

Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes.

Chiara Amorino*, Arnaud Gloter*

November 20, 2020

Abstract

We consider the solution $X = (X_t)_{t \geq 0}$ of a multivariate stochastic differential equation with Levy-type jumps and with unique invariant probability measure with density μ . We assume that a continuous record of observations $X^T = (X_t)_{0 \leq t \leq T}$ is available.

In the case without jumps, Dalalyan and Reiss [9] and Strauch [28] have found convergence rates of invariant density estimators, under respectively isotropic and anisotropic Hölder smoothness constraints, which are considerably faster than those known from standard multivariate density estimation.

We extend the previous works by obtaining, in presence of jumps, some estimators which have the same convergence rates they had in the case without jumps for $d \geq 2$ and a rate which depends on the degree of the jumps in the one-dimensional setting.

We propose moreover a data driven bandwidth selection procedure based on the Goldenshluger and Lepski method [13] which leads us to an adaptive non-parametric kernel estimator of the stationary density μ of the jump diffusion X .

Adaptive bandwidth selection, anisotropic density estimation, ergodic diffusion with jumps, Lévy driven SDE

1 Introduction

Diffusion phenomena arise from a Markovian stochastic modeling and as a solution of SDEs with or without jumps in many areas of applied mathematics. Their investigation concerns different mathematical branches and therefore research interest in questions such as existence and regularity of solutions of stochastic differential equations has constantly grown over the past years.

The study of the statistical properties of diffusion models has emerged since such models are widely used for applications in finance and biology. Diffusion processes with jumps, in particular, have been used in neuroscience for instance in [10] while in finance they have been introduced to model the dynamic of asset prices [15], [22], exchange rates [3], or volatility processes [2].

In this work, we aim at estimating adaptively the invariant density μ associated to the process $(X_t)_{t \geq 0}$, solution of the following multivariate stochastic differential equation with Levy-type jumps:

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

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

Practical concerns raise new questions such as the dependence of statistical features on the observation scheme: it is, for the applications, a subject of interest to consider basic questions in different observation scenarios. From a theoretical point of view, it is however also of substantial interest to work under the assumption that a continuous record of the diffusion considered is available.

In this framework, it belongs to the folklore of the statistics for stochastic processes without jumps that the invariant density can be estimated under standard nonparametric assumptions with a

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

parametric rate (cfr Chapter 4.2 in [17]). The proof relies on the existence of diffusion local time and its properties and so such a result is restricted to the one - dimensional setting.

In the mono -dimensional case Schmisser considers, in [25], a strictly stationary and β mixing process $(X_t)_{t \geq 0}$ observed at discrete times $t = 0, \Delta_n, \dots, n\Delta_n$. The author estimates the successive derivatives $\mu^{(j)}$ of the stationary density either on a compact set or on \mathbb{R} thanks to a penalized least square method. If the derivative $\mu^{(j)}$ belongs to the Besov space $\mathcal{B}_{2,\infty}^\alpha$, then the L^2 risk of the estimator converges with rate $(n\Delta_n)^{\frac{-2\alpha}{2\alpha+2j+1}}$, and the procedure does not require the knowledge of α . When $j = 0$, the invariant density associated to the process $(X_t)_{t \geq 0}$ is estimated with a convergence rate which is the same found by Comte and Merlevède in [6] and [7].

Regarding the literature on statistical properties of multidimensional diffusion processes in the continuous case, an important reference is given by Dalalyan and Reiss in [9], where they show an asymptotic statistical equivalence for inference on the drift in the multidimensional diffusion case. As a by-product of the study they prove, under isotropic Hölder smoothness constraints, convergence rates of invariant density estimators for pointwise estimation which are faster than those known from standard multivariate density estimation. Their result relies on upper bounds on the variance of additive diffusion functionals, proven by an application of the spectral gap inequality in combination with a bound on the transition density of the process.

Still in the continuous case, in a recent paper, Strauch [28] has extended their work by building adaptive estimators in the multidimensional diffusion case which achieve fast rates of convergence over anisotropic Hölder balls.

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 [24] the author takes account of the developments). However, results on minimax rates of convergence in classical statistical models were rare for a lot of time.

The question of optimal bandwidth selection based on i. i. d. observations for density estimation with respect to sup - norm risk was not completely solved until the pretty recent developments gathered in [19]. The methodology detailed in Goldenshluger and Lepski [13] inspired the data-driven selection procedure of the bandwidth of the kernel estimator proposed by many authors such as Strauch in [28] and Comte, Priour and Samson in [8] and provides the starting point for the study of our adaptive procedure as well.

In presence of jumps, we are only aware of a few works which takes place in the non parametric framework. For example, Schmisser investigates in [26] the non parametric adaptive estimation of the coefficients of a jumps diffusion process and in [12], together with Funke, the non parametric adaptive estimation of the drift of an integrated jump diffusion process.

In this paper, we provide a non-parametric estimator of the invariant density μ with a fully data-driven procedure of the bandwidth. We propose to estimate the invariant density μ by means of a kernel estimator, we therefore introduce some kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$. A natural estimator of μ at $x \in \mathbb{R}^d$ in the anisotropic context is given by

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

where $h = (h_1, \dots, h_d)$ is a multi - index bandwidth, which will be chosen through the data-driven selection procedure. We first prove some bounds on the transition semigroup and on the transition density that will be useful to find sharp upper bounds on the variance of integral functionals of the diffusion X . Through them, we find the following convergence rates for the pointwise estimation of the invariant density of our diffusion with jumps:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \begin{cases} \frac{(\log T)^{(2 - \frac{(1+\alpha)}{2}) \vee 1}}{T} & \text{for } d = 1, \\ \frac{\log T}{T} & \text{for } d = 2, \\ T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} & \text{for } d \geq 3, \end{cases}$$

where $\alpha \in (0, 2)$ is the degree of jumps activity of the Lévy process and $\bar{\beta}$ is the harmonic mean smoothness of the invariant density over the d different dimensions.

We remark that the rate we find for $d \geq 3$ is the same Strauch found in [28] in absence of jumps, which is also the rate gathered in [9] up to replacing the mean smoothness with β , the common

smoothness over the d dimensions.

The case $d = 1$ evidences the main difference between what happens with and without jumps. Indeed, if in the continuous case the optimal convergence rate was $\frac{1}{T}$, now the rate we found is

between $\frac{\log T}{T}$ and $\frac{(\log T)^{\frac{3}{2}}}{T}$. It is worth noting here that such a convergence rate is not necessarily the optimal one in the jumps framework. As a matter of fact in the continuous case different approaches, as the diffusion local time, have been used to get the rate $\frac{1}{T}$; we do not exclude the possibility that also in presence of jumps the implementation of other methods could lead to a convergence rate faster than the one presented here above for the mono-dimensional setting.

To complete the comparison to the continuous framework, we recall that in both [9] and [28] the convergence rate found in the case $d = 2$ was $\frac{(\log T)^4}{T}$ and so the convergence of the estimator seems being faster in presence of jumps than without them. The reason why it happens is that, to find the convergence rate, the transition density $(p_t)_{t \in \mathbb{R}^+}$ is needed to be upper bounded. If in [9] the authors assume to have $p_t(x, y) \leq c(t^{-\frac{d}{2}} + t^{\frac{3d}{2}})$ and in [28] Nash and Poincaré inequalities lead Strauch to a bound analogous to the one presented in [9]; Lemma 1 below provides us a different bound which guides us to a different rate. However, in absence of the term $t^{\frac{3d}{2}}$ in the assumption before, which is the case for example considering a bounded drift, also in the continuous setting the convergence rate turns out being, as in the jump -diffusion case, equal to $\frac{\log T}{T}$.

It is moreover worth noting here that, if in [9] and [28] they needed to assume the existence of the transition density and a bound on it, we derive them through Lemma 1: all the assumptions we need are directly on the model (1).

