c|\xi|^{-2},
\end{aligned}$$

which implies (3.8) since $|\varphi(\xi)| \leq 1$. The proof of the proposition is now completed. □

Theorem 2.3 is an application of the following central limit theorem for discrete stationary sequences. Let $Y_n = (Y_{n,i},\, i \in \mathbb{Z})$, $n \geq 1$ be a sequence of strictly stationary discrete time $\mathbb{R}^m$ valued random process. We define

the α-mixing coefficient of Y_n by

$$\alpha_{n,k} := \sup_{A\in\sigma(Y_{n,i},\, i\le 0),\quad B\in\sigma(Y_{n,i},\, i\ge k)} \big(\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)\big)$$

and we set $\alpha_k := \sup_{n\ge 1}\alpha_{n,k}$ (see also Sect. 1 in [25]). We denote by $Y^{(r)}$ the r-th component of an m dimensional random vector Y.

Theorem 3.3 (Thm. 1.1 [13]). *Assume that*

(i) $\mathbb{E}[Y_{n,i}^{(r)}]=0$ *and* $|Y_{n,i}^{(r)}|\le M_n$ *for every* $n\ge 1$, $i\ge 1$ *and* $1\le r\le m$, *where* M_n *is a constant depending only on* n.

(ii)

$$\sup_{i\ge 1, 1\le r\le m}\mathbb{E}[(Y_{n,i}^{(r)})^2]<\infty.$$

(iii) *For every* $1\le r,s\le m$ *and for every sequence* $b_n\to\infty$ *such that* $b_n\le n$ *for every* $n\ge 1$, *we have*

$$\lim_{n\to\infty}\frac{1}{b_n}\mathbb{E}\left[\sum_{i=1}^{b_n}Y_{n,i}^{(r)}\sum_{j=1}^{b_n}Y_{n,j}^{(s)}\right]=\sigma_{r,s}.$$

(iv) *There exists* $\gamma_0\in(1,\infty)$ *such that* $\sum_{k\ge 1}k\alpha_k^{\frac{\gamma_0-1}{\gamma_0}}<\infty$.

(v) *For some constant* $c>0$ *and for every* $n\ge 1$, $M_n\le cn^{\frac{\gamma_0^2}{(3\gamma_0-1)(2\gamma_0-1)}}$.

Then,

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}}\xrightarrow{\mathcal{D}}N(0,\Sigma)\quad \text{as } n\to\infty,$$

where $\Sigma=(\sigma_{r,s})_{1\le r,s\le m}$.

The proof of Theorem 2.5 is based on the following Kullback version of the main theorem on lower bounds in [43], see Lemma C.1 of [41]:

Lemma 3.4. *Fix* $\beta,\mathcal{L}\in(0,\infty)^2$ *and assume that there exists* $f_0\in\mathcal{H}_2(\beta,\mathcal{L})$ *and a finite set* J_T *such that one can find* $\{f_j,\, j\in J_T\}\subset\mathcal{H}_2(\beta,\mathcal{L})$ *satisfying*

$$\|f_j-f_k\|_\infty\ge 2\psi>0\qquad \forall j\ne k\in J_T. \tag{3.10}$$

Moreover, denoting $\mathbb{P}_j^{(T)}$ *the probability measure associated with* f_j, $\forall j\in J_T$, $\mathbb{P}_j^{(T)}\ll\mathbb{P}_0^{(T)}$ *and*

$$\frac{1}{|J_T|}\sum_{j\in J_T}KL(\mathbb{P}_j^{(T)},\mathbb{P}_0^{(T)})=\frac{1}{|J_T|}\sum_{j\in J_T}\mathbb{E}_j^{(T)}\left[\log\left(\frac{d\mathbb{P}_j^{(T)}}{d\mathbb{P}_0^{(T)}}(X^T)\right)\right]\le\delta\log(|J_T|) \tag{3.11}$$

for some $\delta\in(0,\frac{1}{8})$. *Then, for* $q>0$, *we have*

$$\inf_{\tilde{\mu}_T}\sup_{\mu_b\in\mathcal{H}_2(\beta,\mathcal{L})}(\mathbb{E}_b^{(T)}[\psi^{-q}\,\|\tilde{\mu}_T-\mu_b\|_\infty^q])^{1/q}\ge c(\delta)>0,$$

where the infimum is taken over all the possible estimators $\tilde{\mu}_T$ of μ_b.

4. Proof of the main results

4.1. Proof of Theorem 2.1

By the symmetry of the covariance operator and the stationarity of the process,

$$\begin{aligned} T\,\mathrm{Var}(\hat{\mu}_{h,T}(x)) &= \frac{1}{T}\int_0^T\int_0^T \mathrm{Cov}(\mathbb{K}_h(x-X_t),\mathbb{K}_h(x-X_s))\mathrm{d}s\,\mathrm{d}t \\ &= \frac{2}{T}\int_0^T (T-u)\mathrm{Cov}(\mathbb{K}_h(x-X_u),\mathbb{K}_h(x-X_0))\mathrm{d}u \\ &= 2\int_0^T (1-\frac{u}{T})\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{K}_h(x-y)\mathbb{K}_h(x-z)g_u(y,z)\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}u \\ &\leq \int_{\mathbb{R}} |\mathbb{K}_h(x-y)| \sup_{z\in\mathbb{R}}\int_0^\infty |g_u(y,z)|\mathrm{d}u\,\mathrm{d}y \int_{\mathbb{R}} |\mathbb{K}_h(x-z)|\mathrm{d}z. \end{aligned}$$

In the proof of Proposition 3.2 we have shown that

$$\sup_{z\in\mathbb{R}}\int_0^\infty |g_u(y,z)|\mathrm{d}u \leq c(1+\mu(y)(1+f^*(y))).$$

It follows that

$$T\,\mathrm{Var}(\hat{\mu}_{h,T}(x)) \leq c\int_{\mathbb{R}} |\mathbb{K}_h(x-y)|(1+\mu(y)(1+f^*(y)))\mathrm{d}y,$$

since, by the definition of the kernel function,

$$\int_{\mathbb{R}} |\mathbb{K}_h(x-z)|\mathrm{d}z = \int_{x-h}^{x+h} |\mathbb{K}_h(x-z)|\mathrm{d}z \leq \|\mathbb{K}_h\|_\infty\, h \leq \frac{\|K\|_\infty}{h} h = \|K\|_\infty .$$

Then, by the definition of $\mathbb{K}_h$, we get that

$$\begin{aligned} &\int_{\mathbb{R}} |\mathbb{K}_h(x-y)|(1+\mu(y)(1+f^*(y)))\mathrm{d}y \\ &\qquad = \frac{1}{h}\int_{x-h}^{x+h} |K(\frac{x-y}{h})|(1+\mu(y)(1+f^*(y)))\mathrm{d}y \\ &\qquad \leq \|K\|_\infty \int_{-1}^{1} (1+\mu(x-h\tilde{y})(1+f^*(x-h\tilde{y})))d\tilde{y}, \end{aligned}$$

where we have applied the change of variable $\tilde{y} := \frac{x-y}{h}$. Now we observe that, if $|x-h\tilde{y}| \leq 1$, then $f^*(x-h\tilde{y})$ is bounded by construction. Otherwise, for $|x-h\tilde{y}| > 1$, we have

$$f^*(x-h\tilde{y}) = e^{\epsilon|x-h\tilde{y}|} \leq e^{\epsilon|x|}e^{\epsilon h|\tilde{y}|} \leq e^{\epsilon|x|}e^{\epsilon},$$

where in the last inequality we have used the fact that both h and $|\tilde{y}|$ are smaller than 1. Therefore, we have shown that

$$T\,\mathrm{Var}(\hat{\mu}_{h,T}(x)) \leq ce^{\epsilon|x|},$$

where c is independent of T, h and x. Finally, from the bias-variance decomposition (2.5) we obtain (2.3), which concludes the desired proof.

