Minimax rate of estimation for invariant densities associated to continuous stochastic differential equations over anisotropic Holder classes.

Chiara Amorino* Arnaud Gloter †

November 25, 2021

Abstract

We study the problem of the nonparametric estimation for the density π of the stationary distribution of a d-dimensional stochastic differential equation $(X_t)_{t\in[0,T]}$. From the continuous observation of the sampling path on $[0,T]$, we study the estimation of $\pi(x)$ as T goes to infinity. For $d \geq 2$, we characterize the minimax rate for the $\mathbf{L}^2$-risk for the pointwise estimation over a class of anistropic Hölder functions π with regularity $\beta = (\beta_1, ..., \beta_d)$. For $d \geq 3$, our finding is that, having ordered the smoothness such that $\beta_1 \leq ... \leq \beta_d$, the minimax rate depends on the fact that $\beta_2 < \beta_3$ or $\beta_2 = \beta_3$. This rate is $(\frac{\log T}{T})^\gamma$ in the first case and $(\frac{1}{T})^\gamma$ in the second, for an explicit exponent γ depending on the dimension and on $\bar{\beta}_3$, the harmonic mean of the smoothness over the d directions after having removed β_1 and β_2, the smallest ones. We also show that kernel based estimators achieve the optimal minimax rate. Finally we show that, in the bi-dimensional case, the kernel density estimators achieves the rate $\frac{\log T}{T}$ and this is optimal in the minimax sense.

Non-parametric estimation, stationary measure, minimax rate, convergence rate, ergodic diffusion, anisotropic density estimation.

1 Introduction

In this work, we propose kernel density estimators for estimating the invariant density associated to a stochastic differential equation. More precisely, we study the inference for the d-dimensional process

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s, \quad t \in [0,T], \tag{1}$$

with $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $W = (W_t, t \geq 0)$ a d-dimensional Brownian motion.

Stochastic differential equations are very attractive for statisticians nowadays, as they model stochastic evolution as time evolves. These models have a variety of applications in many disciplines and emerge in a natural way in the study of many phenomena. We can found examples of these applications in astronomy [61], mechanics [46], physics [56], geology [28], ecology [38], biology [58] and epidemiology [10]. Some other examples are economics [16], mathematical finance [40], political analysis and social processes [20] as well as neurology [37], genetic analysis [48], cognitive psychology [64] and biomedical sciences [14].

Due to the importance of the stochastic differential equations, inference for such a model has been widely investigated. Authors studied parametric and non-parametric inference starting from continuous or discrete observations, under various asymptotic frameworks such as small diffusions asymptotics on a fixed time interval and long time interval for ergodic models. Some landmarks in the area are the books of Kutoyants [45], Iacus [41], Kessler et al [42] and Höpfner [39]. In the meantime, there have been a big amount of papers on the topic. Among them, we quote Comte et al [22], Dalalyan and Reiss [23], Genon-Catalot [31], Gobet et al [33], Hoffmann [36], Larédo [49]

*Université du Luxembourg, L-4364 Esch-Sur-Alzette, Luxembourg. The author gratefully acknowledges financial support of ERC Consolidator Grant 815703 “STAMFORD: Statistical Methods for High Dimensional Diffusions”.

†Laboratoire de Mathématiques et Modélisation d'Evry, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France.

and Yoshida [66].
The in-depth study on stochastic differential equations lead the way for statistical inference for more complicated models such as SDEs driven by Lévy processes (see for example [53]), diffusions with jumps ([5], [6], [52], [60]), stochastic partial differential equations ([2], [19]), diffusions with mixed effects ([57], [25]) and Hawkes processes ([11], [27], [4]).

In this paper we propose a kernel density estimator and we aim at finding the convergence rates of estimation for the stationary measure π associated to the process X solution to (1). Moreover, we will discuss the optimality of such rates.

The estimation of the invariant measure is a problem already widely faced in many different contexts by many authors, we quote [54], [24], [17], [67], [1] and [15] among them. One of the reasons why nowadays this problem attracts the attention of many statisticians is the huge amount of numerical methods connected to it, the Markov Chain Monte Carlo method above all. In [47] and [55], for example, it is possible to find an approximation algorithm for the computation of the invariant density while the analysis of the invariant distributions is used to to analyze the stability of stochastic differential systems in [35] and [15]. It can also be used in order to estimate the drift coefficient in a non-parametric way (see [51] and [59]).
As it is easy to see, the estimation of the invariant measure is at the same time a long-standing problem and a highly active current topic of research. For this reason, but not only, kernel estimators are widely employed as powerful tools. For example in [15] and [18] some kernel estimators are used to estimate the marginal density of a continuous time process, in [8], [7] and [50] they are used in a jump-diffusion framework where the last is, in particular, an application for the estimation of the volatility.

In a context closer to ours, in [23] and [62] kernel density estimators have been used for the study of the convergence rate for the estimation of the invariant density associated to a reversible diffusion process with unit diffusion part. In particular in [23] the authors prove, as a by-product of their study, some convergence rates for the pointwise estimation of the invariant density under isotropic Hölder smoothness constraints. The convergence rate they found are $\frac{\log T}{T}$ when the dimension d is equal to 2 and $(\frac{1}{T})^{\frac{2\beta}{2\beta+d-2}}$ for $d \geq 3$, where β is the common smoothness over the d different directions. Strauch extended their work in [62], by building adaptive kernel estimators to estimate the invariant density over anisotropic Hölder balls. They achieve the same convergence rates as in [23], up to replace β with $\bar{\beta}$, the harmonic mean over the d different directions. As the smoothness properties of elements of a function space may depend on the chosen direction of $\mathbb{R}^d$, the notion of anisotropy plays an important role.

In this work we aim at estimating the invariant density π by means of the kernel estimator $\hat{\pi}_{h,T}$ assuming to have the continuous record of the process X solution to (1) up to time T. We first of all prove the following upper bound for the mean squared error, for $d \geq 3$

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \lesssim \begin{cases} (\frac{\log T}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}} & \text{for } \beta_2 < \beta_3, \\ (\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}} & \text{for } \beta_2 = \beta_3, \end{cases} \tag{2}$$

where Σ is a class of coefficients for which the stationarity density has some prescribed regularity, $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and $\bar{\beta}_3$ is the harmonic mean over the smoothness after having removed the two smallest. In particular, we have $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2}\sum_{l\geq 3}\frac{1}{\beta_l}$ and so it clearly follows that $\bar{\beta}_3 \geq \bar{\beta}$. It yields that the convergence rate we find is in general faster than the one proposed in [62]. It is quite surprising that the presence of the logarithm in the convergence rate depends on whether or not $\beta_2 = \beta_3$. The reason why it happens is that we have different ways of bounding the variance. It derives in different possible degrees of freedom for the bandwidth $h = (h_1, ..., h_d)$, which allows us to get rid of the logarithm only in some particular cases (see Remark 1).
We also prove the following upper bound, in the bidimensional context:

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \lesssim \frac{\log T}{T}.$$

The convergence rate here above is the same as in [23] and [62]. However, the problem of the optimality of such a rate was still an open question, so far. The only results we are aware of, in this context, are some lower bounds for the sup-norm risk (see for example [62], or [8] for an analogous result in the jump-framework). In this paper we obtain a minimax lower bound on the L^2-risk for the pointwise estimation with the same rate $\frac{\log T}{T}$. To do that, we introduce the minimax risk

$$\mathcal{R}_T := \inf_{\tilde{\pi}_T} \sup_{(a,b) \in \Sigma} \mathbb{E}[(\tilde{\pi}_T(x_0) - \pi_{(a,b)}(x_0))^2],$$

where the infimum is taken over all the possible estimators of the invariant density. Then, we show that

$$\mathcal{R}_T \geq \frac{\log T}{T}.$$

It means that, for $d = 2$, it is not possible to propose an estimator whose convergence rates is better than the one we found for the kernel density estimator we proposed.
Regarding the case $d \geq 3$ we prove that, in general, the following lower bound holds true:

$$\mathcal{R}_T \geq (\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}.$$

Then, when $\beta_2 < \beta_3$, it is possible to improve the result here above by showing

$$\mathcal{R}_T \geq (\frac{\log T}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}.$$

These results highlight how crucial the condition $\beta_2 < \beta_3$ is. They imply that, on a class of diffusions X whose invariant density has the prescribed regularity, it is not possible to find an estimator with rates of estimation faster than in (2). It follows that the kernel density estimator we propose achieves the best possible convergence rates.

It is worth remarking that the upper bound for the mean squared error is based on an upper bound on the transition density and on the mixing properties of the process. Because of the last, the constants involved depend on the coefficients and it is in general very challenging to understand how the constant can be controlled uniformly. However, relying on the theory of Lyapounov-Poincaré as introduced in [12], we prove a mixing inequality uniform on the class of coefficient $(a, b) \in \Sigma$ (see Proposition 1).We remark that the Lyapounov-Poincaré method can be extended to get uniform mixing inequality also in more general frameworks, such as the jump context.
We therefore get a minimax upper bound on the risk of the estimator $\hat{\pi}_{h,T}$. We complement it by obtaining a minimax lower bound on the risk of the estimator $\hat{\pi}_{h,T}$ as well.

The outline of the paper is the following. In Section 2 we introduce the model and we provide the assumptions on it, while in Section 3 we propose the kernel estimator and we state the upper bounds for the mean squared error. Section 4 is devoted to the statements of the lower bounds which complement the result given in the previous section. In Section 5 we prove the upper bounds stated in Section 3 while Section 6 is devoted to the proof of the lower bounds. Some technical results are moreover showed in the appendix.

2 Model Assumptions

We want to estimate in a non parametric way the invariant density associated to a d-dimensional diffusion process X. In the sequel we assume that a continuous record of the process $X^T = \{X_t, 0 \leq t \leq T\}$ up to time T is available. The diffusion is a strong solution of the following stochastic differential equation:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s, \quad t \in [0, T], \tag{3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $W = (W_t, t \geq 0)$ is a d-dimensional Brownian motion. The initial condition X_0 and W and L are independent. We denote $\tilde{a} := a \cdot a^T$.
We denote with $|.|$ and $< ., . >$ respectively the Euclidian norm and the scalar product in $\mathbb{R}^d$, and for a matrix in $\mathbb{R}^d \otimes \mathbb{R}^d$ we denote its operator norm by $|.|$.

A1: *The functions $b(x)$ and $a(x)$ are bounded globally Lipschitz functions of class $\mathcal{C}^1$, such that for all $x \in \mathbb{R}^d$,*

$$|a(x)| \leq a_0, \ |b(x)| \leq b_0, \ |\frac{\partial}{\partial x_i} b(x)| \leq b_1, \ |\frac{\partial}{\partial x_i} a(x)| \leq a_1, \ \textit{for } i \in \{1, \dots, d\},$$

where $a_0 > 0,\ b_0 > 0,\ a_1 > 0,\ b_1 > 0$ *are some constants. Moreover, for some* $a_{min} > 0$,

$$a_{min}^2 \mathbb{I}_{d\times d} \leq \tilde{a}(x)$$

where $\mathbb{I}_{d\times d}$ *denotes the* $d \times d$ *identity matrix.*

A2 (Drift condition) :
There exist $\tilde{C} > 0$ *and* $\tilde{\rho} > 0$ *such that* $< x, b(x) > \leq -\tilde{C}|x|,\ \forall x : |x| \geq \tilde{\rho}$.

The assumptions here above involve the ergodicity of the process and they are needed in order to show the existence of a Lyapounov function. Hence, the process X admits a unique invariant distribution μ and the ergodic theorem holds (see Lemma 2 below). We suppose that the invariant probability measure μ of X is absolutely continuous with respect to the Lebesgue measure and from now on we will denote its density as π: $d\mu = \pi dx$.

As in several cases the regularity of some function $g : \mathbb{R}^d \to \mathbb{R}$ depends on the direction in $\mathbb{R}^d$ chosen, we decide to work under anisotropic smoothness constraints. In particular, we assume that the density π we want to estimate belongs to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ defined below.

Definition 1. *Let* $\beta = (\beta_1, ..., \beta_d)$, $\beta_i \geq 0$, $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$. *A function* $g : \mathbb{R}^d \to \mathbb{R}$ *is said to belong to the anisotropic Hölder class* $\mathcal{H}_d(\beta, \mathcal{L})$ *of functions if, for all* $i \in \{1, ..., d\}$,

$$\left\| D_i^k g \right\|_\infty \leq \mathcal{L}_i \qquad \forall k = 0, 1, ..., \lfloor \beta_i \rfloor,$$

$$\left\| D_i^{\lfloor \beta_i \rfloor} g(. + te_i) - D_i^{\lfloor \beta_i \rfloor} g(.) \right\|_\infty \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \qquad \forall t \in \mathbb{R},$$

for $D_i^k g$ *denoting the* k*-th order partial derivative of* g *with respect to the* i*-th component,* $\lfloor \beta_i \rfloor$ *denoting the largest integer strictly smaller than* β_i *and* $e_1, ..., e_d$ *denoting the canonical basis in* $\mathbb{R}^d$.

This leads us to consider a class of coefficients (a, b) for which the stationary density $\pi = \pi_{(a,b)}$ has some prescribed Hölder regularity.

Definition 2. *Let* $\beta = (\beta_1, ..., \beta_d)$, $\beta_i \geq 0$ *and* $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$, $0 < a_{min} \leq a_0$ *and* $a_1 > 0,\ b_0 > 0,\ b_1 > 0,\ \tilde{C} > 0,\ \tilde{\rho} > 0$.

We define $\Sigma(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$ *the set of couple of functions* (a, b) *where* $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ *and* $b : \mathbb{R}^d \to \mathbb{R}^d$ *are such that*

- *a and b satisfy Assumption A1 with the constants* $(a_{min}, a_0, a_1, b_0, b_1)$,
- *b satisfies Assumption A2 with the constants* $(\tilde{C}, \tilde{\rho})$,
- *the density* $\pi_{(a,b)}$ *of the invariant measure associated to the stochastic differential equation* (3) *belongs to* $\mathcal{H}_d(\beta, 2\mathcal{L})$.

In the sequel we will use repeatedly some known results about the transition density and the ergodicity of diffusion processes. For such a reason we state them.

Lemma 1. *Assume that Suppose that A1-A2 hold true. Then, there exists a transition density* $p_t(x, y)$ *such that, for* $T \geq 0$, *there exist* $c_T > 1$ *such that for any* $t \in [0, T]$ *and any pair of points* $x, y \in \mathbb{R}^d \times \mathbb{R}^d$, *we have*

$$c_T^{-1} t^{-\frac{d}{2}} e^{-\frac{|y-x|^2}{t a_{min}^2}} \leq p_t(x, y) \leq c_T t^{-\frac{d}{2}} e^{-\frac{|y-x|^2}{4 t a_0^2}}. \tag{4}$$

Moreover, the constants c_T *depends only on* $T,\ a_{min},\ a_0,\ a_1,\ b_0,\ b_1$ *and* d.

Lemma 2. *Suppose that A1-A2 hold true. Then, process* X *admits a unique invariant measure* π *and it is exponentially* β*-mixing.*

These results are classical. For the estimate in Lemma 1 see for example Proposition 5.1 in [32] (or Section A.2.3. in [9]), while the ergodic property stated in Lemma 2 can be found in Theorem 1 of [65].

3 Estimator and upper bounds

Given the observation X^T of a diffusion X, solution of (3), we propose to estimate the invariant density $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ by means of a kernel estimator. We therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \infty, \quad \text{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^l dx = 0, \tag{5}$$

for all $l \in \{0, ..., M\}$ with $M \geq \max_i \beta_i$.
For $j \in \{1, ..., d\}$, we denote by X_t^j the j-th component of X_t. A natural estimator of π at $x = (x_1, ..., x_d)^T \in \mathbb{R}^d$ in the anisotropic context is given by

$$\hat{\pi}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m}) du =: \frac{1}{T} \int_0^T \mathbb{K}_h(x - X_u) du. \tag{6}$$

The multi-index $h = (h_1, ..., h_d)$ is small. In particular, we assume $h_i < 1$ for any $i \in \{1, ..., d\}$.

We now want to discuss the convergence rates achieved from the estimator defined in (6) when a continuous record of the process X is available.

In order to get minimax upper bound on the risk of the estimator $\hat{\pi}_{h,T}$ we will need to apply a mixing inequality uniform on the class of coefficients $(a, b) \in \Sigma(\beta, \mathcal{L}, a_{\min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$. The following lemma will be useful in this sense. In the sequel, we will denote as $\mathbb{E}$ the expectation with rapport to the stationary process $(X_t)_{t \geq 0}$, whose invariant probability is $\pi = \pi_{(a,b)}$.

Lemma 3. *Let $\beta = (\beta_1, \ldots, \beta_d)$, $\beta_i \geq 0$, $\mathcal{L} = (\mathcal{L}_1, \ldots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$, $0 < a_{min} \leq a_0$, $a_1 > 0$, $b_0 > 0, b_1 > 0$, $\tilde{C} >$, $\tilde{\rho} > 0$. Let K be a compact subset of $\mathbb{R}^d$. There exists a constant $C > 0$ independent of $(a, b) \in \Sigma := \Sigma(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$ such that, for all measurable function $\varphi : K \to \mathbb{R}$ bounded, we have*

$$\left|\mathbb{E}[\varphi(X_t)\varphi(X_0)] - \pi(\varphi)^2\right| \leq Ce^{-t/C} \|\varphi\|_\infty^2 . \tag{7}$$

As immediate consequence of the fact that $\mathbb{K}$ is compactly supported we get the proposition below, which ensures the uniformity of the constants involved in the mixing properties of the process.

Proposition 1. *Let us consider the same notation as in Lemma 3. Then, there exist constants $\rho_{mix} > 0$ and $C_{mix} > 0$ such that $\forall h = (h_1, \ldots, h_d)$ with $h_i < 1$,*

$$\left|\mathbb{E}\left[\mathbb{K}_h(x - X_t)\mathbb{K}_h(x - X_0)\right] - \pi(\mathbb{K}_h)^2\right| \leq C_{mix} e^{-\rho_{mix} t} \|\mathbb{K}_h\|_\infty^2 . \tag{8}$$

Moreover, the constants C_{mix} and ρ_{mix} are uniform over the set of coefficients $(a, b) \in \Sigma$.

