Rate of estimation for the stationary distribution of jump-processes over anisotropic Holder classes[*]

Chiara Amorino

Unité de Recherche en Mathématiques
Université du Luxembourg
e-mail: chiara.amorino@uni.lu

Abstract: We study the problem of the non-parametric estimation for the density π of the stationary distribution of the multivariate stochastic differential equation with jumps $(X_t)_{0 \leq t \leq T}$, when the dimension d is such that $d \geq 3$. From the continuous observation of the sampling path on $[0, T]$ we show that, under anisotropic Hölder smoothness constraints, kernel based estimators can achieve fast convergence rates. In particular, they are as fast as the ones found by Dalalyan and Reiss [11] for the estimation of the invariant density in the case without jumps under isotropic Hölder smoothness constraints. Moreover, they are faster than the ones found by Strauch [32] for the invariant density estimation of continuous stochastic differential equations, under anisotropic Hölder smoothness constraints. Furthermore, we obtain a minimax lower bound on the L2-risk for pointwise estimation, with the same rate up to a $\log(T)$ term. It implies that, on a class of diffusions whose invariant density belongs to the anisotropic Holder class we are considering, it is impossible to find an estimator with a rate of estimation faster than the one we propose.

MSC2020 subject classifications: Primary 62G07, 62G20; secondary 60J74.

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

Received December 2020.

1. Introduction

Diffusion processes with jumps are recently becoming powerful tools to model various stochastic phenomena in many areas such as physics, biology, medical sciences, social sciences, economics, and so on. In finance, jump-processes were introduced to model the dynamic of exchange rates ([6]), asset prices ([25], [22]), or volatility processes ([5], [15]). Utilization of jump-processes in neuroscience, instead, can be found for instance in [14]. Therefore, stochastic differential equations with jumps are nowadays widely studied by statisticians.

In this work, we aim at estimating the invariant density π associated to the process $(X_t)_{t \geq 0}$, solution of the following multivariate stochastic differential

equation with Levy-type jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds,dz), \quad (1)$$

where W is a d-dimensional Brownian motion and $\tilde{\mu}$ a compensated Poisson random measure with a possible infinite jump activity. We assume that a continuous record of observations $X^T = (X_t)_{0 \le t \le T}$ is available.

The problem of non-parametric estimation of the stationary measure of a continuous mixing process is both a long-standing problem (see for instance N'Guyen [26] and references therein) and a living topic. First of all, because invariant distributions are crucial in the study of the long-run behaviour of diffusions (we refer to Has'minskii [19] and Ethier and Kurtz [16] for background on the stability of stochastic differential systems). Then, because of the huge quantity of numerical methods connected to it (such as the Markov chain Monte Carlo methods). In Lamberton and Pages [24], for example, it has been proposed an approximation algorithm for the computation of the invariant distribution of a continuous Brownian diffusion, extended then in Panloup [29] to a diffusion with Lévy jumps. Recent works on the recursive approximation of the invariant measure can also be found in Honoré, Menozzi [20] for a continuous diffusion and in Gloter, Honoré, Loukianova [18] for a Poisson compound process.

Our goal, in particular, is to find the convergence rate of estimation for the stationary measure π associated to the process X solution to (1). After that, we will discuss the optimality of such a rate.

Considering stochastic differential equations without jumps, some results are known. In the specific context where the continuous time process is a one-dimensional diffusion process, observed continuously on some interval $[0,T]$, it has been shown that the rate of estimation of the stationary measure is $\sqrt{T}$ (see Kutoyants [23]). If the process is a diffusion observed discretely on $[0,T]$, with a sufficiently high frequency, then it is still possible to estimate the stationary measure with rate $\sqrt{T}$ (see [28] [10]). In [30] Schmisser estimates the successive derivatives $\pi^{(j)}$ of the stationary density associated to a strictly stationary and β mixing process $(X_t)_{t \ge 0}$ observed discretely. When $j = 0$, the convergence rate is the same found by Comte and Merlevède in [9] and [10].

Regarding the literature on statistical properties of multidimensional diffusion processes in absence of jumps, an important reference is given by Dalalyan and Reiss in [11] where, as a by-product of the study, they prove some convergence rates for the pointwise estimation of the invariant density, under isotropic Hölder smoothness constraints. In a recent paper [32], Strauch has extended their work by building adaptive estimators in the multidimensional diffusion case which achieve fast rates of convergence over anisotropic Hölder balls. 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 presence of jumps, we are only aware of a few works which take place in the non parametric framework. In [13], for example, the authors estimate in a non-parametric way the drift function of a diffusion with jumps driven by a

Hawkes process while in [3] the estimation of the integrated volatility is considered. Schmisser investigates, in [31], the non parametric adaptive estimation of the coefficients of a jumps diffusion process and together with Funke she also investigates, in [17], the non parametric adaptive estimation of the drift of an integrated jump diffusion process.

Closer to the purpose of this work, in [4] and [2] the convergence rate for the pointwise estimation of the invariant density associated to (1) is considered. The work [4] is devoted to the low-dimensional case, which is for $d = 1$ and $d = 2$. In [2], for $d \geq 3$, it is proved that the mean squared error can be upper bounded by $T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}}$; where $\bar{\beta}$ is the harmonic mean smoothness of the invariant density over the d different dimensions. We remark that the rate here above reported is the same found by Strauch in [32] in the continuous case, which is also the rate proved by Dalalyan and Reiss, up to replacing the mean smoothness $\bar{\beta}$ with β, the common smoothness over the d direction.

In this paper, we want to estimate the invariant density π by means of a kernel estimator, we therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$. A natural estimator of π at some point $x \in \mathbb{R}^d$ in the anisotropic context is given by

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

where $h = (h_1, ..., h_d)$ is a multi - index bandwidth. First of all we extend the previous results by proving the following upper bound for the mean squared error:

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

where $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2} \sum_{l \geq 3} \frac{1}{\beta_l}$. As by construction $\bar{\beta}_3$ is bigger that $\bar{\beta}$, the convergence rate here above is in general faster than the one proposed in [2].

After that, we want to understand if it is possible to improve the convergence rate by using other density estimators and which is the best possible rate of convergence. To answer, the idea is to look for lower bounds for the minimax risk associated to the anisotropic Holder class. For the computation of lower bounds, we introduce a jump-process simpler than (1):

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a \, dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma \, z \, \tilde{\mu}(ds, dz), \qquad (3)$$

where a and γ are constants. We moreover assume the intensity of the jumps to be finite. We anticipate here the definition of the minimax risk that will be given in (14):

$$\mathcal{R}_T(\beta, \mathcal{L}) := \inf_{\tilde{\pi}_T} \sup_{b \in \Sigma(\beta, \mathcal{L})} \mathbb{E}_b^{(T)}[(\tilde{\pi}_T(x_0) - \pi_b(x_0))^2],$$

where the infimum is taken over all possible estimators of the invariant density and $\Sigma(\beta, \mathcal{L})$ gathers the drifts for which the considered process is stationary

and whose stationary measure has the prescribed Holder regularity. In order to prove a lower bound for the minimax risk, the knowledge of the link between b and π_b is crucial. In absence of jumps, considering reversible diffusion processes with unit diffusion part (as in both [11] and [32]), such a connection was explicit:

$$b(x) = -\nabla V(x) = \frac{1}{2}\nabla(\log \pi_b)(x),$$

where $V \in C^2(\mathbb{R}^d)$ is referred to as potential. Adding the jumps, it is no longer true. In our framework, it is challenging to get a relation between b and π_b. The idea is to write the drift b in function of π_b knowing that they must satisfy $A_b^*\pi_b = 0$, where A^* is the adjoint operator of A, the generator of the diffusion X solution to (3) (see Proposition 2 below, a similar argument can also be found in [12]).

We are in this way able to prove the following main result:

$$\inf_{\tilde{\pi}_T} \sup_{b \in \Sigma(\beta, \mathcal{L})} \mathbb{E}_b^{(T)}[(\tilde{\pi}_T(x_0) - \pi_b(x_0))^2] \underset{\sim}{>} T^{-\frac{2\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}},$$

where we recall it is $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2}\sum_{l \geq 3}\frac{1}{\beta_l}$. It follows that, on a class of diffusions X whose invariant density belongs to the anisotropic Holder class we are considering, it is impossible to find an estimator with a rate of estimation better than $T^{-\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}$, for the pointwise L^2 risk. Comparing the lower bound here above with the upper bound in (2) we observe that, up to a logarithmic term, the two convergence rates we found are the same.

Furthermore, we present some numerical results in dimension 3. We show that the variance depends only on the biggest bandwidth. The simulations match with the theory and illustrate we can remove the two smallest bandwidths, which are associates to the smallest smoothness. It implies we get a convergence rate which does not depend on the two smallest smoothness β_1 and β_2.

The outline of the paper is the following. In Section 2 we introduce the model and we give the assumptions, while in Section 3 we propose the kernel estimator for the estimation of the invariant density and we state the upper bound for the mean squared error. In Section 4 we complement them with lower bounds for the minimax risk while in Sections 5 and 6 we provide, respectively, the proofs of the upper and lower bounds. Some technical result are moreover proved in Section 7.

2. Model

We consider the question of nonparametric estimation of the invariant density of a d-dimensional diffusion process X, assuming that a continuous record of the process up to time T is available. The diffusion is given as a strong solution of

the following stochastic differential equations with jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds, dz), \quad t \in [0, T],$$
(4)

where the coefficients are such that $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$. The process $W = (W_t, t \geq 0)$ is a d-dimensional Brownian motion and μ is a Poisson random measure on $[0, T] \times \mathbb{R}^d$ associated to the Lévy process $L = (L_t)_{t \in [0,T]}$, with $L_t := \int_0^t \int_{\mathbb{R}^d} z\tilde{\mu}(ds, dz)$. The compensated measure is $\tilde{\mu} = \mu - \bar{\mu}$. We suppose that the compensator has the following form: $\bar{\mu}(dt, dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.

The initial condition X_0, W and L are independent. In the sequel, we will denote $\tilde{a} := a \cdot a^T$.

2.1. Assumptions

We want first of all to show an upper bound on the mean squared error, as we will see in detail in Section 5. To do that, we need the following assumptions to hold:

A1: *The functions $b(x)$, $\gamma(x)$ and $\tilde{a}(x)$ are globally Lipschitz and, for some $c \geq 1$,*

$$c^{-1}\mathbb{I}_{d \times d} \leq \tilde{a}(x) \leq c\mathbb{I}_{d \times d},$$

where $\mathbb{I}_{d \times d}$ denotes the $d \times d$ identity matrix.
Denoting with $|.|$ and $< ., . >$ respectively the Euclidean norm and the scalar product in $\mathbb{R}^d$, we suppose moreover that there exists a constant $c > 0$ such that, $\forall x \in \mathbb{R}^d$, $|b(x)| \leq c$.

A1 ensures that equation (4) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [1] (Theorems 6.2.9. and 6.4.6.).

A2 (Drift condition) :
There exist $C_1 > 0$ and $\tilde{\rho} > 0$ such that $< x, b(x) > \leq -C_1|x|$, $\forall x : |x| \geq \tilde{\rho}$.

In the following A3 are gathered the assumptions on the jumps:

A3 (Jumps) : *1. The Lévy measure F is absolutely continuous with respect to the Lebesgue measure and we denote $F(z) = \frac{F(dz)}{dz}$.*
2. We suppose that there exist $c > 0$ such that for all $z \in \mathbb{R}^d \setminus \{0\}$, $F(z) \leq \frac{c}{|z|^{d+\alpha}}$, with $\alpha \in (0, 2)$ and that $supp(F) = \mathbb{R}^d \setminus \{0\}$.

3. The jump coefficient γ is upper bounded, i.e. $\sup_{x \in \mathbb{R}^d} |\gamma(x)| := \gamma_{\max} < \infty$. We suppose moreover that there exists a constant c_1 such that, $\forall x \in \mathbb{R}^d$, $Det(\gamma(x)) > c_1$.

4. If $\alpha = 1$, we require for any $0 < r < R < \infty$ $\int_{r < |z| < R} z F(z) dz = 0$.

5. There exists $\epsilon_0 > 0$ and a constant $\hat{c} > 0$ such that $\int_{\mathbb{R}^d \backslash \{0\}} |z|^2 e^{\epsilon_0 |z|} F(z) dz \le \hat{c}$.

As showed in Lemma 2 of [2] A2 ensures, together with the last point of A3, the existence of a Lyapunov function, while the second and the third points of A3 involve the irreducibility of the process. The process X admits therefore a unique invariant distribution μ and the ergodic theorem holds. We assume the invariant probability measure μ of X being absolutely continuous with respect to the Lebesgue measure and from now on we will denote its density as π: $d\mu = \pi dx$.

Our goal is to propose an estimator for the invariant density estimation and to study its convergence rate. We start our analysis by introducing the natural estimator in this context and by analysing upper bounds for the mean squared error. Then, we investigate the existence of a lower bound for the minimax risk.

3. Estimator and upper bound

In this section we introduce the expression for our estimator of the stationary measure π of the stochastic equation with jumps (4) in an anisotropic context. After that, we present the rate of convergence the estimator achieves, depending on the smoothness of π.

The notion of anisotropy plays an important role. Indeed, the smoothness properties of elements of a function space may depend on the chosen direction of $\mathbb{R}^d$. The Russian school considered anisotropic spaces from the beginning of the theory of function spaces in 1950-1960s (in [27] the author takes account of the developments). However, results on minimax rates of convergence in classical statistical models over anisotropic classes were rare for a lot of time. We work under the following anisotropic smoothness constraints.

Definition 1. *Let $\beta = (\beta_1, ..., \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$. 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 $i \in \{1, ..., d\}$,*

$$\left\| D_i^k g \right\|_\infty \le \mathcal{L}_i \qquad \forall k = 0, 1, ..., \lfloor \beta_i \rfloor,$$
$$\left\| D_i^{\lfloor \beta_i \rfloor} g(. + t e_i) - D_i^{\lfloor \beta_i \rfloor} g(.) \right\|_\infty \le \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \qquad \forall t \in \mathbb{R},$$

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 $\mathbb{R}^d$.

We deal with the estimation of the density π belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$. Given the observation X^T of a diffusion X, solution of (4), we propose to estimate the invariant density π by means of a kernel estimator. We introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying

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

for all $l \in \{1, ..., M\}$ with $M \geq \max_i \beta_i$.

Denoting by X_t^j, $j \in \{1, ..., d\}$ the j-th component of X_t, $t \geq 0$, a natural estimator of $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ at $x = (x_1, ..., x_d)^T \in \mathbb{R}^d$ in the anisotropic context is given by

$$\hat{\pi}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K\left(\frac{x_m - X_u^m}{h_m}\right) du =: \frac{1}{T} \int_0^T \mathbb{K}_h(x - X_u) du, \quad (5)$$

where $h = (h_1, ..., h_d)$ is a multi-index bandwidth and it is small. In particular, we assume $h_i < \frac{1}{2}$ for any $i \in \{1, ..., d\}$.

The asymptotic behaviour of the estimator relies on the standard bias variance decomposition. Hence, we need an evaluation for the variance of the estimator, as in next proposition. We prove it in Section 5.

One can remark that in [32], where a continuous reversible diffusion process with unit diffusion is considered, the author formulates implications on the functional inequalities (of Poincaré and Nash-type) to get an upper bound for the variance of the estimator. The main advantage in using functional inequalities is that they allow the constants involved in the upper bound of the variance to be controlled uniformly. However, this approach is restricted only to symmetric diffusion framework and so it can not be applied in our setting. To overcome this difficulty we derive some upper bounds on the variance of our estimator by exploiting the mixing properties of X. In particular, the proof of the proposition below relies on a bound on the transition density (see Lemma 1 in [2]) and on the exponential ergodicity and the exponential β-mixing property of the process X (as established in Lemma 2 of [2]). However, this approach has some disadvantages. One above all the fact that, as the upper bounds relies on mixing properties, the constants depend on the coefficients. Hence, it is very challenging to understand how the constants involved can be controlled uniformly and this is still an open question.

Proposition 1. *Suppose that A1 - A3 hold. If π is bounded and $\hat{\pi}_{h,T}$ is the estimator given in* (5), *then there exists a constant c independent of T such that*

• *If $h_1 h_2 < (\prod_{l \geq 3} h_l)^{\frac{2}{d-2}}$, then*

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \frac{\sum_{j=1}^d |\log(h_j)|}{\prod_{l \geq 3} h_l}. \quad (6)$$

- *If otherwise $h_1 h_2 \geq (\prod_{l \geq 3} h_l)^{\frac{2}{d-2}}$, then*

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \frac{1}{\prod_{l \geq 3} h_l}. \tag{7}$$

We underline that, in the upper bound of the variance here above, it would have been possible to remove, in the denominator, the contribution of no matter which two bandwidths. We arbitrarily choose to remove the contribution of h_1 and h_2 as in the bias term they are associated to β_1 and β_2, which are the smallest values of smoothness (see Theorem 1 below) and so they provide the strongest constraints.

One may wonder about the origin of the logarithmic term in the upper bound (6). We will see in the proof of Proposition 1 that it is possible to estimate the absolute value of the covariance

$$k(s) := Cov(\mathbb{K}_h(x - X_0), \mathbb{K}_h(x - X_s))$$

with $\frac{1}{\prod_{j \geq l+1} h_j} \frac{1}{s^{\frac{l}{2}}}$ for any $l \in \{0, ..., d\}$. Then, we will need to integrate such term over the time s. When s is small the best choice consists in taking $l = 0$, while for s far away from the neighbourhood of 0 is convenient to take $l = d$. As we will see in the proof of Proposition 1, it is possible to make the bound on the variance smaller by considering also the case that stands in between, for which s is not zero but can be arbitrarily small. Here the best choice is to take $l = 2$, which provides the logarithm as in (6).