4.2. Proof of Theorem 2.3

We aim to apply Theorem 3.3. For this, we split the interval $[0,T]$ into n intervals $[t_{i-1},t_i]$, where $t_i = i\Delta$ for any $i \in \{0,\dots,n\}$, $n\Delta = T$, and $n = \lfloor T \rfloor$ with $T \geq 1$, which implies that $1 \leq \Delta < 2$.

For each $n \geq 1$ and $1 \leq r \leq m$, we consider the sequence $(Y_{n,i}^{(r)})_{i\geq 1}$ defined as

$$Y_{n,i}^{(r)} := \frac{1}{\sqrt{\Delta}}\left(\int_{t_{i-1}}^{t_i} \mathbb{K}_h(x_r - X_u)\mathrm{d}u - \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \mathbb{K}_h(x_r - X_u)\mathrm{d}u\right]\right),$$

for $x_r \in \mathbb{R}$. We denote by $Y_{n,i}$ the $\mathbb{R}^m$ valued random vector defined by $Y_{n,i} = (Y_{n,i}^{(1)},\dots,Y_{n,i}^{(m)})$. By construction,

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}} = \sqrt{T}(\hat{\mu}_{h,T}(x) - \mathbb{E}[\hat{\mu}_{h,T}(x)]),$$

where $\hat{\mu}_{h,T}(x) - \mathbb{E}[\hat{\mu}_{h,T}(x)]$ is the vector

$$(\hat{\mu}_{h,T}(x_1) - \mathbb{E}[\hat{\mu}_{h,T}(x_1)],\dots,\hat{\mu}_{h,T}(x_m) - \mathbb{E}[\hat{\mu}_{h,T}(x_m)]).$$

It is clear that $\mathbb{E}[Y_{n,i}] = 0$ for all $n \geq 1$ and $i \geq 1$. Moreover, for all $i \geq 1$, $1 \leq r \leq m$ and $n \geq 1$ we have

$$|Y_{n,i}^{(r)}| \leq \frac{1}{\sqrt{\Delta}}\|\mathbb{K}_h\|_\infty \Delta \leq \frac{\|K\|_\infty}{h(T)}\sqrt{2}.$$

We choose $h(T) := (\frac{1}{T})^{\frac{1}{2}-\epsilon} = (\frac{1}{n\Delta})^{\frac{1}{2}-\epsilon} \geq c(\frac{1}{n})^{(\frac{1}{2}-\epsilon)}$, for some $\epsilon \in (0,\frac{1}{2})$. Hence, assumption (i) holds true with $M_n := cn^{\frac{1}{2}-\epsilon}$. Concerning assumption (ii) we remark that, for any $i \geq 1$ and any $1 \leq r \leq m$,

$$\begin{aligned}\mathbb{E}[(Y_{n,i}^{(r)})^2] &= \mathrm{Var}\left(\frac{1}{\sqrt{\Delta}}\int_0^\Delta \mathbb{K}_h(x_r - X_u)\mathrm{d}u\right) = \mathrm{Var}(\sqrt{\Delta}\hat{\mu}_{h,\Delta}(x_r)) \\ &= \Delta\mathrm{Var}(\hat{\mu}_{h,\Delta}(x_r)) \leq \Delta\frac{c}{\Delta} = c,\end{aligned}$$

where in the last inequality we have used (4.1). We next check condition (iii). Let b_n be a sequence of integers such that $b_n \to \infty$ and $b_n \leq n$ for every n. For every $1 \leq r \leq m$ and $1 \leq s \leq m$, we have

$$\begin{aligned}
&\frac{1}{b_n}\mathbb{E}\left[\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right] = \frac{1}{\Delta b_n}\int_0^{\Delta b_n}\int_0^{\Delta b_n} \mathrm{Cov}(\mathbb{K}_h(x_r - X_u), \mathbb{K}_h(x_s - X_v))\mathrm{d}u\, dv \\
&= 2\int_0^{\Delta b_n}(1 - \frac{u}{\Delta b_n})\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{K}_h(x_r - z_1)\mathbb{K}_h(x_s - z_2)g_u(z_1, z_2)dz_1\, dz_2\, \mathrm{d}u \\
&= 2\int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^{\Delta b_n}(1 - \frac{u}{\Delta b_n})K(w_1)K(w_2)g_u(x_r - h(T)w_1, x_s - h(T)w_2)\mathrm{d}u\, dw_1\, dw_2,
\end{aligned}$$

where we have used Fubini's theorem and the change of variables $w_1 := \frac{x_r - z_1}{h(T)}$, $w_2 := \frac{x_s - z_2}{h(T)}$. Using dominated convergence and the fact that $h(T) \to 0$ for $T \to \infty$ and $\Delta b_n \to \infty$ for $n \to \infty$ as $\Delta \geq 1$, we obtain

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{b_n}\mathbb{E}\left[\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right] &= 2\int_{\mathbb{R}} K(w_1)\int_{\mathbb{R}} K(w_2)\int_0^{\infty} g_u(x_r, x_s)\mathrm{d}u\, dw_2\, dw_1 \\
&= 2\int_0^{\infty} g_u(x_r, x_s)\mathrm{d}u =: \sigma(x_r, x_s),
\end{aligned}$$

which proves (iii). Remark that it is possible to use dominated convergence theorem since we have shown in the proof of Proposition 3.2 that

$$\sup_{y\in\mathbb{R}}|g_u(x, y)| \leq c\left(u^{-1/2}\mathbf{1}_{\{u\leq 2\}} + \mu(x)(1 + f^*(x))e^{-\rho u}\mathbf{1}_{\{u>2\}}\right),$$

for some positive constants c and ρ. In particular, we have

$$\begin{aligned}
&|(1 - \frac{u}{\Delta b_n})K(w_1)K(w_2)g_u(x_r - h(T)w_1, x_s - h(T)w_2)1_{[0,b_n]}(u)1_{\mathbb{R}^2}(w_1, w_2)| \\
&\leq c\left(u^{-1/2}\mathbf{1}_{\{u\leq 2\}} + e^{\epsilon(|x_r|+|w_1|)}e^{-\rho u}\mathbf{1}_{\{u>2\}}\right)|K(w_1)K(w_2)| \in L^1(\mathbb{R}^+ \times \mathbb{R}^2),
\end{aligned}$$

as K has support on $[-1, 1]$.