3.1 Upper bounds

As in low dimension and in high dimension the proposed estimator performs differently, we start considering what happens for $d \geq 3$.
In the sequel, we will denote as $\bar{\beta}_3$ the harmonic mean over the $d - 2$ largest components:

$$\frac{1}{\bar{\beta}_3} := \frac{1}{d-2} \sum_{l \geq 3} \frac{1}{\beta_l}. \tag{9}$$

In the next theorem we provide the convergence rate for the pointwise estimation of the invariant density in high dimension. Its proof can be found in Section 5.

Theorem 1. *Suppose that $d \geq 3$. Consider $0 < a_{min} \leq a_0$ and $a_1 > 0$, $b_0 > 0, b_1 > 0$, $\tilde{C} > 0$, $\tilde{\rho} > 0$ and denote $\Sigma := \Sigma(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$. If $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$, then there exist $c > 0$ and $T_0 > 0$ such that, for $T \geq T_0$, the optimal choice for the multidimensional bandwidth h yields the following convergence rates.*

- *If $\beta_2 < \beta_3$, then*

$$\sup_{(a,b) \in \Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \leq c(\frac{\log T}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}.$$

- *If otherwise $\beta_2 = \beta_3$, then*

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \leq c(\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}.$$

We underline that, with the notation $\mathbb{E}$, we always mean the expectation with respect to the invariant measure π linked to the coefficients (a, b).

In the isotropic context we have, in particular, $\bar{\beta}_3 = \beta$, where β is the common smoothness over the d directions. The convergence rate we derive is therefore the same as in [23], being equal to $(\frac{1}{T})^{\frac{2\beta}{2\beta+d-2}}$.

The asymptotic behavior of the estimator and so the proof of Theorem 1 is based on the standard bias -variance decomposition. Hence, we need an upper bound on the variance, as in next proposition.

Proposition 2. *Suppose that A1 - A2 hold for some a_{min}, a_0, a_1, b_0, b_1, $\tilde{C}$ and $\tilde{\rho}$; that $d \geq 3$ and that $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$. Suppose moreover that $\beta_1 \leq ... \leq \beta_d$, and let $k_0 = \max\{i \in \{1, \dots, d\} \mid \beta_1 = \dots = \beta_i\}$. If $\hat{\pi}_{h,T}$ is the estimator given in (6), then there exist , $c > 0$ and $T_0 > 0$ such that, for $T \geq T_0$, the following hold true.*

- *If $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$, then*

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \frac{\sum_{j=1}^{d} |\log(h_j)|}{\prod_{l\geq 3} h_l}. \tag{10}$$

- *If $k_0 \geq 3$, then*

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}} (\prod_{l\geq k_0+1} h_l)}. \tag{11}$$

- *If otherwise $k_0 = 1$ and $\beta_2 = \beta_3$, then*

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \frac{1}{\sqrt{h_2 h_3} \prod_{l\geq 4} h_l}.$$

Moreover, the constants c and T_0 are uniform over the set of coefficients

$$(a, b) \in \Sigma(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}).$$

We first of all shed light to the fact that, in the right hand side of (10), it would have been possible to remove the contribution of no matter which two bandwidths. We arbitrarily choose to remove the contribution of the first two as h_1 and h_2 are associated, in the bias term, to β_1 and β_2. Indeed, as we order the smoothness, β_1 and β_2 are the smallest ones and so they provide the strongest constraints.

One can remark that the bound on the variance for $\beta_2 < \beta_3$ in the proposition here above is the same as the one for jump diffusion processes (see Proposition 1 in [3]). The reason why it happens is that both propositions rely on the exponential β-mixing of the considered process and on a bound on the transition density. Comparing Lemma 1 with Lemma 1 of [7] it is possible to see that the upper bound for the transition density associated to a jump-diffusion consists in two terms: one derives from the gaussian component while the other is due to the presence of jumps. It is worth noting that in our computations the contribution of the second is always negligible compared to the one coming from the first.

We now study the behaviour of our estimator in low dimension. In particular for $d = 2$ the following theorem holds true.

Theorem 2. *Suppose that $d = 2$. Consider $0 < a_{min} \leq a_0$, $a_1 > 0$, $b_0 > 0, b_1 > 0$, $\tilde{C}$, $\tilde{\rho}$ and $\Sigma := \Sigma(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$. Then, there exist $c > 0$ and $T_0 > 0$ such that, for $T \geq T_0$, the optimal choice for the multidimensional bandwidth h yields the following convergence rates.*

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \lesssim \frac{\log T}{T}.$$

To conclude the part regarding the upper bounds on the mean squared error associated to our estimator (6) when a continuous record of the process is available, we are left to discuss the mono-dimensional case. It is known that, under our hypothesis, the proposed kernel estimator achieves the parametric rate $\frac{1}{T}$ and such a rate is optimal (see for example [44] or Theorem 1 in [43]).

4 Lower bounds

One may wonder if the convergence rates found by using kernel density estimators are optimal or if it is possible to improve them by using other density estimators. We aim at showing that the convergence rates found in Theorems 1 and 2 are optimal. We will focus first on the case $d \geq 3$. In order to do that, we will start computing a lower bound in a general case. After that, we will show it is possible to improve it, up to ask $\beta_2 < \beta_3$.

We can write down the expression of the minimax risk for the estimation, at some point x_0, of an invariant density π belonging to the anisotropic Holder class $\mathcal{H}_d(\beta, 2\mathcal{L})$. Let $x_0 \in \mathbb{R}^d$ and $\Sigma(\beta, \mathcal{L}, a_{\min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$ as in Definition 2 here above. We define the minimax risk

$$\mathcal{R}_T(\beta, \mathcal{L}, a_{\min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) := \inf_{\tilde{\pi}_T} \sup_{(a,b)\in\Sigma(\beta,\mathcal{L},a_{\min},a_0,a_1,b_0,b_1,\tilde{C},\tilde{\rho})} \mathbb{E}[(\tilde{\pi}_T(x_0) - \pi_{(a,b)}(x_0))^2], \quad (12)$$

where the infimum is taken on all possible estimators of the invariant density, based on X_t for $t \in [0, T]$. The following lower bound will be showed in Section 5.

Theorem 3. *Let $\beta = (\beta_1, \dots, \beta_d)$, $1 < \beta_1 \leq \dots \leq \beta_d$, $\beta_2 > 2$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$. Consider $0 < a_{min} \leq a_0$ and $a_1 > 0$, $b_0 > 0, b_1 > 0$. Then, there exist $\tilde{C}$, $\tilde{\rho}$, $c > 0$ and $T_0 > 0$ such that, for $T \geq T_0$,*

$$\mathcal{R}_T(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \geq c\,(\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}.$$

Theorem 3 implies that, on a class of diffusions X whose invariant density belongs to $\mathcal{H}_d(\beta, 2\mathcal{L})$, it is not possible to find an estimator with a rate of estimation better than $T^{-\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$.
Comparing the result here above with the second point of Theorem 1 we observe that the convergence rate we found in the lower bound is the same as the in upper bound, when $\beta_2 = \beta_3$. When $\beta_2 < \beta_3$, instead, it is possible to improve the lower bound previously obtained, as gathered in the following theorem. Its proof can be found in Section 5.

Theorem 4. *Let $\beta = (\beta_1, \dots, \beta_d)$, $0 < \beta_1 \leq \dots \leq \beta_d$, $\beta_2 > 2$, and $\beta_2 < \beta_3$. Moreover, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$. Consider $0 < a_{min} \leq a_0$ and $a_1 > 0$, $b_0 > 0, b_1 > 0$. Then, there exist $\tilde{C}$, $\tilde{\rho}$, $c > 0$ and $T_0 > 0$ such that, for $T \geq T_0$,*

$$\mathcal{R}_T(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \geq c\,(\frac{\log T}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}. \quad (13)$$

We will see that the condition $\beta_2 < \beta_3$ is crucial in order to recover a logarithm in the lower bound. Comparing the lower bounds in Theorems 3 and 4 with the upper bounds in Theorem 1 it follows that the kernel density estimator we proposed in (6) achieves the best possible convergence rate.

It is possible to ensure an analogous lower bound in the bi-dimensional case, which ensure the optimality of the convergence rate found in Theorem 2.

Theorem 5. *Let* $d = 2$, $\beta = (\beta_1, \beta_2)$, $0 < \beta_1 \leq \beta_2$, $\beta_2 > 2$, $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2)$, $\mathcal{L}_i > 0$. *Consider* $0 < a_{min} \leq a_0$ *and* $a_1 > 0$, $b_0 > 0, b_1 > 0$. *Then, there exist* $\tilde{C}$, $\tilde{\rho}$, $c > 0$ *and* $T_0 > 0$ *such that, for* $T \geq T_0$,

$$\mathcal{R}_T(\beta, \mathcal{L}, a_{min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \geq c\,(\frac{\log T}{T}).$$

We deduce that also in the bi-dimensional case the estimator proposed in (6) achieves the best possible convergence rate.

5 Proofs upper bounds stated in Section 3

This section is devoted to the proof of all the results stated in Section 3, about the behaviour of the estimator proposed in (6), assuming that a continuous record of the process X is available. As Theorem 1 is a consequence of Proposition 2, we start proving Proposition 2.

5.1 Proof of Proposition 2

Proof. In the sequel, the constant c may change from line to line and it is independent of T. From the definition (6) and the stationarity of the process we get

$$Var(\hat{\pi}_{h,T}(x)) = \frac{1}{T^2} \int_0^T \int_0^T k(t-s) dt\, ds,$$

where

$$k(u) := Cov(\mathbb{K}_h(x - X_0), \mathbb{K}_h(x - X_u)).$$

We deduce that

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{1}{T} \int_0^T |k(s)| ds.$$

We want to use different way to upper bound the variance, relying on the bound on the transition density gathered in Lemma 1 and on the mixing properties of the process as in Lemma 2. Hence, we split the time interval $[0, T]$ into 4 pieces:

$$[0, T] = [0, \delta_1) \cup [\delta_1, \delta_2) \cup [\delta_2, D) \cup [D, T],$$

where δ_1, δ_2 and D will be chosen later, to obtain an upper bound on the variance as sharp as possible. We will see that the bound on the variance will depend on k_0 only on the interval $[\delta_1, \delta_2)$. For this reason we start considering what happens on $[0, \delta_1)$ without taking into account the fact that k_0 can be larger or smaller than 3.

• For $s \in [0, \delta_1)$, from Cauchy -Schwartz inequality and the stationarity of the process we get

$$|k(s)| \leq Var(\mathbb{K}_h(x - X_0))^{\frac{1}{2}} Var(\mathbb{K}_h(x - X_s))^{\frac{1}{2}} = Var(\mathbb{K}_h(x - X_0)) \leq \int_{\mathbb{R}^d} (\mathbb{K}_h(x - y))^2 \pi(y) dy.$$

As $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, its infinitive norm is bounded. Using also the definition of $\mathbb{K}_h$ given in (6) it follows

$$|k(s)| \leq \frac{c}{\prod_{l=1}^d h_l} \tag{14}$$

which implies

$$\int_0^{\delta_1} |k(s)| \leq \frac{c\,\delta_1}{\prod_{l=1}^d h_l}. \tag{15}$$

• For $s \in [\delta_1, \delta_2)$, taking $\delta_2 < 1$, we use the definition of transition density, for which

$$|k(s)| \leq \int_{\mathbb{R}^d} |\mathbb{K}_h(x - y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x - y')| p_s(y, y') dy' \pi(y) dy.$$

We now act differently according on the value of k_0. If $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$, then we introduce $q_s(y'_3 ... y'_d | y'_1, y'_2, y)$ as below:

$$q_s(y'_3 ... y'_d | y'_1, y'_2, y) := e^{-\lambda_0 \frac{|y_1 - y'_1|^2}{s}} \times e^{-\lambda_0 \frac{|y_2 - y'_2|^2}{s}} \times \frac{1}{\sqrt{s}} e^{-\lambda_0 \frac{|y_3 - y'_3|^2}{s}} \times ... \times \frac{1}{\sqrt{s}} e^{-\lambda_0 \frac{|y_d - y'_d|^2}{s}}.$$

From Lemma 1 we know it is

$$p_s(y,y') \leq \frac{c_0}{s} q_s(y_3'...y_d'|y_1',y_2',y).$$

Let us stress that

$$\sup_{s\in(0,1)} \sup_{y_1',y_2',y\in\mathbb{R}^{d+2}} \int_{\mathbb{R}^{d-2}} q_s(y_3'...y_d'|y_1',y_2',y)dy_3'...dy_d' \leq c < \infty. \tag{16}$$

Then,

$$|k(s)| \leq \frac{c_0}{s} \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| q_s(y_3'...y_d'|y_1',y_2',y)dy'\,\pi(y)dy. \tag{17}$$

Using the definition of $\mathbb{K}_h$ and (16) we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| q_s(y_3'...y_d'|y_1',y_2',y)dy' \\
&\leq \frac{c}{\prod_{j\geq 3} h_j} \int_{\mathbb{R}} \frac{1}{h_1} K(\frac{y_1'-x_1}{h_1}) \int_{\mathbb{R}} \frac{1}{h_2} K(\frac{y_2'-x_2}{h_2})(\int_{\mathbb{R}^{d-2}} q_s(y_3'...y_d'|y_1',y_2',y)dy_3'...dy_d')dy_2'\,dy_1' \\
&\leq \frac{c}{\prod_{j\geq 3} h_j} \int_{\mathbb{R}} \frac{1}{h_1} K(\frac{y_1'-x_1}{h_1}) \int_{\mathbb{R}} \frac{1}{h_2} K(\frac{y_2'-x_2}{h_2})dy_2'\,dy_1' \\
&\leq \frac{c}{\prod_{j\geq 3} h_j}.
\end{aligned}$$

It is worth underlining that, here above, it would have been possible to remove the contribution of no matter which couple of bandwidth. We choose to remove h_1 and h_2 because they are associated, in the bias term, to the smallest values of the smoothness (β_1 and β_2) and so they provide the strongest constraints.
It implies that, when $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$, we get

$$|k(s)| \leq \frac{c}{\prod_{j\geq 3} h_j} \frac{1}{s} \tag{18}$$

and so

$$\int_{\delta_1}^{\delta_2} |k(s)|ds \leq \int_{\delta_1}^{\delta_2} \frac{c}{\prod_{j\geq 3} h_j} \frac{1}{s} ds = c \frac{\log(\delta_2) - \log(\delta_1)}{\prod_{j\geq 3} h_j}, \tag{19}$$

where the constant c does not depend on the coefficient $(a,b) \in \Sigma$. We now consider what happens on $[\delta_1, \delta_2)$ when $k_0 \geq 3$ or $k_0 = 1$ and $\beta_2 = \beta_3$. In analogy to what done before we introduce $q_s(y_{k_0+1}'...y_d'|y_1',...,y_{k_0}',y)$ which is such that

$$q_s(y_{k_0+1}'...y_d'|y_1',...,y_{k_0}',y) := e^{-\lambda_0 \frac{|y_1-y_1'|^2}{s}} \times ... \times e^{-\lambda_0 \frac{|y_{k_0}-y_{k_0}'|^2}{s}} \times \frac{1}{\sqrt{s}} e^{-\lambda_0 \frac{|y_{k_0+1}-y_{k_0+1}'|^2}{s}} \times ... \times \frac{1}{\sqrt{s}} e^{-\lambda_0 \frac{|y_d-y_d'|^2}{s}}.$$

Using again Lemma 1 we can write

$$p_s(y,y') \leq c_0 s^{-\frac{k_0}{2}} q_s(y_{k_0+1}'...y_d'|y_1',...,y_{k_0}',y)$$

and as before we have

$$\sup_{s\in(0,1)} \sup_{y_1',...,y_{k_0}',y\in\mathbb{R}^{d+k_0}} \int_{\mathbb{R}^{d-k_0}} q_s(y_{k_0+1}'...y_d'|y_1',...,y_{k_0}',y)dy_{k_0+1}'...dy_d' \leq c < \infty.$$

Acting as in (17) and below it follows

$$\begin{aligned}
|k(s)| &\leq c_0 s^{-\frac{k_0}{2}} \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| q_s(y_{k_0+1}'...y_d'|y_1',...,y_{k_0}',y)dy'\,\pi(y)dy \qquad (20) \\
&\leq c s^{-\frac{k_0}{2}} \frac{1}{\prod_{l\geq k_0+1} h_l}.
\end{aligned}$$

Hence, when $k_0 \geq 3$, we have

$$\int_{\delta_1}^{\delta_2} |k(s)|ds \leq \frac{c}{\prod_{l \geq k_0+1} h_l} \int_{\delta_1}^{\delta_2} s^{-\frac{k_0}{2}} ds \leq \frac{c}{\prod_{l \geq k_0+1} h_l} \delta_1^{1-\frac{k_0}{2}}. \tag{21}$$

We remark that, for $k_0 = 1$, the reasoning here above still applies. However, as $1 - \frac{k_0}{2} = \frac{1}{2}$ is now positive, we obtain

$$\int_{\delta_1}^{\delta_2} |k(s)|ds \leq \frac{c}{\prod_{l \geq 2} h_l} \delta_2^{\frac{1}{2}}. \tag{22}$$

As the bound no longer depends on δ_1, we can choose to take $\delta_1 = 0$.