To better understand how to choose the bandwidths whose contributions we will remove, let us see more in detail what happens for $d = 3$. In this case, the two strongest constraints are connected to the two smallest bandwidths and so we arbitrarily decide to remove their contributions. It follows that the upper bound on the variance will depend only the largest bandwidth between h_1, h_2 and h_3, up to a logarithmic term. In particular, for $d = 3$, equation (6) in Proposition 1 becomes

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c \sup_{j=1,2,3} |\log(h_j)|}{T} \inf(\frac{1}{h_1}, \frac{1}{h_2}, \frac{1}{h_3}), \qquad \text{for } h_1 h_2 < h_3^2. \tag{8}$$

On the other side, when $h_1 h_2 \geq h_3^2$, we have

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T} \inf(\frac{1}{h_1}, \frac{1}{h_2}, \frac{1}{h_3}). \tag{9}$$

As this final result is quite surprising, we decide to support it by presenting some simulations. Our goal is to illustrate that the variance will depend only on the largest bandwidth. We consider the process X solution of

$$X_t = X_0 + \int_0^t b(X_s)ds + W_t + \int_0^t \int_{\mathbb{R}^3 \setminus \{0\}} z \, \tilde{\mu}(ds, dz),$$
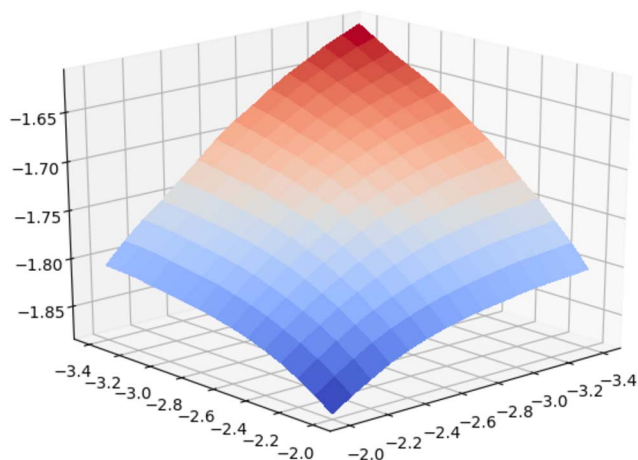
FIG 1. *3d graphic for* $\log_{10}(Var(\hat{\pi}_{h,T}(x)))$ *when* h_3 *is big.*

with $b(x) = -4\frac{x}{|x|}e^{-\frac{1}{4|x|-1}}1_{|x|>\frac{1}{4}}$. The Brownian motion has variance I_3 and the jump process is a compound Poisson, with intensity 1 and Gaussian jump law $N(0, I_3)$. We evaluate the variance of the kernel estimator for different values of the bandwidth h_1, h_2 and h_3 over the interval $[0, T]$, where we choose $T = 100$. The process is simulated by an Euler scheme with discretization step $\Delta_n = 10^{-7}$ and the integral in the definition of the kernel estimator is replaced by a Riemann sum whose discretization step is once again 10^{-7}. We use a Monte Carlo method based on 2000 replications and we provide a 3d graphic, in which on the x and y-axis there are respectively the values of $\log_{10}(h_1)$ and $\log_{10}(h_2)$ while on the z-axis there is the value of $\log_{10}(Var(\hat{\pi}_{h,T}(x)))$. The idea is to fix h_3 bigger than h_1 and h_2 and to see how the variance of our estimator changes in function of h_1 and h_2, in a logarithmic scale.

In particular, we ta take $h_3 = 10^{-0.5}$ and h_1 and h_2 that belong to $[10^{-2}, 10^{-3.4}]$. Therefore, h_3 is much larger than the other bandwidths and so the variance of the estimator should be, according with our results, dependent only on h_3. In particular, as it is $h_1 h_2 < h_3^2$, from (8) we obtain the theoretical variance is upper bounded by $\frac{c}{T}\frac{|\log(h_3)|}{h_3}$. Even if the 3d graphic reported in Figure 1 does not seem to represent a completely constant function, one can easily see looking at the z-axis that the variance is remotely dependent on h_1 and h_2. The minimal variance, indeed, is achieved for $h_1 = h_2 = 10^{-2}$ and its value is $10^{-1.88}$ while its maximal value is $10^{-1.61}$ and it is achieved for $h_1 = h_2 = 10^{-3.4}$. It means that the variance varies a little: for the kernel bandwidths which move from $(h_1, h_2, h_3) = (10^{-2}, 10^{-2}, 10^{-0.5})$ to $(h_1, h_2, h_3) = (10^{-3.4}, 10^{-3.4}, 10^{-0.5})$, the volume of the Kernel support is divided by $10^{1.4} \times 10^{1.4} \simeq 631$ while the variance of the estimator is just multiplied by $10^{0.27} \simeq 1.86$.

Another evidence of the dependence of the variance of the estimator only on h_3 is given by Figure 2 below. To better understand the graphic below,
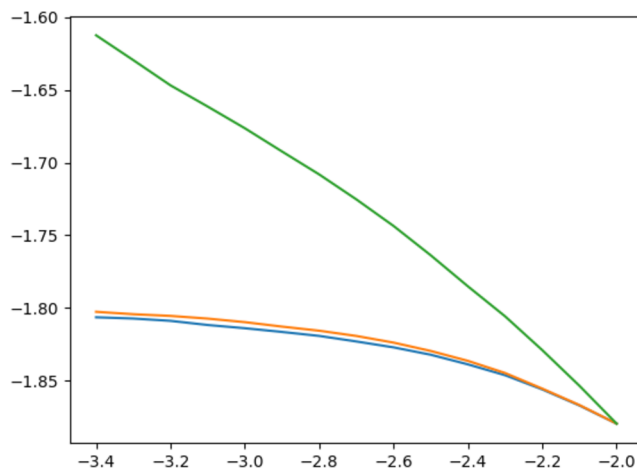
FIG 2. *Slope representation for the edge curves and diagonal curve when h_3 is big.*

we underline that the orange and blue curves correspond to the two edges, respectively. In particular, the orange curve corresponds to the variation of the variance for $h_1 = 10^{-0,5}$ fixed and h_2 which shifts from $10^{-0.5}$ to $10^{-2.4}$, while the blue curve represents the variation of the variance when h_2 is fixed equal to $10^{-0,5}$ and h_1 goes from $10^{-0.5}$ to $10^{-2.4}$.

The green curves corresponds to the diagonal of the 3d graphic and so it represents the variance of the estimator when $h_1 = h_2$ moves from $10^{-0.5}$ to $10^{-2.4}$. We start discussing the behaviour of the green curve. According with the theory we know the variance should not be dependent on $h_1 = h_2$. Therefore, the derivative of the log-variance function with respect to $\log_{10}(h_1) = \log_{10}(h_2)$ should be null. The numerical results match with the theoretical ones, as the slope of the diagonal is quite weak, being equal to -0.186.

Regarding the edge curves, one can easily remark that the results provided by Figure 2 match with the theoretical results. The two edge curves are indeed totally flat: their slopes are -0.048 and -0.051.

Based on the upper bounded on the variance found in Proposition 1 discussed above, we can now state the main result on the asymptotic behaviour of the estimator. Its proof can be found in Section 5.

Theorem 1. *Suppose that A1 - A3 hold. If $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in* (5) *satisfies, for $d \geq 3$, the following risk estimates:*

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \lesssim \sum_{l=1}^{d} h_l^{2\beta_l} + \frac{1}{T} \frac{\sum_{j=1}^{d} |\log(h_j)|}{\prod_{l \geq 3} h_l}. \tag{10}$$

Taken $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and defined $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2} \sum_{l \geq 3} \frac{1}{\beta_l}$, the rate optimal choice

for the bandwidth h provided in ([32](#)) *and* ([33](#)) *below yields the convergence rate*

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

Moreover, in the isotropic context $\beta_1 = \beta_2 = ... = \beta_d =: \beta$, *the following convergence rate holds true:*

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \underset{\sim}{<} (\frac{1}{T})^{\frac{2\beta}{2\beta+d-2}}. \tag{11}$$

We recall that in [2], under the same assumptions, the following convergence rate has been found for the pointwise estimation of the invariant density for $d \geq 3$:

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \underset{\sim}{<} T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}}, \tag{12}$$

where $\bar{\beta}$ is the harmonic mean smoothness of the invariant density over the d different dimensions, such that

$$\frac{1}{\bar{\beta}} := \frac{1}{d} \sum_{l=1}^{d} \frac{1}{\beta_l}.$$

We remark that the rate in ([12](#)) for $d \geq 3$ is the same Strauch found in [32] in absence of jumps, which is also the rate gathered in the isotropic context proposed in [11], up to replacing the mean smoothness $\bar{\beta}$ with β, the common smoothness over the d dimensions, as we did in ([11](#)).
By construction, $\bar{\beta}_3$ is bigger than $\bar{\beta}$ and, therefore, the upper bound found in Theorem [1](#) is faster than the one reported in ([12](#)) in a general anisotropic context.
Now the following two questions arise. Can we improve the rate by using other density estimators? What is the best possible rate of convergence? To answer these questions it is useful to consider the minimax risk $\mathcal{R}_T(\beta, \mathcal{L})$ associated to the anisotropic Holder class $\mathcal{H}_d(\beta, \mathcal{L})$ we defined in Definition [1](#), as we are going to explain in Section [4](#).

4. Lower bounds

In this section, we wonder if it is possible to construct any estimator with a rate better than the one obtained in Theorem [1](#).
For the computation of lower bounds, we introduce the following stochastic differential equation with jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a \, dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma \, z \, \tilde{\mu}(ds, dz), \tag{13}$$

where a and γ are constants, γ is also invertible and b is a Lipschitz and bounded function. We assume that the jump measure satisfies the conditions gathered in

points 1,2,4 and 5 of A3. We moreover suppose that there exists λ_1 such that

$$\int_{\mathbb{R}^d} F(z)dz = \lambda_1 < \infty$$

and, $\forall i \in \{1, ..., d\}$,

$$|\sum_{j \neq i}(aa^T)_{ij}(aa^T)_{jj}^{-1}| \leq \frac{1}{2}.$$

We underline that if the matrix $a \cdot a^T$ is diagonal, then the request here above is always satisfied. If it is not the case, such an assumption implies that the diagonal terms dominate on the others.

As the model satisfies A1, we know that the stochastic differential equation with jumps (13) admits a solution. Moreover, as γ is inversible, A3 is automatically true. If A2 also holds we know from Lemma 2 in [2] that the process admits a unique stationary measure, that we note π_b. We omit in the notations the dependence on a and γ as they will be fixed in the sequel, while the connection between b and π_b will be made explicit in Section 6.1.

If the invariant measure exists, we denote as $\mathbb{P}_b$ the law of a stationary solution $(X_t)_{t \geq 0}$ of (13) and we note be $\mathbb{E}_b$ the corresponding expectation. Moreover we will note by $\mathbb{P}_b^{(T)}$ the law of $(X_t)_{t \in [0,T]}$, solution of (13).

In order to write down an expression for the minimax risk of estimation, we have to consider a set of solutions to the equation (13) which are stationary and whose stationary measure has the prescribed Holder regularity introduced in Definition 1. It leads us to the following definition.

Definition 2. *Let $\beta = (\beta_1, ..., \beta_d)$, $\beta_i > 1$ and $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$. We define $\Sigma(\beta, \mathcal{L})$ the set of the Lipschitz and bounded functions $b : \mathbb{R}^d \to \mathbb{R}^d$ satisfying A2 and for which the density π_b of the invariant measure associated to the stochastic differential equation (13) belongs to $\mathcal{H}_d(\beta, 2\mathcal{L})$.*

We introduce the minimax risk for the estimation at some point. Let $x_0 \in \mathbb{R}^d$ and $\Sigma(\beta, \mathcal{L})$ as in Definition 2 here above. We define the minimax risk

$$\mathcal{R}_T(\beta, \mathcal{L}) := \inf_{\tilde{\pi}_T} \sup_{b \in \Sigma(\beta, \mathcal{L})} \mathbb{E}_b^{(T)}[(\tilde{\pi}_T(x_0) - \pi_b(x_0))^2], \tag{14}$$

where the infimum is taken on all possible estimators of the invariant density. Our main result is a lower bound for the minimax risk here above defined. The proof is based on the two hypotheses method, explained for example in Section 2.3 of [33].

Theorem 2. *There exists $c > 0$ such that, if $\hat{c} < c$ (recall: $\hat{c}$ is defined in the fifth point of A3), then*

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

where we recall it is $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2}\sum_{l \geq 3}\frac{1}{\beta_l}$.

The condition on $\hat{c}$ follows from the fact that, in our approach, the jumps have to be not too big. In this way it is possible to build ergodic processes where, in the analysis of the link between the invariant measure and the drift function, the continuous part of the generator dominates (see Lemma 2).

Regarding the choice of the model, it is worth noticing that our framework does not allow to consider continuous processes as well as jump diffusions simultaneously, as we need the coefficients to be always different from zero to get the mixing properties of our process. Hence, we choose to take into account the case where we have an additional information: we do have the jumps. In particular, we are looking for a lower bound on a class of processes where we know that the jumps really occurred, which is truly challenging. It is interesting to remark that it is possible to follow the schema provided in Section 6 also when one aims at finding a lower bound on a class of continuous diffusion processes. The main difference would be the absence of the discrete part of the generator A_d, which would implies the absence of its adjoint $A_{d,i}^*$ in the definition of the coordinates of b (see Equation (38)). As we will see, in the construction of the priors, the idea will be to provide a first density with the prescribed regularity and then to give the second as the first plus a bump. As we will need to consider the drifts associated to the built priors, we will need to evaluate the adjoint of the generator of the process in the bump. The main difficulty comes from the discrete part of the generator, being a non-local operator (see Points 1 and 2 of Proposition 4: without the jumps the difference between the drifts would be here just zero).

It follows from Theorem 2 that, on a class of diffusions X whose invariant density belongs to $\mathcal{H}_d(\beta, \mathcal{L})$ and starting from the observation of the process $(X_t)_{t \in [0,T]}$, it is impossible to find an estimator with a rate of estimation better than $T^{-\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}$, for the pointwise L^2 risk. Comparing the lower bound here above with the upper bound gathered in Theorem 1 we observe that, up to a logarithmic term, the two convergence rates we found are the same. Hence, the convergence rate we found by means of a kernel estimator is the best possible, but only up to a logarithmic term.

5. Proof upper bound

This section is devoted to the proof of the upper bound gathered in Theorem 1. To do that, we need first of all to prove Proposition 1. Before proving it we recall a result from [2] that will be useful in the sequel.