We now check (iv). We remark that if a process is β-mixing, then it is also α-mixing and the following estimation holds (see Thm. 3 in Sect. 1.2.2 of [25])

$$\alpha_k \leq \beta_{Y_{n,i}}(k) = \beta_X(k) \leq ce^{-\gamma_1 k}.$$

Therefore, it suffices to show that there exists $\gamma_0 \in (1, \infty)$ such that

$$\sum_{k\geq 1} ke^{-k\gamma_1\frac{(\gamma_0-1)}{\gamma_0}} < \infty,$$

which is true for any $\gamma_0 > 1$, so (iv) is satisfied.

We are left to show (v). Set $f(\gamma_0) := \frac{{\gamma_0}^2}{(3\gamma_0-1)(2\gamma_0-1)}$ and observe that $f(1) = \frac{1}{2}$ and for $\gamma_0 > 1$, f is continuous, strictly decreasing, and $\frac{1}{6} < f(\gamma_0) < \frac{1}{2}$. Therefore, given $\epsilon \in (0, \frac{1}{2})$, there always exists $\gamma_0 > 1$ such that for all

$n \geq 1$,

$$n^{\frac{1}{2}-\epsilon} \leq n^{f(\gamma_0)}.$$

Thus, condition (v) is satisfied. We can then apply Theorem 3.3 which directly leads us to (2.6) and concludes the desired proof.

4.3. Proof of Theorem 2.4

The proof of of Theorem 2.4 follows as the proof of the lower bound for $d \geq 3$ obtained in Theorem 3 of [2]. Therefore, we will only explain the main steps and the principal differences.

Step 1 The first step consists in showing that given a density function f, we can always find a drift function b_f such that f is the unique invariant density function of equation (2.7) with drift coefficient $b = b_f$. We give the statement and proof in dimension $d = 1$, as in Propositions 2 and 3 of [2] it is only done for $d \geq 2$.

Proposition 4.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^2$ positive probability density satisfying the following conditions*

1. $\lim_{y\to\pm\infty} f(y) = 0$ *and* $\lim_{y\to\pm\infty} f'(y) = 0$.
2. *There exist $\hat{c}_1 > 0$ and $0 < \epsilon < \frac{\epsilon_0}{|\gamma|}$, where ϵ_0 is as in Assumption* **A4** *such that, for any $y, z \in \mathbb{R}$,*

$$f(y \pm z) \leq \hat{c}_1 e^{\epsilon|z|} f(y).$$

3. *For $\epsilon > 0$ as in 2. there exists $\hat{c}_2(\epsilon) > 0$ such that*

$$\sup_{y<0} \frac{1}{f(y)} \int_{-\infty}^{y} f(w)\mathrm{d}w < \hat{c}_2 \quad \text{and} \quad \sup_{y>0} \frac{1}{f(y)} \int_{y}^{\infty} f(w)\mathrm{d}w < \hat{c}_2.$$

4. *There exists $0 < \tilde{\epsilon} < \frac{a^2}{2\gamma^2 c_4 \hat{c}_2 \hat{c}_4 \hat{c}_1}$ and $R > 0$ such that for any $|y| > R$, $\frac{f'(y)}{f(y)} \leq -\tilde{\epsilon}\mathrm{sgn}(y)$, where c_4 is as in Assumption* **A4**. *Moreover, there exists $\hat{c}_3$ such that for any $y \in \mathbb{R}$, $|f'(y)| \leq \hat{c}_3 f(y)$.*
5. *For any $y \in \mathbb{R}$ and $\tilde{\epsilon}$ as in 4. $|f''(y)| \leq \hat{c}_4 \tilde{\epsilon}^2 f(y)$.*

Then there exists a bounded Lipschitz function b_f which satisfies **A2** *such that f is the unique invariant density to equation (2.7) with drift coefficient $b = b_f$.*

Proof. Let A_d be the discrete part of the generator of the diffusion process X solution of (2.7) and let A_d^* its adjoint. We define b_f as

$$b_f(x) = \begin{cases} \frac{1}{f(x)} \int_{-\infty}^{x} (\frac{1}{2}a^2 f''(w) + A_d^* f(w))\mathrm{d}w, & \text{if } x < 0; \\ -\frac{1}{f(x)} \int_{x}^{\infty} \frac{1}{2}a^2 f''(x)(w) + A_d^* f(w)\mathrm{d}w, & \text{if } x > 0, \end{cases}$$

where

$$A_d^* f(x) = \int_{\mathbb{R}} [f(x - \gamma z) - f(x) + \gamma z f'(x)] F(z)\mathrm{d}z.$$

Then, following Proposition 3 in [2], one can check that b_f is bounded, Lipschitz, and satisfies **A2**. Moreover, if we replace b by b_f in equation (2.7), then f is the unique invariant density. □

Step 2 The second step consists in defining two probability density functions f_0 and f_1 in $\mathcal{H}_1(\beta, \mathcal{L})$.

We first define $f_0(y) = c_\eta f(\eta|y|)$, where $\eta \in (0, \frac{1}{2})$, c_η is such that $\int f_0 = 1$, where f is defined as follows. We first consider the piecewise function

$$g(x) = \begin{cases} e^{-|x|}, & \text{if } |x| \geq 1 \\ e^{-4(|x|-\frac{1}{2})^2}, & \text{if } \frac{1}{2} < |x| < 1 \\ 1, & \text{if } |x| \leq \frac{1}{2}. \end{cases}$$

Observe that g is continuous, satisfies $\frac{1}{2}e^{-|x|} \leq g(x) \leq 2e^{-|x|}$ for all $x \in \mathbb{R}$, and each piece belongs to C^∞ and has bounded derivatives. We define f as a $\mathcal{C}^\infty$ approximation of g, with bounded derivatives of all orders and satisfying

$$\frac{1}{2}e^{-|x|} \leq f(x) \leq 2e^{-|x|}, \quad |f'(|x|)| \leq 5e^{-|x|}, \quad \text{and} \quad |f''(|x|)| \leq 14e^{-|x|}. \tag{4.1}$$

Observe that the two latter inequalities are satisfied by g piecewise.