We no longer need to assume that the drift is of the form $b = -\nabla V$ (where $V \in \mathcal{C}^2$ is referred to as potential) as it was in both [9] and [28].

After having provided the rates of convergence of the estimators we finally propose, in the case $d \geq 3$, a fully data-driven selection procedure of the bandwidth of the kernel estimator, inspired by the methodology detailed in Goldenshluger and Lepski [13]. The method has the decisive advantage of being anisotropic: the bandwidths selected in each direction are in general different, which is coherent with the possibly different regularities with respect to each variable. Finally, we prove that for the selected optimal bandwidth the following estimation holds:

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 \inf_{h \in \mathcal{H}_T} (B(h) + V(h)) + c_1 e^{-c_2 (\log T)^2}, \quad (2)$$

where we have denoted as $\|\cdot\|_A$ the L^2 norm on A , a compact subset of \mathbb{R}^d and as \mathcal{H}_T the set of candidate bandwidths; $B(h)$ is a bias term and $V(h)$ an estimate of the variance bound. We remark that the estimator leads to an automatic trade - off between the bias and the variance: the second term on the right hand side of (2) is indeed negligible compared to the first one.

Moreover, as the rate optimal choice $h(T)$ belongs to the set of candidate bandwidths \mathcal{H}_T , (2) turns out being

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2 (\log T)^2},$$

where $\bar{\beta}$ is the mean smoothness of the invariant density.

The paper is organised as follows. We give in Section 2 the assumptions on the process X . In Section 3 we define the anisotropic Hölder balls and we construct our estimator. Section 4 is devoted to the statements of our main results; which will be proven in the two following sections. In particular, we show how we get the convergence rates for the invariant density estimation in Section 5 while in Section 6 we prove the estimator we find through our bandwidth selection procedure achieves the previously introduced convergence rate. Some technical results are moreover presented in the Appendix.

2 Model Assumptions

We consider the question of nonparametric estimation of the invariant density of a d-dimensional diffusion process X , assuming that a continuous record $X^T = \{X_t, 0 \leq t \leq T\}$ up to time T is observed. This diffusion is given as a strong solution of the following stochastic differential equations with jumps:

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t a(X_s) dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_{s-}) z \tilde{\nu}(ds, dz), \quad t \in [0, T], \quad (3)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $a : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$; $W = (W_t, t \geq 0)$ is a d -dimensional Brownian motion and ν 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 $L_t := \int_0^t \int_{\mathbb{R}^d} z \tilde{\nu}(ds, dz)$. The compensated measure is $\tilde{\nu} = \nu - \bar{\nu}$; we suppose that the compensator has the following form: $\bar{\nu}(dt, dz) := F(dz)dt$, where conditions on the Lévy measure F will be given later.

The initial condition X_0 , W and L are independent. We denote $\tilde{a} := a \cdot a^T$.

In what follows, we suppose the following assumptions 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 $|\cdot|$ and $\langle \cdot, \cdot \rangle$ respectively the Euclidian norm and the scalar product in \mathbb{R}^d , we suppose moreover that there exists a constant $\bar{c} > 0$ such that, $\forall x \in \mathbb{R}^d$, $|b(x)| \leq \bar{c}$.

Under Assumption 1 the equation (3) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [1] (Theorems 6.2.9. and 6.4.6.).

We need the above assumption on the volatility and we need the drift to be bounded in order to prove a bound on the transition density (see Lemma 1), which will be useful in the sequel.

The next assumption involves the ergodicity of the process and it is needed in order to show the existence of a Lyapounov function.

A2 (Drift condition) :

There exist $\tilde{C} > 0$ and $\tilde{\rho} > 0$ such that $\langle x, b(x) \rangle \leq -\tilde{C}|x|$, $\forall x : |x| \geq \tilde{\rho}$.

We furthermore need the following 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 $c_1 > 0$ such that for all $z \in \mathbb{R}^d$, $F(z) \leq \frac{c_1}{|z|^{d+\alpha}}$, with $\alpha \in (0, 2)$ and that $\text{supp}(F) = \mathbb{R}^d$.

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 $c_2 > 0$ such that, $\forall x \in \mathbb{R}^d$, $\text{Det}(\gamma(x)) > c_2$.

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 $c_0 > 0$ such that $\int_{\mathbb{R}^d} |z|^2 e^{\epsilon_0 |z|} F(z) dz \leq c_0$.

As we will see in Lemma 2 below, Assumption 2 ensures, together with the last points of Assumption 3, the existence of a Lyapunov function. The process X admits therefore a unique invariant distribution π and the ergodic theorem holds.

In particular, the third point of Assumption 3 together with $\text{supp}(F) = \mathbb{R}^d$, as asked in the last part of point 2 of A3, ensures the irreducibility of the process. We require it, the drift condition A2 and the fifth point of A3 in order to show that our process is exponentially ergodic according to Definition 1, as proved in Lemma 2. The boundedness of the coefficient, as required in A1, and the fourth point of A3 are needed to show a bound on the transition density, as done in Lemma 1. The second point of A3 says that we are in a framework comparable to an α -stable process.

We assume the invariant probability measure π of X being absolutely continuous with respect to the Lebesgue measure and from now on we will denote its density as μ : $d\pi = \mu dx$.

For any set $S \subset \mathbb{R}^d$ we define $\mu(S) := \int_S \mu(x) dx$ and, by abuse of notation, we will write $\mu(f) := \mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x) \mu(x) dx$ for functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

We define moreover $L^2(\mu) := \{f : \mathbb{R}^d \rightarrow \mathbb{R} : \int_{\mathbb{R}^d} |f(x)|^2 \mu(x) dx < \infty\}$ and

$L^1(\mu) := \{f : \mathbb{R}^d \rightarrow \mathbb{R} : \int_{\mathbb{R}^d} |f(x)| \mu(x) dx < \infty\}$.

For each $g \in L^1(\mu)$ we denote as $\|g\|_{L^1(\mu)} := \mu(|g|)$ the L^1 norm with respect to μ on \mathbb{R}^d .

The transition semigroup of the process X on $L^1(\mu)$ is $P_t f(x) := \mathbb{E}[f(X_t) | X_0 = x]$.

The transition density is denoted by p_t and it is such that $P_t f(x) = \int_{\mathbb{R}^d} f(y) p_t(x, y) dy$; we will see in Lemma 1 that it exists.

The process X is called β - mixing if $\beta_X(t) = o(1)$ for $t \rightarrow \infty$ and exponentially β - mixing if there exists a constant $\gamma > 0$ such that $\beta_X(t) = O(e^{-\gamma t})$ for $t \rightarrow \infty$, where β_X is the β - mixing coefficient of the process X as defined in Section 1.3.2 of [11]. We recall that, for a Markov process X with transition semigroup $(P_t)_{t \in \mathbb{R}^+}$ and $\mathcal{L}(X_0) = \eta$, the β - mixing coefficient of X is given by

$$\beta_X(t) := \sup_{s \in \mathbb{R}^+} \int_{\mathbb{R}^d} \|P_t(x, \cdot) - \eta P_{s+t}(x, \cdot)\|_{TV} \eta P_s(dx, \cdot), \quad (4)$$

where $\eta P_t = \mathcal{L}(X_t)$ and $\|\lambda\|_{TV}$ stands for the total variation norm of a signed measure λ . For the exponential mixing property of general multidimensional diffusions, the reader may consult Theorem 3 of Kusuoka and Yoshida [16] for the α - mixing; Meyn and Tweedie [23], Stramer and Tweedie [27] and Veretennikov [30] for the β - mixing. The mixing property for general diffusions with jumps has been investigated by Masuda in [21]. Now we mention the notion of exponential ergodicity in the sense of [23].

Definition 1. *We say that X is exponentially ergodic if it admits a unique invariant distribution π and additionally if there exist positive constants c and ρ for which, for each f centered under $\mu = \frac{d\pi}{dx}$,*

$$\|P_t f\|_{L^1(\mu)} \leq c e^{-\rho t} \|f\|_{\infty}.$$

We will see in Lemma 2 that both the exponential ergodicity and the exponential β - mixing can be derived from our assumptions.