• For $s \in [\delta_2, D)$ we use the same estimation for any value of k_0, but for $k_0 = 1$, $\beta_2 = \beta_3$. We still consider the bound on the transition density gathered in Lemma 1. Such a bound is not uniform in t big. However, for $t \geq 1$, it is

$$\begin{aligned} p_t(y, y') &= \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, z) p_{\frac{1}{2}}(z, y') dz \\ &\leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, z) e^{-2\lambda_0 |y-z|^2} dz \\ &\leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, z) dz \leq c. \end{aligned}$$

We deduce, for all t,

$$p_t(y, y') \leq c_0 t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-y'|^2}{t}} + c. \tag{23}$$

It follows

$$\begin{aligned} |k(s)| &\leq c \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| (s^{-\frac{d}{2}} + 1) dy' \pi(y) dy \\ &\leq c(s^{-\frac{d}{2}} + 1). \end{aligned} \tag{24}$$

We therefore get

$$\int_{\delta_2}^{D} |k(s)|ds \leq c(\delta_2^{1-\frac{d}{2}} + D). \tag{25}$$

When $k_0 = 1$ and $\beta_2 = \beta_3$, instead, we act analogously we did in the previous interval in order to get (20). We choose to remove, in particular, the contribution of the first three bandwidths. We obtain, $\forall s \in [\delta_2, D)$,

$$|k(s)| \leq c(s^{-\frac{3}{2}} \frac{1}{\prod_{l \geq 4} h_l} + 1), \tag{26}$$

where the final constant comes from (23), as now s can be larger than 1. It follows

$$\int_{\delta_2}^{D} |k(s)|ds \leq c(\frac{1}{\prod_{l \geq 4} h_l} \frac{1}{\delta_2^{\frac{1}{2}}} + D). \tag{27}$$

• For $s \in [D, T]$ we exploit the mixing properties of the process, as stated in Proposition 1. The following control on the covariance holds true:

$$|k(s)| \leq c \left\| \mathbb{K}_h(x - \cdot) \right\|_\infty^2 e^{-\rho s} \leq c(\frac{1}{\prod_{j=1}^d h_j})^2 e^{-\rho s},$$

for ρ and c positive constants which are uniform over the set of coefficients $(a, b) \in \Sigma$. It entails

$$\int_{D}^{T} |k(s)|ds \leq c(\frac{1}{\prod_{j=1}^d h_j})^2 e^{-\rho D}. \tag{28}$$

We now put all the pieces together. For $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$ we collect together (15), (19), (25) and (28). We deduce

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}(\frac{\delta_1}{\prod_{l=1}^d h_l} + \frac{1}{\prod_{j \geq 3} h_j}(|\log(\delta_1)| + |\log(\delta_2)|) + \delta_2^{1-\frac{d}{2}} + D + (\frac{1}{\prod_{j=1}^d h_j})^2 e^{-\rho D}). \tag{29}$$

We now want to choose δ_1, δ_2 and D for which the estimation here above is as sharp as possible. To do that, we take $\delta_1 := h_1 h_2$, $\delta_2 := (\prod_{j\geq 3} h_j)^{\frac{2}{d-2}}$ and $D := [\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1) \wedge T]$. Replacing them in (29) we obtain

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}(\frac{1}{\prod_{j\geq 3} h_j} + \frac{\sum_{j=1}^d |\log(h_j)|}{\prod_{j\geq 3} h_j} + \frac{1}{\prod_{j\geq 3} h_j} + \sum_{j=1}^d |\log(h_j)| + 1)$$
$$\leq \frac{c}{T}\frac{\sum_{j=1}^d |\log(h_j)|}{\prod_{j\geq 3} h_j}.$$

If otherwise $k_0 \geq 3$, we consider (21) instead of (19) which, together with (15), (25) and (28) provides

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}(\frac{\delta_1}{\prod_{l=1}^d h_l} + \frac{\delta_1^{1-\frac{k_0}{2}}}{\prod_{j\geq k_0+1} h_j} + \delta_2^{1-\frac{d}{2}} + D + (\frac{1}{\prod_{j=1}^d h_j})^2 e^{-\rho D}). \tag{30}$$

We observe that the balance between the first two terms is achieved for $\delta_1 := (\prod_{l=1}^{k_0} h_l)^{\frac{2}{k_0}}$. Then, we can choose $\delta_2 = 1$. We moreover take $D := [\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1) \wedge T]$, as before. It yields

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}\frac{(\prod_{l=1}^{k_0} h_l)^{\frac{2}{k_0}}}{\prod_{l\geq 1} h_l} = \frac{1}{T}\frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}}(\prod_{l\geq k_0+1} h_l)},$$

as we wanted.
To conclude, by (22) with $\delta_1 = 0$, (27) and (28) we have proven that when $k_0 = 1$ and $\beta_2 = \beta_3$ we have

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}(\frac{\delta_2^{\frac{1}{2}}}{\prod_{l\geq 2} h_l} + \frac{1}{\delta_2^{\frac{1}{2}}\prod_{l\geq 4} h_l} + D + (\frac{1}{\prod_{j=1}^d h_j})^2 e^{-\rho D})$$
$$\leq \frac{c}{T}(\frac{\sqrt{h_2 h_3}}{\prod_{l\geq 2} h_l} + \frac{1}{\sqrt{h_2 h_3}\prod_{l\geq 4} h_l}) = \frac{c}{T}\frac{1}{\sqrt{h_2 h_3}\prod_{l\geq 4} h_l},$$

where the last estimation follows by having chosen $\delta_2 := h_2 h_3$. The parameter D is moreover chosen as above. Remarking that all the constants are uniform over the set of coefficient $(a, b) \in \Sigma$, the result is proven. □

5.2 Proof of Theorem 1

Proof. From the usual bias-variance decomposition it is

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \leq |\mathbb{E}[\hat{\pi}_{h,T}(x)] - \pi(x)|^2 + Var(\hat{\pi}_{h,T}(x)).$$

An upper bound for the variance is gathered in Proposition 2. Regarding the bias, a standard computation (see for example the proof of Proposition 1.2 of [63] or Proposition 1 of [21]) provides

$$|\mathbb{E}[\hat{\pi}_{h,T}(x)] - \pi(x)|^2 \leq c\sum_{j=1}^d h_j^{\beta_j}, \tag{31}$$

with the constant c that does not depend on x, nor on $(a, b) \in \Sigma$.
The result is then obtained by looking for the balance between the two terms.
We start considering the case where $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$. Thanks to Proposition 2 we know the bound (10) holds true. After simple computations (as in the proof of Theorem 1 of [3]) it is easy to see that the balance is achieved by choosing the rate optimal bandwidth $h^* := ((\frac{\log T}{T})^{a_1}, ..., (\frac{\log T}{T})^{a_d})$, with

$$a_1 > \frac{\bar{\beta}_3}{\beta_1(2\bar{\beta}_3 + d - 2)}, \qquad a_2 > \frac{\bar{\beta}_3}{\beta_2(2\bar{\beta}_3 + d - 2)} \tag{32}$$

and

$$a_l = \frac{\bar{\beta}_3}{\beta_l(2\bar{\beta}_3 + d - 2)} \qquad \forall l \in \{3, ..., d\}. \tag{33}$$

Replacing the value of h^* in the upper bounds of the variance and of the bias we get

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le c(\frac{\log T}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}},$$

with the same constant c for all $(a, b) \in \Sigma$.
When $k_0 = 1$ and $\beta_2 = \beta_3$ the bias-variance decomposition consists in

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le c\sum_{j=1}^{d} h_j^{\beta_j} + \frac{c}{T}\frac{1}{\sqrt{h_2 h_3}\prod_{l\ge 4} h_l}.$$

As $\beta_2 = \beta_3$ we choose $h_2 = h_3$, such that the upper bound on the variance gathered in Proposition 2 becomes simply $\frac{c}{T}\frac{1}{\prod_{l\ge 3} h_l}$. Then, similar computations as above implies that the balance is achieved by choosing $h^*(T) := ((\frac{1}{T})^{a_1}, ..., (\frac{1}{T})^{a_d})$, with a_1 as in (32), $a_2 = a_3$ with a_j as in (33), for $j \ge 3$. It yields

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le c(\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}.$$

Assuming now that $k_0 \ge 3$, we have

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le c\sum_{j=1}^{d} h_j^{\beta_j} + \frac{c}{T}\frac{1}{h_1^{k_0-2}(\prod_{l\ge k_0+1} h_l)},$$

where we have already chosen to take $h_1 = h_2 = ... = h_{k_0}$. We now look for the rate optimal bandwidth by setting $h_l^* := (\frac{1}{T})^{a_l}$ and searching a_1, a_{k_0+1}, ... , a_d such that the bias and variance are balanced. It consists in solving the following system:

$$\begin{cases} a_1\beta_1 = a_l\beta_l \qquad \forall l \in \{k_0+1, ..., d\} \\ 1 - a_1(k_0 - 2) - \sum_{l\ge k_0+1} a_l = 2a_1\beta_1. \end{cases} \tag{34}$$

As a consequence of the first $d - k_0$ equations, we can write

$$a_l = \frac{\beta_1}{\beta_l}a_1, \qquad \forall l \in \{k_0+1, ..., d\}. \tag{35}$$

Hence, the last equation becomes

$$\begin{aligned} 2\beta_1 a_1 &= 1 - a_1(k_0 - 2) - \beta_1 a_1 \sum_{l\ge k_0+1}\frac{1}{\beta_l} \\ &= 1 - a_1(k_0 - 2) - \beta_1 a_1 \frac{d - k_0}{\bar{\beta}_k}, \end{aligned} \tag{36}$$

where $\bar{\beta}_k$ is the mean smoothness over β_{k_0+1}, ... , β_d and it is such that $\frac{1}{\bar{\beta}_k} = \frac{1}{d-k_0}\sum_{l\ge k_0+1}\frac{1}{\beta_l}$. We also observe that, as $\frac{1}{\bar{\beta}_3} = \frac{1}{d-2}\sum_{l\ge 3}\frac{1}{\beta_l}$, it is

$$\frac{k_0 - 2}{\beta_1} + \frac{d - k_0}{\bar{\beta}_k} = \frac{d - 2}{\bar{\beta}_3}. \tag{37}$$

Hence, (36) can be seen as

$$2\beta_1 a_1 = 1 - a_1\beta_1\frac{d-2}{\bar{\beta}_3},$$

which leads to the choice

$$a_1 = \frac{\bar{\beta}_3}{\beta_1(2\bar{\beta}_3 + d - 2)}.$$

Thanks to the first $d-k_0$ equations in the system it follows

$$a_l = \frac{\bar{\beta}_3}{\beta_l(2\bar{\beta}_3+d-2)} \qquad \forall l \in \{k_0+1, ..., d\}\,.$$

Plugging the rate optimal bandwidth h^* in the bound of the mean squared error and recalling that the constant c does not depend on $(a,b) \in \Sigma$, we obtain

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le c(\frac{1}{T})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}},$$

as we wanted. □

Remark 1. *As the situation is complicated, it is worth highlighting how the optimal bandwidths depend on the smoothness.*

- *When $\beta_2 < \beta_3$, h_1 and h_2 are arbitrarily small while h_3, ... , h_d are fixed. In this case there is the logarithm in the rate.*
- *When $\beta_2 = \beta_3$ we have two possibilities for the bandwidths, depending on β_1. In both cases we remove the logarithm in the rate.*
  - *If $\beta_1 < \beta_2$, then h_1 can be arbitrarily small while h_2, ... , h_d are fixed.*
  - *If $\beta_1 = \beta_2$, then all the bandwidths h_1, ... , h_d are fixed.*

It is interesting to see that when we have two degrees of freedom on the bandwidths it is impossible to remove the logarithm in the convergence rate, while having one degree of freedom on the bandwidths is enough to modify the convergence rate: it is the same it would have been without any degree of freedom on the bandwidths.

5.3 Proof of Theorem 2

Proof. The proof relies, as before, on the bias-variance decomposition. Concerning the bias, (31) still holds true. We need to provide an upper bound on the variance, based on Lemma 1 and Proposition 1, as before. The main difference compared to the proof of Proposition 2 is that we now split the integral over $[0,T]$ in three pieces:

$$[0,T] = [0,\delta) \cup [\delta, D) \cup [D,T),$$

where δ and D will be chosen later. On the first and on the last piece we act as in the proof of Proposition 2 and so (15) and (28) keep holding. Regarding the interval $[\delta, D)$, we act on it as we did on $[\delta_2, D)$. From (24), recalling that now $d=2$, we obtain

$$\int_\delta^D |k(s)|ds \le c\int_\delta^D (s^{-1}+1)ds \le c(|\log D| + |\log \delta| + D).$$

Putting all the pieces together we get

$$\begin{aligned} Var(\hat{\pi}_{h,T}(x)) &\le \frac{c}{T}(\frac{\delta}{h_1\,h_2} + |\log D| + |\log \delta| + D + (\frac{1}{h_1\,h_2})^2 e^{-\rho D}) \\ &\le \frac{c}{T}(1 + \sum_{j=1}^{2} |\log h_j| + \sum_{j=1}^{2} |\log(|\log h_j|)|) \\ &\le \frac{c\sum_{j=1}^2 |\log h_j|}{T}, \end{aligned}$$

where we have opportunely chosen $\delta = h_1 h_2$ and $D := [\max(-\frac{2}{\rho}\log(h_1\,h_2), 1) \wedge T]$. It follows

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le ch_1^{2\beta_1} + ch_2^{2\beta_2} + \frac{c\sum_{j=1}^2 |\log h_j|}{T}.$$

To conclude it is enough to observe that the optimal choice for the bandwidth consists in taking $h_l^* = (\frac{\log T}{T})^{a_l}$, with $a_l \ge \frac{1}{2\beta_l}$ for $l=1,2$, which provides the wanted convergence rate. □

6 Proof of the lower bounds stated in Section 4

This section is devoted to the proof of the lower bounds, as stated in Section 4.

6.1 Proof of Theorem 3

The proof of Theorem 3 is based on the two hypothesis method, as explained for example in Section 2.3 of Tsybakov [63] and follow the standard scheme provided in Section 6 of [3]. We start by making explicit link between the drift and the stationary measure. Then, we provide two priors depending on some calibration parameters and, to conclude, we find some conditions on the calibration such that it is possible to prove a lower bound for the minimax risk introduced in (12).

First, remark that if a drift b belongs to the class of coefficients $\Sigma(\beta, \mathcal{L}, a_{\min}, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$ then it is possible to check that the translated drift $x \mapsto b(x+h)$ belongs to $\Sigma(\beta, \mathcal{L}, a_{\min}, a_0, a_1, b_0, b_1, \tilde{C}/2, \tilde{\rho}')$ with $\tilde{\rho}' = \max(\tilde{\rho} + |h|, 2|h|\frac{\tilde{C}+b_0}{\tilde{C}})$. As a consequence, it is sufficient to prove the theorem in the case where $x_0 = (0, \dots, 0)$, and the general case can be deduced by translation. A second remark, is that if X is solution of (3) where the coefficients (a, b) satisfies A1 with the constants $(a_{\min}, a_0, a_1, b_0, b_1)$, then the diffusion $t \mapsto a_{\min}^{-1} X_t$ satisfies the assumption A1 with the constants $(1, a_0/a_{\min}, a_1, b_0/a_{\min}, b_1)$. Hence, it is possible to assume without loss of generality that $a_{\min} = 1$.

In the proof, we will lower bound the risk on the subclass of model (3) given by the following simpler stochastic differential equation, for which the diffusion coefficient $a = \mathbb{I}_{d\times d}$ is constant :

$$dX_t = b(X_t)dt + dW_t. \tag{38}$$

and where b is any drift function such that $(\mathbb{I}_{d\times d}, b) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$.

Proof. • Explicit link between the drift and the stationary measure.
We first of all need to introduce A, the generator of the diffusion X solution of (38):

$$Af(x) := \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{i=1}^{d} b^i(x) \frac{\partial}{\partial x_i} f(x). \tag{39}$$

We now introduce a class of function that will be useful in the sequel:

$$\mathcal{C} := \Big\{ f : \mathbb{R}^d \to \mathbb{R},\ f \in C^2(\mathbb{R}^d) \text{ such that } \forall i \in \{1, ..., d\} \lim_{x_i \to \pm\infty} f(x) = 0,$$
$$\lim_{x_i \to \pm\infty} \frac{\partial}{\partial x_i} f(x) = 0 \text{ and } \int_{\mathbb{R}^d} f(x)dx < \infty \Big\}.$$

We denote furthermore as A^* the adjoint operator of A on $\mathbf{L}^2(\mathbb{R}^d, dx)$ which is such that, for $f, g \in \mathcal{C}$,

$$\int_{\mathbb{R}^d} Af(x)g(x)dx = \int_{\mathbb{R}^d} f(x)A^*g(x)dx.$$

The form of A^* is known (see for example Lemma 2 in [3] remarking that, in our case, the discrete part of the generator A_d is zero and so we do not have its adjoint in A^*):

$$A^*g(x) = \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} g(x) - (\sum_{i=1}^{d} \frac{\partial b^i}{\partial x_i} g(x) + b^i \frac{\partial g}{\partial x_i}(x)).$$

If $g : \mathbb{R}^d \to \mathbb{R}$ is a probability density of class $\mathcal{C}^2$, solution of $A^*g = 0$, then it can be checked by Ito's formula it is an invariant density for the process we are considering. When the stationary distribution π is unique, therefore, it can be computed as solution of the equation $A^*\pi = 0$. As proposed for example in [3] and in [26], we consider π as fixed and b as the unknown variable. We therefore want to compute a function $b = b_g$ solution to $A^*g = 0$. For $g \in \mathcal{C}$ and $g > 0$, we introduce for all $x \in \mathbb{R}^d$ and for all $i \in \{1, ..., d\}$,

$$b_g^i(x) = \frac{1}{g(x)} \frac{1}{2} \frac{\partial g}{\partial x_i}(x) \tag{40}$$

and $b_g(x) = (b_g^1(x), ..., b_g^d(x))$. It is enough to remark that, for $b_g^i(x)$ defined as above it is

$$\frac{\partial\, b_g^i}{\partial x_i}(x) = \frac{1}{g(x)} \frac{1}{2} \frac{\partial^2 g}{\partial x_i^2}(x) - \frac{b_g^i(x)}{g(x)} \frac{\partial\, g}{\partial x_i}(x) \tag{41}$$

to see that the function b_g here above introduced is actually solution of $A^* g(x) = 0$, for any $x \in \mathbb{R}^d$. We know that π is solution to $A^*\pi(x) = 0$ for $b = b_\pi$ and so it is a stationary measure for the process X whose drift is b_π. However, if b_π satisfies A1-A2 then, from Lemma 2, we know there exists a Lyapounov function and that the stationary measure of the equation with drift coefficient b_π is unique. It follows it is equal to π.