It is easy to see that η can be chosen small enough so that $f_0 \in \mathcal{H}_1(\beta, \mathcal{L})$. Indeed, first, it is clear that all the derivatives of f_0 can be bounded by the constant $\mathcal{L}$ for η small enough. Furthermore, the following bounds hold true for any x and t in $\mathbb{R}$

$$\begin{aligned} &|D^{\lfloor\beta\rfloor} f_0(x+t) - D^{\lfloor\beta\rfloor} f_0(x)| \\ &\leq |D^{\lfloor\beta\rfloor} f_0(x+t) - D^{\lfloor\beta\rfloor} f_0(x)|^{\beta-\lfloor\beta\rfloor} (2 \left\| D^{\lfloor\beta\rfloor} f_0 \right\|_\infty)^{1-(\beta-\lfloor\beta\rfloor)} \\ &\leq \left\| D^{\lfloor\beta\rfloor+1} f_0 \right\|_\infty^{\beta-\lfloor\beta\rfloor} (2 \left\| D^{\lfloor\beta\rfloor} f_0 \right\|_\infty)^{1-(\beta-\lfloor\beta\rfloor)} |t|^{\beta-\lfloor\beta\rfloor}. \end{aligned}$$

Again, it suffices to choose η small enough to ensure that

$$\left\| D^{\lfloor\beta\rfloor+1} f_0 \right\|_\infty^{\beta-\lfloor\beta\rfloor} (2 \left\| D^{\lfloor\beta\rfloor} f_0 \right\|_\infty)^{1-(\beta-\lfloor\beta\rfloor)} \leq \mathcal{L},$$

which shows that $f_0 \in \mathcal{H}_1(\beta, \mathcal{L}) \subset \mathcal{H}_1(\beta, 2\mathcal{L})$.

We also ask that the constant c_4 in Assumption **A4** is such that

$$c_4 < \frac{a^2}{2\gamma^2 4^2 28}. \tag{4.2}$$

This means that the jumps have to integrate an exponential function. The bound depends on the coefficients a and γ and so it depends only on the model.

Under the conditions above it is easy to see that f_0 satisfies the assumptions of Proposition 4.1 with $\hat{c}_1 = 4$, $\epsilon = \eta$, $\hat{c}_2 = \frac{4}{\eta}$, $R = \frac{1}{\eta}$, $\tilde{\epsilon} = \eta$ $\hat{c}_3 = 28\eta$, and $\hat{c}_4 = 28$. Indeed, point 1 of Proposition 4.1 clearly holds true from the definition of f_0. To show the second point we observe that, thanks to (4.1), we have

$$f_0(y \pm z) = c_n f(\eta|y \pm z|) \leq 2c_n e^{-\eta|y|} e^{\eta|z|} \leq 4 f_0(y) e^{\eta|z|},$$

which implies point 2 with $\hat{c}_1 = 4$ and $\epsilon = \eta$, since we can choose η small enough to make the condition on ϵ satisfied. In order to prove point 3 we use again (4.1). It follows that, for any $y < 0$,

$$\frac{1}{f_0(y)} \int_{-\infty}^{y} f_0(w) \mathrm{d}w = \frac{1}{c_n f(\eta|y|)} \int_{-\infty}^{y} c_n f(\eta|w|) \mathrm{d}w$$

$$\leq 2e^{\eta|y|}\int_{-\infty}^{y} 2e^{-\eta w}\mathrm{d}w = 4e^{\eta|y|}\frac{e^{-\eta|y|}}{\eta} = \frac{4}{\eta}.$$

For $y > 0$ an analogous reasoning applies. Thus, f_0 satisfies the third point with $\hat{c}_2(\epsilon) = \hat{c}_2(\eta) = \frac{4}{\eta}$. For the fourth point, we observe that, for $|y| > \frac{1}{\eta}$,

$$f_0(y) = -\eta\,\mathrm{sgn}(y) f_0(y).$$

That is, the first part of point 4 holds true for $|y| > R$, taking $R = \frac{1}{\eta}$ and $\tilde{\epsilon} = \eta$. Moreover, we observe that using (4.1) we have, for $k = 1, 2$,

$$|f_0^{(k)}(y)| = |c_n f^{(k)}(\eta|y|)| \leq 14 c_n \eta^k e^{-\eta|y|} \leq 28\eta^k f_0(y).$$

This shows that both the fourth and the fifth points hold true, with $\hat{c}_3(\eta) = 28\eta$ and $\hat{c}_4 = 28$. Finally, we need to check that the condition on $\tilde{\epsilon}$ given in the fourth point which writes as

$$\tilde{\epsilon} = \eta < \frac{a^2}{2\gamma^2 c_4 \hat{c}_2 \hat{c}_4 \hat{c}_1} = \frac{a^2\,\eta}{2\gamma^2 c_4\, 4\, 28\, 4},$$

which is equivalent to (4.2). Hence, f_0 satisfies all the assumptions in Proposition 4.1.

Therefore, $b_0 := b_{f_0}$ belongs to $\Sigma(\beta, \mathcal{L})$. Recall that b_0 belongs to $\Sigma(\beta, \mathcal{L})$ if and only if f_0 belongs to $\mathcal{H}_1(\beta, 2\mathcal{L})$ and b_0 is bounded, Lipschitz and satisfies the drift condition **A2**.

We next define

$$f_1(x) = f_0(x) + \frac{1}{M_T}\hat{K}\left(\frac{x - x_0}{H}\right), \tag{4.3}$$

where $x_0 \in \mathbb{R}$ is fixed and $\hat{K} : \mathbb{R} \to \mathbb{R}$ is a C^∞ function with support on $[-1, 1]$ such that

$$\hat{K}(0) = 1, \quad \int_{-1}^{1} \hat{K}(z)\mathrm{d}z = 0.$$

Here H is a constant and M_T will be calibrated later and satisfies that $M_T \to \infty$ as $T \to \infty$. Observe that in the proof of the lower bound for the case $d \geq 3$ presented in [2], H is a function of T converging to 0 as $T \to \infty$. For the case $d = 1$, it suffices to chose it constant and we will see below that the same computations done in [2] will work in this case and it suffices to calibrate M_T.

Then it can be shown as in Lemma 3 of [2] that if for all $\epsilon > 0$ and T sufficiently large,

$$\frac{1}{M_T} \leq \epsilon H^\beta \qquad \text{and} \qquad \frac{1}{H} = o(M_T) \tag{4.4}$$

as $T \to \infty$, then if $\epsilon > 0$ is small enough we have that $b_1 := b_{f_1}$ belongs to $\Sigma(\beta, \mathcal{L})$ for T sufficiently large. Indeed, on one hand, (4.4) is clearly true when H is a constant. On the other hand, the same argument used in Lemma 3 of [2] applies to show that f_1 belongs to $\mathcal{H}_1(\beta, 2\mathcal{L})$ when H is a constant, up to choose ϵ in (4.4) smaller than a constant depending on $\mathcal{L}$ and H.

Step 3 As $b_0, b_1 \in \Sigma(\beta, \mathcal{L})$, we can write

$$R(\tilde{\mu}_T(x_0)) \geq \frac{1}{2}\mathbb{E}_1^{(T)}[(\tilde{\mu}_T(x_0) - f_1(x_0))^2] + \frac{1}{2}\mathbb{E}_0^{(T)}[(\tilde{\mu}_T(x_0) - f_0(x_0))^2],$$

where $\mathbb{E}_i^{(T)}$ denotes the expectation with respect to b_i. Then, following as in [2], using Girsanov's formula, we can show that if

$$\sup_{T\geq 0} T\frac{1}{M_T^2 H} < \infty, \tag{4.5}$$

then for sufficiently large T,

$$R(\tilde{\mu}_T(x_0)) \geq \frac{C}{8\lambda}\frac{1}{M_T^2}, \tag{4.6}$$

where the constants C and λ are as in Lemma 4 of [2] and they do not depend on the point x_0. We finally look for the larger choice of $\frac{1}{M_T^2}$ for which both (4.4) and (4.5) hold true. It suffices to choose $M_T = \sqrt{T}$ to conclude the proof of Theorem 2.4.