In Lemmas 2 and 1 below we will prove some bounds on the transition semigroup and on the transition density that will be useful to establish tight upper bounds on the variance

$$\text{Var}\left(\int_0^T f(X_s) ds\right), \quad f \in L^2(\mu)$$

of integral functionals of the diffusion X .

Bounds of this type were proven before, in [9] (cf. their Proposition 1), by combining estimates based on the spectral gap inequality and on upper bounds on the transition densities of X . Through them they prove, under isotropic Hölder smoothness constraints, convergence rates of invariant density estimators for pointwise estimation which are considerably faster than those known from standard multivariate density estimation.

We replace the spectral gap inequality with a control from L^1 to L^∞ given by the exponential ergodicity. Moreover, contrary to [9], we don't need to assume that such controls hold true since we get them as consequence of Lemma 1 and 2 below, having required some assumptions only directly on the model (3).

In the next section we will construct adaptive estimators for the density in the multidimensional diffusion case with jumps, which achieve fast rates of convergence over anisotropic Hölder balls.

3 Construction of the estimator

In several cases, the regularity of some function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ depends on the direction in \mathbb{R}^d chosen. We thus work under the following anisotropic smoothness constraints.

Definition 2. *Let $\beta = (\beta_1, \dots, \beta_d)$, $\beta_i \geq 1$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$. A function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ of functions if, for all $i \in \{1, \dots, d\}$,*

$$\|D_i^k g\|_{\infty} \leq \mathcal{L}_i \quad \forall k = 0, 1, \dots, \lfloor \beta_i \rfloor,$$

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

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, \dots, e_d denoting the canonical basis in \mathbb{R}^d .

From now on we deal with the estimation of the density μ belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$.

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

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

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

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

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

As we will see in Section 4.2, a main question concerns the choice of the multi-index bandwidth $h = (h_1, \dots, h_d)^T$.

4 Main results

4.1 Convergence rates for invariant density estimation

We want to investigate on the convergence rates for invariant density estimation. In order to determine the asymptotic behaviour of our estimator for $T \rightarrow \infty$, we study the variance of general additive functionals of X in d dimension. To do so, we need some properties as the exponential ergodicity of the process and a bound on the transition density. Such properties will be derived from our assumptions through the following lemmas, that we will prove in the appendix. The following bounds on the transition density and on the transition semigroup hold true.

Lemma 1. *Suppose that A1 - A3 hold. Then there exists a transition density $p_t(x, y)$ such that, for $T \geq 0$, there exist $c_0 > 0$ and $\lambda_0 > 0$ such that for any $t \in [0, T]$ and any pair of points $x, y \in \mathbb{R}^d$, we have*

$$p_t(x, y) \leq c_0(t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{t}} + \frac{t}{(t^{\frac{1}{2}} + |y-x|)^{d+\alpha}}).$$

Lemma 2. *Suppose that A1 - A3 hold. Then the process X is exponentially ergodic and exponentially β - mixing.*

On the basis of the two previous lemmas we can prove the following bound on the variance, which is the heart of the study on the convergence rate.

Proposition 1. *Suppose that A1 - A3 hold and let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded, measurable function with support \mathcal{S} satisfying $|\mathcal{S}| < 1$. Then, there exists a constant C independent of f such that*

- $\text{Var}(\int_0^T f(X_t) dt) \leq CT \|f\|_{\infty}^2 |\mathcal{S}|^2 (1 + (\log(\frac{1}{|\mathcal{S}|}))^{2 - \frac{(1+\alpha)}{2}} + \log(\frac{1}{|\mathcal{S}|}))$ for $d = 1$,
- $\text{Var}(\int_0^T f(X_t) dt) \leq CT \|f\|_{\infty}^2 |\mathcal{S}|^2 (1 + \log(\frac{1}{|\mathcal{S}|}))$ for $d = 2$,
- $\text{Var}(\int_0^T f(X_t) dt) \leq CT \|f\|_{\infty}^2 |\mathcal{S}|^{1 + \frac{2}{d}}$ for $d \geq 3$.

From the bias - variance decomposition in the anisotropic case (see Proposition 1 in [5]) we get the following bound

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^d h_l^{2\beta_l} + T^{-2} \text{Var}\left(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K\left(\frac{x_m - X_t^m}{h_m}\right) dt\right).$$

We want to bound the variance here above using Proposition 1 on the function $f(y) := \frac{1}{\prod_{l=1}^d h_l} \prod_{m=1}^d K(\frac{x_m - y_m}{h_m})$. As it will be explained in the proof of Proposition 2 in Section 5, for $d \geq 3$ it is

$$\text{Var}(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_t^m}{h_m}) dt) \leq cT(\prod_{l=1}^d h_l)^{\frac{2}{d}-1},$$

which leads us to the following convergence rate.

Proposition 2. *Suppose that A1 - A3 hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in (5) satisfies, for $d \geq 3$, the following risk estimates:*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^d h_l^{2\beta_l} + T^{-1}(\prod_{l=1}^d h_l)^{\frac{2}{d}-1}. \quad (6)$$

Defining

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

the rate optimal choice $h_l = h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}}$ yields the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}}.$$

We underline that, in the continuous case, the convergence rate found by Strauch in [28] for the estimation of the invariant density μ belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta + 1, \mathcal{L})$ is $T^{-\frac{2(\bar{\beta}+1)}{2(\bar{\beta}+1)+d-2}}$, for $d \geq 3$.

In Proposition 2 we estimate μ over anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ and we therefore extend [28] to the jumps - diffusion case: the convergence rate we obtain is the same it was in the case without jumps, which is also analogous to the rate first obtained by Dalalyan and Reiss in [9] for the estimation of the invariant density μ over isotropic Hölder class $\mathcal{H}_d(\beta + 1, \mathcal{L})$, up to replacing the mean smoothness $\bar{\beta} + 1$ with $\beta + 1$, the common smoothness over the d different dimensions.

For $d = 1$ and $d = 2$, the bound on the variance changes. Therefore, the rate optimal choice h will be different as well, as explained in following two propositions.

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

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h^{2\beta} + \frac{1}{T}(1 + (\log(\frac{1}{h}))^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{h})). \quad (7)$$

The rate optimal choice for h yields to the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{(\log T)^{(2-\frac{(1+\alpha)}{2}) \vee 1}}{T}.$$

It is worth remarking that, in Proposition 3, it is stated the main difference between the case with and without jumps. Indeed, if in the continuous case the convergence rate was $\frac{1}{T}$, now it depends on the degree of the jumps α and it is between $\frac{\log T}{T}$ and $\frac{(\log T)^{\frac{3}{2}}}{T}$.

We need to say that the convergence rate we have found here above for the estimation of the invariant density of a stochastic differential equations with jumps in the one dimensional setting is not necessarily the optimal one. In the continuous case other methods have been explored for such an estimation when $d = 1$, as the use of diffusion local time to get the optimal rate $\frac{1}{T}$. We do not rule out the possibility to get a sharper bound through the exploitation of other approaches also for the jumps case, finding therefore a convergence rate faster than the one presented in the previous proposition.

Proposition 4. *Suppose that A1 - A3 hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in (5) satisfies, for $d = 2$, the following risk estimates:*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h_1^{2\beta_1} + h_2^{2\beta_2} + \frac{1}{T}(1 + \log(\frac{1}{h_1 h_2})). \quad (8)$$

The rate optimal choice for h yields to the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{\log T}{T}.$$

From the pointwise estimation of the invariant density gathered in the three previous propositions we move to the estimation on $L^2(A)$, where A is a compact set of \mathbb{R}^d .

In the sequel, for $A \subset \mathbb{R}^d$ compact and for $g \in L^2(A)$, $\|g\|_A^2 := \int_A |g(x)|^2 dx$ denotes the L^2 norm with respect to Lebesgue on A .