From Lemma 1 in [2], which heavily relies on the first point of Theorem 1.1 in [7], we know that the following upper bound on the transition density holds true for $t \in [0, 1]$:

$$p_t(y, y') \leq c_0 \left(t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-y'|^2}{t}} + \frac{t}{(t^{\frac{1}{2}} + |y - y'|)^{d+\alpha}} \right) =: p_t^G(y, y') + p_t^J(y, y').$$

Such a bound is not uniform in t big. Nevertheless, for $t \geq 1$, we have

$$p_t(y, y') = \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, \zeta) p_{\frac{1}{2}}(\zeta, y') d\zeta \leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, \zeta)(e^{-\lambda_0 (y'-\zeta)^2 \frac{1}{2}}$$

$$+ \frac{1}{(\sqrt{\frac{1}{2}} + |y' - \zeta|)^{d+\alpha}}) d\zeta \leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(y, \zeta) d\zeta \leq c.$$

We deduce, for all t,

$$p_t(y, y') \leq p_t^G(y, y') + p_t^J(y, y') + c. \tag{15}$$

Proof. Proposition 1
In the sequel, the constant c may change from line to line and it is independent of T.
From the definition (5) 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

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

In order to find an upper bound for the integral in the right hand side here above we will 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 which is as sharp as possible.
• For $s \in [0, \delta_1)$, from Cauchy-Schwartz inequality and the stationarity of the process we get

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

The variance is smaller than

$$\int_{\mathbb{R}^d} (\mathbb{K}_h(x - y))^2 \pi(y) dy.$$

Using the boundedness of π and the definition of $\mathbb{K}_h$ given in (5) it follows

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

which implies

$$\int_0^{\delta_1} |k(s)| \leq \frac{c\, \delta_1}{\prod_{l=1}^d h_l}. \tag{16}$$

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

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

From (15) it follows

$$|k(s)| \le k_1(s) + k_2(s) + c, \qquad (17)$$

with

$$k_1(s) := \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| p_s^G(y,y') dy' \pi(y) dy, \qquad (18)$$

$$k_2(s) := \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| p_s^J(y,y') dy' \pi(y) dy.$$

We now study $k_1(s)$. To this end we observe that, for $y' = (y_1', ...y_d')$, it is

$$p_s^G(y,y') \le \frac{c}{s} q_s^G(y_3'...y_d'|y_1', y_2', y),$$

where

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

Let us stress that

$$\sup_{s \in (0,1)} \sup_{y_1', y_2', y \in \mathbb{R}^{d+2}} \int_{\mathbb{R}^{d-2}} q_s^G(y_3'...y_d'|y_1', y_2', y) dy_3'...dy_d' \le c < \infty. \qquad (19)$$

Then, from the definition of $k_1(s)$ given in (18), we get

$$k_1(s) \le \frac{c}{s} \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| (\int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| q_s^G(y_3'...y_d'|y_1', y_2', y) dy') \pi(y) dy. \qquad (20)$$

Using the definition of $\mathbb{K}_h$ and (19) we obtain

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

We remark that in the reasoning 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. Replacing the result here above in (20) and as $\int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)|\pi(y)dy < c$, it implies

$$\int_{\delta_1}^{\delta_2} k_1(s)ds \leq \int_{\delta_1}^{\delta_2} \frac{c}{\prod_{j\geq 3} h_j} \frac{1}{s} ds = c\frac{\log(\delta_2) - \log(\delta_1)}{\prod_{j\geq 3} h_j}. \qquad (21)$$

We want to act in the same way on $k_2(s)$. We observe it is

$$
\begin{aligned}
p_t^J(y,y') &= c_0 t \frac{1}{(t^{\frac{1}{2}} + |y - y'|)^{1+\frac{\alpha}{d}}} \times ... \times \frac{1}{(t^{\frac{1}{2}} + |y - y'|)^{1+\frac{\alpha}{d}}} \\
&\leq c_0 t \frac{1}{(t^{\frac{1}{2}} + |y_1 - y_1'|)^{1+\frac{\alpha}{d}}} \times ... \times \frac{1}{(t^{\frac{1}{2}} + |y_d - y_d'|)^{1+\frac{\alpha}{d}}} \\
&\leq c_0 t\, t^{-\frac{1}{2}(1+\frac{\alpha}{d})}\, t^{-\frac{1}{2}(1+\frac{\alpha}{d})} \frac{1}{(t^{\frac{1}{2}} + |y_3 - y_3'|)^{1+\frac{\alpha}{d}}} \times ... \times \frac{1}{(t^{\frac{1}{2}} + |y_d - y_d'|)^{1+\frac{\alpha}{d}}} \\
&=: c_0 t^{-\frac{\alpha}{d}} q_t^J(y_3'...y_d'|y_1', y_2', y).
\end{aligned}
$$

We remark that

$$\sup_{y_1', y_2', y \in \mathbb{R}^{d+2}} \int_{\mathbb{R}^{d-2}} q_s^J(y_3'...y_d'|y_1', y_2', y)dy_3'...dy_d' \leq s^{-\frac{\alpha}{2d}(d-2)}, \qquad (22)$$

as each of the $d-2$ multiplication factors can be seen as $\frac{1}{s^{(1+\frac{\alpha}{d})\frac{1}{2}}} \frac{1}{(1+\frac{|y_j - y_j'|}{\sqrt{s}})^{1+\frac{\alpha}{d}}}$

and we applied the change of variable $\frac{y_j - y_j'}{\sqrt{s}} =: z$.

From the definition of $k_2(s)$ we have

$$k_2(s) \leq c_0 s^{-\frac{\alpha}{d}} \int_{\mathbb{R}^d} |\mathbb{K}_h(x-y)| \Big(\int_{\mathbb{R}^d} |\mathbb{K}_h(x-y')| q_s^J(y_3'...y_d'|y_1', y_2', y)dy'\Big)\pi(y)dy.$$

Again, acting as on $k_1(s)$, we use the definition of the kernel function $\mathbb{K}_h$ and the integrability of q_s^J gathered in (22) to obtain

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

Hence, as $1 - \frac{\alpha}{2} > 0$,

$$\int_{\delta_1}^{\delta_2} k_2(s)ds \leq c\int_{\delta_1}^{\delta_2} s^{-\frac{\alpha}{d}} \frac{cs^{-\frac{\alpha}{2d}(d-2)}}{\prod_{j\geq 3} h_j} ds = \frac{c}{\prod_{j\geq 3} h_j} \int_{\delta_1}^{\delta_2} s^{-\frac{\alpha}{2}} ds = \frac{c\delta_2^{1-\frac{\alpha}{2}}}{\prod_{j\geq 3} h_j}.$$
$$(23)$$

From (17), (21) and (23) it follows

$$\int_{\delta_1}^{\delta_2} |k(s)| ds \leq \frac{c}{\prod_{j \geq 3} h_j}(|\log(\delta_1)| + |\log(\delta_2)|) + \frac{c\delta_2^{1-\frac{\alpha}{2}}}{\prod_{j \geq 3} h_j} + c\delta_2$$
$$\leq \frac{c}{\prod_{j \geq 3} h_j}(|\log(\delta_1)| + |\log(\delta_2)|) + c\delta_2, \tag{24}$$

as $1 - \frac{\alpha}{2} > 0$ and so the term coming from k_2 is negligible compared to $\frac{|\log(\delta_2)|}{\prod_{j \geq 3} h_j}$, for δ_2 small enough.
• For $s \in [\delta_2, D)$ we still use (15) observing that, in particular,

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

We therefore get

$$\int_{\delta_2}^{D} |k(s)| ds \leq c \int_{\delta_2}^{D} (s^{-\frac{d}{2}} + s^{1-\frac{d+\alpha}{2}} + 1)ds$$
$$\leq c(\delta_2^{1-\frac{d}{2}} + \delta_2^{2-\frac{d+\alpha}{2}}1_{\{d>4-\alpha\}} + D^{2-\frac{d+\alpha}{2}}1_{\{d<4-\alpha\}}$$
$$+ (|\log D| + |\log \delta_2|)1_{\{d=4-\alpha\}} + D)$$
$$\leq c(\delta_2^{1-\frac{d}{2}} + D^{2-\frac{d+\alpha}{2}}1_{\{d<4-\alpha\}} + D) \tag{25}$$

where we have used that, as $d \geq 3$, $1 - \frac{d}{2} < 0$. The exponent of the second term in the integral here above, after having integrated, is $2 - \frac{d+\alpha}{2}$. It is more than zero if $d < 4 - \alpha$, which is possible only if $\alpha \in (0,1)$ and $d = 3$, less then zero otherwise. Moreover, $\delta_2^{2-\frac{d+\alpha}{2}}1_{\{d\geq4-\alpha\}}$ is negligible compared to $\delta_2^{1-\frac{d}{2}}$ as $\alpha < 2$ and so $2 - \frac{d+\alpha}{2} > 1 - \frac{d}{2}$. Moreover, the logarithmic terms are negligible compared to the others, for δ_2 small enough and D large enough.
• For $s \in [D, T]$ our main tool is Lemma 2 in [2]. As the process X is exponentially β- mixing, indeed, the following control on the covariance holds true:

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

for ρ and c positive constants as given in Definition 1 of exponential ergodicity in [2]. It entails

$$\int_{D}^{T} |k(s)| ds \leq c(\frac{1}{\prod_{j=1}^{d} h_j})^2 e^{-\rho D}. \tag{26}$$

Collecting together (16), (24), (25) and (26) we deduce

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}(\frac{\delta_1}{\prod_{l=1}^{d} h_l} + \frac{1}{\prod_{j \geq 3} h_j}(|\log(\delta_1)| + |\log(\delta_2)|) + \delta_2 + \delta_2^{1-\frac{d}{2}}$$
$$\tag{27}$$

$$+ D + \Big(\frac{1}{\prod_{j=1}^{d} h_j}\Big)^2 e^{-\rho D}\Big),$$

where we have also used that $D^{2-\frac{d+\alpha}{2}} 1_{\{d<4-\alpha\}} \leq D$. Indeed, since $d \geq 3$ and $\alpha \in (0,2)$, we always have $2 - \frac{d+\alpha}{2} \leq 1$. Moreover we know that $D \geq 1$ by definition. When $d < 4 - \alpha$, the power is positive, thus $D^{2-\frac{d+\alpha}{2}} 1_{\{d<4-\alpha\}} \leq D$. 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, if $h_1 h_2 < (\prod_{j\geq 3} h_j)^{\frac{2}{d-2}}$ 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 (27) we obtain

$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T}\Big(\frac{1}{\prod_{j\geq 3} h_j} + \frac{\sum_{j=1}^{d} |\log(h_j)|}{\prod_{j\geq 3} h_j} + \Big(\prod_{j\geq 3} h_j\Big)^{\frac{2}{d-2}} + \frac{1}{\prod_{j\geq 3} h_j}$$

$$+ \sum_{j=1}^{d} |\log(h_j)| + 1\Big) \leq \frac{c}{T} \frac{\sum_{j=1}^{d} |\log(h_j)|}{\prod_{j\geq 3} h_j},$$

where the last inequality is a consequence of the fact that, as for any $j \in \{1, ..., d\}$ h_j is small and in particular they are smaller than $\frac{1}{2}$, all the other terms are bounded by $c \frac{\sum_{j=1}^{d} |\log(h_j)|}{\prod_{j\geq 3} h_j}$ and (6) is therefore proved.

If otherwise $h_1 h_2 \geq (\prod_{j\geq 3} h_j)^{\frac{2}{d-2}}$, we estimate directly $|k(s)|$ as in (16) between 0 and δ_2. Using also (25) and (26) we get

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

Choosing once again $\delta_2 := (\prod_{j\geq 3} h_j)^{\frac{2}{d-2}}$ and $D := [\max(-\frac{2}{\rho} \log(\prod_{j=1}^{d} h_j), 1) \wedge T]$ and recalling also that $\delta_2 = (\prod_{j\geq 3} h_j)^{\frac{2}{d-2}} \leq h_1 h_2$, we get

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

as we wanted. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.1. Proof of Theorem 1

Proof. We write the usual bias-variance decomposition

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \leq |\mathbb{E}[\hat{\pi}_{h,T}(x)] - \pi(x)|^2 + Var(\hat{\pi}_{h,T}(x)). \qquad (28)$$

Regarding the bias, a standard computation (see for example the proof of Proposition 2 of [2]) provides

$$|\mathbb{E}[\hat{\pi}_{h,T}(x)] - \pi(x)|^2 \leq c \sum_{j=1}^{d} h_j^{\beta_j}. \qquad (29)$$

An analogous computation can be found in Proposition 1.2 of [33] or in Proposition 1 of [8].

It is here important to remark that the constant c does not depend on x. For $h_1 h_2 < (\prod_{l>3} h_l)^{\frac{2}{d-2}}$, the estimation (29) here above together with the decomposition (28) and the upper bound on the variance gathered in (6) of Proposition 1, gives us (10).

In order to choose the rate optimal bandwidth, we define $h_l(T) := (\frac{\log T}{T})^{a_l}$ for $l \in \{1, ..., d\}$ and we look for $a_1,... a_d$ such that the upper bound of the mean-squared error in the right hand side of (10) is as small as possible. We remark that

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

$$\leq \frac{c}{T} \log T (\frac{T}{\log T})^{\sum_{l\geq 3} a_l}$$

$$= c(\frac{\log T}{T})^{1-\sum_{l\geq 3} a_l}$$

Therefore, after having replaced $h_l(T)$, the right hand side of (10) is

$$c\sum_{l=1}^d (\frac{\log T}{T})^{2a_l\beta_l} + c(\frac{\log T}{T})^{1-\sum_{l\geq 3} a_l}. \tag{30}$$

To get the balance we have to solve the following system in $a_3,..., a_d$:

$$\begin{cases} \beta_i a_i = \beta_{i+1} a_{i+1} & \forall i \in \{3, ..., d-1\} \\ 2\beta_d a_d = 1 - \sum_{l\geq 3} a_l, \end{cases}$$

while a_1 and a_2 have to be big enough to ensure that both $(\frac{1}{T})^{2\beta_1 a_1}$ and $(\frac{1}{T})^{2\beta_2 a_2}$ are negligible compared to the other terms. We observe that, as a consequence of the first $d-3$ equations, we can write

$$a_l = \frac{\beta_d}{\beta_l} a_d, \qquad \forall l \in \{3, ..., d-1\}. \tag{31}$$

Hence, the last equation becomes

$$2\beta_d a_d = 1 - \beta_d a_d \sum_{l\geq 3} \frac{1}{\beta_l} = 1 - \beta_d a_d \frac{d-2}{\bar{\beta}_3},$$

where $\bar{\beta}_3$ is the mean smoothness over $\beta_3,..., \beta_d$ and it is such that $\frac{1}{\bar{\beta}_3} = \frac{1}{d-2} \sum_{l\geq 3} \frac{1}{\beta_l}$. It follows

$$a_l = \frac{\bar{\beta}_3}{\beta_l(2\bar{\beta}_3 + d - 2)} \qquad \forall l \in \{3, ..., d-1\}. \tag{32}$$

Regarding a_1 and a_2, we take them big enough to ensure that

$$a_1 > \frac{\bar{\beta}_3}{\beta_1(2\bar{\beta}_3 + d - 2)}, \qquad a_2 > \frac{\bar{\beta}_3}{\beta_2(2\bar{\beta}_3 + d - 2)}. \tag{33}$$

Plugging them in (30) we get

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

as we wanted.

We now observe that, in the anisotropic case, the multi bandwidth h always satisfies $h_1 h_2 < (\prod_{l \ge 3} h_l)^{\frac{2}{d-2}}$ while it is possible to improve the convergence rate in the isotropic case, by removing the logarithm. Indeed, $h_1 h_2 < (\prod_{l \ge 3} h_l)^{\frac{2}{d-2}}$ holds true if and only if

$$(\frac{1}{T})^{a_1 + a_2} < (\frac{1}{T})^{(a_3 + \dots + a_d)\frac{2}{d-2}}.$$

Because of the choice of $a_1, ..., a_d$ gathered in (32) and (33), it holds true if

$$\frac{1}{2}(\frac{1}{\beta_1} + \frac{1}{\beta_2}) > \frac{1}{\bar{\beta}_3}. \tag{34}$$

As $\beta_1 \le \beta_2 \le ... \le \beta_d$, equation (34) always holds true, in the anisotropic context.

However, in the isotropic context, we have $h_1 h_2 = (\prod_{l \ge 3} h_l)^{\frac{2}{d-2}} = h^2$. Here estimation (29) together with decomposition (28) and the upper bound on the variance gathered in (7) of Proposition 1 gives us, remarking also that $\beta_1 = ... = \beta_d =: \beta$,

$$\mathbb{E}[|\hat{\pi}_{h,T}(x) - \pi(x)|^2] \le ch^{2\beta} + \frac{c}{T}\frac{1}{h^{d-2}}.$$

It leads us to the rate optimal choice $h(T) = (\frac{1}{T})^{\frac{1}{2\beta + d - 2}}$, which yields

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

as we wanted. $\qquad\qquad\square$

6. Proof lower bound

We want to prove Theorem 2 using the two hypothesis method, as explained for example in Section 2.3 of Tsybakov [33]. The idea is to introduce two drift functions b_0 and b_1 which belong to $\Sigma(\beta, \mathcal{L})$ and for which the laws $\mathbb{P}_{b_0}$ and $\mathbb{P}_{b_1}$ are close. To do it, the knowledge of the link between b and π_b is crucial. In particular, we will study in detail the above mentioned link in Section 6.1 while we will provide two priors in Section 6.2. In Section 6.3 we will use these preliminaries in order to prove the lower bound for the pointwise minimax risk gathered in Theorem 2.

6.1. Explicit link between the drift and the stationary measure

In absence of jumps, most of the times, reversible diffusion processes with unit diffusion processes are considered in order to estimate the invariant density (see [11] and [32]). In this case, the connection between the drift function and the invariant measure is explicit:

$$b(x) = -\nabla V(x) = \frac{1}{2} \nabla (\log \pi_b)(x),$$

where V is a $C^2(\mathbb{R}^d)$ function, which we refer to as potential. Adding the jumps, it is no longer true and so, in our framework, it is challenging to get a relation between b and π_b. We need to introduce A, the generator of the diffusion X solution of (13). It is composed by a continuous part and a discrete one: $A = A_c + A_d$, with

$$A_c f(x) := \frac{1}{2} \sum_{i,j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{i=1}^{d} b^i(x) \frac{\partial}{\partial x_i} f(x), \qquad (35)$$

$$A_d f(x) := \int_{\mathbb{R}^d} [f(x + \gamma \cdot z) - f(x) - \gamma \cdot z \cdot \nabla f(x)] F(z) dz.$$

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

$$\mathcal{C} := \left\{ f : \mathbb{R}^d \to \mathbb{R}, \ f \in C^2(\mathbb{R}^d) \text{ such that } \forall i \in \{1, ..., d\} \lim_{x_i \to \pm\infty} f(x) = 0, \right.$$

$$\left. \lim_{x_i \to \pm\infty} \frac{\partial}{\partial x_i} f(x) = 0 \text{ and } \int_{\mathbb{R}^d} f(x) dx < \infty \right\}.$$

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

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

The following lemma, that will be proven in Section 7, makes explicit the form of A_b^*.