Remark 4.2. The two hypothesis method used above does not work to prove the 2-dimensional lower bound of Theorem 2.5. Indeed, following as above, we can define

$$f_1(x) = f_0(x) + \frac{1}{M_T}\hat{K}\left(\frac{x-x_0}{H_1(T)}\right)\hat{K}\left(\frac{x-x_0}{H_2(T)}\right).$$

Then, it is possible to show that (4.6) still holds and, therefore, we should take M_T such that $\frac{1}{M_T^2} = \frac{\log T}{T}$. On the other hand, condition (4.5) now becomes

$$\sup_{T\geq 0} T\frac{1}{M_T^2}\left(\frac{H_2(T)}{H_1(T)} + \frac{H_1(T)}{H_2(T)}\right) < \infty.$$

The optimal choice of the bandwidth is achieved for $H_2(T) = H_1(T)$ which yields to $\sup_{T\geq 0} T\frac{1}{M_T^2} < \infty$, which is clearly not satisfied when $\frac{1}{M_T^2} = \frac{\log T}{T}$.

4.4. Proof of Theorem 2.5

We will apply Lemma 3.4 with $\psi := v\sqrt{\frac{\log T}{T}}$, where $v > 0$ is fixed. As above we divide the proof into three steps.

Step 1 As in the one-dimensional case, the first step consists in showing that given a density function f, we can always find a drift function b_f such that f is the unique invariant density function of equation (2.7) with drift coefficient $b = b_f$, which is proved in Propositions 2 and 3 of [2]. We remark that condition (2.9) is needed in Proposition 4.3 to ensure that the terms on the diagonal of the volatility coefficient a dominate on the others, which is crucial to get that b_f satisfies the drift condition **A2**.

Step 2 We next define the probability density $f_0 \in \mathcal{H}_2(\beta, \mathcal{L})$, the finite set J_T, and the set of probability densities $\{f_j,\, j \in J_T\} \subset \mathcal{H}_2(\beta, \mathcal{L})$ needed in order to apply Lemma 3.4.

We first define f_0 as π_0 in Section 7.2 of [2], which is the two-dimensional version of f_0 defined in the proof of Theorem 2.4, that is,

$$f_0(x) = c_\eta f(\eta(aa^T)_{11}^{-1}|x_1|)f(\eta(aa^T)_{22}^{-1}|x_2|), \quad x = (x_1, x_2) \in \mathbb{R}^2, \tag{4.7}$$

where f is as in Step 2 of the proof of Proposition 4.1. The density f_0 belongs to $\mathcal{H}_2(\beta, \mathcal{L})$ by construction.

We then set

$$J_T := \left\{1, \dots, \lfloor \frac{1}{\sqrt{H_1}} \rfloor\right\} \times \left\{1, \dots, \lfloor \frac{1}{\sqrt{H_2}} \rfloor\right\}, \tag{4.8}$$

where in order to lighten the notation we will write H_1 and H_2 for $H_1(T)$ and $H_2(T)$, respectively, which are two quantities that converge to 0 as $T \to \infty$ and need to be calibrated.

Finally, for $j := (j_1, j_2) \in J_T$, we define $x_j := (x_{j,1}, x_{j,2}) = (2j_1H_1, 2j_2H_2)$ and we set

$$f_j(x) := f_0(x) + 2v\sqrt{\frac{\log T}{T}}\hat{K}\left(\frac{x_1 - x_{j,1}}{H_1}\right)\hat{K}\left(\frac{x_2 - x_{j,2}}{H_2}\right),$$

where recall that $v > 0$ is fixed and $\hat{K}$ is as in (4.3).

Acting as in Lemma 3 of [2], recalling that the rate $\frac{1}{M_T}$ therein is now replaced by $\sqrt{\frac{\log T}{T}}$ (see also points 1. and 3. in the proof of Prop. 4.3 below), it is easy to see that if there exists $\epsilon > 0$ sufficiently small such that for large T,

$$\sqrt{\frac{\log T}{T}} \le \epsilon H_1^{\beta_1}, \qquad \sqrt{\frac{\log T}{T}} \le \epsilon H_2^{\beta_2}, \tag{4.9}$$

then, for any $j \in J_T$ and large T, $b_j \in \Sigma(\beta, \mathcal{L})$. In particular, $f_j \in \mathcal{H}_2(\beta, \mathcal{L})$. Therefore, $\{f_j,\, j \in J_T\} \subset \mathcal{H}_2(\beta, \mathcal{L})$.

In order to evaluate the difference between f_j and f_k we remark first of all that, as $\hat{K}$ has support on $[-1, 1]$, $\prod_{l=1}^2 \hat{K}(\frac{x_l - x_{j,l}}{H_l})$ is different from 0 only if $|\frac{x_l - x_{j,l}}{H_l}| \le 1$ for any $l \in \{1, 2\}$. Then,

$$\begin{aligned}
\|f_j - f_k\|_\infty &\ge |f_j(x_j) - f_k(x_j)| \\
&= 2v\sqrt{\frac{\log T}{T}}\Big[\prod_{l=1}^2 \hat{K}\Big(\frac{x_{j,l} - x_{j,l}}{H_l}\Big) - \prod_{l=1}^2 \hat{K}\Big(\frac{x_{j,l} - x_{k,l}}{H_l}\Big)\Big] \\
&= 2v\sqrt{\frac{\log T}{T}}\prod_{l=1}^2 \hat{K}(0) = 2v\sqrt{\frac{\log T}{T}} = 2\psi,
\end{aligned}$$

where we have used that, as $j \neq k$, there is a $l_0 \in \{1, 2\}$ such that $l_0 \neq k_0$ and so in particular, by construction, $|j_{l_0} - k_{l_0}| \ge 1$. It follows that

$$|\frac{x_{j,l_0} - x_{k,l_0}}{H_{l_0}}| = |\frac{2j_{l_0}H_{l_0} - 2k_{l_0}H_{l_0}}{h_{l_0}}| \ge 2$$

and so the kernel evaluated in this point is null. This proves the first condition of Lemma 3.4.

Step 3 We are left to show the remaining conditions of Lemma 3.4. The absolute continuity $\mathbb{P}_j^{(T)} \ll \mathbb{P}_0^{(T)}$ and the expression for $\frac{d\mathbb{P}_j^{(T)}}{d\mathbb{P}_0^{(T)}}(X^T)$ are both obtained by Girsanov formula, as in Lemma 4 of [2]. We have,

$$KL(\mathbb{P}_j^{(T)}, \mathbb{P}_0^{(T)}) = \mathbb{E}_j^{(T)}\left[\log\left(\frac{f_j}{f_0}(X^T)\right)\right] + \frac{1}{2}\mathbb{E}_j^{(T)}\left[\int_0^T |a^{-1}(b_0(X_u) - b_j(X_u))|^2 \mathrm{d}u\right],$$

where the law of $X^T = (X_t)_{t\in[0,T]}$ under $\mathbb{P}_j^{(T)}$ is the one of the solution to equation (2.7) with $b = b_0$.