As a consequence of Propositions 2, 3 and 4 and the fact that the constants which turn out in the proofs do not depend on x , the following corollary holds true:

Corollary 1. *If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then for the rate optimal choice for $h = h(T)$ provided in Propositions 2, 3 and 4 we have the following risk estimates:*

$$\mathbb{E}[\|\hat{\mu}_{h,T} - \mu\|_A^2] \lesssim V_d(T) := \begin{cases} \frac{(\log T)^{(2 - \frac{(1+\alpha)}{2}) \vee 1}}{T} & \text{for } d = 1 \\ \frac{\log T}{T} & \text{for } d = 2 \\ T^{-\frac{2\beta}{2\beta+d-2}} & \text{for } d \geq 3. \end{cases}$$

The proof of Corollary 1 will be given in Section 5.

One may wonder if it is possible to improve the convergence rate by using other density estimators or what is the best possible rate of convergence. To answer, the idea would be to consider the minimax risk associated to the anisotropic Holder class $\mathcal{H}_d(\beta, \mathcal{L})$, as done in the continuous framework by Strauch [28] (see Theorem 3.6). To achieve such a goal, the knowledge of the link between b and the density, that we denote as μ_b to make explicit the dependence on the drift, is crucial. In absence of jumps the connection between the drift function and the invariant measure was explicit:

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

but, adding the jumps, it is no longer true. In our framework, it is challenging to get a relation between b and μ_b and it is therefore challenging to discuss the optimality of the proposed estimator. As the optimality of the proposed estimator is still an open question and it seems to us an interesting issue which deserves further investigation, a future perspective consists in looking for a lower bound for the minimax risk associated to the anisotropic Holder class $\mathcal{H}_d(\beta, \mathcal{L})$ in our jump-diffusion framework.

4.2 Adaptive procedure

The question of density estimation belongs to the canonical framework of nonparametric statistics. As detailed in Propositions 3 and 4, both the bandwidth and the upper bound on the rate of convergence appearing on the right hand side of (7) and (8) do not depend on the unknown smoothness of the invariant density μ and so there is no gain in implementing a data-driven bandwidth selection procedure for density estimation in the framework of continuous observations of a one or two dimensional diffusion process with jumps. Hence, throughout the sequel we restrict to the case $d \geq 3$.

It is clear from the previous section that for $d \geq 3$, instead, the proposed bandwidth choice depends on the regularity of the density μ , which is unknown. This is why we study a data-driven bandwidth selection device.

We emphasize that the d selected bandwidths are different, and this anisotropy property is important in our setting: the regularity in each direction can be various. The bandwidth selection procedure has to be able to provide such different choices for h_1, h_2, \dots, h_d .

To select h adequately, we propose the following method, inspired from Goldenshluger and Lepski [13].

We define the set of candidate bandwidths \mathcal{H}_T as

$$\mathcal{H}_T \subset \left\{ h = (h_1, \dots, h_d)^T \in (0, 1]^d : \frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d h_l \leq \left(\frac{1}{\log T} \right)^{\frac{3d}{d-2}} \right\}, \quad (9)$$

The conditions on $\prod_{l=1}^d h_l$ we have just given are needed to use Talagrand inequality, on the basis of which we show our adaptive result.

We suppose moreover that the growth of $|\mathcal{H}_T|$ is at most polynomial in T , which is there exists $c > 0$ for which $|\mathcal{H}_T| \leq cT^c$.

An example of \mathcal{H}_T is the following set of candidate bandwidths:

$$\mathcal{H}_T := \left\{ h = (h_1, \dots, h_d)^T \in (0, 1]^d : h_i = \frac{1}{k_i} \text{ with } k_i \in \mathbb{N}, \frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d \frac{1}{k_l} \leq \left(\frac{1}{\log T} \right)^{\frac{3d}{d-2}} \right\}. \quad (10)$$

In correspondence of the variation of $h \in \mathcal{H}_T$, we have the following family of estimators, defined as in (5)

$$\mathcal{F}(\mathcal{H}_T) := \left\{ \hat{\mu}_h(x) = \frac{1}{T} \int_0^T \mathbb{K}_h(X_u - x) du : x \in \mathbb{R}^d, h \in \mathcal{H}_T \right\}$$

where, as defined in (5), for $y \in \mathbb{R}^d$, it is

$$\mathbb{K}_h(y) := \prod_{l=1}^d \frac{1}{h_l} \prod_{m=1}^d K\left(\frac{y_m}{h_m}\right). \quad (11)$$

We aim at selecting an estimator from the family $\mathcal{F}(\mathcal{H}_T)$ in a completely data-driven way, based only on the observation of the continuous trajectory of the process X solution of (3).

We now turn to describing the selection procedure from $\mathcal{F}(\mathcal{H}_T)$, which is based on auxiliary estimators relying on the convolution operator. According to our records, it was introduced in [20] as a device to circumvent the lack of ordering among a set of estimators in anisotropic case, where the increase of the variance of an estimator does not imply a decrease of its bias.

For any bandwidths $h = (h_1, \dots, h_d)^T$, $\eta = (\eta_1, \dots, \eta_d)^T \in \mathcal{H}_T$ and $x \in \mathbb{R}^d$, we define

$$\mathbb{K}_h * \mathbb{K}_\eta(x) := \prod_{j=1}^d (K_{h_j} * K_{\eta_j})(x_j) = \prod_{j=1}^d \int_{\mathbb{R}} K_{h_j}(u - x_j) K_{\eta_j}(u) du.$$

We moreover define the kernel estimators

$$\hat{\mu}_{h,\eta}(x) := \frac{1}{T} \int_0^T (\mathbb{K}_h * \mathbb{K}_\eta)(X_u - x) du, \quad x \in \mathbb{R}^d.$$

We remark that for how we have defined the kernel estimators, since the convolution is commutative, it is $\hat{\mu}_{h,\eta} = \hat{\mu}_{\eta,h}$.

The proposed selection procedure relies on comparing the differences $\hat{\mu}_{h,\eta} - \hat{\mu}_\eta$.

We define

$$A(h) := \sup_{\eta \in \mathcal{H}_T} (\|\hat{\mu}_{h,\eta} - \hat{\mu}_\eta\|_A^2 - V(\eta))_+, \quad (12)$$

with

$$V(h) := \frac{k}{T} \left(\prod_{l=1}^d h_l \right)^{\frac{2}{d}-1},$$

where k is a numerical constant which is large. In particular, it is sufficient to choose it bigger than the constants $2k_0^*$ and $2k_0$ which appear in Lemma 4. Even if k is not explicit, it can be calibrated by simulations as done for example in Section 5 of [8] through the implementation of a method inspired by Lacour, Massart and Rivoirard in [18].

Heuristically, $A(h)$ is an estimate of the squared bias and $V(h)$ of the variance bound. It is worth noticing that the penalty term $V(h)$ which is used here comes from Proposition 1 for the function

f being the Kernel function.
Thus, the selection is done by setting

$$\tilde{h} := \arg \min_{h \in \mathcal{H}_T} (A(h) + V(h)). \quad (13)$$

We introduce the following notation: $\mu_h := \mathbb{K}_h * \mu$, which is the function that is estimated without bias by $\hat{\mu}_h$, i.e. $\mathbb{E}[\hat{\mu}_h(x)] = \mu_h(x)$. Moreover we define $\mu_{h,\eta} := \mathbb{K}_h * \mathbb{K}_\eta * \mu$ and a bias term $B(h) := \|\mu_h - \mu\|_{\tilde{A}}^2$, where we have denoted as $\|\cdot\|_{\tilde{A}}$ the L^2 -norm on \tilde{A} , a compact set in \mathbb{R}^d which contains A . It is $\tilde{A} := \{\zeta \in \mathbb{R}^d : d(\zeta, A) \leq 2\sqrt{d}\}$, where $d(\zeta, A) := \min_{x \in A} |\zeta - x|$. The following result holds.