Lemma 1. *Let A_b^* the adjoint operator on $L^2(\mathbb{R}^d)$ of A, generator of the diffusion X solution of (13), where the subscript b is to underline its dependence on the drift function. Then, for $g \in \mathcal{C}$, it is*

$$A_b^* g(x) = \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (a \, a^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} g(x) - (\sum_{i=1}^{d} \frac{\partial \, b^i}{\partial x_i} g(x) + b^i \frac{\partial \, g}{\partial x_i}(x)) +$$

$$+ \int_{\mathbb{R}^d} [g(x - \gamma \cdot z) - g(x) + \gamma \cdot z \cdot \nabla g(x)] F(z) dz.$$

If $g : \mathbb{R}^d \to \mathbb{R}$ is a probability density of class $\mathcal{C}^2$, solution of $A_b^* g = 0$, then it is an invariant density for the process we are considering. When the stationary distribution π_b is unique, therefore, it can be computed as solution of the equation $A_b^* \pi_b = 0$. As one can see from Lemma 1, the adjoint operator has a pretty complicate form. Hence, it seems impossible to find explicit solutions g of $A_b^* g = 0$ for any b and consequently it seems impossible to write π_b as an explicit function of b.

However, it can be seen that if one consider π_b as fixed and b as the unknown variable, then finding solutions in b is simpler. Moreover, the adjoint of the discrete part of the generator does not depend on b and therefore the solution in b is the same it would have been in absence of jumps, plus a second term which derives from the contribution of the jumps. In order to compute a function $b = b_g$ solution of $A_b^* g = 0$, we need to introduce some notations.

For $g \in \mathcal{C}$ we denote as $A_d^* g$ the adjoint operator of $A_d g$ which is, for all $x \in \mathbb{R}^d$,

$$A_d^* g(x) = \int_{\mathbb{R}^d} [g(x - \gamma \cdot z) - g(x) + \gamma \cdot z \cdot \nabla g(x)] F(z) dz.$$

Moreover, we introduce the following quantity, that will be useful in the sequel:

$$A_{d,i}^* g(x) = \int_{\mathbb{R}^d} [g(x_1 - (\gamma \cdot z)_1, ..., x_i - (\gamma \cdot z)_i, x_{i+1}, ..., x_d) \tag{36}$$

$$-g(x_1 - (\gamma \cdot z)_1, ..., x_{i-1} - (\gamma \cdot z)_{i-1}, x_i, ..., x_d) + (\gamma \cdot z)_i \frac{\partial g(x)}{\partial x_i}] F(z) dz.$$

To make easier the notation here above, we denote as $\bar{x}_i$ the vector $(x_1 - (\gamma \cdot z)_1, ..., x_i - (\gamma \cdot z)_i, x_{i+1}, ..., x_d)$ for $i \in \{1, ..., d\}$ while $\bar{x}_0$ is simply x. Clearly, it implies that $\bar{x}_d = x - (\gamma \cdot z)$ and so it is easy to prove that the sum of $A_{d,i}^* g(x)$ on i is $A_d^* g(x)$:

$$\sum_{i=1}^{d} A_{d,i}^* g(x) = \sum_{i=1}^{d} \int_{\mathbb{R}^d} [g(\bar{x}_i) - g(\bar{x}_{i-1}) + (\gamma \cdot z)_i \frac{\partial g(x)}{\partial x_i}] F(z) dz \tag{37}$$

$$= \int_{\mathbb{R}^d} [g(\bar{x}_d) - g(\bar{x}_0) + \sum_{i=1}^{d} (\gamma \cdot z)_i \frac{\partial g(x)}{\partial x_i}] F(z) dz$$

$$= \int_{\mathbb{R}^d} [g(x - \gamma \cdot z) - g(x) + \gamma \cdot z \cdot \nabla g(x)] F(z) dz = A_d^* g(x).$$

Then, for $g \in \mathcal{C}$ and $g > 0$, we introduce for all $x \in \mathbb{R}^d$ and for all $i \in \{1, ..., d\}$,

$$b_g^i(x) = \frac{1}{g(x)} \int_{-\infty}^{x_i} \left(\frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(w_i) + A_{d,i}^* g(w_i) \right) dw = \tag{38}$$

$$= \frac{1}{g(x)} \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial g}{\partial x_j}(x) + \frac{1}{g(x)} \int_{-\infty}^{x_i} A_{d,i}^* g(w_i) dw, \quad \text{if } x_i < 0;$$

$$b_g^i(x) = -\frac{1}{g(x)} \int_{x_i}^{\infty} \left(\frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(w_i) + A_{d,i}^* g(w_i)\right) dw$$

$$= \frac{1}{g(x)} \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial g}{\partial x_j}(x) - \frac{1}{g(x)} \int_{x_i}^{\infty} A_{d,i}^* g(w_i) dw, \quad \text{if } x_i \geq 0;$$

where $w_i = (x_1, ..., x_{i-1}, w, x_{i+1}, ..., x_d)$. We observe that, by the definition of $A_{d,i}^*$ and the fact that the function g is integrable, b^i is well defined. Moreover, as both g and its derivatives goes to zero at infinity and using that the Lebesgue measure is invariant on $\mathbb{R}$, it is

$$\int_{-\infty}^{\infty} \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(w_i) + A_{d,i}^* g(w_i) dw = 0.$$

Hence, the two definitions of b^i given here above are equivalent on $\mathbb{R}$. We finally denote as $b_g : \mathbb{R}^d \to \mathbb{R}^d$ the function such that, for all $x \in \mathbb{R}^d$, $b_g(x) = (b_g^1(x), ..., b_g^d(x))$.

We show that the function b_g here above introduced is actually solution of $A_b^* g(x) = 0$.

Proposition 2. *1. Let g a positive function in $\mathcal{C}$. Then,*

$$A_{b_g}^* g(x) = 0, \qquad \forall x \in \mathbb{R}^d.$$

2. Let $\pi : \mathbb{R}^d \to \mathbb{R}$ a probability density such that $\pi \in \mathcal{C}$ and $\pi > 0$. If b_π, defined as in (38), is a bounded Lipschitz function which satisfies A2, then π is the unique stationary probability of the stochastic differential equation (13) with drift coefficient $b = b_\pi$.

Proof. 1. For $b_g^i(x)$ defined as in (38), we get

$$\frac{\partial b_g^i}{\partial x_i}(x) = \frac{1}{g(x)} \left(\frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(x) + A_{d,i}^* g(x)\right) - \frac{b_g^i(x)}{g(x)} \frac{\partial g}{\partial x_i}(x).$$

Replacing $b_g^i(x)$ and $\frac{\partial b_g^i}{\partial x_i}(x)$ in $A_{b_g}^* g(x)$ given by Lemma 1 and using (37), we easily obtain $A_{b_g}^* g(x) = 0$.

2. From Ito's formula, one can check that any π solution of $A_b^* \pi(x) = 0$ is a stationary measure for the process X solution of (13). From point 1 we know that π is solution to $A_{b_\pi}^* \pi(x) = 0$ and so it is a stationary measure for the process X whose drift is b_π. However, we have assumed b_π to be a bounded Lipschitz function which satisfies A2 and, from Lemma 2 of [2], we know it is enough to ensure the existence of a Lyapounov and to show that the stationary measure of the equation with drift coefficient b_π is unique. It follows it is equal to π. $\qquad \square$

We recall that our purpose, in this section, is to clarify the link between the drift coefficient b_π of the stochastic differential equation (13) and the unique stationary distribution π. As a consequence of the second point of Proposition 2, it is achieved when b_π is a bounded Lipschitz function which satisfies A2. We now introduce some assumptions on π for which the associated drift b_π has the wanted properties.

Ad: 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. We denote $k_1 := \max_h ((a \cdot a^T)^{-1})_{hh}$. There exists $\epsilon > 0$ such that $\epsilon < \frac{\epsilon_0}{|\gamma_j| d\, k_1}$, (with $|\gamma_j|$ the euclidean norm of the j-th line of the matrix γ and ϵ_0 the value appearing in the fifth point of Assumption A3), for which for any $y, z \in \mathbb{R}$,

$$\pi_j(y \pm z) \le c_2 e^{\epsilon((a \cdot a^T)^{-1})_{jj}|z|} \pi_j(y),$$

   where c_2 is some constant > 0.
3. For $\epsilon > 0$ as in point 2 there exists $c_3(\epsilon) > 0$ such that

$$\sup_{y<0} \frac{1}{\pi_j(y)} \int_{-\infty}^y \pi_j(w) dw < c_3(\epsilon),$$

$$\sup_{y>0} \frac{1}{\pi_j(y)} \int_y^\infty \pi_j(w) dw < c_3(\epsilon).$$

4. We denote $k_2 := \max_{l,h \in \{1,...d\}} |(\gamma^T \cdot \gamma)_{lh}|$ and we recall that $\hat{c}$ is the constant appearing in the fifth point of A3. There exists $0 < \tilde\epsilon < \frac{1}{4\, k_1^2\, k_2\, \hat{c}\, c_3(\epsilon) c_2^d c_5}$, where c_5 is the constant that will be introduced below, in the fifth point of Ad, and there exists R such that, for any $y: |y| > \frac{R}{\sqrt{d}}$,

$$\frac{\pi_j'(y)}{\pi_j(y)} \le -\tilde\epsilon((a \cdot a^T)^{-1})_{jj}\, \mathrm{sgn}(y).$$

   Moreover, there exists a constant c_4 such that, for any $y \in \mathbb{R}$,

$$|\frac{\pi_j'(y)}{\pi_j(y)}| \le c_4.$$

5. For each $i \in \{1, ..., d\}$, for any $x \in \mathbb{R}^d$ and for $\tilde\epsilon$ as in point 4 there exists a constant c_5 such that

$$|\frac{\partial^2 \pi(x)}{\partial x_i \partial x_j}| \le c_5 (a \cdot a^T)^{-1})_{jj}(a \cdot a^T)^{-1})_{ii}\, \tilde\epsilon^2 \pi(x).$$

Moreover, there exists a constant $\tilde{c}_5$ such that, for any $x \in \mathbb{R}^d$, it is

$$|\frac{\partial^2 \pi(x)}{\partial x_i \partial x_j}| \le \tilde{c}_5.$$

Even though the just listed properties do not seem very natural and they have been introduced especially to make the associated drift function such that we can use the second point of Proposition 2, they are all satisfied by choosing a probability density in an exponential form, as we will see better in Lemma 2.

The proof of the following proposition will be given in Section 7.

Proposition 3. *Suppose that π satisfies Ad. Then b_π, defined as in* (38)*, is a bounded Lipschitz function which satisfies A2.*

From Proposition 3 here above and the second point of Proposition 2 it follows that, if we choose carefully the probability density such that all the properties gathered in Assumption Ad hold true, then π is the unique stationary probability of the stochastic differential equation (13) with drift coefficient b_π.
The next subsection is devoted to the building of two densities which satisfy the properties listed in Ad.

6.2. Construction of the priors

The proof of the lower bound is made by a comparison between the minimax risk introduced in (14) and some Bayesian risk where the Bayesian prior is supported on a set of two elements. 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 Ad 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 π_0 a probability measure. For any $y \in \mathbb{R}$ we define $\pi_{k,0}(y) := f(\eta(a \cdot a^T)_{kk}^{-1}|y|)$, where

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

and η is a constant in $(0, \frac{1}{2})$ which plays the same role as ϵ and $\tilde{\epsilon}$ did in Ad, 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, f is a C^∞ function and it is such that, for any $x \in \mathbb{R}$,

- $\frac{1}{2}e^{-|x|} \le f(x) \le 2e^{-|x|}$,
- $|f'(x)| \le 2e^{-|x|}$,
- $|f''(x)| \le 2e^{-|x|}$.

The function f has been introduced with the purpose of making $\pi_{k,0}(y)$ a $\mathcal{C}^\infty$ function for which all the conditions in Ad are satisfied. We state the following lemma, which will be proven in Section 7.

Lemma 2. *Let $\eta \in (0, \frac{1}{2})$. We suppose that the constant $\hat{c}$ introduced in the fifth point of A3 satisfies*

$$\hat{c} \leq \frac{k_3}{4^{d+4} k_1^2 k_2}, \tag{39}$$

where $k_3 := \min_h ((a \cdot a^T)^{-1})_{hh}$ and k_1 and k_2 are as defined in Ad. Then, taking $\eta = \epsilon = \tilde{\epsilon}$, the probability density π_0 satisfies Assumption Ad.

We remark that, as a consequence of the fifth point of A3 and of (39) here above, the assumption required on the jumps in order to make π_0 satisfy Ad is that there exists ϵ_0 such that

$$\int_{\mathbb{R}^d \setminus \{0\}} |z|^2 e^{\epsilon_0 |z|} F(z) dz \leq \frac{k_3}{4^{d+4} k_1^2 k_2}.$$

It means that the jumps have to behave well. In particular, they have to integrate an exponential function and such an integral has to be upper bounded by a constant which depends on the model.

From Proposition 3, we know that b_{π_0} is a bounded lipschitz function which satisfies A2 and, using also the second point of Proposition 2 it follows that π_0 is the unique stationary probability of $X^{(0)}$ solution of

$$X_t^{(0)} = X_0^{(0)} + \int_0^t b_{\pi_0}(X_s^{(0)}) ds + \int_0^t a \, dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma \, z \, \tilde{\mu}(ds, dz). \tag{40}$$

It yields $b_{\pi_0} \in \Sigma(\beta, \mathcal{L})$, according to Definition 2. To provide the second drift function belonging to $\Sigma(\beta, \mathcal{L})$ on which we want to apply the two hypothesis method, we introduce the probability measure π_1. We are given it as π_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

$$K(0) = 1, \qquad \int_{-1}^1 K(z) dz = 0. \tag{41}$$

We set

$$\pi_1(x) := \pi_0(x) + \frac{1}{M_T} \prod_{l=1}^d K\left(\frac{x_l - x_0^l}{h_l(T)}\right), \tag{42}$$

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 (14), 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 (41) 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 π_0 and π_1 belong to the anisotropic Holder class $\mathcal{H}_d(\beta, 2\mathcal{L})$ (according with Definition 2 of $\Sigma(\beta, \mathcal{L})$) and the laws $\mathbb{P}_{b_{\pi_0}}$ and $\mathbb{P}_{b_{\pi_1}}$ are close. It will provide us some constraints, under which we will choose M_T and $h_l(T)$ such that the lower bound on the minimax risk is as large as possible. In order to make the here above mentioned constraints explicit, we first of all need to evaluate how the two proposed drift functions differ in a neighbourhood of $x_0 \in \mathbb{R}^d$, as stated in the next proposition. Its proof will be given in Section 7.

Proposition 4. *Let us define the compact 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* $\forall i \in \{1, ..., d\}$: $|b_{\pi_1}^i(x) - b_{\pi_0}^i(x)| \le \frac{c}{M_T}$.
2. *For any* $i \in \{1, ..., d\}$: $\int_{K_T^c} |b_{\pi_1}^i(x) - b_{\pi_0}^i(x)| \pi_0(x) dx \le \frac{c}{M_T} \prod_{l=1}^{d} h_l(T)$.
3. *For any* $x \in K_T$ *and* $\forall i \in \{1, ..., d\}$: $|b_{\pi_1}^i(x) - b_{\pi_0}^i(x)| \le \frac{c}{M_T} \sum_{j=1}^{d} \frac{1}{h_j(T)}$, *where c is a constant independent of T.*

Using Proposition 4 it is possible to show that also b_{π_1} belongs to $\Sigma(\beta, \mathcal{L})$, up to calibrate properly M_T and $h_i(T)$, for $i \in \{1, ..., d\}$.

Lemma 3. *Let* $\epsilon > 0$ *and assume that, for all T large,*

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

We suppose moreover that $\sum_{j=1}^{d} \frac{1}{h_j(T)} = o(M_T)$ *as* $T \to \infty$. *Then, if* $\epsilon > 0$ *is small enough, we have*

$$b_{\pi_1} \in \Sigma(\beta, \mathcal{L}),$$

for all T sufficiently large.