By the definition of the f_j's it is easy to see that the first term is $o(1)$ as $T \to \infty$. In fact, as $\hat{K}$ is supported in $[-1,1]$,

$$\mathbb{E}_j^{(T)}\left[\log\left(\frac{f_j}{f_0}(X^T)\right)\right] = \int_{\mathbb{R}^2} \log\left(1 + \frac{2v\sqrt{\frac{\log T}{T}}\hat{K}\left(\frac{x_1 - x_{j,1}}{H_1}\right)\hat{K}\left(\frac{x_2 - x_{j,2}}{H_2}\right)}{f_0(x)}\right) f_0(x)\mathrm{d}x$$
$$\leq \left|\log\left(1 + c_* v\sqrt{\frac{\log T}{T}}\|\hat{K}\|_\infty^2\right)\right|,$$

which tends to zero as $T \to \infty$, where $c_* := \frac{8}{c_\eta}e^{4\eta k}$, c_η is the constant of normalization introduced in the definition of f_0, and $k := \max_{i=1,2}(aa^T)^{-1}_{ii}$. In fact, this follows from the definition of f_0 in (4.7). Since $f(x) \geq \frac{1}{2}e^{-|x|}$, we obtain

$$\frac{1}{f_0(x)} \leq \frac{1}{c_\eta}\frac{2}{e^{-\eta(aa^T)^{-1}_{11}|x_1|}}\frac{2}{e^{-\eta(aa^T)^{-1}_{22}|x_2|}} \leq \frac{4}{c_\eta}e^{\eta k(|H_1|+|x_{j,1}|+|H_2|+|x_{j,2}|)},$$

where we have also used the fact that, as $\hat{K}$ is supported in $[-1,1]$, we have $x \in [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2]$. Finally, by the definition of x_j and the fact that $H_i \to 0$ as $T \to \infty$ for $i = 1,2$ (and so for T large enough they are smaller than 1), we get

$$\frac{1}{f_0(x)} \leq \frac{4}{c_\eta}e^{4\eta k} \quad \text{for any } x \in [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2]. \tag{4.10}$$

Regarding the second term, using the stationarity of the process X^T, we have

$$\mathbb{E}_j^{(T)}\left[\int_0^T |a^{-1}(b_0(X_u) - b_j(X_u))|^2\mathrm{d}u\right] = T\int_{\mathbb{R}^2}|a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)\mathrm{d}x.$$

Then, the following asymptotic bound will be proved at the end of this Section.

Proposition 4.3. *For T large enough,*

$$\int_{\mathbb{R}^2}|a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)\mathrm{d}x \leq 64\frac{e^{8\eta k}}{c_\eta^2}k^2v^2H_1H_2\left(\frac{1}{H_1} + \frac{1}{H_2}\right)^2\frac{\log T}{T}.$$

Taking the optimal choice for the bandwidth in Proposition 4.3, which is $H_1 = H_2$, we get that

$$\int_{\mathbb{R}^2}|a^{-1}(b_0(x) - b_j(x))|^2 f_0(x)\mathrm{d}x \leq 64\frac{e^{8\eta k}}{c_\eta^2}k^2v^2 4\frac{\log T}{T}.$$

In particular, after having ordered $\beta_1 \leq \beta_2$, we choose $H_1 = H_2 = (\frac{\log T}{T})^\alpha$ with $\alpha \leq \frac{1}{2\beta_2} = (\frac{1}{2\beta_1} \wedge \frac{1}{2\beta_2})$ so that condition (4.9) is satisfied. We therefore get

$$KL(\mathbb{P}_j^{(T)}, \mathbb{P}_0^{(T)}) \leq 128\frac{e^{8\eta k}}{c_\eta^2}k^2\,v^2\log T \leq 128\frac{e^{8\eta k}}{c_\eta^2\alpha}k^2\,v^2\log(|J_T|),$$

being the last estimation a consequence of the fact that, by construction,

$$\log(|J_T|) \geq \alpha \log\left(\frac{T}{\log T}\right) = \alpha \log(T)(1 + o(1)).$$

It is therefore enough to choose v such that $128\frac{e^{8\eta k}}{c_\eta^2 \alpha} k^2 v^2 < \frac{1}{8}$ (ie $v^2 < \frac{c_\eta^2 \alpha}{1024\, k^2 e^{8\eta k}}$) and apply Lemma 3.4 to conclude the proof of Theorem 2.5.

4.5. Proof of Proposition 4.3

The proof of Proposition 4.3 follows similarly as Proposition 4 of [2]. Indeed, we first define the set

$$K_T^j := [x_{j,1} - H_1, x_{j,1} + H_1] \times [x_{j,2} - H_2, x_{j,2} + H_2],$$

where we recall that we write H_1 and H_2 for $H_1(T)$ and $H_2(T)$, respectively, in order to simplify the notation. Then we show the following points for T large enough:

1. There exists a constant $c > 0$ such that, for any x in the complementary set of K_T, that we denote as $K_T^{j\,c}$, and for any $i \in \{1, 2\}$,

$$|b_j^i(x) - b_0^i(x)| \leq c\, v \sqrt{\frac{\log T}{T}}.$$

2. There exists a constant $c > 0$ such that, for any $i \in \{1, 2\}$,

$$\int_{K_T^{j\,c}} |b_j^i(x) - b_0^i(x)| f_0(x) \mathrm{d}x \leq c\, v \sqrt{\frac{\log T}{T}} H_1 H_2.$$

3. For any $x \in K_T^j$ and $i \in \{1, 2\}$,

$$|b_j^i(x) - b_0^i(x)| \leq \frac{8}{c_\eta} e^{4\eta k} k v \sqrt{\frac{\log T}{T}} \left(\frac{1}{H_1} + \frac{1}{H_2}\right).$$

The proof of the first two points follows exactly the one in Proposition 4 of [2], remarking that

$$d_T(x) := \pi_1(x) - \pi_0(x) = \frac{1}{M_T} \prod_{l=1}^{d} K\left(\frac{x_l - x_0^l}{h_l(T)}\right)$$

in [2] is now replaced by

$$d_T^j(x) := f_j(x) - f_0(x) = 2v \sqrt{\frac{\log T}{T}} \hat{K}\left(\frac{x_1 - x_{j,1}}{H_1}\right) \hat{K}\left(\frac{x_2 - x_{j,2}}{H_2}\right),$$

and the set

$$K_T := [x_0^1 - h_1(T), x_0^1 + h_1(T)] \times \cdots \times [x_0^d - h_d(T), x_0^d + h_d(T)]$$

introduced in [2] is now replaced by K_T^j. We recall that K and $\hat{K}$ are exactly the same kernel function. The proof of Proposition 4 of [2] is based on the fact that $d_T(x)$ and its derivatives are null for $x \in K_T^c$. In the same

way, $d_T^j(x)$ and its derivatives are null for $x \in K_T^{j\,c}$. Then, acting as in [2], it is easy to see that the first two points above hold true.