Theorem 1. *Suppose that assumptions A1 - A3 hold. Then, we have*

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 \inf_{h \in \mathcal{H}_T} (B(h) + V(h)) + c_1 e^{-c_2 (\log T)^2},$$

for c_1 and c_2 positive constants.

The bound stated in Theorem 1 shows that the estimator leads to an automatic trade-off between the bias $\|\mu_h - \mu\|_A^2$ and the variance $V(h)$, up to a multiplicative constant c_1 . The last term is indeed negligible. The proof of Theorem 1 is postponed to Section 6.

We recall that Proposition 2 provides us the rate optimal choice $h(T)$ for $d \geq 3$, which is $h_l(T) = (\frac{1}{T})^{\frac{\tilde{\beta}}{\beta_l(2\tilde{\beta}+d-2)}}$.

Using such a bandwidth we will prove in Section 6 the following theorem.

Theorem 2. *Suppose that assumptions A1 - A3 hold and let \mathcal{H}_T be defined by (10). Then, we have*

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 (\frac{1}{T})^{\frac{2\tilde{\beta}}{2\tilde{\beta}+d-2}} + c_1 e^{-c_2 (\log T)^2},$$

for c_1 and c_2 positive constants.

Underlining once again that the second term in the right hand side of the equation here above is negligible compared to the first, we have that the risk estimates we get using the bandwidth provided by our selection procedure converges to zero fast. In particular, its convergence rate coincides to the optimal one provided by both [9] and [28] in the case without jumps.

5 Proof convergence rates for invariant density estimation

In this section we prove Propositions 2, 3 and 4, which gives us the convergence rate for the estimation of the invariant density $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ in the three different situation: $d = 1$, $d = 2$ and $d \geq 3$.

We emphasize that all the constants will appear in the proofs do not depend on the point x considered.

We start showing the bound on the variance gathered in Proposition 1.

5.1 Proof of Proposition 1

Proof. We consider first of all the case $d \geq 3$. By the symmetry of the covariance operator it is

$$\text{Var}(\int_0^T f(X_s) ds) = \int_0^T \int_0^T \text{Cov}(f(X_s), f(X_t)) dt ds = 2 \int_0^T \int_0^s \text{Cov}(f(X_s), f(X_t)) dt ds.$$

We introduce the centered function $f_c := f - \mu(f)$. It follows

$$\text{Cov}(f(X_s), f(X_t)) = \mathbb{E}[f_c(X_s) f_c(X_t)].$$

We now observe that, by the definition of transition density, it is

$$\mathbb{E}[f_c(X_s) f_c(X_t)] = \int_{\mathbb{R}^d} f_c(x) \int_{\mathbb{R}^d} f_c(y) p_{s-t}(x, y) \mu(y) dy dx = \mathbb{E}[f_c(X_{s-t}) f_c(X_0)].$$

Therefore,

$$\text{Var}\left(\int_0^T f(X_s)ds\right) = 2 \int_0^T \int_0^s \mathbb{E}[f_c(X_{s-t})f_c(X_0)]dt ds.$$

Applying the change of variable $u := s - t$ and using Fubini theorem we have the quantity here above is equal to

$$2 \int_0^T \int_0^{T-u} \mathbb{E}[f_c(X_0)f_c(X_u)]ds du = 2 \int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du.$$

Let now $0 < \delta < D \leq T$, where the specific choice of δ and D will be given later. The idea is to deal with the integral here above in different way for u which is in different intervals. For this reason we see $\int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du$ as

$$\int_0^\delta (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du + \int_\delta^D (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du + \int_D^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du. \quad (14)$$

We now observe that

$$\int_0^\delta (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du = \int_0^\delta (T-u)(\mathbb{E}[f(X_0)f(X_u)] - (\mu(f))^2)du \leq cT \int_0^\delta |< P_u f, f >_\mu| du, \quad (15)$$

where we have denoted as $< \cdot, \cdot >_\mu$ the scalar product deriving by the norm with respect to the measure μ , for which $< g, h >_\mu := \int_{\mathbb{R}^d} g(x)h(x)\mu(x)dx$, each $g, h \in L^2(\mu)$. In the last inequality we have moreover used that $(\mu(f))^2$ is always more than 0.

Now we use Cauchy-Schwartz inequality and the fact that $P_u f$ is a contraction map in $L^2(\mu)$ to get

$$\int_0^\delta |< P_u f, f >_\mu| du \leq \int_0^\delta \sqrt{\|P_u f\|_\mu^2 \|f\|_\mu^2} du \leq \int_0^\delta \sqrt{\|f\|_\mu^4} du \leq \|f\|_\infty^2 \mu(\mathcal{S})\delta, \quad (16)$$

where in the last inequality we have used the estimation

$$\|f\|_\mu^2 = \int_{\mathcal{S}} |f(x)|^2 \mu(x)dx \leq \|f\|_\infty^2 \mu(\mathcal{S}).$$

Concerning the second integral in (14), we remark that (15) still holds on $[\delta, D]$. We then estimate it through the definition of transition semigroup. It is

$$\int_\delta^D |< P_u f, f >_\mu| du \leq \int_\delta^D \int_{\mathbb{R}^d} |f(x)| \int_{\mathbb{R}^d} |f(y)| p_u(x, y) dy \mu(x) dx du. \quad (17)$$

We want to use the bound on the transition density given in Lemma 1 which holds for $t \in [0, T]$ but it is not uniform in t big. Nevertheless, for $t \geq 1$, we have

$$\begin{aligned} p_t(x, y) &= \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x, \zeta) p_{\frac{1}{2}}(\zeta, y) d\zeta \leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x, \zeta) (e^{-\lambda_0(y-\zeta)^2 \frac{1}{2}} + \frac{1}{(\sqrt{\frac{1}{2}} + |y-\zeta|)^{d+\alpha}}) d\zeta \leq \\ &\leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x, \zeta) d\zeta \leq c, \end{aligned}$$

where the constant c changes from line to line. The right hand side of (17) is therefore upper bounded by

$$\begin{aligned} &\int_\delta^D \int_{\mathbb{R}^d} |f(x)| c \int_{\mathbb{R}^d} |f(y)| (u^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{u}} + \frac{u}{(u^{\frac{1}{2}} + |y-x|)^{d+\alpha}} + 1) dy \mu(x) dx du \leq \\ &\leq \int_\delta^D \int_{\mathcal{S}} |f(x)| c \int_{\mathcal{S}} |f(y)| (u^{-\frac{d}{2}} + u^{1-\frac{(d+\alpha)}{2}} + 1) dy \mu(x) dx du \leq c \|f\|_\infty^2 \mu(\mathcal{S}) |\mathcal{S}| \int_\delta^D (u^{-\frac{d}{2}} + u^{1-\frac{(d+\alpha)}{2}} + 1) du, \end{aligned}$$

where we have bounded in both integrals the absolute value of f with its infinity norm.

Now we want to calculate the integral with respect to the variable u . We observe that, since $d \geq 3$, $1 - \frac{d}{2} < 0$. 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. For $d = 4 - \alpha$ an extra logarithmic term appears. Therefore, we have to consider the two different possibilities, according to the fact that the exponent would be positive or negative. It follows

$$\begin{aligned} \int_{\delta}^D | \langle P_u f, f \rangle_{\mu} | du &\leq c \|f\|_{\infty}^2 \mu(\mathcal{S}) |\mathcal{S}| (\delta^{1-\frac{d}{2}} + \delta^{2-\frac{d+\alpha}{2}} 1_{\{d \geq 4-\alpha\}} \\ &\quad + (|\log D| + |\log \delta|) 1_{\{d=4-\alpha\}} + D^{2-\frac{d+\alpha}{2}} 1_{\{d < 4-\alpha\}} + D). \end{aligned} \quad (18)$$

We are now left to estimate the third integral of (14). From Lemma 2 it follows it is upper bounded by

$$cT \int_D^T | \langle P_u f_c, f_c \rangle_{\mu} | du \leq cT \int_D^T \sqrt{\|P_u f_c\|_{L^1(\mu)}^2 \|f_c\|_{\infty}^2} du \leq cT \int_D^T \sqrt{(e^{-\rho u} \|f_c\|_{\infty})^2 \|f_c\|_{\infty}^2} du. \quad (19)$$

We recall it is $f_c(x) = f(x) - \mu(f)$, where $\mu(f) = \int_{\mathcal{S}} f(x) \mu(x) dx$.