Hence, we need b_π to be a function satisfying A1-A2. We introduce some assumptions on π for which the associated drift b_π has the wanted properties.

A3: Let $\pi : \mathbb{R}^d \to \mathbb{R}$ a probability density with regularity $\mathcal{C}^2$ such that, for any $x = (x_1, ..., x_d) \in \mathbb{R}^d$, $\pi(x) = c_n \prod_{j=1}^d \pi_j(x_j) > 0$, where c_n is a normalization constant. We suppose moreover that the following holds true for each $j \in \{1, ..., d\}$:

1. $\lim_{y \to \pm\infty} \pi_j(y) = 0$ and $\lim_{y \to \pm\infty} \pi_j'(y) = 0$.
2. There exists $\tilde{\epsilon} > 0$ and $R > 0$ such that, for any $y : |y| > \frac{R}{\sqrt{d}}$,
$$\frac{\pi_j'(y)}{\pi_j(y)} \le -\tilde{\epsilon}\, \mathrm{sgn}(y).$$
3. There exists a constant c_1 such that, for any $y \in \mathbb{R}$,
$$|\frac{\pi_j'(y)}{\pi_j(y)}| \le c_1.$$
4. There exists a constant c_2 such that, for any $x \in \mathbb{R}^d$,
$$|\frac{\partial^2 \pi}{\partial x_i^2}(x)| \le c_2 \pi(x)$$

The properties listed here above have been introduced in order to make the associated drift function satisfying A1-A2 so that, up to know that $\pi_{(a,b)} \in \mathcal{H}_d(\beta, 2\mathcal{L})$, it would follow $(I_{d\times d}, b) \in \Sigma$. It is easy to see from the definition of b_π given in (40) that, having π in a multiplicative form, we get

$$|b_\pi^i(x)| = |\frac{1}{2} \frac{\pi_i'(x_i)}{\pi_i(x_i)}| \le \frac{c_1}{2} =: \tilde{b}_0,$$

where we have used the third point of A3. The drift function is also clearly lipschitz. Moreover from (41), the third and the fourth points of A3 and the just proven boundedness of b_π we have

$$|\frac{\partial b_\pi^i}{\partial x_i}(x)| \le \frac{c_2}{2} + c_1 \tilde{b}_0 =: \tilde{b}_1.$$

In order to show that also the A2 holds true we need to investigate the behaviour of $x_i b_\pi^i(x)$. From the second point of A3, which holds true for any x_i such that $|x_i| > \frac{R}{\sqrt{d}}$, it is

$$x_i b_\pi^i(x) = \frac{x_i}{2} \frac{\pi_i'(x_i)}{\pi_i(x_i)} \le -\frac{x_i}{2} \tilde{\epsilon}\, \mathrm{sgn}(x_i) = -\frac{\tilde{\epsilon}}{2} |x_i|.$$

Using also the boundedness of b_π^i showed before, it follows

$$x \cdot b_\pi(x) = \sum_{i=1}^d x_i b_\pi^i(x) = \sum_{x_i : |x_i| > \frac{R}{\sqrt{d}}}^d x_i b_\pi^i(x) + \sum_{x_i : |x_i| \le \frac{R}{\sqrt{d}}}^d x_i b_\pi^i(x)$$

$$\le -\tilde{c} \sum_{x_i : |x_i| > \frac{R}{\sqrt{d}}}^d |x_i| + \frac{R}{\sqrt{d}} c \le -c_1|x| + c_2,$$

where the last inequality is a consequence of the fact that, for $|x| > R$, there has to be at least a component x_i such that $|x_i| > \frac{R}{\sqrt{d}}$. Hence, we can use the sup norm and compare it with the euclidean one. Moreover, as $|x|$ is lower bounded by R, it exists a constant $C_1 > 0$ such that

$$-c_1|x| + c_2 \leq -C_1|x|.$$

The proposed drift b_π is therefore a bounded lipschitz function that satisfies A2, up to know that on the linked invariant density the properties gathered in A3 hold true. In the next step we propose two priors with the prescribed properties.

• Construction of the priors.
We want to provide two drift functions belonging to $\Sigma(\beta, \mathcal{L})$ and, to do it, we introduce two probability densities defined on the purpose to make A3 hold true. We set $\pi^{(0)}(x) := c_\eta \prod_{k=1}^d \pi_{k,0}(x_k)$, where c_η is the constant that makes $\pi^{(0)}$ a probability measure. For any $y \in \mathbb{R}$ we define $\pi_{k,0}(y) := f(\eta|y|)$, where

$$f(x) := \begin{cases} e^{-|x|} & \text{if } |x| \geq 1 \\ \in [1, e^{-1}] & \text{if } \frac{1}{2} < |x| < 1 \\ 1 & \text{if } |x| \leq \frac{1}{2} \end{cases}$$

and η is a constant in $(0, \frac{1}{2})$ which plays the same role as $\tilde{\epsilon}$ did in A3, as it can be chosen as small as we want. In particular we choose η small enough to get $\pi^{(0)} \in \mathcal{H}_d(\beta, \mathcal{L})$. Moreover, we assume f to be a C^∞ function that satisfies $|f^{(k)}(x)| \leq 2e^{-|x|}$ for $k = 1, 2$.
It is easy to see that π_0 satisfies A3, as it is clearly positive and in a multiplicative form. The first point holds true by construction. We observe then that, for y such that $|y| > \frac{1}{\eta}$, we have $\pi'_{j,0}(y) = -\eta \operatorname{sgn}(y) \pi_{j,0}(y)$ for any $j \in \{1, ..., d\}$. Therefore, point 2 of A3 is satisfied for y such that $|y| > \frac{R}{\sqrt{d}}$, up to take $R := \frac{\sqrt{d}}{\eta}$.
Regarding point 3 of A3, it clearly holds true for y such that $|y| > \frac{1}{\eta}$ and $|y| < \frac{1}{2\eta}$ for what said before and as the derivative is zero, respectively. When $\frac{1}{2\eta} \leq |y| \leq \frac{1}{\eta}$, point 3 of A3 is satisfied thanks to the condition $|f'(x)| \leq 2e^{-|x|}$. An analogous reasoning can be applied to ensure the validity of the fourth point of A3

As $\pi^{(0)}$ is an invariant density belonging to $\mathcal{H}_d(\beta, 2\mathcal{L})$ and satisfying A3, the associated coefficients are such that $(I_{d\times d}, b^{(0)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, \tilde{b}_0, \tilde{b}_1, C_1, R) \subset \Sigma$ for some $\tilde{C}$, $\tilde{\rho}$, with $\tilde{b}_0$, $\tilde{b}_1$, C_1 and R as above. To provide the second hypothesis, we introduce the probability measure $\pi^{(1)}$. We are given it as $\pi^{(0)}$ to which we add a bump: let $K : \mathbb{R} \to \mathbb{R}$ be a C^∞ function with support on $[-1, 1]$ and such that

$$K(0) = 1, \qquad \int_{-1}^{1} K(z)dz = 0. \tag{42}$$

We set

$$\pi^{(1)}(x) := \pi^{(0)}(x) + \frac{1}{M_T} \prod_{l=1}^{d} K\left(\frac{x_l - x_0^l}{h_l(T)}\right), \tag{43}$$

where $x_0 = (x_0^1, ..., x_0^d) \in \mathbb{R}^d$ is the point in which we are evaluating the minimax risk, as defined in (12), M_T and $h_l(T)$ will be calibrated later and satisfy $M_T \to \infty$ and, $\forall l \in \{1, ..., d\}$, $h_l(T) \to 0$ as $T \to \infty$. From the properties of the kernel function given in (42) we obtain

$$\int_{\mathbb{R}^d} \pi^{(1)}(x)dx = \int_{\mathbb{R}^d} \pi^{(0)}(x)dx = 1.$$

Moreover, as $\pi^{(0)} > 0$, K has support compact and $\frac{1}{M_T} \to 0$, for T big enough we can say that $\pi^{(1)} > 0$ as well. The fact of the matter consists of calibrating M_T and $h_l(T)$ such that both the densities $\pi^{(0)}$ and $\pi^{(1)}$ belong to the anisotropic Holder class $\mathcal{H}_d(\beta, 2\mathcal{L})$ (according with Definition 2) and the laws $\mathbb{P}^{(0)}$ and $\mathbb{P}^{(1)}$ are close.
To do that, we need to evaluate the difference between the two proposed drifts. We introduce the following set of $\mathbb{R}^d$:

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

Then, for T large enough,

1. For any $x \in K_T^c$ and $\forall i \in \{1, ..., d\}$: $b_i^{(1)}(x) = b_i^{(0)}(x)$.
2. For any $x \in K_T$ and $\forall i \in \{1, ..., d\}$: $|b_i^{(1)}(x) - b_i^{(0)}(x)| \le \frac{c}{M_T} \sum_{j=1}^d \frac{1}{h_j(T)}$, where c is a constant independent of T.

Before proceeding with the proof of these two points we introduce some notations:

$$I^i[\pi](x) := \frac{1}{2} \frac{\partial \pi}{\partial x_i}(x), \qquad d_T(x) := \pi^{(1)}(x) - \pi^{(0)}(x) = \frac{1}{M_T} \prod_{l=1}^d K(\frac{x_l - x_0^l}{h_l(T)}). \tag{44}$$

As the support of K is in $[-1, 1]$, for any $x \in K_T^c$ both $d_T(x)$ and its derivatives are 0 and, in particular, $\pi^{(0)}(x) = \pi^{(1)}(x)$. We can therefore write, using the linearity of the operator I^i,

$$b_i^{(1)}(x) = \frac{1}{\pi^{(0)}(x)} \tilde{I}^i[\pi^{(0)}](x) + \frac{1}{\pi^{(0)}(x)} \tilde{I}^i[d_T](x) = b_i^{(0)}(x) + \frac{1}{\pi^{(0)}(x)} \tilde{I}^i[d_T](x) = b_i^{(0)}(x).$$

Regarding the second point here above, we observe that on K_T we have

$$b_i^{(1)} - b_i^{(0)} = (\frac{1}{\pi^{(1)}} - \frac{1}{\pi^{(0)}}) \tilde{I}^i[\pi^{(0)}] + \frac{1}{\pi^{(1)}} \tilde{I}^i[d_T] = \frac{\pi^{(0)} - \pi^{(1)}}{\pi^{(1)}} \frac{1}{\pi^{(0)}} \tilde{I}^i[\pi^{(0)}] + \frac{1}{\pi^{(1)}} \tilde{I}^i[d_T] = \frac{d_T}{\pi^{(1)}} b_i^{(0)} + \frac{1}{\pi^{(1)}} \tilde{I}^i[d_T].$$

For how we have defined $\pi^{(1)} = \pi^{(0)} + d_T$, we see first of all it is lower bounded away from 0. Moreover we know that $\pi^{(0)}$ satisfies Assumption A3 and so $b_i^{(0)}$ is bounded. Furthermore, we have the following controls on d_T:

$$\|d_T\|_\infty \le \frac{c}{M_T}, \qquad \left\| \frac{\partial d_T}{\partial x_j} \right\|_\infty \le \frac{c}{M_T} \frac{1}{h_j(T)}.$$

It follows that, for any $x \in K_T$,

$$|b_i^{(1)} - b_i^{(0)}| \le \frac{c}{M_T} (1 + \sum_{j=1}^d \frac{1}{h_j(T)}) \le \frac{c}{M_T} \sum_{j=1}^d \frac{1}{h_j(T)},$$

where the last inequality is a consequence of the fact that, $\forall j \in \{1, ..., d\}$, $h_j(T) \to 0$ for $T \to \infty$ and so, if compared with the second term in the equation here above, all the other terms are negligible.

Then, it is possible to show that also $(I_{d\times d}, b^{(1)})$ belongs to Σ, up to calibrate properly M_T and $h_i(T)$, for $i \in \{1, ..., d\}$. We recall that we already know that $b^{(0)}$ satisfies A1-A2. Due to points 1 and 2 above also $b^{(1)}$ is bounded, up to ask that $\sum_{j=1}^d \frac{1}{h_j(T)} = o(M_T)$. Remark also that we have $< b^{(1)}(x), x > \le -\tilde{C}|x|$ for $|x| \ge \tilde{\rho}$ for T large enough, using that $b^{(1)}$ and $b^{(0)}$ coincide on K_T^c. Moreover, after some computations we have

$$\begin{aligned}\left\| \nabla b^{(1)} - \nabla b^{(0)} \right\|_\infty &\le \frac{c}{M_T} \Big[\|d_T\|_\infty + \sum_{y \in \{x_1, \dots, x_d\}} \|d_T\|_\infty \left\| \frac{\partial d_T}{\partial y} \right\|_\infty \\ &+ \frac{1}{M_T} \sum_{y, y' \in \{x_1, \dots, x_d\}} \left\| \frac{\partial d_T}{\partial y} \right\|_\infty \left\| \frac{\partial d_T}{\partial y'} \right\|_\infty + \sum_{y, y' \in \{x_1, \dots, x_d\}} \left\| \frac{\partial^2 d_T}{\partial y \partial y'} \right\|_\infty \Big] \\ &\le \frac{1}{M_T} \sum_{i,j=1}^d \frac{c}{h_i h_j} \le \frac{1}{M_T} \frac{c}{h_1^2}.\end{aligned}$$

Hence, to get that $\left\| \nabla b^{(1)} \right\| \le b_1$ it is sufficient that

$$\frac{c}{M_T} \frac{1}{h_1^2} \to 0 \tag{45}$$

for T going to ∞. Furthermore, requiring that

$$\frac{1}{M_T} \le \epsilon h_i(T)^{\beta_i} \qquad \forall i \in \{1, ..., d\}, \tag{46}$$

it is easy to derive the Holder regularity of $\pi^{(1)}$ starting from the regularity of $\pi^{(0)}$, as proved for example in Lemma 3 of [3]. It follows that, under condition (46), (which implies also $\sum_{j=1}^d \frac{1}{h_j(T)} = o(M_T)$, as $\beta_j > 1$ for any j) and (45), both $(I_{d\times d}, b^{(0)})$ and $(I_{d\times d}, b^{(1)})$ belong to Σ.

• Choice of the calibration

Before we keep proceeding, we introduce some notations. We denote as $\mathbb{P}_0$ (respectively $\mathbb{P}_1$) the law of a stationary solution $(X_t)_{t\geq 0}$ of (38) whose drift coefficient is $b^{(0)}$ (respectively $b^{(1)}$). Moreover we will note $\mathbb{P}_0^{(T)}$ the law of $(X_t)_{t\in[0,T]}$, solution of the same stochastic differential equation as here above. The corresponding expectation will be denoted as $\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}$ (respectively $\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(1)})}$).
To find a lower bound for the risk we will need to use that there exist C and $\lambda > 0$ such that, for all T large enough,

$$\mathbb{P}_0^{(T)}(Z^{(T)} \geq \frac{1}{\lambda}) \geq C, \tag{47}$$

where we have introduced the notation $Z^{(T)} := \frac{d\mathbb{P}_1^{(T)}}{d\mathbb{P}_0^{(T)}}$. To ensure its validity it is enough to remark that the proof of Lemma 4 in [3] is the same even in absence of jumps. Hence, we know (47) holds true if

$$\sup_{T\geq 0} T \int_{\mathbb{R}^d} |b^{(1)}(x) - b^{(0)}(x)|^2 \pi^{(0)}(x)\, dx < \infty.$$

From points 1 and 2 above, it is equivalent to ask

$$\sup_{T\geq 0} T \frac{c}{M_T^2} (\sum_{j=1}^d \frac{1}{h_j^2(T)}) |K_T| = \sup_{T\geq 0} T \frac{c}{M_T^2} (\prod_{l=1}^d h_l(T)) (\sum_{j=1}^d \frac{1}{h_j^2(T)}) < \infty. \tag{48}$$

Then, as $(I_{d\times d}, b^{(0)})$ and $(I_{d\times d}, b^{(1)})$ belong to Σ, we have

$$\begin{aligned}
R(\tilde{\pi}_T(x_0)) &\geq \frac{1}{2}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(1)})}[(\tilde{\pi}_T(x_0) - \pi^{(1)}(x_0))^2] + \frac{1}{2}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}[(\tilde{\pi}_T(x_0) - \pi^{(0)}(x_0))^2] \\
&\geq \frac{1}{2}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}[(\tilde{\pi}_T(x_0) - \pi^{(1)}(x_0))^2 Z^{(T)}] + \frac{1}{2}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}[(\tilde{\pi}_T(x_0) - \pi^{(0)}(x_0))^2] \\
&\geq \frac{1}{2\lambda}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}[(\tilde{\pi}_T(x_0) - \pi^{(1)}(x_0))^2 1_{\{Z^{(T)} \geq \frac{1}{\lambda}\}}] + \frac{1}{2}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}[(\tilde{\pi}_T(x_0) - \pi^{(0)}(x_0))^2 1_{\{Z^{(T)} \geq \frac{1}{\lambda}\}}] \\
&= \frac{1}{2\lambda}\mathbb{E}^{(T)}_{(I_{d\times d}, b^{(0)})}\big[[(\tilde{\pi}_T(x_0) - \pi^{(1)}(x_0))^2 + (\tilde{\pi}_T(x_0) - \pi^{(0)}(x_0))^2] 1_{\{Z^{(T)} \geq \frac{1}{\lambda}\}}\big],
\end{aligned}$$

for all $\lambda > 1$. We remark it is

$$(\tilde{\pi}_T(x_0) - \pi^{(1)}(x_0))^2 + (\tilde{\pi}_T(x_0) - \pi^{(0)}(x_0))^2 \geq (\frac{\pi^{(1)}(x_0) - \pi^{(0)}(x_0)}{2})^2$$

and so we obtain

$$R(\tilde{\pi}_T(x_0)) \geq \frac{1}{8\lambda}(\pi^{(1)}(x_0) - \pi^{(0)}(x_0))^2 \mathbb{P}_0^{(T)}(Z^{(T)} \geq \frac{1}{\lambda}) \geq \frac{c}{M_T^2}, \tag{49}$$

where we have used (47) and that, by construction, $\pi^{(1)}(x_0) - \pi^{(0)}(x_0) = \frac{1}{M_T}\prod_{l=1}^d K(0) = \frac{1}{M_T}$. Hence, we have to find the largest choice for $\frac{1}{M_T^2}$, subject to the constraints (46) and (48). To do that, we suppose at the beginning to saturate (46) for any $j \in \{1, ..., d\}$. From the order of β we obtain

$$h_1(T) = h_2(T)^{\frac{\beta_2}{\beta_1}} \leq h_2(T) \leq ... \leq h_d(T). \tag{50}$$

We plug it in (48) and we observe that the biggest term in the sum is $\frac{\prod_{l\neq 1} h_l(T)}{h_1(T)}$. In order to make it as small as possible, we decide to increment $h_1(T)$ up to get $h_1(T) = h_2(T)$, remarking that it is not an improvement to take $h_1(T)$ also bigger than $h_2(T)$ because otherwise $\frac{\prod_{l\neq 2} h_l(T)}{h_2(T)}$ would be the biggest term, and it would be larger than $\frac{\prod_{l\neq 1} h_l(T)}{h_1(T)}$ for $h_1(T) = h_2(T)$. Therefore, we take $h_1(T) = h_2(T)$ and $h_l(T) = (\frac{1}{M_T})^{\frac{1}{\beta_l}}$ for $l \geq 2$. With this choice (45) is always satisfied as $\beta_2 > 2$. Moreover, we have

$$\prod_{l\geq 3} h_l(T) = (\frac{1}{M_T})^{\sum_{l\geq 3}\frac{1}{\beta_l}} = (\frac{1}{M_T})^{\frac{d-2}{\bar{\beta}_3}}.$$

And so condition (48) turns out being

$$\sup_T T \frac{1}{M_T^2} \prod_{l \geq 3} h_l(T) = \sup_T T \frac{1}{M_T^2} (\frac{1}{M_T})^{\frac{d-2}{\bar{\beta}_3}} \leq c.$$

It leads us to the choice $M_T = T^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}$. It implies

$$R(\tilde{\pi}_T(x_0)) \geq (\frac{1}{T})^{\frac{2\,\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}},$$

as we wanted.