Proof. From the first point of Proposition 4 here above we know that, $\forall i \in \{1, ..., d\}$,

$$b_{\pi_1}^i = b_{\pi_0}^i + O(\frac{1}{M_T}) = b_{\pi_0}^i + o(1) \qquad \forall x \in K_T^c, \tag{43}$$

while the third point of Proposition 4 provides us,

$$b_{\pi_1}^i = b_{\pi_0}^i + O(\frac{1}{M_T} \sum_{j=1}^{d} \frac{1}{h_j(T)}) = b_{\pi_0}^i + o(1) \qquad \forall x \in K_T, \tag{44}$$

being the last equality a consequence of $\sum_{j=1}^{d} \frac{1}{h_j(T)} = o(M_T)$, for T going to ∞. We recall that we have built the density π_0 especially to apply Proposition 3 on b_{π_0}. Therefore, b_{π_0} is a bounded Lipschitz function which satisfies A2 and, as a

consequence of (44) and (43), the same goes for b_{π_1}. From the second point of Proposition 2 it follows that π_1 is the unique stationary measure associated to b_{π_1}. As $\Sigma(\beta, \mathcal{L})$ is defined as in Definition 2, the proof of the lemma is complete as soon as $\pi_1 \in \mathcal{H}_d(\beta, 2\mathcal{L})$. Let us check the Hölder condition with respect to the i-th component. We first of all introduce the following notation: $\forall x \in \mathbb{R}^d$

$$d_T(x) := \pi_1(x) - \pi_0(x) = \frac{1}{M_T} \prod_{l=1}^{d} K\left(\frac{x_l - x_0^l}{h_l(T)}\right). \tag{45}$$

For all $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$ it is

$$|D_i^{\lfloor \beta_i \rfloor} \pi_1(x + te_i) - D_i^{\lfloor \beta_i \rfloor} \pi_1(x)|$$
$$\leq |D_i^{\lfloor \beta_i \rfloor} \pi_0(x + te_i) - D_i^{\lfloor \beta_i \rfloor} \pi_0(x)| + |D_i^{\lfloor \beta_i \rfloor} d_T(x + te_i) - D_i^{\lfloor \beta_i \rfloor} d_T(x)|$$
$$\leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} + \frac{\|K\|_\infty^{d-1}}{M_T (h_i(T))^{\lfloor \beta_i \rfloor}} |K^{\lfloor \beta_i \rfloor}\left(\frac{x_i + t - x_0^i}{h_i(T)}\right) - K^{\lfloor \beta_i \rfloor}\left(\frac{x_i - x_0^i}{h_i(T)}\right)|,$$

where we have used the definition of d_T and the fact that $\pi_0 \in \mathcal{H}_d(\beta, \mathcal{L})$. We now observe that

$$|K^{\lfloor \beta_i \rfloor}\left(\frac{x_i + t - x_0^i}{h_i(T)}\right) - K^{\lfloor \beta_i \rfloor}\left(\frac{x_i - x_0^i}{h_i(T)}\right)| \leq$$

$$\leq |K^{\lfloor \beta_i \rfloor}\left(\frac{x_i + t - x_0^i}{h_i(T)}\right) - K^{\lfloor \beta_i \rfloor}\left(\frac{x_i - x_0^i}{h_i(T)}\right)|^{\beta_i - \lfloor \beta_i \rfloor} (2\left\|K^{\lfloor \beta_i \rfloor}\right\|_\infty)^{1 - (\beta_i - \lfloor \beta_i \rfloor)} \leq$$

$$\leq \left\|K^{\lfloor \beta_i \rfloor + 1}\right\|_\infty^{\beta_i - \lfloor \beta_i \rfloor} \left(\frac{|t|}{|h_i(T)|}\right)^{\beta_i - \lfloor \beta_i \rfloor} (2\left\|K^{\lfloor \beta_i \rfloor}\right\|_\infty)^{1 - (\beta_i - \lfloor \beta_i \rfloor)}.$$

Therefore, defining

$$c_K := \|K\|_\infty^{d-1} \left\|K^{\lfloor \beta_i \rfloor + 1}\right\|_\infty^{\beta_i - \lfloor \beta_i \rfloor} (2\left\|K^{\lfloor \beta_i \rfloor}\right\|_\infty)^{1 - (\beta_i - \lfloor \beta_i \rfloor)},$$

it follows

$$|D_i^{\lfloor \beta_i \rfloor} \pi_1(x + te_i) - D_i^{\lfloor \beta_i \rfloor} \pi_1(x)| \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} + \frac{c_K}{M_T (h_i(T))^{\beta_i}} |t|^{\beta_i - \lfloor \beta_i \rfloor}.$$

We have assumed that, $\forall i \in \{1, ..., d\}$, $\frac{1}{M_T} \leq \epsilon h_i(T)^{\beta_i}$. We can choose ϵ small enough to ensure that $\epsilon \leq \frac{\mathcal{L}_i}{c_K}$, obtaining

$$|D_i^{\lfloor \beta_i \rfloor} \pi_1(x + te_i) - D_i^{\lfloor \beta_i \rfloor} \pi_1(x)| \leq 2\mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor}.$$

Moreover, from the definition of π_1 and the fact that $\pi_0 \in \mathcal{H}_d(\beta, \mathcal{L})$ we also get, for any $x \in \mathbb{R}^d$ and $k = 0, ..., \lfloor \beta_i \rfloor$

$$|D_i^k \pi_1(x)| \leq \mathcal{L}_i + \frac{\|K\|_\infty^{d-1} \left\|K^k\right\|_\infty}{M_T (h_i(T))^k} =: \mathcal{L}_i + \frac{c_K'}{M_T (h_i(T))^k}.$$

As $k \leq \lfloor \beta_i \rfloor \leq \beta_i$ it follows $h_i(T)^k \geq h_i(T)^{\beta_i}$ and so

$$\frac{1}{M_T} \leq \epsilon h_i(T)^{\beta_i} \leq \epsilon h_i(T)^k.$$

Again, it is enough to choose ϵ such that $\epsilon \leq \frac{\mathcal{L}_i}{c_K'}$ to get

$$|D_i^k \pi_1(x)| \leq 2\mathcal{L}_i.$$

We have proven the required Hölder controls on the derivatives of π_1, the lemma follows. $\qquad \Box$

We remark that the two conditions on the calibration parameters provide

$$\sum_{j=1}^{d} \frac{h_j^{\beta_j}(T)}{h_j(T)} = o(1),$$

which is always true as we have asked $\beta_j > 1$ for any $j \in \{1, ..., d\}$ in Definition 2.

6.3. Proof of Theorem 2

Proof. We first of all recall the notations previously introduced. We denote $\mathbb{P}_b$ the law of the stationary solution of (13) on the canonical space $C([0, \infty), \mathbb{R}^d)$ and $\mathbb{E}_b$ the corresponding expectation; we also denote as $\mathbb{P}_b^{(T)}$ and $\mathbb{E}_b^{(T)}$ their restrictions on $C([0, T], \mathbb{R}^d)$. For any measurable function $\tilde{\pi}_T(x_0) : C([0, T], \mathbb{R}^d) \to \mathbb{R}$ we will estimate by below, for T large, its risk

$$R(\tilde{\pi}_T(x_0)) := \sup_{b \in \Sigma(\beta, \mathcal{L})} \mathbb{E}_b^{(T)} [(\tilde{\pi}_T(x_0) - \pi_b(x_0))^2].$$

We want to use the two hypothesis method based on the two drift functions b_{π_0} and b_{π_1} which, therefore, have to belong to $\Sigma(\beta, \mathcal{L})$. It is $b_{\pi_0} \in \Sigma(\beta, \mathcal{L})$ by construction. Moreover, from Lemma 3 we know in detail the constraints required on the calibrations M_T and $h_i(T)$ in order to get b_{π_1} belonging to $\Sigma(\beta, \mathcal{L})$. We therefore assume that the following conditions hold true:

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

$$\sum_{j=1}^{d} \frac{1}{h_j(T)} = o(M_T), \quad \text{as } T \to \infty. \tag{47}$$

As $b_{\pi_0}, b_{\pi_1} \in \Sigma(\beta, \mathcal{L})$ we have

$$R(\tilde{\pi}_T(x_0)) \geq \frac{1}{2} \mathbb{E}_{b_{\pi_1}}^{(T)} [(\tilde{\pi}_T(x_0) - \pi_1(x_0))^2] + \frac{1}{2} \mathbb{E}_{b_{\pi_0}}^{(T)} [(\tilde{\pi}_T(x_0) - \pi_0(x_0))^2].$$

In order to lower bound the right hand side we need the following lemma, which will be showed in Section 7.

Lemma 4.   *1. The measure $\mathbb{P}^{(T)}_{b_{\pi_1}}$ is absolutely continuous with respect to $\mathbb{P}^{(T)}_{b_{\pi_0}}$.*

2. We denote $Z^{(T)} := \frac{d\mathbb{P}^{(T)}_{b_{\pi_1}}}{d\mathbb{P}^{(T)}_{b_{\pi_0}}}$ and we assume that

$$\sup_{T \geq 0} T \int_{\mathbb{R}^d} |b_{\pi_1}(x) - b_{\pi_0}(x)|^2 \pi_0(x)\, dx < \infty. \tag{48}$$

Then, there exist C and $\lambda > 0$ such that, for all T large enough,

$$\mathbb{P}^{(T)}_{b_{\pi_0}}(Z^{(T)} \geq \tfrac{1}{\lambda}) \geq C.$$

From (48) it turns out another condition on the calibration quantities. Indeed, using all the three points of Proposition 4, it is

$$\int_{\mathbb{R}^d} |b_{\pi_1}(x) - b_{\pi_0}(x)|^2 \pi_0(x)\, dx$$

$$= \int_{K_T} |b_{\pi_1}(x) - b_{\pi_0}(x)|^2 \pi_0(x)\, dx + \int_{K_T^c} |b_{\pi_1}(x) - b_{\pi_0}(x)|^2 \pi_0(x)\, dx$$

$$\leq \frac{c}{M_T^2}(\prod_{l=1}^{d} h_l(T))(\sum_{j=1}^{d} \frac{1}{h_j^2(T)}) + \frac{c}{M_T} \frac{1}{M_T} \prod_{l=1}^{d} h_l(T)$$

$$\leq \frac{c}{M_T^2}(\prod_{l=1}^{d} h_l(T))(\sum_{j=1}^{d} \frac{1}{h_j^2(T)}),$$

as $h_j(T)$ goes to 0 for T going to infinity and so the second term here above is negligible compared to the first one. It provides us the constraints on the calibration

$$\sup_{T \geq 0} T \frac{c}{M_T^2}(\prod_{l=1}^{d} h_l(T))(\sum_{j=1}^{d} \frac{1}{h_j^2(T)}) < \infty, \tag{49}$$

that we need to require in order to apply Lemma 4 here above. From Lemma 4, as $Z^{(T)}$ exists, we can write

$$R(\tilde{\pi}_T(x_0)) \geq \frac{1}{2}\mathbb{E}^{(T)}_{b_{\pi_0}}[(\tilde{\pi}_T(x_0) - \pi_1(x_0))^2 Z^{(T)}] + \frac{1}{2}\mathbb{E}^{(T)}_{b_{\pi_0}}[(\tilde{\pi}_T(x_0) - \pi_0(x_0))^2]$$

$$\geq \frac{1}{2\lambda}\mathbb{E}^{(T)}_{b_{\pi_0}}[(\tilde{\pi}_T(x_0) - \pi_1(x_0))^2 1_{\left\{Z^{(T)} \geq \frac{1}{\lambda}\right\}}] + \frac{1}{2}\mathbb{E}^{(T)}_{b_{\pi_0}}[(\tilde{\pi}_T(x_0) - \pi_0(x_0))^2 1_{\left\{Z^{(T)} \geq \frac{1}{\lambda}\right\}}]$$

$$= \frac{1}{2\lambda}\mathbb{E}^{(T)}_{b_{\pi_0}}\big[[(\tilde{\pi}_T(x_0) - \pi_1(x_0))^2 + (\tilde{\pi}_T(x_0) - \pi_0(x_0))^2]1_{\left\{Z^{(T)} \geq \frac{1}{\lambda}\right\}}\big],$$

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

$$R(\tilde{\pi}_T(x_0)) \geq \frac{1}{8\lambda} (\pi_1(x_0) - \pi_0(x_0))^2 \mathbb{P}_{b_{\pi_0}}^{(T)} (Z^{(T)} \geq \frac{1}{\lambda}).$$

We recall that π_0 and π_1 have been built in Section 6.2 and in particular, since π_1 has been defined as below (41), it is

$$\pi_1(x_0) - \pi_0(x_0) = \frac{1}{M_T} \prod_{l=1}^{d} K(0) = \frac{1}{M_T},$$

where we have also used that $K(0) = 1$, as stated in (41). Moreover from Lemma 4 we know that for some λ, as soon as (49) holds,

$$\mathbb{P}_{b_{\pi_0}}^{(T)} (Z^{(T)} \geq \frac{1}{\lambda}) > 0.$$

We deduce that, if (46), (47) and (49) are satisfied, then

$$R(\tilde{\pi}_T(x_0)) \geq \frac{c}{M_T^2}, \tag{50}$$

for $c > 0$. Hence, we have to find the largest choice for $\frac{1}{M_T^2}$, subject to the constraints (46), (47) and (49).

We observe that (47) can be seen as $\sum_{j=1}^{d} \frac{h_j(T)^{\beta_j}}{h_j(T)} = o(1)$, which holds true as in Definition 1 of Hölder space we have asked $\beta_j > 1 \ \forall j \in \{1, ..., d\}$. Regarding the other conditions, we suppose at the beginning to saturate (46) for any $j \in \{1, ..., d\}$. From the order of β we obtain

$$h_1(T) = h_2(T)^{\frac{\beta_2}{\beta_1}} \leq h_2(T) \leq ... \leq h_d(T). \tag{51}$$

We plug it in (49) 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 $h_1(T)$, such that condition (46) is no longer saturated for $j = 1$. In particular, we increase $h_1(T)$ 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 $h_1(T) = h_2(T)$. Then, we have the possibility to no longer saturate condition (46) also for other j, which means to increase some $h_j(T)$. However, it implies the worse term $\frac{\prod_{l \neq 1} h_l(T)}{h_1(T)}$ to be bigger, and so it does not consist in a good choice. Finally, we take $h_j(T)$ which saturates (46) for any $j \neq 1$ and $h_1(T) = h_2(T)$. Replacing them in (49), we get the following condition:

$$\sup_T T \frac{1}{M_T^2} \prod_{l \geq 3} h_l(T) \leq c. \tag{52}$$

Now it is

$$\prod_{l \geq 3} h_l(T) = \left(\frac{1}{M_T}\right)^{\sum_{l \geq 3} \frac{1}{\bar{\beta}_l}} = \left(\frac{1}{M_T}\right)^{\frac{d-2}{\bar{\beta}_3}}. \tag{53}$$

Replacing (53) in (52), it leads us to the choice

$$\frac{1}{M_T} = \left(\frac{1}{T}\right)^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}.$$

Plugging the value of M_T in (50) we obtain that, for any possible estimator $\tilde{\pi}_T$ of the invariant density, it is

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

The wanted lower bound on the minimax risk defined in (14) follows. □

7. Proofs

This section is devoted to the proofs of the technical results we have introduced in the previous section.

7.1. Proof of Lemma 1

Proof. We aim at making explicit the adjoint operator of A, the generator of the process solution to (13), on $L^2(\mathbb{R}^d)$. It is such that, for f, g belonging to the set $\mathcal{C}$ as introduced in Section 6.1,

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

We start analysing the continuous part of the generator of (13), A. From (35), a repeated use of integration by parts and the fact that the function g vanishes for x_i going to $\pm\infty$ for any $i \in \{1, ..., d\}$ we get

$$\int_{\mathbb{R}^d} A_c f(x)g(x)dx$$

$$= \frac{1}{2} \sum_{i,j=1}^{d} (a \cdot a^T)_{ij} \int_{\mathbb{R}^d} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} g(x)dx + \sum_{i=1}^{d} \int_{\mathbb{R}^d} b^i(x) \frac{\partial f(x)}{\partial x_i} g(x)dx$$

$$= \frac{1}{2} \sum_{i,j=1}^{d} (a \cdot a^T)_{ij} \int_{\mathbb{R}^d} f(x) \frac{\partial^2 g(x)}{\partial x_i \partial x_j} dx - \sum_{i=1}^{d} \int_{\mathbb{R}^d} f(x) \frac{\partial}{\partial x_i} (b^i g)(x)dx$$

$$= \int_{\mathbb{R}^d} f(x) \Big[\frac{1}{2} \sum_{i,j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial^2 g(x)}{\partial x_i \partial x_j} dx - \sum_{i=1}^{d} \big(g(x) \frac{\partial b^i(x)}{\partial x_i} + \frac{\partial g(x)}{\partial x_i} b^i(x) \big) \Big] dx$$

$$=: \int_{\mathbb{R}^d} f(x) A_{c,b}^* g(x)dx. \tag{54}$$

We now look for the adjoint operator of the discrete part of the generator A_d as defined in (35). It is

$$\int_{\mathbb{R}^d} A_d f(x)g(x)dx = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} f(x + \gamma \cdot z)F(z)dz \Big) g(x)dx$$

$$- \int_{\mathbb{R}^d} f(x) \Big(\int_{\mathbb{R}^d} F(z)dz \Big) g(x)dx - \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \gamma \cdot z \cdot \nabla f(x)F(z)dz \Big) g(x)dx$$

$$=: I_1 + I_2 + I_3$$

We evaluate first of all I_1, on which we operate the change of variable $u := x + \gamma \cdot z$. It provides us

$$I_1 = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |\gamma^{-1}| f(u) F(\gamma^{-1}(u - x)) du \Big) g(x) dx$$

$$= \int_{\mathbb{R}^d} f(u) \Big(\int_{\mathbb{R}^d} |\gamma^{-1}| F(\gamma^{-1}(u - x)) g(x) dx \Big) du,$$

with the last equality which follows from Fubini theorem. We recall that $|\gamma^{-1}|$ stands for the absolute value of the determinant of the matrix γ^{-1}. Regarding I_2, one can clearly isolate the adjoint part without further computations as $(-\int_{\mathbb{R}^d} F(z)dz)g(x)$. The last term left to deal with is I_3. From integration by parts and once again the fact that g vanishes for x_i going to $\pm\infty$ we obtain

$$I_3 = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \gamma \cdot z \cdot \nabla g(x) F(z) dz \Big) f(x) dx.$$

Therefore

$$A_d^* g(x) = \int_{\mathbb{R}^d} |\gamma^{-1}| F(\gamma^{-1} \cdot (x - y)) g(y) dy - \int_{\mathbb{R}^d} F(z) dz\, g(x)$$

$$+ \int_{\mathbb{R}^d} \gamma \cdot z \cdot \nabla g(x) F(z) dz$$

$$= \int_{\mathbb{R}^d} [g(x - \gamma \cdot z) - g(x) + \gamma \cdot z \cdot \nabla g(x)] F(z) dz, \tag{55}$$

where we have also changed the variable $\gamma^{-1} \cdot (x - y) = z$ in the first integral. From (54) and (55) the lemma follows. $\quad\square$