Therefore we have

$$|f_c(x)| \leq |f(x)| + |\mu(f)| \leq |f(x)| + \|f\|_{\infty} \mu(\mathcal{S})$$

and so

$$\|f_c\|_{\infty} \leq \|f\|_{\infty} + \|f\|_{\infty} \mu(\mathcal{S}) \leq c \|f\|_{\infty},$$

where in the last inequality we have used the following estimation

$$|\mu(\mathcal{S})| \leq \|\mu\|_{\infty} |\mathcal{S}| \leq c |\mathcal{S}| \quad (20)$$

and the fact that $|\mathcal{S}| < 1$.

Therefore we get

$$cT \int_D^T | \langle P_u f_c, f_c \rangle_{\mu} | du \leq cT \|f\|_{\infty}^2 \int_D^T e^{-\rho u} du \leq cT \|f\|_{\infty}^2 e^{-\rho D}. \quad (21)$$

Replacing (16), (18) and (21) in (14) we have that

$$\begin{aligned} \left| \int_0^T (T-u) \mathbb{E}[f_c(X_0) f_c(X_u)] du \right| &\leq cT \|f\|_{\infty}^2 |\mathcal{S}| (\delta + |\mathcal{S}| \delta^{1-\frac{d}{2}} + |\mathcal{S}| \delta^{2-\frac{d+\alpha}{2}} 1_{\{d \geq 4-\alpha\}} + |\mathcal{S}| D^{2-\frac{d+\alpha}{2}} 1_{\{d < 4-\alpha\}} + \\ &\quad + |\mathcal{S}| (|\log D| + |\log \delta|) 1_{\{d=4-\alpha\}} + |\mathcal{S}| D) + cT \|f\|_{\infty}^2 e^{-\rho D}. \end{aligned} \quad (22)$$

We now want to choose δ and D for which the estimation here above is as sharp as possible. Recalling that the exponent on δ are less than zero in the second and the third terms of the right hand side of (22), we have that for a small choice of δ would correspond the smallness of the first term while the second and the third would be big, the opposite would hold for a big δ . In the same way, the behaviour of the last two terms of the right hand side of (22) relies on the choice D . Aiming at balancing them, we define $\delta := |\mathcal{S}|^{\frac{2}{d}}$ and $D := [\max(-\frac{2}{\rho} \log(|\mathcal{S}|), 1)] \wedge T$. Replacing them in (22) if $T > (-\frac{2}{\rho} \log(|\mathcal{S}|))$ we get

$$\begin{aligned} \left| \int_0^T (T-u) \mathbb{E}[f_c(X_0) f_c(X_u)] du \right| &\leq \\ &\leq cT \|f\|_{\infty}^2 (|\mathcal{S}|^{1+\frac{2}{d}} + |\mathcal{S}|^{1+\frac{4-\alpha}{d}} + |\mathcal{S}|^2 [(\log |\mathcal{S}|)^{2-\frac{d+\alpha}{2}} + |\log |\mathcal{S}|| + |\log |\log |\mathcal{S}||| + 1]), \end{aligned}$$

which give us the result we wanted remarking that both 2 and $1 + \frac{4-\alpha}{d}$ are always more than $1 + \frac{2}{d}$ for $d \geq 3$ and $\alpha \in (0, 2)$.

Otherwise, if $T \leq (-\frac{2}{\rho} \log(|\mathcal{S}|))$, by the definition of D we obtain $D = T$. We still have $|\mathcal{S}|^2 D^{2-\frac{d+\alpha}{2}} 1_{\{d < 4-\alpha\}} \leq c |\mathcal{S}|^2 (\log |\mathcal{S}|)^{2-\frac{d+\alpha}{2}}$ and, moreover, the last integral which we dealt with in (21) is in this case between T and T and so its contribution is null. Hence, the result still holds true.

We now consider the case $d = 1$. We can act exactly like we did in the case $d \geq 3$, splitting

the integral in three parts. Estimations (16) and (21) still holds while, using again Lemma 1 on the interval $[\delta, D]$, we have

$$\begin{aligned} |\int_{\delta}^D (T-u) \mathbb{E}[f_c(X_0)f_c(X_u)] du| &\leq cT \|f\|_{\infty}^2 \mu(\mathcal{S}) |\mathcal{S}| \int_{\delta}^D (u^{-\frac{1}{2}} + u^{1-\frac{(1+\alpha)}{2}} + 1) du \leq \\ &\leq cT \|f\|_{\infty}^2 \mu(\mathcal{S}) |\mathcal{S}| (D^{\frac{1}{2}} + D^{2-\frac{(1+\alpha)}{2}} + D), \end{aligned}$$

where we have used that now, integrating, both the exponent we get are positive. In total in the case $d = 1$, using also (20), we therefore have

$$\text{Var}(\int_0^T f(X_s) ds) \leq cT \|f\|_{\infty}^2 (|\mathcal{S}| \delta + |\mathcal{S}|^2 (D + D^{2-\frac{(1+\alpha)}{2}}) + e^{-\rho D}).$$

As we have already done, we want to make the estimation here above as sharp as possible. This time there isn't any constraint on the smallness of δ and so we can choose directly $\delta := 0$. Regarding D we observe that, if $\alpha > 1$, then $2 - \frac{(1+\alpha)}{2} < 1$; otherwise $2 - \frac{(1+\alpha)}{2} > 1$. In each case we have the same trade off we had in the case $d \geq 3$ and so we keep taking $D := [\max(-\frac{2}{\rho} \log(|\mathcal{S}|), 1)] \wedge T$; it follows $\text{Var}(\int_0^T f(X_s) ds) \leq cT \|f\|_{\infty}^2 |\mathcal{S}|^2 (1 + (\log |\mathcal{S}|)^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{|\mathcal{S}|}))$.

In the case $d = 2$ estimations (16) and (21) keep holding true. The bound on the transition density gathered in Lemma 1 for $d = 2$ yields

$$\begin{aligned} |\int_{\delta}^D (T-u) \mathbb{E}[f_c(X_0)f_c(X_u)] du| &\leq cT \|f\|_{\infty}^2 \mu(\mathcal{S}) |\mathcal{S}| \int_{\delta}^D (u^{-1} + u^{1-\frac{(2+\alpha)}{2}} + 1) du \leq \\ &\leq cT \|f\|_{\infty}^2 \mu(\mathcal{S}) |\mathcal{S}| (\log(\frac{D}{\delta}) + D^{2-\frac{(2+\alpha)}{2}} + D), \end{aligned}$$

having remarked that $2 - \frac{(2+\alpha)}{2} = 1 - \frac{\alpha}{2} > 0$ because $\alpha \in (0, 2)$. This entails, using also (20),

$$\text{Var}(\int_0^T f(X_s) ds) \leq cT \|f\|_{\infty}^2 (|\mathcal{S}| \delta + |\mathcal{S}|^2 (\log(\frac{D}{\delta}) + D^{2-\frac{(2+\alpha)}{2}} + D) + e^{-\rho D}).$$