□

6.2 Proof of Theorem 4

Proof. Following the same ideas as in the proof of Theorem 3 we construct a prior supported by two different models corresponding to two drift functions $b^{(0)}$ and $b^{(1)}$, and with associated stationary probabilities $\pi^{(0)}$ and $\pi^{(1)}$. However, the shape of the bump used in the construction of $\pi^{(1)}$, given below by (54), is different with the one used in the proof of Theorem 3 and defined in (44)

In the construction of the prior, the first two components are treated differently than the other ones. For this reason, we introduce cylindrical coordinates $(r, \theta, x_3, \dots, x_d)$ that translates into the Cartesian coordinates $(r\cos(\theta), r\sin(\theta), x_3, \dots, x_d)$. We recall the expression of the Laplacian in cylindrical coordinates $\Delta\pi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \pi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \pi}{\partial \theta^2} + \sum_{k=3}^d \frac{\partial^2 \pi}{\partial x_k^2}$ and of the divergence operator $\nabla \cdot \xi = \frac{1}{r}\frac{\partial(r\xi_r)}{\partial r} + \frac{1}{r}\frac{\partial \xi_\theta}{\partial \theta} + \sum_{k=3}^d \frac{\partial \xi_k}{\partial x_k}$ where ξ is the vector field $\xi = \xi_r \vec{e}_r + \xi_\theta \vec{e}_\theta + \sum_{k=3}^d \xi_k \vec{e}_k$ with $\vec{e}_r = (\cos(\theta), \sin(\theta), 0, \dots, 0)^T$, $\vec{e}_\theta = (-\sin(\theta), \cos(\theta), 0, \dots, 0)^T$, and $(\vec{e}_k)_{k=1,\dots,d}$ is the canonical Cartesian basis of $\mathbb{R}^d$. For π smooth stationary probability of the diffusion $dX_t = b(X_t)dt + dW_t$ the condition $A_b^* \pi = 0$ can be written as

$$\frac{1}{2}\Delta\pi - \nabla \cdot (\pi b) = 0 \tag{51}$$

which yields in cylindrical coordinates to

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \pi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \pi}{\partial \theta^2} + \sum_{k=3}^d \frac{\partial^2 \pi}{\partial x_k^2} = \frac{1}{r}\frac{\partial\,(r\pi b_r)}{\partial r} + \frac{1}{r}\frac{\partial(\pi b_\theta)}{\partial \theta} + \sum_{k=3}^d \frac{\partial(\pi b_k)}{\partial x_k}. \tag{52}$$

Given a stationary probability $\pi = \pi(r, x_3, \dots, x_d) > 0$ independent of θ, we see that the drift b solution of (52) is given by $b = b_r \vec{e}_r + b_\theta \vec{e}_\theta + \sum_{k=3}^d b_k \vec{e}_k$ with

$$\begin{aligned} b_r &= \frac{1}{\pi}\frac{\partial \pi}{\partial r}, \quad b_\theta = 0, \\ b_k &= \frac{1}{\pi}\frac{\partial \pi}{\partial x_k}, \ \forall k \geq 3. \end{aligned} \tag{53}$$

• Construction of the priors. As in the proof of Theorem 3, the prior is based on two points $(\pi^{(0)}, \pi^{(T)})$. First, we define $\pi^{(0)}$. Let $\psi : [0, \infty) \to [0, \infty)$ be a smooth function, vanishing on $[0, 1/2]$ and satisfying $\psi(x) = x$ for $x \geq 1$. We let

$$\pi^{(0)}(r, x_3, \dots, x_d) = c_\eta e^{-\eta\psi(r)} \prod_{k=3}^d e^{-\eta\psi(|x_k|)},$$

where $\eta > 0$ and c_η is such that $\int_{[0,\infty)\times[0,2\pi)\times\mathbb{R}^{d-2}} \pi^{(0)}(r, x_3, \dots, x_d) r \, dr d\theta dx_3 \dots dx_d = 1$. Remark that for $\eta \to 0$, we have $c_\eta = O(\eta^d)$. Using that ψ has bounded derivatives and vanishes near 0 one can check that $(x_1, \dots, x_d) \mapsto \pi^{(0)}(\sqrt{x_1^2 + x_2^2}, x_3, \dots, x_d)$ is a smooth function and that $\left\| \frac{\partial^l}{\partial x_k^l}\pi^{(0)}(\sqrt{x_1^2 + x_2^2}, x_3, \dots, x_d) \right\|_\infty = O(c_\eta) = O(\eta^d)$ for all $1 \leq k \leq d$ and $l \geq 0$. Hence, we can choose η small enough such that $\pi^{(0)} \in \mathcal{H}_d(\beta, \mathcal{L})$. The drift function $b^{(0)}$ associated to $\pi^{(0)}$ given by (53) is such that $b_r^{(0)}(r, x_3, \dots, x_d) = -\eta\psi'(r) = -\eta$ if $r > 1$ and $b_k^{(0)}(r, x_3, \dots, x_d) = -\eta\,\mathrm{sgn}(x_k)$ for $|x_k| > 1$. By computation analogous to the ones before, there exists $\tilde{C} > 0$ and $\tilde{\rho} > 0$ such that

$< b^{(0)}(x), x >\leq -\tilde{C}|x|$ for $x \geq \tilde{\rho}$. Moreover, $\left\|b^{(0)}\right\|_\infty + \left\|\nabla b^{(0)}\right\|_\infty = O(\eta)$. Hence, if η is chosen small enough, $(\mathbb{I}_{d\times d}, b^{(0)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0/2, b_1/2, \tilde{C}, \tilde{\rho}) \subset \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$.

The construction of $\pi^{(T)}$ is more elaborate. We add a bump centered at 0 to $\pi^{(0)}$. Let $K : \mathbb{R} \to \mathbb{R}$ be a smooth function with support on $[-1, 1]$ and satisfying (42). We set $\pi^{(1)} = \pi^{(0)} + \frac{1}{M_T} d_T$ where

$$d_T(r, x_3, \dots, x_d) = J_{r_{\min}(T), r_{\max}(T)}(r) \times \prod_{i=3}^{d} K(\frac{x_i}{h_i(T)}), \tag{54}$$

where $0 < r_{\min}(T) < r_{\max}(T)/4 < r_{\max}(T) \leq h_3(T) \leq \dots \leq h_d(T)$ will be calibrated later and goes to zero with a rate polynomial in $1/T$. In the following we will suppress the dependence in T of $r_{\min}, r_{\max}, (h_j)_{3\leq j\leq d}$ in order to lighten the notations. The function $J_{r_{\min}, r_{\max}} : [0, \infty) \to \mathbb{R}$ is a smooth function with support on $[0, r_{\max}]$ satisfying the following properties:

$$J_{r_{\min}, r_{\max}} \text{is constant on } [0, r_{\min}/4] \text{ and decreasing on } [r_{\min}/4, r_{\max}] \text{ with } J_{r_{\min}, r_{\max}}(0) \geq 1 \tag{55}$$

$$0 \leq J_{r_{\min}, r_{\max}}(r) \leq c \left(\frac{\ln(r_{\max}/r) + 1}{\ln(r_{\max}/r_{\min})} \wedge 1 \right), \tag{56}$$

$$\left| \frac{\partial J_{r_{\min}, r_{\max}}(r)}{\partial r} \right| \leq \frac{c}{(\ln(r_{\max}/r_{\min}))(r \wedge r_{\min})}, \tag{57}$$

$$\left| \frac{\partial^k J_{r_{\min}, r_{\max}}(r)}{\partial r^k} \right| \leq \frac{c(k)}{(\ln(r_{\max}/r_{\min}))r_{\min}^k}, \ \forall k \geq 1, \tag{58}$$

where the constants c and $c(k)$ are independent of $r_{\min}$, $r_{\max}$. The existence of such function $J_{r_{\min}, r_{\max}}$ is postponed to Lemma 5, and we give in Remark 2 some hints underneath its construction. Remark that $\pi^{(1)}$ is a probability measure for T large enough, as the mean of d_T is zero and $\pi^{(1)}$ is positive if T is large enough, as we assume $1/M_T \to 0$.

We denote by $b^{(1)}$ the drift function associated to $\pi^{(1)}$ through the relations (53). We now prove the following lemma.

Lemma 4. *We have,*

$$\int_{\mathbb{R}^d} \left| b^{(1)}(\sqrt{x_1^2 + x_2^2}, x_3, \dots, x_d) - b^{(0)}(\sqrt{x_1^2 + x_2^2}, x_3, \dots, x_d) \right|^2 dx_1 \dots dx_d \leq c \frac{\prod_{i=3}^d h_i}{M_T^2 \ln(\frac{r_{max}}{r_{min}})}, \tag{59}$$

$$\left\| b^{(1)} - b^{(0)} \right\|_\infty \leq \frac{c}{M_T} \left[\frac{1}{r_{min} \ln(\frac{r_{max}}{r_{max}})} + \frac{1}{h_3} \right], \tag{60}$$

$$\left\| \nabla b^{(1)} - \nabla b^{(0)} \right\|_\infty \leq \frac{c}{M_T} \left[\frac{1}{r_{min}^2 \ln(\frac{r_{max}}{r_{max}})} + \frac{1}{h_3^2} \right]. \tag{61}$$

Proof. • We start with the proof of (59). Recalling $\pi^{(1)} = \pi^{(0)} + \frac{d_T}{M_T}$ and using (53), we deduce

$$b^{(1)} - b^{(0)} = \frac{d_T}{M_T \pi^{(1)}} b^{(0)} + \frac{1}{M_T \pi^{(1)}} \sum_{k=3}^{d} \frac{\partial d_T}{\partial x_k} \vec{e}_k + \frac{1}{M_T \pi^{(1)}} \frac{\partial d_T}{\partial r} \vec{e}_r. \tag{62}$$

In the evaluation of the $\mathbf{L}^2$ norm of $b^{(1)} - b^{(0)}$, we start by the contribution of the radial component. From (62), we have

$$b_r^{(1)} - b_r^{(0)} = b_r^{(0)} \left(\frac{d_T}{M_T \pi^{(1)}} \right) + \frac{1}{M_T \pi^{(1)}} \frac{\partial d_T}{\partial r}.$$

Remarking that $\pi^{(1)}$ and $\pi^{(0)}$ coincides out of the compact set $K_T = \{r \leq r_{\max}\} \times [0, 2\pi] \times \prod_{i=3}^d [-h_i, h_i]$, we deduce that $b^{(1)}$ and $b^{(0)}$ are equal on K_T^c. Thus,

$$\begin{aligned}
\left\| b_r^{(1)} - b_r^{(0)} \right\|_2^2 &= \int_{K_T} |b_r^{(1)}(r, x_3, \dots, x_d) - b_r^{(0)}(r, x_3, \dots, x_d)|^2 r dr d\theta dx_3 \dots dx_d \\
&\leq \frac{c}{M_T^2} \int_{K_T} \left[|d_T(r, x_3, \dots, x_d)|^2 + \left| \frac{\partial d_T(r, x_3, \dots, x_d)}{\partial r} \right|^2 \right] r dr d\theta dx_3 \dots dx_d \\
&=: \frac{c}{M_T^2} (I_1(T) + I_2(T)),
\end{aligned} \tag{63}$$

where we used that $\pi^{(1)}$ is lower bounded independently of T on the compact set K_T, for T large enough, and the boundedness of $b^{(0)}$. Using the definition of d_T given by (54), and (56), we have

$$\begin{aligned}
I_1(T) &= \int_0^{r_{\max}} |J_{r_{\min},r_{\max}}(r)|^2 r dr \int_{\prod_{i=3}^d [-h_i,h_i]} \prod_{i=3}^d K(\frac{x_i}{h_i})^2 dx_3 \dots dx_d \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \int_0^{r_{\max}} \left(\frac{\ln(r_{\max}/r)+1}{\ln(r_{\max}/r_{\min})} \wedge 1 \right)^2 r dr \left(\prod_{i=3}^d h_i \right) \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \left[\int_0^{r_{\max}/\sqrt{\ln(r_{\max}/r_{\min})}} r dr + \int_{r_{\max}/\sqrt{\ln(r_{\max}/r_{\min})}}^{r_{\max}} \left(\frac{\ln(r_{\max}/r)+1}{\ln(r_{\max}/r_{\min})} \right)^2 r dr \right] \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \left[\frac{r_{\max}^2}{\ln(r_{\max}/r_{\min})} + \int_{r_{\max}/\sqrt{\ln(r_{\max}/r_{\min})}}^{r_{\max}} \left(\frac{\ln(\sqrt{\ln(r_{\max}/r_{\min})})+1}{\ln(r_{\max}/r_{\min})} \right)^2 r dr \right] \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \left[\frac{r_{\max}^2}{\ln(r_{\max}/r_{\min})} + \int_{r_{\max}/\sqrt{\ln(r_{\max}/r_{\min})}}^{r_{\max}} \frac{c}{\ln(r_{\max}/r_{\min})} r dr \right] \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \frac{r_{\max}^2}{\ln(r_{\max}/r_{\min})}. \qquad (64)
\end{aligned}$$

Using now (57),

$$\begin{aligned}
I_2(T) &= \int_0^{r_{\max}} |\frac{\partial J_{r_{\min},r_{\max}}(r)}{\partial r}|^2 r dr \int_{\prod_{i=3}^d [-h_i,h_i]} \prod_{i=3}^d K(\frac{x_l}{h_l(T)})^2 dx_3 \dots dx_d \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \int_0^{r_{\max}} \left(\frac{1}{(\ln(r_{\max}/r_{\min}))(r \wedge r_{\min})} \right)^2 r dr \left(\prod_{i=3}^d h_i \right) \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \frac{1}{(\ln(r_{\max}/r_{\min}))^2} \left[\int_0^{r_{\min}} \frac{1}{r_{\min}^2} r dr + \int_{r_{\min}}^{r_{\max}} \frac{r}{r^2} dr \right] \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \frac{1}{(\ln(r_{\max}/r_{\min}))^2} \left[1 + \ln(\frac{r_{\max}}{r_{\min}}) \right] \\
&\leq c \left\|K\right\|_\infty^{2(d-3)} \left(\prod_{i=3}^d h_i \right) \frac{1}{(\ln(r_{\max}/r_{\min}))} \qquad (65)
\end{aligned}$$

Collecting (63), (64) and (65), we deduce

$$\left\| b_r^{(1)} - b_r^{(0)} \right\|_2^2 \leq c \frac{\prod_{i=3}^d h_i}{M_T^2 \ln(\frac{r_{\max}}{r_{\min}})}. \qquad (66)$$

We now compute the contribution of b_k for $k \geq 3$ in the $\mathbf{L}^2$ norms of $b^{(1)} - b^{(0)}$. Using (62), we have

$$\begin{aligned}
\left\| b_k^{(1)} - b_k^{(0)} \right\|_2^2 &\leq \frac{c}{M_T^2} \int_{K_T} \left[|d_T(r,x_3,\dots,x_d)|^2 + \left| \frac{\partial d_T(r,x_3,\dots,x_d)}{\partial x_k} \right|^2 \right] r dr d\theta dx_3 \dots dx_d \\
&:= \frac{c}{M_T^2} (I_1(T) + I_3(T)). \qquad (67)
\end{aligned}$$

We have to upper bound $I_3(T)$ which is the new term. From the definition of d_T, we have

$$\begin{aligned}
I_3 &= \int_0^{r_{\max}} |J_{r_{\min},r_{\max}}(r)|^2 r dr \int_{\prod_{i=3}^d [-h_i,h_i]} \left(\prod_{\substack{i=3 \\ i \neq k}}^d K(\frac{x_i}{h_i})^2 \right) |K'(\frac{x_k}{h_k})|^2 \frac{1}{h_k^2} dx_3 \dots dx_d \\
&\leq \left\|K\right\|_\infty^{2(d-4)} \left\|K'\right\|_\infty^2 \int_0^{r_{\max}} \left(\frac{\ln(r_{\max}/r)+1}{\ln(r_{\max}/r_{\min})} \wedge 1 \right)^2 r dr \left(\frac{\prod_{i=3}^d h_i}{h_k^2} \right)
\end{aligned}$$

where we used (57). Now by exactly the same computation yielding to (64), we have,

$$I_3 \leq c\,\|K\|_\infty^{2(d-4)}\,\|K'\|_\infty^2\,\frac{\prod_{i=3}^d h_i}{h_k^2}\frac{r_{\max}^2}{\ln(r_{\max}/r_{\min})}.$$

Using that $r_{\max} \leq h_k$, we deduce

$$I_3 \leq c\,\|K\|_\infty^{2(d-4)}\,\|K'\|_\infty^2\,\frac{\prod_{i=3}^d h_i}{\ln(r_{\max}/r_{\min})} \tag{68}$$

Collecting (67), (64) and (68) we get

$$\left\|b_k^{(1)} - b_k^{(0)}\right\|_2^2 \leq c\frac{\prod_{i=3}^d h_i}{M_T^2 \ln(\frac{r_{\max}}{r_{\min}})}. \tag{69}$$

Eventually, the upper bound (59) is a consequence of (66), (69) and $b_\theta^{(1)} = b_\theta^{(0)} = 0$.