7.2. *Proof of Proposition 3*

Proof. We start proving that b_π is bounded. We can assume WLOG $x_i < 0$, if $x_i > 0$ an analogous reasoning applies. As π is in a multiplicative form, we can compute

$$\frac{1}{\pi(x)} \int_{-\infty}^{x_i} \frac{1}{2} \sum_{j=1}^d (a \cdot a^T)_{ij} \frac{\partial^2 \pi}{\partial x_i \partial x_j}(w_i) dw$$

$$= \frac{1}{\pi(x)} \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \left(\frac{\partial \pi}{\partial x_j}(x) - \frac{\partial \pi}{\partial x_j}(x_1, ..., -\infty, .., x_d) \right)$$

$$= \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \left(\frac{\pi'_j(x_j)}{\pi_j(x_j)} - \frac{\pi'_j(-\infty)}{\pi_j(x_j)} \right) = \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\pi'_j(x_j)}{\pi_j(x_j)},$$

where we have also used that, for the first point of Ad, $\pi'_j(-\infty) = 0$. Comparing the equation here above with the definition (38) of b^i one can see that, for all $x \in \mathbb{R}^d$, $x_i < 0$,

$$b^i_\pi(x) = \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\pi'_j(x_j)}{\pi_j(x_j)} + \frac{1}{\pi(x)} \int_{-\infty}^{x_i} A^*_{d,i}\pi(w_i)dw =: I_1[\pi] + I_2[\pi].$$

From the fourth point of Ad it easily follows that there exists a constant $c > 0$ for which

$$|I_1[\pi]| \leq c. \tag{56}$$

Regarding $I_2[\pi]$, we start evaluating, for any $x \in \mathbb{R}^d$

$$A^*_{d,i}\pi(x) = \int_{\mathbb{R}^d} [\pi(\bar{x}_i) - \pi(\bar{x}_{i-1}) + (\gamma \cdot z)_i \frac{\partial}{\partial x_i}\pi(x)]F(z)dz,$$

where $\bar{x}_i = (x_1 - (\gamma \cdot z)_1, ..., x_i - (\gamma \cdot z)_i, x_{i+1}, ..., x_d)$. From intermediate value theorem we have

$$|A^*_{d,i}\pi(x)| \leq \int_{\mathbb{R}^d} \int_0^1 (\gamma \cdot z)_i^2 |\frac{\partial^2}{\partial x_i^2}\pi(x_1 - (\gamma \cdot z)_1, ..., x_i - s(\gamma \cdot z)_i, x_{i+1}, ..., x_d)F(z)|dsdz.$$

The fifth point of Ad (with the notation introduced in the fouth one) provides us an upper bound on the second derivative of π which yields, using also that π is in a multiplicative form,

$$|A^*_{d,i}\pi(x)| \leq c_n k_1^2 \tilde{\epsilon}^2 c_5 \int_{\mathbb{R}^d} \int_0^1 (\gamma \cdot z)_i^2 \prod_{j<i} \pi_j(x_j - (\gamma \cdot z)_j) \pi_i(x_i - s(\gamma \cdot z)_i)$$

$$\times \prod_{j>i} \pi_j(x_j) ds F(z) dz.$$

$$\leq c_n d\, k_1^2\, k_2\, c_5\, \tilde{\epsilon}^2 \int_{\mathbb{R}^d} \int_0^1 |z|^2 \prod_{j<i} \pi_j(x_j - (\gamma \cdot z)_j) \pi_i(x_i - s(\gamma \cdot z)_i)$$

$$\times \prod_{j>i} \pi_j(x_j) ds F(z) dz. \tag{57}$$

The second point of Ad yields

$$\pi_i(x_i - s(\gamma \cdot z)_i) \leq c_2 e^{\epsilon(a \cdot a^T)_{ii}^{-1}|s \sum_{k=1}^{d} \gamma_{ik} z_k|} \pi_i(x_i).$$

We apply exactly the same on $\pi_j(x_j - (\gamma \cdot z)_j)$, for $j < i$ and we replace them in the right hand side of (57), recalling that $|s| < 1$. We obtain it is upper bounded by

$$d\, k_1^2\, k_2\, c_2^i\, c_n\, c_5 \tilde{\epsilon}^2 \prod_{j=1}^{d} \pi_j(x_j) \int_{\mathbb{R}^d} |z|^2 e^{\epsilon(a \cdot a^T)_{jj}^{-1}|\sum_{j \leq i} \sum_{k=1}^{d} \gamma_{jk} z_k|} F(z) dz.$$

Now, by the definition of ϵ given in the third point of Ad, we know it satisfies $\epsilon(a \cdot a^T)_{jj}^{-1}|\sum_{j \leq i} \sum_{k=1}^{d} \gamma_{jk} z_k| \leq \epsilon_0 d|z|\frac{1}{d} = \epsilon_0|z|$. It follows that the integral in z is bounded by $\hat{c}$ and so

$$|A_{d,i}^* \pi(x)| \leq d\, k_1^2\, k_2\, c_2^i\, c_n\, c_5\, \hat{c}\tilde{\epsilon}^2 \prod_{j=1}^{d} \pi_j(x_j).$$

We plug it in $I_2[\pi]$, getting

$$|I_2[\pi]| \leq \frac{k_1^2\, k_2\, \hat{c}\, c_2^d\, c_5 \tilde{\epsilon}^2}{\pi_i(x_i)} \int_{-\infty}^{x_i} \pi_i(w) dw \leq k_1^2\, k_2\, \hat{c}\, c_3(\epsilon)\, c_2^d\, c_5\, \tilde{\epsilon}^2, \qquad (58)$$

as the integral on w is upper bounded by $c_3(\epsilon)$, from the third point of Ad. We have proved (56) and (58) and, therefore, b^i is clearly bounded.

We now want to prove the drift condition A2 on b_π^i. To do it, we investigate the behavior of $x_i b_\pi^i(x)$. From the fourth point of Ad, which holds true for any x_i such that $|x_i| > \frac{R}{\sqrt{d}}$, it is

$$x_i I_1[\pi] \leq \frac{x_i}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij}(-\tilde{\epsilon}(a \cdot a^T)_{jj}^{-1} \operatorname{sgn}(x_j))$$

$$= -\frac{\tilde{\epsilon}}{2} x_i \operatorname{sgn}(x_i) - \frac{\tilde{\epsilon}}{2} x_i \sum_{j \neq i} (a \cdot a^T)_{ij}(a \cdot a^T)_{jj}^{-1} \operatorname{sgn}(x_j).$$

As we have assumed $|\sum_{j \neq i} (a \cdot a^T)_{ij}(a \cdot a^T)_{jj}^{-1}| < \frac{1}{2}$, it is

$$\left| -\frac{\tilde{\epsilon}}{2} x_i \sum_{j \neq i} (a \cdot a^T)_{ij}(a \cdot a^T)_{jj}^{-1} \operatorname{sgn}(x_j) \right| < \frac{\tilde{\epsilon}}{4} |x_i|.$$

It follows

$$x_i I_1[\pi] \leq -\frac{\tilde{\epsilon}}{2} |x_i| + \frac{\tilde{\epsilon}}{4} |x_i| = -\frac{\tilde{\epsilon}}{4} |x_i|.$$

Using (58) we also get

$$|x_i I_2[\pi]| \leq k_1^2\, k_2\, \hat{c}\, c_3(\epsilon)\, c_2^d\, c_5\, \tilde{\epsilon}^2 |x_i|.$$

Hence, for x_i such that $|x_i| > \frac{R}{\sqrt{d}}$, there exists $\tilde{c} > 0$ such that

$$x_i b_\pi^i(x) \leq x_i I_1[\pi] + |x_i I_2[\pi]| \leq (-\frac{1}{4} + k_1^2\, k_2\, \hat{c}\, c_3(\epsilon)\, c_2^d\, c_5\, \tilde{\epsilon})\tilde{\epsilon}|x_i| \leq -\tilde{c}|x_i|, \quad (59)$$

where the last inequality is a consequence of the fact we have assumed $\tilde{\epsilon} < \frac{1}{4\,k_1^2\,k_2\,\hat{c}\,c_3(\epsilon)\,c_2^d\,c_5}$. From (59), using also the boundedness of b_π^i showed before, it follows

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

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

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 drift condition on b_π clearly holds.
As b is also Lipschitz, the result follows.   □

7.3. Proof of Lemma 2

Proof. We recall that π_0 has been defined as

$$\pi_0(x) := c_\eta \prod_{k=1}^d \pi_{k,0}(x_k),$$

with $\pi_{k,0}(y) := f(\eta(a \cdot a^T)_{kk}^{-1}|y|)$ and

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

By construction, π_0 is clearly in a multiplicative form and always positive. Moreover, point 1 of Ad directly hold true from the definition of $\pi_{j,0}(y)$. To show the second point to hold, we observe it is

$$\pi_{k,0}(y \pm z) = f(\eta(a \cdot a^T)_{kk}^{-1}|y \pm z|) \leq 2e^{-\eta(a \cdot a^T)_{kk}^{-1}|y \pm z|}$$

$$\leq 2e^{-\eta(a \cdot a^T)_{kk}^{-1}|y|} e^{\eta(a \cdot a^T)_{kk}^{-1}|z|} \leq 4\pi_{k,0}(y) e^{\eta(a \cdot a^T)_{kk}^{-1}|z|}.$$

It implies that point 2 of Ad holds with $c_2 = 4$ and $\epsilon = \eta$, as we can choose η small enough to make also the condition in the definition of ϵ satisfied.
In order to prove that the third point Ad holds true, we need to show that, for any $y < 0$,

$$\frac{1}{\pi_{k,0}(y)} \int_{-\infty}^y \pi_{k,0}(w) dw < \infty. \tag{60}$$

By the lower and upper bounds on $\pi_{k,0}$ provided through the first property of f we know it is

$$\frac{1}{\pi_{k,0}(y)} \int_{-\infty}^{y} \pi_{k,0}(w)dw \le 2e^{\eta(a\cdot a^T)_{kk}^{-1}|y|} \int_{-\infty}^{y} 2e^{-\eta(a\cdot a^T)_{kk}^{-1}|w|}dw$$

$$= 4e^{\eta(a\cdot a^T)_{kk}^{-1}|y|} \frac{e^{-\eta(a\cdot a^T)_{kk}^{-1}|y|}}{\eta(a\cdot a^T)_{kk}^{-1}} \le \frac{4}{\eta k_3} =: c_3(\eta).$$

For $y > 0$ an analogous reasoning applies, thus the third point of Ad follows with $c_3(\epsilon) = c_3(\eta) = \frac{4}{k_3\eta}$. It is easy to check that also the fourth point of Ad hold true as, for $|y| > \frac{1}{\eta(a\cdot a^T)_{kk}^{-1}}$, it is

$$\pi'_{k,0}(y) = -\eta(a\cdot a^T)_{kk}^{-1}\operatorname{sgn}(y)\pi_{k,0}(y) \quad \forall k \in \{1,..,d\}.$$

It means that the fourth point of Ad holds true for $|y| > \frac{R}{\sqrt{d}}$, up to take $R = \frac{\sqrt{d}}{\eta k_3}$. Moreover, in order to prove that also the fifth point of Ad holds, we observe it is

$$\frac{\partial^2 \pi_0(x)}{\partial x_i \partial x_j} = c_\eta \prod_{l \ne i,j} \pi_{l,0}(x_l)(\pi'_{i,0}(x_i)\pi'_{j,0}(x_j)1_{i\ne j} + \pi''_{j,0}(x_j)1_{i=j}).$$

From the definition of π_0 and the properties of f we have that, for any $j \in \{1,..,d\}$ and for $k = 1,2$,

$$|\pi_{j,0}^{(k)}(x_j)| = |f^{(k)}(\eta(a\cdot a^T)_{jj}^{-1}|x_j|)| \le 2(\eta(a\cdot a^T)_{jj}^{-1})^k e^{-\eta(a\cdot a^T)_{jj}^{-1}|x_j|}$$

$$\le 4(\eta(a\cdot a^T)_{jj}^{-1})^k \pi_{j,0}(x_j).$$

It follows

$$|\frac{\partial^2 \pi_0(x)}{\partial x_i \partial x_j}| \le 16\eta^2(a\cdot a^T)_{ii}^{-1}(a\cdot a^T)_{jj}^{-1}\pi_0(x).$$

It provides us that condition five of Ad holds true with $c_5 = 16$ and $\tilde\epsilon = \eta$. Finally, we have to check that, according with the definition of $\tilde\epsilon$ given in the fifth point of Ad, it is

$$\tilde\epsilon = \eta < \frac{1}{4k_1^2 k_2 \hat c c_3(\eta)c_2^d c_5} = \frac{\eta k_3}{4k_1^2 k_2 \hat c\, 4\, 4^d 4^2},$$

where we have also replaced the values of the constants we have found. It holds true of and only if

$$1 < \frac{k_3}{4^{d+4}k_1^2 k_2 \hat c},$$

which is equivalent to ask

$$\hat c < \frac{k_3}{4^{d+4}k_1^2 k_2}.$$

Being it exactly the condition assumed in the statement of this lemma, all the points gathered in Ad are satisfied. $\qquad\square$

7.4. Proof of Proposition [4]

Proof. Point 1

We suppose $x_i < 0$. If otherwise it is $x_i \geq 0$ it is enough to act in the same way on the integral between x_i and ∞ to get the same result. We first of all introduce the following quantities

$$\tilde{I}_1^i[\pi_0](x) := \frac{1}{2} \sum_{j=1}^{d} (a \cdot a^T)_{ij} \frac{\partial \pi_0}{\partial x_j}(x),$$

$$\tilde{I}_2^i[\pi_0](x) = \int_{-\infty}^{x_i} A_{d,i}^* \pi_0(w_i) dw.$$

We moreover introduce the notation

$$\tilde{I}^i[\pi_0](x) = \tilde{I}_1^i[\pi_0](x) + \tilde{I}_2^i[\pi_0](x).$$

According with the definition (38), we have

$$b_{\pi_0}^i(x) = \frac{1}{\pi_0(x)} \tilde{I}^i[\pi_0](x), \qquad b_{\pi_1}^i(x) = \frac{1}{\pi_1(x)} \tilde{I}^i[\pi_1](x).$$

Let us also recall the notation presented in (45) for which

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

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

$$b_{\pi_1}^i(x) = \frac{1}{\pi_1(x)} \tilde{I}^i[\pi_1](x) = \frac{1}{\pi_1(x)} \tilde{I}^i[\pi_0](x) + \frac{1}{\pi_1(x)} \tilde{I}^i[d_T](x). \qquad (61)$$

As the support of K is included in $[-1, 1]$, if

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

then

$$d_T(x) = \frac{1}{M_T} \prod_{l=1}^{d} K(\frac{x_l - x_0^l}{h_l(T)}) = 0.$$

It implies that, on K_T^c, $\pi_0(x) = \pi_1(x)$ which, together with (61), provides us

$$b_{\pi_1}^i(x) = \frac{1}{\pi_0(x)} \tilde{I}^i[\pi_0](x) + \frac{1}{\pi_0(x)} \tilde{I}^i[d_T](x) = b_{\pi_0}^i(x) + \frac{1}{\pi_0(x)} \tilde{I}^i[d_T](x). \qquad (62)$$

It follows that the first point of proposition will be proven as soon as we show that, for

$$x \in K_T^c, \qquad |\frac{1}{\pi_0} \tilde{I}^i[d_T](x)| \leq \frac{c}{M_T}.$$

We start considering $\tilde{I}_1^i[d_T](x)$:

$$\forall x \in K_T^c \qquad \tilde{I}_1^i[d_T](x) = \frac{1}{2}\sum_{j=1}^d (a \cdot a^T)_{ij} \frac{\partial d_T}{\partial x_j}(x) = 0. \tag{63}$$

Indeed, by its definition, the function d_T and all its derivatives vanish outside of the compact set K_T. Regarding $\tilde{I}_2^i[d_T](x)$, by the form of $A_{d,i}^*$ as defined in (36) it is

$$\tilde{I}_2^i[d_T](x) = \int_{-\infty}^{x_i} [\int_{\mathbb{R}^d} \left(d_T(\bar{w}_i) - d_T(\bar{w}_{i-1}) + (\gamma \cdot z)_i \frac{\partial}{\partial x_i} d_T(w_i)\right) F(z) dz] dw \tag{64}$$

$$= \tilde{I}_{2,1}^i[d_T](x) + \tilde{I}_{2,2}^i[d_T](x) + \tilde{I}_{2,3}^i[d_T](x),$$

where it is $\bar{w}_i = (x_1 - (\gamma \cdot z)_1, ..., x_{i-1} - (\gamma \cdot z)_{i-1}, w - (\gamma \cdot z)_i, x_{i+1}, ..., x_d)$, $\bar{w}_{i-1} = (x_1 - (\gamma \cdot z)_1, ..., x_{i-1} - (\gamma \cdot z)_{i-1}, w, x_{i+1}, ..., x_d)$ and $w_i = (x_1, ..., x_{i-1}, w, x_{i+1}, ..., x_d)$.