Aiming at balancing the terms, we choose again $\delta := |\mathcal{S}|$ and $D := [\max(-\frac{2}{\rho} \log(|\mathcal{S}|), 1)] \wedge T$. It follows that $\log(\frac{D}{\delta}) \leq c |\log |\log |\mathcal{S}|| + c |\log(\frac{1}{|\mathcal{S}|})|$ and so, observing that $\log |\log(|\mathcal{S}|)|$ is negligible compared to $\log(\frac{1}{|\mathcal{S}|})$, the bound on the variance becomes

$$\text{Var}(\int_0^T f(X_s) ds) \leq cT \|f\|_{\infty}^2 |\mathcal{S}|^2 (1 + \log(\frac{1}{|\mathcal{S}|})),$$

where we have also used that $2 - \frac{(2+\alpha)}{2}$ is always less than 1 and so $(\log(\frac{1}{|\mathcal{S}|}))^{2-\frac{(2+\alpha)}{2}} < \log(\frac{1}{|\mathcal{S}|})$. The proposition is therefore proved. \square

5.2 Proof of Proposition 2

Proof. In order to prove (6) we write the usual bias-variance decomposition

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \leq |\mathbb{E}[\hat{\mu}_{h,T}(x)] - \mu(x)|^2 + \text{Var}(\hat{\mu}_{h,T}(x)). \quad (23)$$

Using the stationarity of the process and the definition (5) of the estimator, it is

$$\begin{aligned} |\mathbb{E}[\hat{\mu}_{h,T}(x)] - \mu(x)| &= |\int_{\mathbb{R}^d} \mathbb{K}_h(x-y) \mu(y) dy - \mu(x)| \\ &= |\frac{1}{\prod_{l=1}^m h_l} \int_{\mathbb{R}^d} \prod_{m=1}^d K(\frac{x_m - y_m}{h_m}) \mu(y) dy - \mu(x) dx| \\ &= |\int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) [\mu(x_1 - h_1 z_1, \dots, x_d - h_d z_d) - \mu(x)] dz|, \end{aligned} \quad (24)$$

where the last inequality is a consequence of the change of variable $\frac{x_m - y_m}{h_m} =: z_m$ and of the property $\int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) dz = 1$ of the Kernel function. The use of Taylor formula to the partial functions $t \mapsto \mu(x_1 - h_1 z_1, \dots, x_{d-1} - h_{d-1} z_{d-1}, t), \dots, t \mapsto \mu(x_1 - h_1 z_1, \dots, x_{i-1} - h_{i-1} z_{i-1}, t, x_{i+1}, \dots, x_d)$, implemented iteratively, yields

$$\begin{aligned} & \mu(x_1 - h_1 z_1, \dots, x_d - h_d z_d) \\ &= \mu(x) + \sum_{j=1}^d \left(\sum_{k=1}^{\lfloor \beta_j \rfloor - 1} \frac{D_j^k \mu(\tilde{x}_j)}{k!} (h_j z_j)^k + \frac{D_j^{\lfloor \beta_j \rfloor} \mu(x_1 - h_1 z_1, \dots, x_j - \tau_j h_j z_j, x_{j+1}, \dots, x_d)}{\lfloor \beta_j \rfloor!} (h_j z_j)^{\lfloor \beta_j \rfloor} \right), \end{aligned} \quad (25)$$

where $0 \leq \tau_j \leq 1$ and $\tilde{x}_j = (x_1 - h_1 z_1, \dots, x_{j-1} - h_{j-1} z_{j-1}, x_j, \dots, x_d)$. Replacing it in (24) and using that our Kernel function is of order $M \geq \max_i \beta_i$, it follows (24) is equal to

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) \left[\sum_{j=1}^d \left(\sum_{k=1}^{\lfloor \beta_j \rfloor - 1} \frac{D_j^k \mu(\tilde{x}_j)}{k!} (h_j z_j)^k + \frac{D_j^{\lfloor \beta_j \rfloor} \mu(x_1 - h_1 z_1, \dots, x_j - \tau_j h_j z_j, x_{j+1}, \dots, x_d)}{\lfloor \beta_j \rfloor!} (h_j z_j)^{\lfloor \beta_j \rfloor} \right) \right] dz \right| \\ &= \left| \int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) \left[\sum_{j=1}^d \frac{(h_j z_j)^{\lfloor \beta_j \rfloor}}{\lfloor \beta_j \rfloor!} D_j^{\lfloor \beta_j \rfloor} \mu(x_1 - h_1 z_1, \dots, x_j - \tau_j h_j z_j, x_{j+1}, \dots, x_d) \right] dz \right|. \end{aligned}$$

The first term is indeed equal to zero because, even if \tilde{x}_j depends on z , it depends in particular only on z_l for $l < j$. Therefore, for any $j \in \{1..d\}$,

$$\int_{\mathbb{R}} K(z_j) \sum_{k=1}^{\lfloor \beta_j \rfloor - 1} \frac{D_j^k \mu(\tilde{x}_j)}{k!} (h_j z_j)^k dz_j = 0.$$

We now want to evaluate the rest term in Taylor formula. Using the fact that

$$\int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) \left[\sum_{j=1}^d \frac{(h_j z_j)^{\lfloor \beta_j \rfloor}}{\lfloor \beta_j \rfloor!} D_j^{\lfloor \beta_j \rfloor} \mu(x) \right] dz = 0,$$

we get it is equal to

$$\left| \int_{\mathbb{R}^d} \prod_{m=1}^d K(z_m) \left[\sum_{j=1}^d \frac{(h_j z_j)^{\lfloor \beta_j \rfloor}}{\lfloor \beta_j \rfloor!} (D_j^{\lfloor \beta_j \rfloor} \mu(x_1 - h_1 z_1, \dots, x_j - \tau_j h_j z_j, x_{j+1}, \dots, x_d) - D_j^{\lfloor \beta_j \rfloor} \mu(x)) \right] dz \right|, \quad (26)$$

As $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, we get

$$|D_j^{\lfloor \beta_j \rfloor} \mu(x_1 - h_1 z_1, \dots, x_j - \tau_j h_j z_j, x_{j+1}, \dots, x_d) - D_j^{\lfloor \beta_j \rfloor} \mu(x)| \leq \mathcal{L}_j |\tau_j h_j z_j|^{\beta_j - \lfloor \beta_j \rfloor}.$$

Therefore, using also that $|\tau_j| \leq 1$, we have that (26) is upper bounded by

$$\int_{\mathbb{R}^d} \left| \prod_{m=1}^d K(z_m) \right| \sum_{j=1}^d \frac{\mathcal{L}_j}{\lfloor \beta_j \rfloor!} |h_j z_j|^{\beta_j} \leq c \sum_{j=1}^d h_j^{\beta_j}.$$

It is here important to remark that the constant c does not depend on x . An analogous computation can be found in Proposition 1.2 of [29] or in Proposition 1 of [5]. The estimation here above, together with the decomposition (23), gives us

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

Estimation (6) is now a straightforward consequence of Proposition 1 applied to

$$f(y) := \frac{1}{\prod_{l=1}^d h_l} \prod_{m=1}^d K\left(\frac{x_m - y_m}{h_m}\right),$$

whose support \mathcal{S} is such that $|\mathcal{S}| \leq c \prod_{l=1}^d h_l$ and which is by construction such that $\|f\|_\infty \leq c(\prod_{l=1}^d h_l)^{-1}$.

To find the optimal choice of h we define $h_l(T) := (\frac{1}{T})^{a_l}$ for $l \in \{1, \dots, d\}$ and we look for a_1, \dots, a_d such that the upper bound of the mean-squared error in the right hand side of (6) would result as small as possible.