Comparing the third point above with the third point of Proposition 4 of [2], it is clear that our goal is to show that the constant c that appears in the third point of Proposition 4 of [2] is explicit and equal to $\frac{8}{c_\eta}e^{4\eta k}k$ when $d = 2$. Keeping the notation in [2], we first introduce the following quantities:

$$\tilde{I}_1^i[f_0](x) := \frac{1}{2}\sum_{j=1}^{2}(aa^T)_{ij}\frac{\partial f_0}{\partial x_j}(x), \qquad \tilde{I}_2^i[f_0](x) = \int_{-\infty}^{x_i} A_{d,i}^* f_0(w_i)\mathrm{d}w.$$

We moreover introduce the notation

$$\tilde{I}^i[f_0](x) = \tilde{I}_1^i[f_0](x) + \tilde{I}_2^i[f_0](x).$$

According with the definition of b, we have

$$b_0^i(x) = \frac{1}{f_0(x)}\tilde{I}^i[f_0](x), \qquad b_j^i(x) = \frac{1}{f_j(x)}\tilde{I}^i[f_j](x).$$

Since the operator $f \to \tilde{I}^i[f]$ is linear, we deduce that

$$b_j^i(x) = \frac{1}{f_j(x)}\tilde{I}^i[f_j](x) = \frac{1}{f_j(x)}\tilde{I}^i[f_0](x) + \frac{1}{f_j(x)}\tilde{I}^i[d_T^j](x). \tag{4.11}$$

Therefore,

$$b_j^i - b_0^i = (\frac{1}{f_j} - \frac{1}{f_0})\tilde{I}^i[f_0] + \frac{1}{f_j}\tilde{I}^i[d_T^j] = \frac{f_0 - f_j}{f_j}\frac{1}{f_0}\tilde{I}^i[f_0] + \frac{1}{f_j}\tilde{I}^i[d_T^j] = \frac{d_T^j}{f_j}b_0^i + \frac{1}{f_j}\tilde{I}^i[d_T^j].$$

We need to evaluate such a difference on the compact set K_T^j. For this, we will use that fact that $f_j = f_0 + d_T^j$, and obtain a lower bound away from 0. Specifically, from the definition of d_T^j, we get

$$\left\|d_T^j\right\|_\infty \leq 2v\sqrt{\frac{\log T}{T}}\|\hat{K}\|_\infty^2 = 2v\sqrt{\frac{\log T}{T}}. \tag{4.12}$$

In particular,

$$f_j \geq f_0 - |d_T^j| \geq f_0 - 2v\sqrt{\frac{\log T}{T}} \geq \frac{f_0}{2},$$

since $\sqrt{\frac{\log T}{T}} \to 0$ as $T \to \infty$, so for T large enough we have $2v\sqrt{\frac{\log T}{T}} \leq \frac{f_0}{2}$. Then, for any $x \in K_T^j$, using (4.10) we have

$$\frac{1}{f_j(x)} \leq \frac{2}{f_0} \leq \frac{8}{c_\eta}e^{4\eta k}.$$

Moreover, as b_0 is bounded, we deduce that for all $x \in K_T^j$,

$$|b_j^i(x) - b_0^i(x)| \leq \frac{16v}{c_\eta} e^{4\eta k} \left\|b_0^i\right\|_\infty \sqrt{\frac{\log T}{T}} + \frac{8e^{4\eta k}}{c_\eta} \tilde{I}^i[d_T^j](x). \tag{4.13}$$

We therefore need to evaluate $\tilde{I}^i[d_T^j](x) = \tilde{I}_1^i[d_T^j](x) + \tilde{I}_2^i[d_T^j](x)$ on K_T^j. As

$$\left\|\frac{\partial d_T^j}{\partial x_j}\right\|_\infty \leq \frac{2v}{H_j}\sqrt{\frac{\log T}{T}}, \tag{4.14}$$

it clearly follows that

$$\tilde{I}_1^i[d_T]^j(x) \leq 2kv\sqrt{\frac{\log T}{T}}\left(\frac{1}{H_1} + \frac{1}{H_2}\right). \tag{4.15}$$

Regarding $\tilde{I}_2^i[d_T^j](x)$, we can act exactly as in the third point of Proposition 4 of [2]. As $x \in K_T^j$, $x_i \in [x_{j,i} - H_i, x_{j,i} + H_i]$ for $i = 1, 2$. Therefore, using also the definition of d_T^j, the first integral is between $x_{j,i} - H_i$ and x_i. We enlarge the domain of integration to $[x_{j,i} - H_i, x_{j,i} + H_i]$ and then, appealing to (4.12) and (4.14) and the fact that the intensity of the jumps is finite, we get

$$\begin{aligned}
|\tilde{I}_2^i[d_T^j](x)| &\leq \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \int_{\mathbb{R}^2} |d_T^j(\tilde{w}_i) - d_T^j(\tilde{w}_{i-1}) + (\gamma \cdot z)_i \frac{\partial}{\partial x_i} d_T^j(w_i)| F(z)\mathrm{d}z\mathrm{d}w \\
&\leq 2\left(\int_{\mathbb{R}^2} F(z)\mathrm{d}z\right) \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \left\|d_T^j\right\|_\infty \mathrm{d}w \\
&\quad + \int_{x_{j,i}-H_i}^{x_{j,i}+H_i} \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} |(\gamma \cdot z)_i| \left\|\frac{\partial d_T^j}{\partial x_i}\right\|_\infty F(z)\mathrm{d}z\mathrm{d}w \\
&\leq cH_i\sqrt{\frac{\log T}{T}} + \frac{cH_i}{H_i}\sqrt{\frac{\log T}{T}},
\end{aligned}$$

for some $c > 0$. Using this together with (4.13) and (4.15) it follows that, for any $x \in K_T^j$,

$$\begin{aligned}
|b_j(x) - b_0(x)| &\leq c\sqrt{\frac{\log T}{T}} + \frac{8e^{4\eta k}}{c_\eta} kv\sqrt{\frac{\log T}{T}}\left(\frac{1}{H_1} + \frac{1}{H_2}\right) \\
&\quad + cH_i\sqrt{\frac{\log T}{T}} + c\sqrt{\frac{\log T}{T}} \\
&\leq \frac{8e^{4\eta k}}{c_\eta} kv\sqrt{\frac{\log T}{T}}\left(\frac{1}{H_1} + \frac{1}{H_2}\right),
\end{aligned}$$

where the last inequality is a consequence of the fact that, $\forall i \in \{1, 2\}$, $H_i \to 0$ as $T \to \infty$ and so, for T large enough, all the terms are negligible when compared to the second one. Hence, the three points listed at the beginning of the proof hold true. We deduce that

$$\int_{\mathbb{R}^2} |b_0(x) - b_j(x)|^2 f_0(x)\mathrm{d}x$$

$$= \int_{K_T^j} |b_0(x) - b_j(x)|^2 f_0(x)\mathrm{d}x + \int_{K_T^{j\,c}} |b_0(x) - b_j(x)|^2 f_0(x)\mathrm{d}x$$
$$\leq c\, v^2 \frac{\log T}{T} H_1 H_2 + \frac{64 e^{8\eta k}}{c_\eta^2} k^2 v^2 \frac{\log T}{T} \left(\frac{1}{H_1} + \frac{1}{H_2}\right)^2 |K_T^j|.$$

We recall that $|K_T^j| = H_1 H_2$ and that, as $T \to \infty$, $H_i \to 0$. Thus, the first term is negligible compared to the second one. The desired result follows.