- We now prove (60). Using again $\pi^{(1)} = \pi^{(0)} + \frac{d_T}{M_T}$ and (53), we have

$$\left\|b^{(1)} - b^{(0)}\right\|_\infty \leq \frac{c}{M_T}\left[\|d_T\|_\infty + \left\|\frac{\partial d_T}{\partial r}\right\|_\infty + \sum_{k=3}^d \left\|\frac{\partial d_T}{\partial x_k}\right\|_\infty\right].$$

By the definition (54) of d_T with (55), we have $\|d_T\|_\infty \leq c$ and $\left\|\frac{\partial d_T}{\partial x_k}\right\|_\infty \leq \frac{c}{h_k}$ for $k \geq 3$. Using (58), we have $\left\|\frac{\partial d_T}{\partial r}\right\|_\infty \leq \frac{c}{r_{\min}\ln(\frac{r_{\max}}{r_{\min}})}$. The upper bound (60) follows.

- Finally, we prove (61). By (53) and $\pi^{(1)} = \pi^{(0)} + \frac{d_T}{M_T}$, we have after some computations

$$\left\|\nabla b^{(1)} - \nabla b^{(0)}\right\|_\infty \leq \frac{c}{M_T}\Big[\|d_T\|_\infty + \sum_{y\in\{r,x_3,\dots,x_d\}} \|d_T\|_\infty \left\|\frac{\partial d_T}{\partial y}\right\|_\infty +$$
$$\frac{1}{M_T}\sum_{y,y'\in\{r,x_3,\dots,x_d\}} \left\|\frac{\partial d_T}{\partial y}\right\|_\infty \left\|\frac{\partial d_T}{\partial y'}\right\|_\infty + \sum_{y,y'\in\{r,x_3,\dots,x_d\}} \left\|\frac{\partial^2 d_T}{\partial y \partial y'}\right\|_\infty\Big].$$

The upperbound (61) follows from $\left\|\frac{\partial^2 d_T}{\partial x_i \partial x_j}\right\|_\infty \leq c/(h_i h_j)$, $\left\|\frac{\partial^2 d_T}{\partial x_i \partial r}\right\|_\infty \leq c/(h_i r_{\min}\ln(\frac{r_{\max}}{r_{\min}}))$ and $\left\|\frac{\partial^2 d_T}{\partial r^2}\right\|_\infty \leq c/(r_{\min}^2 \ln(\frac{r_{\max}}{r_{\min}}))$. □

We can now discuss the conditions ensuring that $(\mathbb{I}_{d\times d}, b^{(1)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$. Using that $J_{r_{\min},r_{\max}}$ is constant in a neighbourhood of zero, we deduce that $(x_1,\dots,x_d) \mapsto \pi^{(1)}(\sqrt{x_1^2+x_2^2}, x_3,\dots,x_d)$ is a smooth function and by (54) with (58) we have

$$\left|\frac{\partial^k}{\partial x_1^k}\pi^{(1)}(\sqrt{x_1^2+x_2^2}, x_3,\dots,x_d)\right| + \left|\frac{\partial^k}{\partial x_2^k}\pi^{(1)}(\sqrt{x_1^2+x_2^2}, x_3,\dots,x_d)\right| \leq \frac{c(k)}{M_T \ln(\frac{r_{\max}}{r_{\min}}) r_{\min}^k},$$
$$\left|\frac{\partial^k}{\partial x_l^k}\pi^{(1)}(\sqrt{x_1^2+x_2^2}, x_3,\dots,x_d)\right| \leq \frac{c(k)}{M_T h_l^k}, \quad \forall k\geq 1, \forall l \in \{3,\dots,d\}.$$

Then, the following condition is sufficient to ensure that $\pi^{(1)} \in \mathcal{H}(\beta, 2\mathcal{L})$:

$$\frac{1}{M_T} \leq \varepsilon_0 r_{\min}^{\beta_l}\ln(r_{\max}/r_{\min}), \text{for } l = 1,2 \text{ and}, \quad \frac{1}{M_T} \leq \varepsilon_0 h_l^{\beta_l},\ \forall l \in \{3,\dots,d\}, \tag{70}$$

for some constant $\varepsilon_0 > 0$ small enough. Remark also that we have $< b^{(1)}(x), x >\leq -\tilde{C}|x|$ for $|x| \geq \tilde{\rho}$ for T large enough, using that $b^{(1)}$ and $b^{(0)}$ coincide on K_T^c. Moreover, to get that for T large enough, $\left\|b^{(1)}\right\|_\infty \leq b_0$ and $\left\|\nabla b^{(1)}\right\| \leq b_1$ it is sufficient that

$$\frac{c}{M_T}\left[\frac{1}{r_{\min}^2 \ln(\frac{r_{\max}}{r_{\max}})} + \frac{1}{h_3^2}\right] \xrightarrow{T\to\infty} 0. \tag{71}$$

Thus, we have $(\mathbb{I}_{d\times d}, b^{(1)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$ as soon as (70) and (71) are valid.

• Choice of the calibration and proof of (13).
Repeating the proof of (49) we have

$$\mathcal{R}_T(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \gtrsim \frac{1}{M_T^2},$$

as soon as $\sup_{T\geq 0} T \int_{\mathbb{R}^d} |b^{(1)}(x) - b^{(0)}(x)|^2 \pi^{(0)}(x)\, dx < \infty$. From (59) this condition is implied by

$$\sup_{T\geq 0} T \frac{\prod_{i=3}^d h_i}{M_T^2 \ln(\frac{r_{\max}}{r_{\min}})} < \infty. \tag{72}$$

Now, we search $1/M_T \to 0$ tending to zero as slow as possible subject to the existence of $0 < r_{\min} < r_{\max}/4 < r_{\max} \leq h_3 \leq \cdots \leq h_d$ satisfying (70)–(71) and (72). For $l \geq 3$, we set $h_l = (\frac{1}{M_T \varepsilon_0})^{1/\beta_l}$, $r_{\max} = h_3$ and $r_{\min} = (\frac{1}{M_T \varepsilon_0})^{1/\beta_2}$. We have $\log(r_{\max}/r_{\min}) \sim_{T\to\infty} (\frac{1}{\beta_2} - \frac{1}{\beta_3}) \ln(M_T) \to \infty$ using $\beta_2 < \beta_3$. Since $\beta_1 \leq \beta_2$, we deduce that the conditions $\frac{1}{M_T} \leq \varepsilon_0 r_{\min}^{\beta_l} \ln(r_{\max}/r_{\min})$, for $l = 1, 2$ hold true when T is large enough. Hence, (70) is satisfied, and with these choices (71) becomes a consequence of $\beta_3 > \beta_2 > 2$. Replacing h_i, $r_{\min}$, $r_{\max}$ by their expression in function of M_T, the condition (72) writes,

$$\sup_{T\geq 0} T \frac{1}{M_T^{2+(d-2)/\overline{\beta}_3} \ln(M_T)} < \infty,$$

and is satisfied for the choice $M_T = (T/\ln(T))^{\frac{\overline{\beta}_3}{2\overline{\beta}_3+(d-2)}}$. This yields to

$$\mathcal{R}_T(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \gtrsim (T/\ln(T))^{-\frac{2\overline{\beta}_3}{2\overline{\beta}_3+(d-2)}}.$$

To complete the proof Theorem 4, it remains to show the existence of the function $J_{r_{\min}, r_{\max}}$. This is done in the next lemma. □

Lemma 5. *There exists a smooth function* $J_{r_{min}, r_{max}} : \mathbb{R} \to [0, \infty)$ *with compact support on* $[0, r_{\max}]$ *such that* (55)–(58) *hold true.*

Proof. We let $\varphi : [0, \infty) \to [0, 1]$ be a smooth function with $\varphi_{|[0,1/4]} = 0$ and $\varphi_{|[1/2,\infty)} = 1$. We define piecewise the function $J_{r_{\min}, r_{\max}}$ as follows,

$$J_{r_{\min},r_{\max}}(r) = 0, \quad \text{for } r \geq r_{\max},$$
$$J_{r_{\min},r_{\max}}(r) = \frac{1}{\ln(r_{\max}/r_{\min})} \int_r^{r_{\max}} \varphi(1 - r/r_{\max}) \frac{dr}{r}, \quad \text{for } r_{\max}/2 \leq r \leq r_{\max},$$
$$J_{r_{\min},r_{\max}}(r) = J_{r_{\min},r_{\max}}(r_{\max}/2) + \frac{1}{\ln(r_{\max}/r_{\min})} \int_r^{r_{\max}/2} \frac{dr}{r}, \quad \text{for } r_{\min}/2 \leq r \leq r_{\max}/2,$$
$$J_{r_{\min},r_{\max}}(r) = J_{r_{\min},r_{\max}}(r_{\min}/2) + \frac{1}{\ln(r_{\max}/r_{\min})} \int_r^{r_{\min}/2} \varphi(r/r_{\min}) \frac{dr}{r}, \quad \text{for } 0 \leq r \leq r_{\min}/2.$$

Using the definition of φ, one can check that the derivative of the function J is smooth and that J is decreasing with $J'(r) = 0$ for $|r| \leq r_{\min}/4$. By simple computation we have,

$$J_{r_{\min},r_{\max}}(r) = \frac{\Phi_1(r/r_{\max})}{\ln(r_{\max}/r_{\min})}, \quad \text{for } r_{\max}/2 \leq r \leq r_{\max}, \tag{73}$$
$$J_{r_{\min},r_{\max}}(r) = \frac{\Phi_1(1/2) + \ln(\frac{r_{\max}}{2r})}{\ln(r_{\max}/r_{\min})}, \quad \text{for } r_{\min}/2 \leq r \leq r_{\max}/2, \tag{74}$$
$$J_{r_{\min},r_{\max}}(r) = 1 + \frac{\Phi_1(1/2) + \Phi_2(r/r_{\min})}{\ln(r_{\max}/r_{\min})}, \quad \text{for } 0 \leq r \leq r_{\min}/2, \tag{75}$$

where $\Phi_1(u) = \int_u^1 \varphi(1-s) \frac{ds}{s}$ and $\Phi_2(u) = \int_u^{1/2} \varphi(s) \frac{ds}{s}$ are smooth functions on $[0, 1]$.

Thus, using monotonocity of $J_{r_{\min},r_{\max}}$ and $\varphi \geq 0$, we have $J_{r_{\min},r_{\max}}(0) \geq J_{r_{\min},r_{\max}}(r_{\min}/2) \geq 1$. We deduce that (55) is true.

Using (73)–(75) and that Φ_1 and Φ_2 are bounded on $[0, 1]$, we deduce (56). The condition (57) follows again from (73)–(75) and the boundedness of Φ_1' and Φ_2'. The condition (58) is shown by differentiating k times the representations (73)–(75) and using $r_{\min} \leq r_{\max}$. □

Remark 2. *The idea underneath the construction of the bump* $(r,\theta) \mapsto J_{r_{min},r_{max}}(r)$ *in Lemma 5 is the following. Having in mind* (51), *we have constructed a smooth bump, with radius* r_{max}*, height greater than* 1*, and solution of* $\Delta J_{r_{min},r_{max}} = 0$ *on the torus* $r \in [r_{min}/2, r_{max}/2]$.

6.3 Proof of Theorem 5

Proof. The proof of Theorem 5 heavily relies on the proof of Theorem 4. As done above, we prove the theorem in the case where $x_0 = (0,0)$ and $a_{min} = 1$. We will lower bound the risk on the subclass of model (3) given by

$$dX_t = b(X_t)dt + dW_t,$$

where b is any drift function such that $(\mathbb{I}_{d\times d}, b) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$. Following the proof of Theorem 4 we introduce cylindrical coordinates (r,θ) that translates into the Cartesian coordinates $(r\cos(\theta), r\sin(\theta))$. In two dimensions (52) is written without the sum over k and the drift b solution of (52) is given by $b = b_r\vec{e}_r + b_\theta\vec{e}_\theta$ with $\vec{e}_r = (\cos(\theta), \sin(\theta))^T$, $\vec{e}_\theta = (-\sin(\theta), \cos(\theta),)^T$ and

$$b_r = \frac{1}{\pi}\frac{\partial \pi}{\partial r}, \quad b_\theta = 0.$$

• Construction of the priors. Let $\psi : [0,\infty) \to [0,\infty)$ be a smooth function, vanishing on $[0, 1/2]$ and satisfying $\psi(x) = x$ for $x \geq 1$ as in the proof of Theorem 4. We let

$$\pi^{(0)}(r) = c_\eta e^{-\eta\psi(r)},$$

where $\eta > 0$ and c_η is such that $\int_{[0,\infty)\times[0,2\pi)} \pi^{(0)}(r) r \, dr d\theta = 1$. Again, for η small enough we have $\pi^{(0)} \in \mathcal{H}_2(\beta, \mathcal{L})$. Moreover, acting as in the proof of Theorem 4, it is easy to see that $(\mathbb{I}_{d\times d}, b^{(0)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0/2, b_1/2, \tilde{C}, \tilde{\rho}) \subset \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$.

Regarding the construction of $\pi^{(1)}$ is more elaborate, as before we add a bump centered at 0 to $\pi^{(0)}$. We set

$$\pi^{(1)} = \pi^{(0)} + \frac{1}{M_T} J_{r_{\min}(T), r_{\max}(T)}(r),$$

where $J_{r_{\min}(T), r_{\max}(T)}$ is the function introduced in the proof of Theorem 4 which satisfies (55), (56), (57) and (58) and whose existence has been proven in Lemma 5. We recall that $0 < r_{\min}(T) < r_{\max}(T)/4 < r_{\max}(T)$ will be calibrated later and go to zero with a rate polynomial in $1/T$. We denote by $b^{(1)}$ the drift function associated to $\pi^{(1)}$.
Following the proof of Lemma 4 in the bi-dimensional context it is easy to see that the following bounds hold true.

$$\int_{\mathbb{R}^d} \left| b^{(1)}(\sqrt{x_1^2 + x_2^2}) - b^{(0)}(\sqrt{x_1^2 + x_2^2}) \right|^2 dx_1 dx_2 \leq c \frac{1}{M_T^2 \ln(\frac{r_{\max}}{r_{\min}})},$$

$$\left\| b^{(1)} - b^{(0)} \right\|_\infty \leq \frac{c}{M_T} \frac{1}{r_{\min} \ln(\frac{r_{\max}}{r_{\max}})},$$

$$\left\| \nabla b^{(1)} - \nabla b^{(0)} \right\|_\infty \leq \frac{c}{M_T} \frac{1}{r_{\min}^2 \ln(\frac{r_{\max}}{r_{\max}})}.$$

Moreover we still have, for any $k \geq 1$,

$$\left| \frac{\partial^k}{\partial x_1^k} \pi^{(1)}(\sqrt{x_1^2 + x_2^2}) \right| + \left| \frac{\partial^k}{\partial x_2^k} \pi^{(1)}(\sqrt{x_1^2 + x_2^2}) \right| \leq \frac{c(k)}{M_T \ln(\frac{r_{\max}}{r_{\min}}) r_{\min}^k}.$$

Then, the following condition is sufficient to ensure that $\pi^{(1)} \in \mathcal{H}_2(\beta, 2\mathcal{L})$:

$$\frac{1}{M_T} \leq \varepsilon_0 r_{\min}^{\beta_l} \ln(r_{\max}/r_{\min}), \text{for } l = 1, 2 \text{ and} \tag{76}$$

for some constant $\varepsilon_0 > 0$ small enough. As before, if (76) holds true it is sufficient that

$$\frac{c}{M_T} \frac{1}{r_{\min}^2 \ln(\frac{r_{\max}}{r_{\max}})} \xrightarrow{T\to\infty} 0 \tag{77}$$

to obtain $(\mathbb{I}_{d\times d}, b^{(1)}) \in \Sigma(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho})$.

• Choice of the calibration and proof of (13).

Repeating the proof of (49) we clearly have

$$\mathcal{R}_T(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \gtrsim \frac{1}{M_T^2},$$

as soon as $\sup_{T\geq 0} T \int_{\mathbb{R}^2} |b^{(1)}(x) - b^{(0)}(x)|^2 \pi^{(0)}(x)\, dx < \infty$, which is implied by

$$\sup_{T\geq 0} T \frac{1}{M_T^2 \ln(\frac{r_{\max}}{r_{\min}})} < \infty. \tag{78}$$

We look for $1/M_T$ tending to zero as slow as possible subject to the existence of $0 < r_{\min} < r_{\max}/4 < r_{\max}$ satisfying (76)–(77) and (78). We set $r_{\min} = (\frac{1}{M_T \varepsilon_0})^{1/\beta_2}$ and $r_{\max} = (\frac{1}{M_T \varepsilon_0})^{1/\beta_2 - \gamma}$ for any arbitrary $\gamma > 0$. It follows $\log(r_{\max}/r_{\min}) \sim_{T\to\infty} (\frac{1}{\beta_2} - (\frac{1}{\beta_2} - \gamma)) \ln(M_T) \to \infty$ as $\gamma > 0$. Since $\beta_1 \leq \beta_2$, we clearly have from the definition of $r_{\min}$ and $r_{\max}$ that the conditions $\frac{1}{M_T} \leq \varepsilon_0 r_{\min}^{\beta_l} \ln(r_{\max}/r_{\min})$, for $l = 1, 2$ hold true when T is large enough. Hence, (76) is satisfied. The same can be said about (77), recalling we have assumed $\beta_2 > 2$. Replacing $r_{\min}$ and $r_{\max}$ by their expression in function of M_T, condition (78) writes,

$$\sup_{T\geq 0} T \frac{1}{M_T^2 \ln(M_T)} < \infty,$$

and is satisfied for the choice $M_T = (\frac{T}{\log T})^{\frac{1}{2}}$. It provides

$$\mathcal{R}_T(\beta, \mathcal{L}, 1, a_0, a_1, b_0, b_1, \tilde{C}, \tilde{\rho}) \gtrsim \frac{\log T}{T},$$

as we wanted.