We are going to show that for any $x \in K_T^c$, $\tilde{I}_{2,3}^i[d_T](x) = 0$ while $|\frac{1}{\pi_0}\tilde{I}_{2,1}^i[d_T](x)| + |\frac{1}{\pi_0}\tilde{I}_{2,2}^i[d_T](x)| \leq \frac{c}{M_T}$. By the definition (45) of d_T and the fact that its support is included in K_T it is, for any $x \in \mathbb{R}^d$,

$$\tilde{I}_{2,1}^i[d_T](x)$$

$$= \frac{1}{M_T} \int_{-\infty}^{x_i} \int_{\mathbb{R}^d} \prod_{l<i} K(\frac{x_l - (\gamma \cdot z)_l - x_0^l}{h_l(T)}) K(\frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)})$$

$$\times \prod_{l>i} K(\frac{x_l - x_0^l}{h_l(T)}) F(z) dz dw$$

$$= \frac{1}{M_T} \int_{-\infty}^{x_i} \int_{\{z: (\bar{w}_i - \gamma \cdot z) \in K_T\}} \prod_{l<i} K(\frac{x_l - (\gamma \cdot z)_l - x_0^l}{h_l(T)}) K(\frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)})$$

$$\times \prod_{l>i} K(\frac{x_l - x_0^l}{h_l(T)}) F(z) dz dw. \tag{65}$$

With the purpose to use Fubini theorem, we analyse more in detail the condition $(\bar{w}_i - \gamma \cdot z) \in K_T$. It means that,

$$\forall j < i, \quad x_0^j - h_j(T) \leq x_j - (\gamma \cdot z)_j \leq x_0^j + h_j(T),$$

$$\forall j > i, \quad x_0^j - h_j(T) \leq x_j \leq x_0^j + h_j(T), \tag{66}$$

and for $j = i$

$$x_0^i - h_i(T) \leq w - (\gamma \cdot z)_i \leq x_0^i + h_i(T), \tag{67}$$

which gives us

$$x_0^i - h_i(T) + (\gamma \cdot z)_i \leq w \leq x_0^i + h_i(T) + (\gamma \cdot z)_i.$$

Moreover, for $-\infty < w < x_i$, (67) also provides

$$-\infty < w - x_0^i - h_i(T) \leq (\gamma \cdot z)_i \leq w - x_0^i + h_i(T) < x_i - x_0^i + h_i(T).$$

We define the set

$$G_z^i(x) := \Big\{ z \in \mathbb{R}^d : \quad \forall j < i \quad x_j - x_0^j - h_j(T) \leq (\gamma \cdot z)_j \leq x_j - x_0^j + h_j(T),$$
$$-\infty < (\gamma \cdot z)_i < x_i - x_0^i + h_i(T) \text{ and,} \quad \forall j > i, \ -\infty < (\gamma \cdot z)_j < \infty \Big\}.$$

The use of Fubini theorem on (65) provides us

$$\frac{\prod_{l>i} K(\frac{x_l - x_0^l}{h_l(T)})}{M_T} \int_{z \in G_z^i(x)} \prod_{l<i} K(\frac{x_l - (\gamma \cdot z)_l - x_0^l}{h_l(T)})$$
$$\times (\int_{x_0^i - h_i(T) + (\gamma \cdot z)_i}^{x_i} K(\frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)}) dw) F(z) dz. \quad (68)$$

We observe that the supremum value in the innermost integral should have been $\min(x_i, \ x_0^i + h_i(T) + (\gamma \cdot z)_i)$ but when $x_i > x_0^i + h_i(T) + (\gamma \cdot z)_i$, the integral is

$$\int_{x_0^i - h_i(T) + (\gamma \cdot z)_i}^{x_0^i + h_i(T) + (\gamma \cdot z)_i} K(\frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)}) dw = h_i(T) \int_{-1}^{1} K(u) du = 0,$$

where we have used the change of variable $u := \frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)}$ and the property (41) of the kernel function K. When $x_i < x_0^i + h_i(T) + (\gamma \cdot z)_i$ it is $(\gamma \cdot z)_i > x_i - x_0^i - h_i(T)$. We can introduce such a constraint in the set G_z^i, which becomes

$$\tilde{G}_z^i(x) := \Big\{ z \in \mathbb{R}^d : \forall j \leq i \quad x_j - x_0^j - h_j(T) \leq (\gamma \cdot z)_j \leq x_j - x_0^j + h_j(T),$$
$$\forall j > i \quad -\infty < (\gamma \cdot z)_j < \infty \Big\}.$$

We now observe that, defining

$$\mathcal{J} := \Big\{ j \in \{1, ..., d\} : \quad x_j \notin [x_0^j - h_j(T), x_0^j + h_j(T)] \Big\}, \quad (69)$$

and, remarking that for $k \notin \mathcal{J}$ the density $\pi_k(x_k)$ is lower bounded away from zero, it is

$$\frac{1}{\pi_0(x)} \leq \frac{c}{\prod_{j \in \mathcal{J}}^d \pi_j(x_j)} = c e^{\eta \sum_{j \in \mathcal{J}} (a \cdot a^T)_{jj}^{-1} |x_j|}. \quad (70)$$

We recall that $\mathcal{J}$ is not empty as $x \in K_T^c$ but if $j > i$ is such that $j \in \mathcal{J}$, then $\tilde{I}_{2,1}^i[d_T](x) = 0$ as a consequence of (66), which derives from the definition of d_T and the properties of the kernel function K. Moreover, from the definition of $\tilde{G}_z^i(x)$, if $j \leq i$ is such that $j \in \mathcal{J}$, then it must be $x_j \in [(\gamma \cdot z)_j + x_0^j - h_J(T), (\gamma \cdot z)_j + x_0^j + h_j(T)]$. Therefore,

$$e^{\eta(a \cdot a^T)_{jj}^{-1} |x_j|} \leq e^{\eta(a \cdot a^T)_{jj}^{-1} \max(|(\gamma \cdot z)_j + x_0^j - h_J(T)|, |(\gamma \cdot z)_j + x_0^j + h_j(T)|)}$$

$$\leq c\,e^{\eta(a\cdot a^T)_{jj}^{-1}|(\gamma\cdot z)_j|}\leq c\,e^{\eta(a\cdot a^T)_{jj}^{-1}|\gamma||z|}.$$

We define $\bar{j}:=\arg\max_{j\in\mathcal{J}}(a\cdot a^T)_{jj}^{-1}$ and so we get

$$\frac{1}{\pi_0(x)}\leq ce^{\eta|\mathcal{J}|(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma||z|}\leq ce^{\eta d(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma||z|}.$$

Using also the form of $\tilde{I}_{2,1}^i[d_T](x)$ given by (68), but on $\tilde{G}_z^i(x)$ instead of on $G_z^i(x)$ as now we are considering $x_i < x_0^i + h_i(T) + (\gamma\cdot z)_i$, we get

$$|\frac{1}{\pi_0(x)}\tilde{I}_{2,1}^i[d_T](x)|$$

$$\leq \frac{c\,\|K\|_\infty^d}{M_T}\int_{z\in\tilde{G}_z^i(x)}e^{\eta d(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma||z|}|\int_{x_0^i-h_i(T)+(\gamma\cdot z)_i}^{x_i}dw|F(z)dz$$

$$\leq \frac{c}{M_T}\int_{z\in\tilde{G}_z^i}e^{\eta d(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma||z|}|x_i - x_0^i + h_i(T) - (\gamma\cdot z)_i|F(z)dz$$

$$\leq \frac{c}{M_T}\int_{z\in\tilde{G}_z^i(x)}e^{\eta d(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma||z|}(1+|h_i(T)|+|(\gamma\cdot z)_i|)F(z)dz.$$

The integral in z here above is upper bounded. Indeed, we enlarge the integration domain from $\tilde{G}_z^i(x)$ to $\mathbb{R}^d$ and we use the fifth point of A3, taking from the beginning η small enough to guarantee $\eta d(a\cdot a^T)_{\bar{j}\bar{j}}^{-1}|\gamma|\leq\epsilon_0$. It follows

$$\forall x\in K_T^c\qquad |\frac{1}{\pi_0(x)}\tilde{I}_{2,1}^i[d_T](x)|\leq\frac{c}{M_T}. \tag{71}$$

In order to prove the first point of the proposition we are left to evaluate $\tilde{I}_{2,2}^i[d_T](x)$ and $\tilde{I}_{2,3}^i[d_T](x)$ on K_T^c. On $\tilde{I}_{2,2}^i[d_T](x)$ a reasoning analogous to the one on $\tilde{I}_{2,1}^i[d_T](x)$ applies. It is

$$\tilde{I}_{2,2}^i[d_T](x)=\frac{1}{M_T}\int_{-\infty}^{x_i}\int_{\{z:\,(\bar{w}_{i-1}-\gamma\cdot z)\in K_T\}}\prod_{l<i}K(\frac{x_l-(\gamma\cdot z)_l-x_0^l}{h_l(T)})K(\frac{w-x_0^i}{h_i(T)})$$

$$\times\prod_{l>i}K(\frac{x_l-x_0^l}{h_l(T)})F(z)dzdw.$$

We can now act on $\tilde{I}_{2,2}^i[d_T](x)$ as we did above, on $\tilde{I}_{2,1}^i[d_T](x)$. To use Fubini theorem the only difference is that in this case (67) becomes

$$x_0^i - h_i(T)\leq w\leq x_0^i + h_i(T). \tag{72}$$

Defining

$$G_{2,z}^i(x):=\Big\{z\in\mathbb{R}^d:\forall j<i\;\;x_j-x_0^j-h_j(T)\leq(\gamma\cdot z)_j\leq x_j-x_0^j+h_j(T),$$

$$\forall j\geq i\;\;-\infty<(\gamma\cdot z)_j<\infty\Big\},$$

$$\tag{73}$$

Fubini theorem provides $\tilde{I}^i_{2,2}[d_T](x)$ is equal to

$$\frac{\prod_{l>i} K(\frac{x_l - x_0^l}{h_l(T)})}{M_T} \int_{z \in G^i_{2,z}(x)} \prod_{l<i} K(\frac{x_l - (\gamma \cdot z)_l - x_0^l}{h_l(T)})$$
$$\times (\int_{x_0^i - h_i(T)}^{x_i} K(\frac{w - x_0^i}{h_i(T)})dw)F(z)dz. \quad (74)$$

We remark first of all that, if $x_i < x_0^i - h_i(T)$, then the innermost integral is empty. Again, we observe that the supremum value in the integral should have been $\min(x_i, x_0^i + h_i(T))$ but when $x_i > x_0^i + h_i(T)$, the integral is

$$\int_{x_0^i - h_i(T)}^{x_0^i + h_i(T)} K(\frac{w - x_0^i}{h_i(T)})dw = h_i(T) \int_{-1}^{1} K(u)du = 0, \quad (75)$$

using the change of variable $u := \frac{w - x_0^i}{h_i(T)}$ and the property (41) of the kernel function K. We want to evaluate $|\frac{1}{\pi_0(x)}\tilde{I}^i_{2,2}[d_T](x)|$ and so we need to consider once again the set $\mathcal{J}$, as defined in (69). We observe that (70) still holds and that, as a consequence of (66) and of (75), if $j \leq i$ is such that $j \in \mathcal{J}$, then $\tilde{I}^i_{2,2}[d_T](x) = 0$. We are left to study the case in which $j > i$ is such that $j \in \mathcal{J}$. By the definition of $G^i_{2,z}(x)$ it must be $x_j \in [(\gamma \cdot z)_j + x_0^j - h_J(T), (\gamma \cdot z)_j + x_0^j + h_j(T)]$. Therefore, acting exactly as we did in order to prove (71), we easily get

$$|\frac{1}{\pi_0(x)}\tilde{I}^i_{2,2}[d_T](x)| \leq \frac{c}{M_T} \int_{z \in G^i_{2,z}(x)} e^{\eta d(a \cdot a^T)^{-1}_{\bar{j}\bar{j}}|\gamma||z|}(1 + |h_i(T)|)F(z)dz.$$

It follows

$$\forall x \in K_T^c \qquad |\frac{1}{\pi_0(x)}\tilde{I}^i_{2,2}[d_T](x)| \leq \frac{c}{M_T}. \quad (76)$$

Regarding $\tilde{I}^i_{2,3}[d_T](x)$, we want to show it is null for $x \in K_T^c$. As $x \in K_T^c$, there must be a $j \in \{1, ..., d\}$ for which

$$x_j \notin [x_0^j - h_j(T), x_0^j + h_j(T)].$$

If $j \neq i$ then, as $w_i = (x_1, ..., w, ..., x_d)$, it clearly follows that also w_i does not belong to K_T. Therefore, since the function d_T and its derivatives vanish outside the compact set K_T, it yields $\tilde{I}^i_{2,3}[d_T](x) = 0$.

If otherwise $j = i$, then we have to distinguish two different cases: $x_i < x_0^i - h_i(T)$ and $x_i > x_0^i + h_i(T)$. In the first case, as $w \in (-\infty, x_i)$, we have that also w is always less than $x_0^i - h_i(T)$ and so, again, $w_i \notin K_T$. It implies $\tilde{I}^i_{2,3}[d_T](x) = 0$. We are left to study the case in which $x_i > x_0^i + h_i(T)$. We observe that the set $[x_0^i - h_i(T), x_0^i + h_i(T)]$ is now necessarily included in $(-\infty, x_i)$ and outside it the function d_T and its derivatives are null. Therefore, we can in this case see

$\tilde{I}^i_{2,3}[d_T](x)$ as

$$\frac{1}{M_T}\int_{x_0^i-h_i(T)}^{x_0^i+h_i(T)}\int_{\mathbb{R}^d}(\gamma\cdot z)_i\frac{\partial}{\partial x_i}(\prod_{l\neq i}K(\frac{x_l-x_0^l}{h_l(T)})K(\frac{x_i-x_0^i}{h_i(T)}))\big|_{x_i=w}F(z)dzdw.$$
(77)

We observe that

$$\frac{\partial}{\partial x_i}(\prod_{l\neq i}K(\frac{x_l-x_0^l}{h_l(T)})K(\frac{x_i-x_0^i}{h_i(T)}))\big|_{x_i=w}=\frac{1}{h_i(T)}\prod_{l\neq i}K(\frac{x_l-x_0^l}{h_l(T)})K'(\frac{w-x_0^i}{h_i(T)}).$$

Hence, replacing it in (77), we get $\tilde{I}^i_{2,3}[d_T](x)$ is equal to

$$\frac{1}{M_T}\prod_{l\neq i}K(\frac{x_l-x_0^l}{h_l(T)})(\int_{x_0^i-h_i(T)}^{x_0^i+h_i(T)}\frac{1}{h_i(T)}K'(\frac{w-x_0^i}{h_i(T)})dw)\int_{\mathbb{R}^d}(\gamma\cdot z)_iF(z)dz. \quad (78)$$

With the change of variable $u:=\frac{w-x_0^i}{h_i(T)}$ we obtain

$$\int_{x_0^i-h_i(T)}^{x_0^i+h_i(T)}\frac{1}{h_i(T)}K'(\frac{w-x_0^i}{h_i(T)})dw=\int_{-1}^1 K'(u)du=K(1)-K(-1)=0,$$

the last being null as K is a C^∞ function whose support is included in $[-1,1]$. Replacing the last equation in (78), it yields

$$\forall x\in K_T^c\qquad\tilde{I}^i_{2,3}[d_T](x)=0.\tag{79}$$

From (63), (71), (76) and (79) it follows that, for any $x\in K_T^c$, $|\frac{1}{\pi_0(x)}\tilde{I}^i[d_T](x)|\leq\frac{c}{M_T}$ and so, as a consequence of (62), $|b^i_{\pi_1}(x)-b^i_{\pi_0}(x)|\leq\frac{c}{M_T}$.