Acknowledgements. The authors would like to thank the anonymous referees for their helpful remarks that helped to improve the first version of the paper.

References

[1] C. Amorino and A. Gloter, Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. *J. Stat. Plan. Inference* **213** (2021) 106–129.
[2] C. Amorino, Rate of estimation for the stationary distribution of jump-processes over anisotropic Hölder classes. To appear *Electr. J. Stat.* (2022).
[3] C. Amorino, C. Dion, A. Gloter and S. Lemler, On the nonparametric inference of coefficients of self-exciting jump-diffusion. Preprint arXiv:2011.12387 (2020).
[4] D. Applebaum, Lévy processes and stochastic calculus. Cambridge University Press (2009).
[5] G. Banon, Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16** (1978) 380–395.
[6] G. Banon and H.T. Nguyen, Recursive estimation in diffusion model. *SIAM J. Control Optim.* **19** (1981) 676–685.
[7] O.E. Barndorff-Nielsen and N. Shephard, Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B, Stat. Methodol.* **63** (2001) 167–241.
[8] D. Blanke, Estimation de la densité pour des trajectoires non directement observables. *Publ. Inst. Statist. Univ. Paris* **40** (1996) 21–36.
[9] D. Bosq, Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Statist.* **25** (1997) 982–1000.
[10] D. Bosq, Nonparametric Statistics for Stochastic Processes (Second edition), *Lecture Notes Statist.*, 110. Springer-Verlag, New York (1998).
[11] D. Bosq, Minimax rates of density estimators for continuous time processes. *Sankhya. Ser. A* **60** (1998) 18–28.
[12] D. Bosq and Yu. Davydov, Local time and density estimation in continuous time. *Math. Methods Stat.* **8** (1999) 22–45.
[13] D. Bosq, F. de Merlevè and M. Peligrad, Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivar. Anal.* **68** (1999) 78–95.
[14] J.V. Castellana and M.R. Leadbetter, On smoothed probability density estimation for stationary processes. *Stoch. Process Appl.* **21** (1986) 179–193.
[15] Z.Q. Chen, E. Hu, L. Xie and X. Zhang, Heat kernels for non-symmetric diffusion operators with jumps. *J Differ Equ.* **263** (2017) 6576–6634.
[16] N. Cheze-Payaud, Nonparametric regression and prediction for continuous-time processes. *Publ. Inst. Statist. Univ. Paris* **38** (1994) 37–58.
[17] F. Comte and C. Lacour, Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré, Probab. Statis.* **49** (2013) 569–609.
[18] F. Comte and F. Merlevède, Super optimal rates for nonparametric density estimation via projection estimators. *Stoch. Process Appl.* **115** (2005) 797–826.
[19] S. Delattre, A. Gloter and N. Yoshida, Rate of Estimation for the Stationary Distribution of Stochastic Damping Hamiltonian Systems with Continuous Observations. Preprint arXiv:2001.10423 (2020).
[20] M. Delecroix, Sur l'estimation des densités d'un processus stationnaire á temps continu. *Publications de l'ISUP* **XXV** (1980) 17–39.
[21] Y.A. Kutoyants, Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* **1** (1998) 131–155.
[22] N. Dexheimer, C. Strauch and L. Trottner, Mixing it up: A general framework for Markovian statistics beyond reversibility and the minimax paradigm. Preprint arXiv:2011.00308 (2020).
[23] C. Dion and S. Lemler, Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. *Stat. Inference Stochastic Processes* **23** (2020) 489–515.
[24] S. Ditlevsen and P. Greenwood, The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. *J. Math. Biol.* **67** (2013) 239–259.
[25] P. Doukhan, Mixing: properties and examples (Vol. 85). Springer Science and Business Media (2012).
[26] A. Friedman, Partial Differential Equations of Parabolic Type. Prentice-Hall, Englewood Cliffs, N.J. (1964).
[27] B. Funke and E. Schmisser, Adaptive nonparametric drift estimation of an integrated jump diffusion process. *ESAIM: PS* **22** (2018) 236–260.

[28] R.Z. Has'minskii, Stability of differential equations. Sijthoff and Noordhoff, Germantown, MD (1980).
[29] R. Höpfner, M. Hoffmann and E. Löcherbach, Non-parametric estimation of the death rate in branching diffusions. *Scand. J. Stat.* **29** (2002) 665–692.
[30] A. Juditsky and A. Nemirovski, On nonparametric tests of positivity/monotonicity/convexity. *Ann. Stat.* (2002) 498–527.
[31] S.G. Kou, A jump-diffusion model for option pricing. *Manag. Sci.* **48** (2002) 1086–1101.
[32] Y.A. Kutoyants, Some problems of nonparametric estimation by observations of ergodic diffusion process. *Statist. Probab. Lett.* **32** (1997) 311–320.
[33] D. Lamberton and G. Pages, Recursive computation of the invariant distribution of a diffusion. *Bernoulli* **8** (2002) 367–405.
[34] F. Leblanc, Density estimation for a class of continuous time processes. *Math. Methods Stat.* **6** (1997) 171—-199.
[35] O.V. Lepski and V.G. Spokoiny, Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Stat.* (1997) 2512–2546.
[36] H. Masuda, Ergodicity and exponential beta - mixing bounds for multidimensional diffusions with jumps. *Stoch. Process. Appl.* **117** (2007) 35–56.
[37] R.C. Merton, Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.* **3** (1976) 125–144.
[38] H.T. Nguyen, Density estimation in a continuous-time Markov processes. *Ann. Statist.* **7** (1979) 341–348.
[39] F. Panloup, Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *Ann. Appl. Probab.* **18** (2008) 379–426.
[40] E. Schmisser, Non parametric estimation of the diffusion coefficients of a diffusion with jumps. *Stoch. Process. Appl.* **129** (2019) 5364–5405.
[41] C. Strauch, Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Stat.* **46** (2018) 3451–3480.
[42] A.B. Tsybakov, On nonparametric estimation of density level sets. *Ann. Stat.* **25** (1997) 948–969.
[43] A.B. Tsybakov, Introduction to nonparametric estimation. Springer Science and Business Media (2008).
[44] H. Van Zanten, Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. *Nonparametric Statist.* **13** (2001) 833–850.
[45] A.Yu. Veretennikov, On Castellana-Leadbetter's condition for diffusion density estimation. *Stat. Inference Stoch. Process.* **2** (1999) 1–9.