□

A Appendix: proof of technical results

This section is devoted to the proof of the results which are more technical and for which some preliminaries are needed.

A.1 Proof of Lemma 3

Recalling Lemma 2, the main novelty is to prove that the control (7) is uniform on the class of coefficients $(a, b) \in \Sigma$. To simplify the proof, we first strengthen the hypothesis on the coefficients of the diffusion by assuming that they are infinitely differentiable with bounded derivatives or any order, together with the condition $(a, b) \in \Sigma$.

The proof of (7) relies on the theory of Lyapunov-Poincaré inequalities introduced in [12]. We recall some definitions related to these functional inequalities. First, if $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ we define $\Gamma(f) = A(f^2) - 2fA(f)$ where A is the generator of the diffusion, and we have $\Gamma(f) = (\nabla f)^T \tilde{a} \nabla f = |a^T \nabla f|^2$.

Definition 3. *(Lyapunov-Poincaré Inequality [12]) Let W be a $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ function such that $W \geq 1$, $W \in \mathbf{L}^2(\pi)$, $AW \in \mathbf{L}^2(\pi)$. The probability π satisfies a W-Lyapunov-Poincaré inequality if there exists $C_{LP} > 0$ such that for all $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$, with f bounded and ∇f bounded, and $\int_{\mathbb{R}^d} f(x)\pi(x)dx = 0$ we have*

$$\int_{\mathbb{R}^d} f^2(x) W(x)\pi(x)dx \leq C_{LP} \int_{\mathbb{R}^d} [W(x)\Gamma(f)(x) - f(x)^2 AW(x)]\pi(x)dx. \tag{79}$$

We can now state a result of [12]. For the sake of completeness, we give a detailed proof adapted to our context of this result.

Lemma 6 (Bakry et al. [12]). *Assume that the W-Lyapunov-Poincaré inequality holds true with some constant $C_{LP} > 0$. Then, for any $\varphi : \mathbb{R}^d \to \mathbb{R}$ bounded function with $\int_{\mathbb{R}^d} \varphi(x)\pi(x)dx = 0$, we have*

$$\int_{\mathbb{R}^d} [P_t(\varphi)(x)]^2 \pi(x)dx \le e^{-t/C_{LP}} \int_{\mathbb{R}^d} \varphi^2(x)W(x)\pi(x)dx, \tag{80}$$

where $(P_t)_t$ is the semi-group associated to the process X.

Proof. • First, we establish some properties on the stationary density π that will be useful in the sequel. Using Friedman [30], we know that,

$$\left| \frac{\partial^{i+j}}{\partial x^i y^j} p_t(x,y) \right| \le \frac{c}{t^{(d+i+j)/2}} e^{-\frac{|x-y|^2}{ct}} \tag{81}$$

for $i+j \le 2$, for all $T > 0$, $t \in (0,T)$ for some $c > 0$ depending on T. Using the invariance of π, $\pi(y) = \int_{\mathbb{R}^d} \pi(x)p_t(x,y)dx$, we deduce that $\pi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$, and that π is bounded, with bounded derivatives.

From Lemma 1, we know that for $T > 0$, and $t \in (0,T)$, we have $p_t(x,y) \ge \frac{c_T}{t^{d/2}} e^{-\frac{|x-y|^2}{ct}}$, where c_T depends on T, and c depends only on the coefficients of the stochastic differential equation. Hence, we deduce from the invariance of π again that

$$\pi(y) = \int_{\mathbb{R}^d} \pi(x)p_T(x,y)dx \ge \int_{\mathbb{R}^d} \frac{c_T}{T^{d/2}} e^{-\frac{|x-y|^2}{cT}} \pi(x)dx \ge \frac{c_T}{T^{d/2}} \int_{\{|x|\le 1\}} e^{\frac{-2|x|^2}{cT}} \pi(x)dx e^{-\frac{2|y|^2}{cT}}.$$

As T can be chosen arbitrarily large and using $\pi > 0$, we deduce the lower bound

$$\forall \varepsilon > 0, \exists C_\varepsilon > 0, \forall y \in \mathbb{R}^d, \pi(y) \ge C_\varepsilon e^{-\varepsilon|y|^2}. \tag{82}$$

• We now prove (80). Remark that by a density argument we can assume that φ is $\mathcal{C}^2$ and compactly supported. We let $P_t^{*,\pi}$ be the adjoint in $\mathbf{L}^2(\mathbb{R}^d, \pi)$ of the operator P_t, which is given by the expression

$$P_t^{*,\pi}(f)(x) = \frac{1}{\pi(x)} \int_{\mathbb{R}^d} f(y)\pi(y)p_t(x,y)dy, \tag{83}$$

and $A^{*,\pi}$ the generator of the semi group $(P_t^{*,\pi})_{t\ge 0}$ (details can be found e.g. in [34]). If f is $\mathcal{C}^2$ then

$$A^{*,\pi} f = \frac{1}{2} \sum_{1\le i,j\le d} \tilde{a}_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i^* \frac{\partial f}{\partial x_i} \tag{84}$$

with $b_i^* = -b_i + \sum_{j=1}^d \frac{\partial \tilde{a}_{j,i}}{\partial x_j} + \frac{1}{\pi} \sum_{i=1}^d (\frac{\partial \pi}{\partial x_j})\tilde{a}_{j,i}$ as soon as the expression in the right hand side of (84) belongs to $\mathbf{L}^2(\pi)$. We know that if f is in the domain of $A^{*,\pi}$ then $t \mapsto P_t^{*,\pi}(f)$ is differentiable in $\mathbf{L}^2(\pi)$ and $\frac{\partial}{\partial t} P_t^{*,\pi}(f) = P_t^{*,\pi} A^{*,\pi}(f) = A^{*,\pi} P_t^{*,\pi}(f)$.

As in [12], we define $I_t = \int_{\mathbb{R}^d} (P_t^{*,\pi}\varphi(x))^2 W(x)\pi(x)dx$. By formal differentiation, we get for $t > 0$,

$$\frac{\partial I_t}{\partial t} = 2\int_{\mathbb{R}^d} P_t^{*,\pi}\varphi(x) A^{*,\pi} P_t^{*,\pi}\varphi(x)W(x)\pi(x)dx. \tag{85}$$

This differentiation can be justified by proving that $\frac{\partial}{\partial t}(P_t^{*,\pi}(\varphi))^2 = 2P_t^{*,\pi}\varphi \times A^{*,\pi}P_t^{*,\pi}\varphi$ is dominated by a function in $\mathbf{L}^1(W\pi)$. Using (83) and that π is invariant we get $|P_t^{*,\pi}(\varphi)| \le \|\varphi\|_\infty$. Also, $|A^{*,\pi}P_t^{*,\pi}\varphi| = |P_t^{*,\pi}A^{*,\pi}\varphi|$ is bounded by $\|A^{*,\pi}\varphi\|_\infty < \infty$, using (84) and the fact that φ is compactly supported. Since $W \in \mathbf{L}^1(\pi)$, we deduce (85). Then, using that $A^{*,\pi}$ is a differential operator by (84), we have $A^{*,\pi}((P_t^{*,\pi}(\varphi))^2) = 2P_t^{*,\pi}(\varphi)A^{*,\pi}(P_t^{*,\pi}(\varphi)) + \Gamma(P_t^{*,\pi}(\varphi))$. Hence,

$$\frac{\partial I_t}{\partial t} = \int_{\mathbb{R}^d} [A^{*,\pi}((P_t^{*,\pi}\varphi)^2)(x) - \Gamma(P_t^{*,\pi}(\varphi))]W(x)\pi(x)dx.$$

Using (81) with $i = 1$ and $j = 0$, and the fact that φ is compactly supported we deduce that from the differentiation of (83), $|\nabla_x P_t^{*,\pi}(\varphi)(x)| \le \frac{c}{t^{(d+1)/2}} \frac{1}{\pi(x)} \|\pi\|_\infty \|\varphi\|_\infty \int_{y\in \mathrm{supp}(\varphi)} e^{-\frac{|x-y|^2}{ct}} dy \le \frac{c'}{t^{(d+1)/2}} \frac{1}{\pi(x)} e^{-\frac{|x|^2}{ct}} \le \frac{c'}{t^{(d+1)/2}} e^{-\frac{|x|^2}{2ct}}$, where we used (82) with $\varepsilon > 0$ small enough. Thus, $\Gamma(P_t^{*,\pi}(\varphi)) \in$

$\mathbf{L}^1(W\pi)$ and we can write

$$\begin{aligned}\frac{\partial I_t}{\partial t} &= \int_{\mathbb{R}^d} A^{*,\pi}((P_t^{*,\pi}\varphi)^2)(x)W(x)\pi(x)dx - \int_{\mathbb{R}^d} \Gamma(P_t^{*,\pi}(\varphi))W(x)\pi(x)dx \\ &= \int_{\mathbb{R}^d} (P_t^{*,\pi}\varphi)^2(x)AW(x)\pi(x)dx - \int_{\mathbb{R}^d} \Gamma(P_t^{*,\pi}(\varphi))W(x)\pi(x)dx \\ &= \int_{\mathbb{R}^d} [(P_t^{*,\pi}\varphi)^2(x)AW(x) - \Gamma(P_t^{*,\pi}(\varphi))W(x)]\pi(x)dx.\end{aligned}$$

Now we use the Lyapunov-Poincaré inequality with $f = P_t^{*,\pi}(\varphi)$ and get $\frac{\partial I_t}{\partial t} \leq -(1/C_{LP})I_t$. From Gronwall's lemma it yields $I_t \leq e^{-t/C_{LP}} I_0$ and since $W \geq 1$,

$$\int_{\mathbb{R}^d} (P_t^{*,\pi}(\varphi)(x))^2\pi(x)dx \leq I_t \leq e^{-t/C_{LP}} I_0 = e^{-t/C_{LP}} \int_{\mathbb{R}^d} \varphi(x)^2 W(x)\pi(x)dx.$$

By duality between P_t and $P_t^{*,\pi}$ in $\mathbf{L}^2(\pi)$, we deduce (80). □

Lemma 7. *Assume that a and b are $\mathcal{C}^2$ with bounded derivatives and $(a,b) \in \Sigma$. Then, for φ bounded with $\int_{\mathbb{R}^d} \varphi(x)\pi(x)dx = 0$, and all $t > 0$,*

$$\int_{\mathbb{R}^d} [P_t(\varphi)(x)]^2\pi(x)dx \leq ce^{-t/c} \|\varphi\|_\infty^2, \tag{86}$$

where the constant $c > 0$ is uniform over the class Σ.

Proof. The main idea of the proof is that it is possible to show that for all $(a,b) \in \Sigma$, the stationary probability π satisfies a W-Lyapunov-Poincaré inequality with the same function W and same constant C_{LP}. Following [12], the existence of W-Lyapunov-Poincaré inequality is related to the existence of classical Lyapunov functions.

• First, we construct a Lyapunov function V independent of $(a,b) \in \Sigma$. We let $\chi \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ such that $0 \leq \chi(x) \leq |x|$ and $\chi(x) = |x|$ for $|x| \geq 1$ and we set $V(x) = e^{\varepsilon_0 \chi(x)}$ for some $\varepsilon_0 > 0$ which will be calibrated later. We have for $|x| \geq 1$, $\nabla V(x) = \varepsilon_0 \frac{x}{|x|} V(x)$ and thus,

$$AV(x) = \frac{1}{2} \sum_{1 \leq i,j \leq d} \tilde{a}_{i,j}(x) \frac{\partial^2 V}{\partial x_i \partial x_j}(x) + \epsilon_0 V(x) < \frac{x}{|x|}, b(x) > .$$

Using that for $|x| \geq 1$, $|\frac{\partial^2 V}{\partial x_i \partial x_j}| = |e^{\varepsilon_0|x|}\{\varepsilon_0^2 \frac{x_i x_j}{|x|^2} - \varepsilon_0 \frac{x_i x_j}{|x|^3} + \frac{\varepsilon_0}{|x|} 1_{\{i=j\}}\}| \leq e^{\varepsilon_0|x|}\{\varepsilon_0^2 + \frac{2\varepsilon_0}{|x|}\}$, and that $< x, b(x) > \leq -\tilde{C}|x|$ for $|x| \geq \tilde{\rho}$ by Assumption **A2**, we get,

$$AV(x) \leq -\tilde{C}\varepsilon_0 V(x) + \frac{\varepsilon_0}{2} V(x) \sum_{i,j} |\tilde{a}_{i,j}(x)|\{\varepsilon_0 + \frac{2}{|x|}\},$$

for all $|x| \geq \tilde{\rho} \vee 1$. Using $|\tilde{a}_{i,j}(x)| = |(aa^T)_{i,j}(x)| \leq |a(x)|^2 \leq a_0^2$ by Assumption **A1**, we deduce

$$AV(x) \leq -\tilde{C}\varepsilon_0 V(x)[1 - \frac{d^2 a_0^2}{2\tilde{C}}\{\varepsilon_0 + \frac{2}{|x|}\}].$$

Now, we set $\varepsilon_0 = \frac{\tilde{C}}{2d^2 a_0^2}$, and deduce for $|x| \geq \tilde{\rho} \vee 1 \vee \frac{4d^2 a_0^2}{\tilde{C}}$:

$$AV(x) \leq -\frac{\tilde{C}_b \varepsilon_0}{2} V(x).$$

We then define $\alpha = \tilde{C}\varepsilon_0/2$, $R = \tilde{\rho} \vee 1 \vee \frac{4d^2 a_0^2}{\tilde{C}}$ and $\beta = e^{\varepsilon_0 R}[\frac{1}{2} d^2 a_0^2 \varepsilon_0^2 \|\nabla\chi\|_\infty^2 + \frac{1}{2} d^2 a_0^2 \varepsilon_0 \times \sup_{1 \leq i,j \leq d} \left\| \partial^2_{x_i,x_j} \chi \right\|_\infty + b_0 \varepsilon_0 \|\nabla\chi\|_\infty + \alpha]$ where the notation b_0 is introduced in Assumption **A1**. We can check that $|AV(x) + \alpha V(x)| \leq \beta$ for $|x| \leq R$, and in turn V is a Lyapunov function :

$$AV(x) \leq -\alpha V(x) + \beta 1_{\{|x| \leq R\}}, \quad \forall x \in \mathbb{R}^d. \tag{87}$$

• We now prove that (79) holds true with the same function W and constant C_{LP} for all coefficients $(a,b) \in \Sigma$. Using Proposition 3.6 in [12] with (87), it is sufficient to find $R' > 0$ large enough such that

$$\{V \leq 2\beta/\alpha\} \subset B(0,R') \tag{88}$$
$$\pi(B(0,R')) > 1/2 \tag{89}$$

and the local Poincaré inequality is valid on $B(0,R')$ with some constant $\kappa_{R'}$: for all $f \in \mathcal{C}^1$ with bounded derivative,

$$\int_{B(0,R')} f^2(x)\pi(x)dx \leq \kappa_{R'} \int_{\mathbb{R}^d} \Gamma(f)\pi(x)dx + \frac{1}{\pi(B(0,R'))} \left(\int_{B(0,R')} f(x)\pi(x)dx \right)^2. \tag{90}$$

If the conditions (88)–(90) are valid, then by Proposition 3.6 in [12] we deduce the Lyapunov-Poincaré Inequality (79) with $W = V + (\beta\kappa_{R'} - 1)_+$ and $1/C_{LP} = 2\alpha(1 - \frac{\pi(B(0,R')^c)}{\pi(B(0,R'))}) \times (1 + (\beta\kappa_{R'} - 1)_+)^{-1}$.

To get (88), using the expression of V it is sufficient to take $R' \geq (\frac{1}{\varepsilon_0}\ln(2\beta/\alpha)) \vee 1$. Considering (89), we take the expectation with respect to π in (87) and obtain

$$0 = \int_{\mathbb{R}^d} AV(x)\pi(x)dx \leq -\alpha \int_{\mathbb{R}^d} V(x)\pi(x)dx + \beta \int_{\mathbb{R}^d} \pi(x)dx,$$

and thus

$$\int_{\mathbb{R}^d} V(x)\pi(x)dx \leq \frac{\beta}{\alpha}. \tag{91}$$

We deduce $\int_{|x|\geq 1} e^{\varepsilon_0|x|}\pi(x)dx \leq \int_{\mathbb{R}^d} V(x)\pi(x)dx \leq \frac{\beta}{\alpha}$. Using the Markov inequality, this yields, for any $R' \geq 1$, $\pi(\{x \mid |x| \geq R'\}) \leq e^{-\varepsilon_0 R'}\frac{\beta}{\alpha}$. It entails that $R' > \frac{1}{\varepsilon_0}\ln(2\beta/\alpha)$ is sufficient for the condition (89) to hold true. We set $R' = (\frac{1}{\varepsilon_0}\ln(2\beta/\alpha)) \vee 1$, and now we have to check the condition (90). From the Poincaré inequality on the ball $B(0,R')$ endowed with the Lebesgue measure (see e.g. Theorem 4.9 in [29]) and the Proposition 4.2.7 in [13], we know that if

$$1/c_\pi \leq \pi(x) \leq c_\pi, \quad \forall x \in B(0,R'), \tag{92}$$

for some $c_\pi > 0$ then,

$$\int_{B(0,R')} f^2(x)\pi(x)dx \leq Cc_\pi^3 R' \int_{B(0,R')} |\nabla f|^2\pi(x)dx + \frac{1}{\pi(B(0,R'))} \left(\int_{B(0,R')} f(x)\pi(x)dx \right)^2,$$

where C is some universal constant. As the matrix $\tilde{a}$ is lower bounded by Assumption **A1**, we have $|\nabla f|^2 \leq a_{min}^{-1}\Gamma(f)$, and we deduce that (90) holds true with $\kappa_R' = CR'c_\pi^3$. Consequently, by Proposition 3.6 in [12], the Lyapunov-Poincaré inequality holds true. Moreover, the constant C_{LP} in (79) is independent of $(a,b) \in \Sigma$, as soon as we can find a constant c_π in (92) independent of (a,b). Using the invariance of π and (4) with $t = 1$, gives for $x \in B(0,R')$,

$$c_1^{-1} \int_{B(0,R')} \pi(y)e^{-|x-y|^2/a_{\min}^2}dy \leq \pi(x) \leq c_1 \int_{\mathbb{R}^d} \pi(y)e^{-|x-y|^2/(4a_0^2)}dy$$
$$c_1^{-1} \int_{B(0,R')} \pi(y)e^{-4R'^2/a_{\min}^2}dy \leq \pi(x) \leq c_1 \int_{\mathbb{R}^d} \pi(y)dy$$
$$c_1^{-1}\frac{1}{2}e^{-4R'^2/a_{\min}^2} \leq \pi(x) \leq c_1, \tag{93}$$

where in the last line we used (89).