Proof point 2
From (63) and (79), it turns out that our goal is to show that

$$\int_{K_T^c}|\tilde{I}^i_{2,1}[d_T](x)+\tilde{I}^i_{2,2}[d_T](x)|dx\leq\frac{c}{M_T}\prod_{l=1}^d h_l(T).\tag{80}$$

We now recall that, through Fubini theorem, $\tilde{I}^i_{2,1}[d_T](x)$ can be seen as in (68):

$$\int_{K_T^c}|\tilde{I}^i_{2,1}[d_T](x)|dx$$
$$\leq\int_{K_T^c}\frac{\|K\|_\infty^{d-i}}{M_T}\int_{z\in\tilde{G}_z^i(x)}|\prod_{l<i}K(\frac{x_l-(\gamma\cdot z)_l-x_0^l}{h_l(T)})|$$
$$\times(\int_{x_0^i-h_i(T)+(\gamma\cdot z)_i}^{x_i}|K(\frac{w-(\gamma\cdot z)_i-x_0^i}{h_i(T)})|dw)F(z)dz\,dx,$$

where $\tilde{G}_z^i(x)$ as defined below (68):

$$\tilde{G}_z^i(x) := \Big\{ z \in \mathbb{R}^d : \forall j \leq i \quad x_j - x_0^j - h_j(T) \leq (\gamma \cdot z)_j \leq x_j - x_0^j + h_j(T),$$
$$\forall j > i \quad -\infty < (\gamma \cdot z)_j < \infty \Big\}.$$

We know moreover from (66) that, for $j > i$, $x_0^j - h_j(T) \leq x_j \leq x_0^j + h_j(T)$. We observe first of all that

$$\int_{x_0^i - h_i(T) + (\gamma \cdot z)_i}^{x_i} |K(\frac{w - (\gamma \cdot z)_i - x_0^i}{h_i(T)})| dw \leq h_i(T) \int_{\mathbb{R}} |K(u)| du \leq c.$$

Therefore, using also Fubini theorem once again, we get

$$\int_{K_T^c} |\tilde{I}_{2,1}^i[d_T](x)| dx \leq c \int_{\mathbb{R}^d} \int_{x \in G_x(z)} dx F(z) dz,$$

where the set $G_x(z)$ derives from $\tilde{G}_z^i(x)$ and from (66) directly, writing the constraint on the components of x instead of on the components of z:

$$G_x(z) := \Big\{ x \in \mathbb{R}^d : \forall j \leq i \quad (\gamma \cdot z)_j + x_0^j - h_j(T) \leq x_j \leq (\gamma \cdot z)_j + x_0^j + h_j(T),$$
$$\forall j > i \quad x_0^j - h_j(T) \leq x_j \leq x_0^j + h_j(T) \Big\}.$$

Clearly, by its definition, $|G_x(z)| \leq c \prod_{j=1}^d h_j(T)$ and so, as the jump intensity is finite, it follows

$$\int_{K_T^c} |\tilde{I}_{2,1}^i[d_T](x)| dx \leq \frac{c}{M_T} \prod_{j=1}^d h_j(T),$$

as we wanted.
We act in the same way on $\tilde{I}_{2,2}^i[d_T](x)$, remarking that from (74) it is

$$\int_{K_T^c} |\tilde{I}_{2,2}^i[d_T](x)| dx$$
$$\leq \int_{K_T^c} \frac{\|K\|_\infty^{d-i}}{M_T} \int_{z \in G_{2,z}^i(x)} |\prod_{l < i} K(\frac{x_l - (\gamma \cdot z)_l - x_0^l}{h_l(T)})|$$
$$\times (\int_{x_0^i - h_i(T)}^{x_i} |K(\frac{w - x_0^i}{h_i(T)})| dw) F(z) dz \, dx,$$

with $G_{2,z}^i(x)$ as in (73). As before, the integral in dw is bounded and we can apply once again Fubini theorem, getting

$$\int_{K_T^c} |\tilde{I}_{2,2}^i[d_T](x)| dx \leq c \int_{\mathbb{R}^d} \int_{x \in G_{2,x}(z)} dx F(z) dz,$$

where

$$G_{2,x}(z) := \left\{ x \in \mathbb{R}^d : \forall j < i \quad (\gamma \cdot z)_j + x_0^j - h_j(T) \le x_j \le (\gamma \cdot z)_j + x_0^j + h_j(T), \right.$$

$$\left. \forall j \ge i \quad x_0^j - h_j(T) \le x_j \le x_0^j + h_j(T) \right\}.$$

The set $G_{2,x}(z)$ derives from $G_{2,z}^i(x)$, writing the constraint on the components of x instead of on the components of z, and from the fact that $x_0^j - h_j(T) \le x_j \le x_0^j + h_j(T)$ for $j \ge i$, as a consequence of (66), (72) and (75). By its definition, it is $|G_{2,x}(z)| \le c \prod_{j=1}^d h_j(T)$ and so, as the jump intensity is finite, it follows

$$\int_{K_T^c} |\tilde{I}_{2,2}^i[d_T](x)| dx \le \frac{c}{M_T} \prod_{j=1}^d h_j(T).$$

It implies (80).

Proof Point 3

We now want to investigate how different b_{π_0} and b_{π_1} are on the compact set K_T. From (61) we obtain

$$b_{\pi_1}^i - b_{\pi_0}^i = \left(\frac{1}{\pi_1} - \frac{1}{\pi_0}\right)\tilde{I}^i[\pi_0] + \frac{1}{\pi_1}\tilde{I}^i[d_T]$$

$$= \frac{\pi_0 - \pi_1}{\pi_1}\frac{1}{\pi_0}\tilde{I}^i[\pi_0] + \frac{1}{\pi_1}\tilde{I}^i[d_T] = \frac{d_T}{\pi_1}b_{\pi_0}^i + \frac{1}{\pi_1}\tilde{I}^i[d_T].$$

We have to evaluate such a difference on the compact set K_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 from Lemma 2 that π_0 satisfies Assumption Ad and so, using Proposition 3, b_{π_0} is bounded. Furthermore, as d_T has been defined as in (45), we get

$$\|d_T\|_\infty \le \frac{c}{M_T}. \tag{81}$$

Hence, we deduce that

$$\forall x \in K_T \qquad |b_{\pi_1}^i - b_{\pi_0}^i| \le c\left(\frac{1}{M_T} + \tilde{I}^i[d_T](x)\right). \tag{82}$$

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

$$\left\|\frac{\partial d_T}{\partial x_j}\right\|_\infty \le \frac{c}{M_T}\frac{1}{h_j(T)}, \tag{83}$$

it clearly follows

$$\tilde{I}_1^i[d_T](x) \le \frac{c}{M_T}\sum_{j=1}^d \frac{1}{h_j(T)}. \tag{84}$$

Regarding $\tilde{I}_2^i[d_T](x)$, according with (64) we see it as the sum of $\tilde{I}_{2,1}^i[d_T](x)$, $\tilde{I}_{2,2}^i[d_T](x)$ and $\tilde{I}_{2,3}^i[d_T](x)$. As $x \in K_T$, $x_i \in [x_0^i - h_i(T), x_0^i + h_i(T)]$. Therefore, using also the definition of d_T as function of K, the first integral should be between $x_0^i - h_i(T)$ and x_i. We enlarge the domain of integration to $[x_0^i - h_i(T), x_0^i + h_i(T)]$ and so, using also (81) and (83), we get

$$|\tilde{I}_2^i[d_T](x)| \leq \int_{x_0^i - h_i(T)}^{x_0^i + h_i(T)} \int_{\mathbb{R}^d} |d_T(\tilde{w}_i) - d_T(\tilde{w}_{i-1}) + (\gamma \cdot z)_i \frac{\partial}{\partial x_i} d_T(w_i)| F(z) dz dw$$

$$\leq 2\lambda \int_{x_0^i - h_i(T)}^{x_0^i + h_i(T)} \|d_T\|_\infty \, dw + \int_{x_0^i - h_i(T)}^{x_0^i + h_i(T)} \int_{\mathbb{R}^d} |(\gamma \cdot z)_i| \left\| \frac{\partial d_T}{\partial x_i} \right\|_\infty F(z) dz dw$$

$$\leq \frac{ch_i(T)}{M_T} + \frac{ch_i(T)}{M_T} \frac{1}{h_i(T)}. \tag{85}$$

Replacing (84) and (85) in (82) we obtain that, for any $x \in K_T$,

$$|b_{\pi_1}^i - b_{\pi_0}^i| \leq \frac{c}{M_T}(1 + \sum_{j=1}^{d} \frac{1}{h_j(T)} + h_i(T) + 1) \leq \frac{c}{M_T} \sum_{j=1}^{d} \frac{1}{h_j(T)},$$

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. □

7.5. Proof of Lemma 4

Proof. Point 1

The absolute continuity $\mathbb{P}_{b_{\pi_1}}^{(T)} \ll \mathbb{P}_{b_{\pi_0}}^{(T)}$ and the expression for $Z^{(T)} := \frac{d\mathbb{P}_{b_{\pi_1}}^{(T)}}{d\mathbb{P}_{b_{\pi_0}}^{(T)}}$ are both obtained by Girsanov formula, changing the drift b_{π_0} of $X^{(0)}$ solution of (40) to the drift b_{π_1}, appearing in $X^{(1)}$ solution of

$$X_t^{(1)} = X_0^{(1)} + \int_0^t b_{\pi_1}(X_s^{(1)}) ds + \int_0^t a \, dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma \, z \, \tilde{\mu}(ds, dz).$$

It is

$$Z^{(T)} := \frac{d\mathbb{P}_{b_{\pi_1}}^{(T)}}{d\mathbb{P}_{b_{\pi_0}}^{(T)}} ((X_s)_{0 \leq s \leq T})$$

$$= \frac{\pi_1}{\pi_0}(X^{(0)}) \exp \Big[\int_0^T (b_{\pi_1}^t(X_s) - b_{\pi_0}^t(X_s))(a^t a)^{-1} dX_s \tag{86}$$

$$- \frac{1}{2} \int_0^T [b_{\pi_1}^t(X_s)(a^t a)^{-1} b_{\pi_1}(X_s) - b_{\pi_0}^t(X_s)(a^t a)^{-1} b_{\pi_0}(X_s)] ds \Big].$$

An analogous computation for diffusion processes, in absence of jumps and for $d = 1$, can be found in Theorem 1.12 of [23]. In our situation we have the

absolute continuity $\mathbb{P}_{b_{\pi_1}}^{(T)} \ll \mathbb{P}_{b_{\pi_0}}^{(T)}$ and the expression for $Z^{(T)}$ as above thanks to Theorems III.5.19 and IV.4.39 in [21].

We underline the fact that in the expression of $Z^{(T)}$ there is the ratio $\frac{\pi_1}{\pi_0}(X^{(0)})$ because the two diffusions $(X^{(0)})_t$ and $(X^{(1)})_t$ are both stationary, but with different stationary laws.

Point 2

We aim at controlling by below the quantity $\mathbb{P}_{b_{\pi_0}}^{(T)}(Z^{(T)} \geq \frac{1}{\lambda})$, for $\lambda > 0$. We remark that, by the definitions of π_0 and π_1 given in Section 6.2, the ratio $\frac{\pi_1}{\pi_0}$ is equal to 1 outside some compact set that can be chosen independent of T. On the compact set, instead, it converges uniformly to 1. The ratio is therefore bounded away from 0 if T is large and so we have

$$\frac{\pi_1}{\pi_0}(X^{(0)}) \geq c > 0.$$

We therefore focus on the exponential part in (86), that we denote as $\mathcal{E}^{(T)}$. Since under the law $\mathbb{P}_{b_{\pi_0}}^{(T)}$ the process $(X_t)_t$ has the same law as $(X_t^{(0)})_t$, solution of the stochastic differential equation proposed in (40), the law of $\log(\mathcal{E}^{(T)})$ is, after having replaced the dynamic of X, the law of the random variable

$$\frac{1}{2} \int_0^T < (b_{\pi_1}^t(X_s^{(0)}) - b_{\pi_0}^t(X_s^{(0)}))a^{-1}, (b_{\pi_1}^t(X_s^{(0)}) - b_{\pi_0}^t(X_s^{(0)}))a^{-1} > ds$$

$$+ \int_0^T (b_{\pi_1}^t(X_s^{(0)}) - b_{\pi_0}^t(X_s^{(0)}))a^{-1} dW_s$$

$$=: I_T + M_T.$$

Hence we can write that, for T large enough,

$$\mathbb{P}_{b_{\pi_0}}^{(T)}(Z^{(T)} \geq \frac{1}{\lambda}) \geq \mathbb{P}_{b_{\pi_0}}^{(T)}(\mathcal{E}^{(T)} \geq \frac{1}{c\lambda}) = \mathbb{P}_{b_{\pi_0}}^{(T)}(-\log(\mathcal{E}^{(T)}) \leq \log(c\lambda))$$

$$\geq 1 - \mathbb{P}_{b_{\pi_0}}^{(T)}(|\log(\mathcal{E}^{(T)})| > \log(c\lambda)) = 1 - \mathbb{P}(|M_T| + |I_T| > \log(c\lambda)),$$

where in the last equality we have used that, as explained here above, the law of $\log(\mathcal{E}^{(T)})$ under $\mathbb{P}_{b_{\pi_0}}^{(T)}$ is the law of $M_T + I_T$. We now assume that $\lambda > \frac{1}{c}$, such as $\lambda c > 1$. From Markov inequality it follows

$$\mathbb{P}(|M_T| + |I_T| > \log(c\lambda)) \leq \mathbb{P}(|M_T| > \frac{1}{2} \log(c\lambda)) + \mathbb{P}(|I_T| > \frac{1}{2} \log(c\lambda)) \leq$$

$$\leq \frac{4}{(\log(c\lambda))^2} \mathbb{E}[M_T^2] + \frac{2}{\log(c\lambda)} \mathbb{E}[|I_T|] = (\frac{8}{(\log(c\lambda))^2} + \frac{2}{\log(c\lambda)}) \mathbb{E}[I_T],$$

where the last equality is a consequence of the positivity of I_T and of Ito's isometry, which gives us $\mathbb{E}[M_T^2] = 2\mathbb{E}[I_T]$. It remains to evaluate $\mathbb{E}[I_T]$. To do that we remark that

$$|< (b_{\pi_1}^t(X_s^{(0)}) - b_{\pi_0}^t(X_s^{(0)}))a^{-1}, (b_{\pi_1}^t(X_s^{(0)}) - b_{\pi_0}^t(X_s^{(0)}))a^{-1} > |$$

$$\leq \left\|a^{-1}\right\|_{op}^2 \left\|b_{\pi_1}(X_s^{(0)}) - b_{\pi_0}(X_s^{(0)})\right\|_{L^2}^2,$$

where we have introduced the operator norm $\|\cdot\|_{op}$. Then, as the process $(X_t^{(0)})_t$ is stationary with invariant law π_0, it is

$$\mathbb{E}[I_T] \leq T \frac{\left\|a^{-1}\right\|_{op}^2}{2} \int_{\mathbb{R}^d} |b_{\pi_1}(x) - b_{\pi_0}(x)|^2 \pi_0(x)dx.$$

From the assumption (48) in the statement of the lemma, it follows that

$$\sup_{T \geq 0} \mathbb{E}[I_T] < \infty,$$

which is sufficient to ensure that there exists λ_0 such that, for any T large enough,

$$\mathbb{P}_{b_{\pi_0}}^{(T)}(Z^{(T)} \geq \frac{1}{\lambda_0}) \geq \frac{1}{2},$$

as we wanted. $\qquad\qquad\square$

Acknowledgments

References

[1] Applebaum, David. (2009). Lévy processes and stochastic calculus. Cambridge university press. MR2512800

[2] 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. MR4184676

[3] 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. MR4140022

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

[5] Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc., Ser. B, Stat. Methodol., 63, 167-241. MR1841412

[6] Bates, D.S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark. The Review of Financial Studies, 9(1), 69-107.

[7] Chen, Z. Q., Hu, E., Xie, L., & Zhang, X. (2017). Heat kernels for non-symmetric diffusion operators with jumps. Journal of Differential Equations, 263(10), 6576-6634. MR3693184

[8] Comte, F., Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. In Annales de l'IHP Probabilités et statistiques (Vol. 49, No. 2, pp. 569-609). MR3088382

[9] Comte F, and Merlevède, F. (2001). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM Probab. Statist., 6:211238, 2002. New directions in time series analysis. MR1943148

[10] Comte, F. and Merlèvede, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. Stochastic processes and their applications, 115(5), pp.797-826. MR2132599

[11] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probab. Theory Relat. Fields, 137(1), 25–47. MR2278451

[12] 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.

[13] Dion, C., Lemler, S. (2019). Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. Statistical Inference for Stochastic Processes, 1-27. MR4136702

[14] Ditlevsen, S. and Greenwood, P. (2013). The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. Journal of Mathematical Biology 67 239-259. MR3071157

[15] Eraker, B., Johannes, M. and Polson N. (2003). The Impact of Jumps in Volatility and Returns. J. Finance, 58(3), 1269-1300.

[16] Ethier, S.N. and Kurtz, T.G., (1986). Markov Processes: Characterization and Convergence. A John Wiley and Sons. Inc., Publication, 1. MR0838085

[17] Funke, B., Schmisser, E. (2018). Adaptive nonparametric drift estimation of an integrated jump diffusion process. ESAIM: Probability and Statistics, 22, 236-260. MR3903643

[18] Gloter, A., Honoré, I. and Loukianova, D., (2018). Non-asymptotic concentration inequality for an approximation of the invariant distribution of a diffusion driven by compound poisson process. MR4074695

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

[20] Honoré, I. and Menozzi, S., (2016). Non-asymptotic Gaussian estimates for the recursive approximation of the invariant measure of a diffusion. arXiv preprint arXiv:1605.08525. MR4116702

[21] Jacod, J., & Shiryaev, A. N. (1987). Limit theorems for stochastic processes. Springer. MR0959133

[22] Kou, S.G. (2002). A Jump-Diffusion Model for Option Pricing. Management Science, 48, 1086-1101.

[23] Kutoyants, Y. A. (2004). Statistical inference for ergodic diffusion processes. Springer Science & Business Media. MR2144185

[24] Lamberton, D. and Pages, G. (2002). Recursive computation of the invariant distribution of a diffusion. Bernoulli, 8(3), pp.367-405. MR1913112

[25] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3, 125-144.

[26] Nguyen, H-T. (1979). Density estimation in a continuous-time stationary Markov process. Ann. Statist., 7(2):341348. MR0520244

[27] Nikolskii, S. M. (1975). Approximation of Functions of Several Variables and Embedding Theorems (Springer, Berlin). MR0374877

[28] Nishiyama, Y. Estimation for the invariant law of an ergodic diffusion process based on high-frequency data, Journal of Nonparametric Statistics, 23:4, 909-915, DOI:10.1080/10485252.2011.591397 MR2854246

[29] 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. MR2398761

[30] Schmisser, E. (2013). Nonparametric estimation of the derivatives of the stationary density for stationary processes. ESAIM: Probability and Statistics, 17, 33-69. MR3002995

[31] Schmisser, E. (2019). Non parametric estimation of the diffusion coefficients of a diffusion with jumps. Stochastic Processes and their Applications, 129(12), 5364-5405. MR4025707

[32] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics, 46(6B), 3451-3480. MR3852658

[33] Tsybakov, A. B. (2008). Introduction to nonparametric estimation. Springer Science & Business Media. MR2724359