Hence, we have proved the W-Lyapunov-Poincaré inequality (79) with the same function W and constant C_{LP} for all $(a,b) \in \Sigma$.

• Eventually, we deduce (86) by applying Lemma 6, together with the upper bound

$$\int_{\mathbb{R}^d} W(x)\pi(x)dx \leq \int_{\mathbb{R}^d} V(x)\pi(x)dx + (\beta\kappa_{R'} - 1)_+ \leq \beta/\alpha + (\beta\kappa_{R'} - 1)_+$$

where in the last inequality we used (91). □

We now prove Lemma 3.

Proof. First, remark that by a density argument, we can assume in the proof that φ is a smooth function supported on K. The inequality (7) is a consequence of (86) in Lemma 3, but the latter requires that the coefficients of the S.D.E. are of class $\mathcal{C}^2$. Hence, an approximation of the initial S.D.E. by one with smoother coefficients is required. For (a,b) in $\Sigma = \Sigma(a_0, a_1, b_0, b_1, a_{\min}, \tilde{C}_{\mathrm{b}}, \tilde{\rho}_{\mathrm{b}})$, we introduce the following smooth approximations of a and b. Let η be a smooth function supported on the unit ball of $\mathbb{R}^d$ with $\int_{\mathbb{R}} \eta(x)dx = 1$, and we set $a_n = a \star \eta_n$, $b_n = b \star \eta_n$, where $\star$ is the convolution operator and $\eta_n(\cdot) = n\eta(n\cdot)$. As $(a,b) \in \Sigma$, we deduce that

$$\|a - a_n\|_\infty + \|b - b_n\|_\infty \leq \frac{c}{n}. \tag{94}$$

Moreover, for all n large enough, $(a_n, b_n) \in \Sigma' = \Sigma(2a_0, 2a_1, 2b_0, 2b_1, a_{\min}/2, \tilde{C}_{\mathrm{b}}/2, 2\tilde{\rho'}_{\mathrm{b}})$.

We denote by X^n the solution of the S.D.E. (3) with the coefficients (a_n, b_n) in place of (a,b), and by π_n the unique stationary distribution of X^n. In the sequel of the proof, we also emphasize the dependence on the initial condition of the process (3) by denoting as $\mathbb{E}_x$ (resp. $\mathbb{E}_\pi$) the expectation computed when the process starts with the initial condition $X_0 = x$ (resp. $X_0 \overset{\text{law}}{=} \pi$). Since a_n and b_n are smooth coefficients, by Lemma 7, we have

$$\int_{\mathbb{R}^d} \left[\mathbb{E}_x[\varphi(X_t^n)] - \pi_n(\varphi)\right]^2 \pi_n(x)dx \leq ce^{-t/c} \|\varphi\|_\infty^2 \tag{95}$$

where the constant $c = c_{\Sigma'}$ is independent of n and $(a,b) \in \Sigma$. Using classical estimates for solutions of stochastic differential equations with (94), it is possible to show

$$\sup_{x \in \mathbb{R}^d} \mathbb{E}_x \left[|X_t^n - X_t|\right] \leq \frac{c' e^{c't}}{n}, \forall t > 0, \forall n \geq 1, \tag{96}$$

and with some constant $c' > 0$.

Now, we write

$$\begin{aligned}
\mathbb{E}_\pi \left[\varphi(X_t)\varphi(X_0)\right] &= \int_{\mathbb{R}^d} P_t(\varphi)(x)\varphi(x)\pi(x)dx \\
&= \int_{\mathbb{R}^d} E_x[\varphi(X_t^n)]\varphi(x)\pi(x)dx + O(\frac{c'e^{c't}}{n} \|\varphi\|_\infty \|\varphi'\|_\infty),
\end{aligned}$$

where we used (96). On the compact K we can find a constant c_K such that $c_K^{-1} \leq \pi \leq c_K$, $c_K^{-1} \leq \pi_n \leq c_K$ (in the same way as we obtained (93)), using (95), we deduce

$$\int_K \left|\mathbb{E}_x[\varphi(X_t^n)] - \pi_n(\varphi)\right|^2 \pi(x)dx \leq c_K^2 ce^{-t/c} \|\varphi\|_\infty^2. \tag{97}$$

We deduce

$$\begin{aligned}
\mathbb{E}_\pi \left[\varphi(X_t)\varphi(X_0)\right] &= \int_K \pi_n(\varphi)\varphi(x)\pi(x)dx \\
&\quad + O\left(\int_K |\mathbb{E}_x[\varphi(X_t^n)] - \pi_n(\varphi)|\varphi(x)\pi(x)dx\right) + O(\frac{c'e^{c't}}{n} \|\varphi\|_\infty \|\varphi'\|_\infty) \\
&= \pi_n(\varphi)\pi(\varphi) + O(c_K \|\varphi\|_\infty^2 \sqrt{c}e^{-t/(2c)}) + O(\frac{c'e^{c't}}{n} \|\varphi\|_\infty \|\varphi'\|_\infty),
\end{aligned} \tag{98}$$

where in the last line we used (97). It remains to find some upper bound on $|\pi_n(\varphi) - \pi(\varphi)|$. We write for $t > 0$

$$\pi_n(\varphi) - \pi(\varphi) = \pi_n(\varphi) - \mathbb{E}_x(\varphi(X_t^n)) + \mathbb{E}_x(\varphi(X_t^n)) - P_t(\varphi)(x) + P_t(\varphi)(x) - \pi(\varphi)$$

and integrate on the compact set K with respect to the Lebesgue measure to find

$$\begin{aligned}
\mathrm{vol}(K)|\pi_n(\varphi) - \pi(\varphi)| &\leq c_K \|\pi_n(\varphi) - \mathbb{E}_x(\varphi(X_t^n))\|_{\mathbf{L}^1(\pi_n)} + \\
&\qquad \int_K |\mathbb{E}_x(\varphi(X_t^n)) - P_t(\varphi)(x)|\,dx + c_K \|P_t(\varphi)(x) - \pi(\varphi)\|_{\mathbf{L}^1(\pi)},
\end{aligned}$$

where we used that on K, π and π_n are lower bounded by $1/c_K$. From (96), we deduce

$$\mathrm{vol}(K)|\pi_n(\varphi) - \pi(\varphi)| \leq c_K \left\|\pi_n(\varphi) - \mathbb{E}_x(\varphi(X_t^n))\right\|_{\mathbf{L}^1(\pi_n)} + \mathrm{vol}(K)\frac{c'e^{c't}}{n} + c_K \left\|P_t(\varphi)(x) - \pi(\varphi)\right\|_{\mathbf{L}^1(\pi)},$$

In the last equation, we specify $t = \sqrt{\log(n)}$. The first term on the right hand side goes to zero by (95), the second one goes to zero immediately, while the last one goes to zero by the mixing property of the process X (see Lemma 2). We deduce that

$$|\pi_n(\varphi) - \pi(\varphi)| \leq \varepsilon_n(\varphi, K, a, b), \tag{99}$$

for some sequence $\varepsilon_n(\varphi, K, a, b) \xrightarrow{n \to \infty} 0$ (let us stress that this convergence is not uniform with respect to $(a, b) \in \Sigma$, K, or the function φ).

Gathering (98) and (99), we have

$$\left|\mathbb{E}_\pi\left[\varphi(X_t)\varphi(X_0)\right] - \pi(\varphi)^2\right| \leq c_K^2 \left\|\varphi\right\|_\infty^2 \sqrt{c}e^{-ct/2} + \frac{c'e^{c't}}{n}\left\|\varphi\right\|_\infty \left\|\varphi'\right\|_\infty + \left\|\varphi\right\|_\infty \varepsilon_n(\varphi, K, a, b)$$

where the constant c does not depend on n. Letting $n \to \infty$, we deduce (7).

□

References

[1] Aeckerle-Willems, C., & Strauch, C. (2021). Concentration of scalar ergodic diffusions and some statistical implications. In Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (Vol. 57, No. 4, pp. 1857-1887). Institut Henri Poincaré.

[2] Altmeyer, R. and Reiss, M. (2020). Nonparametric estimation for linear SPDEs from local measurements. Annals of Applied Probabilty to appear.

[3] Amorino, C. (2020). Rate of estimation for the stationary distribution of jump-processes over anisotropic Holder classes. arXiv preprint arXiv:2011.11994.

[4] Amorino, C., Dion, C., Gloter, A., & Lemler, S. (2020). On the nonparametric inference of coefficients of self-exciting jump-diffusion. arXiv preprint arXiv:2011.12387.

[5] Amorino, C., & Gloter, A. (2020). Contrast function estimation for the drift parameter of ergodic jump diffusion process. Scandinavian Journal of Statistics, 47(2), 279-346.

[6] Amorino, C., & Gloter, A. (2020). Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. Stochastic Processes and their Applications, 130(10), 5888-5939.

[7] Amorino, C., & Gloter, A. (2021). Invariant density adaptive estimation for ergodic jump–diffusion processes over anisotropic classes. Journal of Statistical Planning and Inference, 213, 106-129.

[8] Amorino, C., & Nualart, E. (2021). Optimal convergence rates for the invariant density estimation of jump-diffusion processes. arXiv preprint arXiv:2101.08548.

[9] Azencott, R. (1984) Densite des diffusions en temps petit: Developpements asymptotiques. In J. Azema and M. Yor (eds), Seminaire de Probabilites XVIII, Lecture Notes in Math. 1059. Berlin: Springer-Verlag.

[10] Bailey, N.T.J. (1957) The Mathematical Theory of Epidemics, Griffin, London.

[11] Bacry, E., Delattre, S., Hoffmann, M., Muzy, J.-F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. Stochastic Processes and their Applications 123, 2475-2499.

[12] Bakry, D., Cattiaux, P., Guillin, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapounov versus Poincaré. Journal of Functional Analysis, 254, 727–759.

[13] Bakry, D., Gentil, I., & Ledoux, M. (2014). Analysis and geometry of Markov diffusion operators (Vol. 103). Cham: Springer.

[14] Banks, H.T. (1975) Modeling and Control in the Biological Sciences, Lecture Notes in Biomathematics, 6, Springer-Verlag, Berlin.

[15] Banon, G. (1978). Nonparametric identification for diffusion processes, SIAM J. Control Optim. 16, 380–395.

[16] Bergstrom, A.R. (1990) Continuous Time Econometric Modeling, Oxford University Press, Oxford.

[17] Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. (Second edition), Lecture Notes Statist., 110, New York: Springer-Verlag.

[18] Bosq, D. (1998) Minimax rates of density estimators for continuous time processes, Sankhya. Ser. A 60, 18–28.

[19] Cialenco, I. (2018). Statistical Inference for SPDEs: an Overview. Statistical Inference for Stochastic Processes, 21, 309-329.

[20] Cobb, L. (1981) Stochastic differential equations for the social sciences, in Mathematical Frontiers of the Social and Policy Sciences.

[21] Comte, F., Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. Annales de l'IHP Probabilités et statistiques, 49(2), 569-609.

[22] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli, 13, 514-543.

[23] Dalalyan, A., & Reiß, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probability theory and related fields, 137(1), 25-47.

[24] Delecroix, M. (1980). Sur l'estimation des densités d'un processus stationnaire á temps continu. Publications de l'ISUP, XXV, 1-2, 17-39.

[25] Delattre M., Genon-Catalot V. and Larédo, C. (2018). Parametric inference for discrete observations of diffusion processes with mixed effects. Stochastic processes and their Applications, 128, 1929-1957.

[26] Delattre, S., Gloter, A., Yoshida, N. (2020). Rate of Estimation for the Stationary Distribution of Stochastic Damping Hamiltonian Systems with Continuous Observations. arXiv preprint arXiv:2001.10423.

[27] Dion, C., Lemler, S. (2019). Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. Statistical Inference for Stochastic Processes, 1-27.

[28] Ditlevsen, P.D., Ditlevsen, S., Andersen, K.K. (2002) The fast climate fluctuations during the stadial and interstadial climate states, Ann. Glaciol., 35, 457–462.

[29] Evans, Lawrence C. ; Gariepy, Ronald F. Measure theory and fine properties of functions. Revised edition. Textbooks in Mathematics. CRC Press, Boca Raton, FL, 2015. xiv+299 pp. ISBN: 978-1-4822-4238-6

[30] Friedman A., (1964). Partial differential equations of parabolic type. Prentice-Hall Inc., Englewood Cliffs, N.J.

[31] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, Statistics, 21, 99-116.

[32] Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. Bernoulli, 7(6), 899-912.

[33] Gobet, E., Hoffmann, M. and Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. The Annals of Statistics, 32, 2223-2253.

[34] Hairer, M. (2009). An Introduction to Stochastic PDEs. arXiv: Probability.

[35] Has'minskii, R. Z. (1980). Stability of differential equations. Germantown, MD: Sijthoff and Noordhoff.

[36] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. Stoch. Proc. and Appl. 79, 135-163.

[37] Holden, A.V. (1976) Models for Stochastic Activity of Neurones, Springer-Verlag, New York.

[38] Holmes, E.E. (2004) Beyond theory to application and evaluation: diffusion approximations for population viability analysis, Ecol. Appl., 14(4), 1272–1293.

[39] Höpfner, R., (2014). Asymptotic Statistics with a view to stochastic processes. Walter de Gruyter, Berlin/Boston.

[40] Hull, J. (2000) Options, Futures and Other Derivatives, Prentice-Hall, Englewood Cliffs, NJ.

[41] Iacus, S. M., (2010). Simulation and inference for stochastic differential equations. With R examples. Springer.

[42] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). Statistical methods for stochastic differential equations. CRC press. Taylor & Francis Group. Boca Raton.

[43] Kutoyants, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. Statistical Inference for Stochastic Processes, 1(2), 131-155.

[44] Kutoyants, Y. A. (2004). On invariant density estimation for ergodic diffusion processes. SORT: statistics and operations research transactions, 28(2), 0111-124.

[45] Kutoyants, Y.A., (2004). Statistical inference for ergodic diffusion processes. Springer, London.

[46] Kushner, H.J. (1967) Stochastic Stability and Control, Academic Press, New York.

[47] Lamberton, D., Pages, G., (2002). Recursive computation of the invariant distribution of a diffusion. Bernoulli 8(3), pp.367-405.

[48] Lange, K. (2002) Mathematical and Statistical Methods for Genetic Analysis, Springer-Verlag, New York.

[49] Larédo, C. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. The Annals of Statistics 18, 1158-1171.

[50] Mancini, C., & Renò, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. Journal of Econometrics, 160(1), 77-92.

[51] Marie, N., & Rosier, A. (2021). Nadaraya-Watson Estimator for IID Paths of Diffusion Processes. arXiv preprint arXiv:2105.06884.

[52] Masuda, H. (2007). Ergodicity and exponential beta-mixing for multidimensional diffusions with jumps. Stoch. Proc. and Appl., 117, 35-56.

[53] Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. Stoch. Proc. Appl. 129, 1013-1059.

[54] Nguyen, H. T. (1979). Density estimation in a continuous-time Markov processes. Ann. Statist. 7, 341-348.

[55] Panloup, F. (2008). Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. The Annals of Applied Probability 18(2), 379-426.

[56] Papanicolaou, G. (1995) Diffusions in random media, in Surveys in Applied Mathematics, Keller, J.B., McLaughin, D., Papanicolaou , G., eds, Plenum Press, New York, 205–255.

[57] Piccini, U., De Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. Scand. J. Statist. 37, 67-90.

[58] Ricciardi, L.M. (1977) Diffusion Processes and Related Topics in Biology, Lecture Notes in Biomathematics, Springer, New York.

[59] Schmisser, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. Statistics, 47(1), 61-84.

[60] Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. Stoch. Proc. Appl., 124, 883-914.

[61] Schuecker, P., B¨ohringer, H., Arzner, K., Reiprich, T.H. (2001) Cosmic mass functions from Gaussian stochastic diffusion processes, Astron. Astrophys., 370, 715–728.

[62] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics, 46(6B), 3451-3480.

[63] Tsybakov, A. B. (2008). Introduction to nonparametric estimation. Springer Science & Business Media.

[64] Tuerlink, F., Maris, E., Ratcliff, R., De Boeck, P. (2001) A comparison of four methods for simulating the diffusion process, Behav. Res. Methods Instrum. Comput., 33(4), 443–456.

[65] Veretennikov, A. Y. (1988). Bounds for the mixing rate in the theory of stochastic equations. Theory of Probability & Its Applications, 32(2), 273-281.

[66] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. J. Multivariate Analysis, 41, 220-242.

[67] Van Zanten, H. (2001). Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. Nonparametric Statist. 13 (6), 833-850.