Replacing the definition of $h_l(T)$ in the bias - variance decomposition, it means searching for a_1, \dots, a_d for which we get the balance and so we have to resolve the following system:

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

We observe that, as a consequence of the first $d-1$ equations, we can write a_l as $\frac{\beta_d}{\beta_l} a_d$ for each $l \in \{1, \dots, d-1\}$. Therefore, the last equation becomes

$$2\beta_d a_d = 1 + (\frac{2}{d} - 1) \beta_d a_d \sum_{l=1}^d \frac{1}{\beta_l}.$$

Defining $\frac{1}{\bar{\beta}} := \frac{1}{d} \sum_{l=1}^d \frac{1}{\beta_l}$, it follows $2\beta_d a_d = 1 + (\frac{2}{d} - 1) \beta_d a_d \frac{d}{\bar{\beta}}$, which yields

$$a_d = \frac{\bar{\beta}}{\beta_d(2\bar{\beta} + (d-2))}$$

and, replacing it in the system, we have

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

Taking in the right hand side of (6) the rate optimal choice $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta} + (d-2))}}$ we get the convergence rate wanted. \square

5.3 Proof of Proposition 3

Proof. The upper bound of the mean-squared error follows from the decomposition bias - variance and from Proposition 1, recalling that for $f(X_t) := \frac{1}{h} K(\frac{x-X_t}{h})$ we have $\|f\|_\infty \leq ch^{-1}$ and its support \mathcal{S} is such that $|\mathcal{S}| \leq ch$.

Now, aiming at balancing the terms, we take $h := (\frac{1}{T})^a$; getting the mean-squared error is upper bounded by

$$(\frac{1}{T})^{2a\beta} + \frac{1}{T} + \frac{(a \log T)^{2 - \frac{(1+\alpha)}{2}}}{T} + \frac{a \log T}{T}.$$

If a gets bigger clearly h gets smaller; it is enough to take a such that $2a\beta > 1$ to obtain the first two terms here above are negligible compared to the last ones, which gives us the convergence rate $\frac{(\log T)^{2 - \frac{(1+\alpha)}{2}}}{T}$ for $\alpha \leq 1$, $\frac{\log T}{T}$ for $\alpha > 1$. In particular, as $\beta \geq 1$, it is enough to take $a > \frac{1}{2}$ to ensure that $2a\beta$ is bigger than 1 for any $\beta \geq 1$ and so the chosen bandwidth does not depend on the smoothness β of μ . \square

5.4 Proof of Proposition 4

Proof. Again, (8) follows naturally from the bias - variance decomposition and Proposition 1.

Regarding the convergence rate, we take again $h_l := (\frac{1}{T})^{a_l}$ for $l = 1, 2$.

It follows $\log(\frac{1}{h_1 h_2}) = a_1 \log T + a_2 \log T$ and so the mean-squared error is upper bounded by

$$(\frac{1}{T})^{2a_1\beta_1} + (\frac{1}{T})^{2a_2\beta_2} + \frac{c \log T}{T}.$$

One may observe that taking both a_1 and a_2 bigger than $\frac{1}{2}$ it clearly follows that $2a_1\beta_1$ and $2a_2\beta_2$ are bigger than 1 for any $\beta_1, \beta_2 \geq 1$. In this way the first two terms here above result negligible compared to the third, which provides the convergence rate $\frac{\log T}{T}$. \square

5.5 Proof of Corollary 1

Proof. Let us stress that the constant appearing in the upper bound on the variance does not depend on the point x considered. The same can be said about the constant deriving from the bias term (see (25)). Therefore, as all the constants which turn out in all the previous propositions do not depend on the point x considered, the corollary is a straightforward consequence of Propositions 2, 3 and 4. Indeed, such propositions yield

$$\mathbb{E}[\|\hat{\mu}_{h,T} - \mu\|_A^2] = \mathbb{E}\left[\int_A |\hat{\mu}_{h,T}(x) - \mu(x)|^2 dx\right] \leq c \int_A cV_d(T) dx \leq c|A|V_d(T).$$

□

6 Proof of the adaptive procedure

The heart of the proof of Theorem 1 consist of finding an upper bound for the expected value of $A(h)$, which is gathered in the following proposition.

Proposition 5. *Suppose that assumptions A1 - A3 hold. Then, $\forall h \in \mathcal{H}_T$,*

$$\mathbb{E}[A(h)] \leq c_1 B(h) + c_1 e^{-c_2 (\log T)^2}.$$

Proposition 5 will be proven after the proofs of Theorems 1 and 2.

6.1 Proof of Theorem 1.

Proof. From triangular inequality it follows $\forall h \in \mathcal{H}_T$

$$\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2 \leq c(\|\hat{\mu}_{\tilde{h}} - \hat{\mu}_{h,\tilde{h}}\|_A^2 + \|\hat{\mu}_{h,\tilde{h}} - \hat{\mu}_h\|_A^2 + \|\hat{\mu}_h - \mu\|_A^2) \quad (27)$$

By the definition (12) of $A(h)$ it follows that the first and the second term of (27) are respectively upper bounded by $A(h) + V(\tilde{h})$ and $A(\tilde{h}) + V(h)$, having also used on the second term that $\hat{\mu}_{h,\tilde{h}} = \hat{\mu}_{\tilde{h},h}$. Then, since \tilde{h} has been defined in (13) as the $h \in \mathcal{H}_T$ for which $A(h) + V(h)$ is minimal, we clearly have that $A(\tilde{h}) + V(\tilde{h}) \leq A(h) + V(h)$.

Hence, for any $h \in \mathcal{H}_T$, we get

$$\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2 \leq c(A(h) + V(h) + \|\hat{\mu}_h - \mu\|_A^2). \quad (28)$$

We want an upper bound for the expected value of the left hand side of the equation (28) and so we need to evaluate $\mathbb{E}[\|\hat{\mu}_h - \mu\|_A^2] = \mathbb{E}[\int_A |\hat{\mu}_h(x) - \mu(x)|^2 dx]$.

From the standard bias variance decomposition, recalling that $\mu_h = \mathbb{K}_h * \mu$ is such that $\mathbb{E}[\hat{\mu}_h(x)] = \mu_h(x)$, we get

$$\mathbb{E}\left[\int_A |\hat{\mu}_h(x) - \mu(x)|^2 dx\right] = \int_A |\mu_h(x) - \mu(x)|^2 dx + \int_A \mathbb{E}[|\hat{\mu}_h(x) - \mu_h(x)|^2] dx.$$

Now, we can upper bound the first term of the right hand side here above by enlarging the integration domain getting

$$\int_A |\mu_h(x) - \mu(x)|^2 dx \leq \int_{\tilde{A}} |\mu_h(x) - \mu(x)|^2 dx = B(h). \quad (29)$$

Moreover, as consequence of Proposition 1 in the case $d \geq 3$ we obtain, as it was in Proposition 2,

$$\int_A \mathbb{E}[|\hat{\mu}_h(x) - \mu_h(x)|^2] dx = \int_A \text{Var}\left(\frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K\left(\frac{x_m - X_u^m}{h_m}\right) du\right) dx \leq |A| \frac{c}{T} \left(\prod_{l=1}^d h_l\right)^{\frac{2}{d}-1}. \quad (30)$$

Comparing the upper bound given in (30) with the definition of $V(h)$ and using also (28) and (29) we get, for each $h \in \mathcal{H}_T$,

$$\mathbb{E}[\|\hat{\mu}_h - \mu\|_A^2] \leq c(B(h) + V(h)) + \mathbb{E}[A(h)].$$

Now from Proposition 5 and the arbitrariness of the bandwidth h we are considering it follows

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 \inf_{h' \in \mathcal{H}_T} (B(h') + V(h')) + c_1 e^{-c_2 (\log T)^2},$$

as we wanted. \square

As a consequence of Theorem 1 we get, considering the rate optimal choice $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\bar{\beta}_l(2\bar{\beta}+d-2)}}$ provided by Proposition 2, the estimation gathered in Theorem 2. Its proof relies on the fact that, for how we have found it in Proposition 2, if the rate optimal bandwidth belongs to \mathcal{H}_T then the $\inf_{h \in \mathcal{H}_T} (B(h) + V(h))$ is clearly realized by it.

6.2 Proof of Theorem 2

Proof. We observe that, for the rate optimal choice $h(T)$ of the bandwidth, the conditions gathered in the right hand side of (9), which are $\frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d h_l \leq (\frac{1}{\log T})^{\frac{3d}{d-2}}$, hold true. Indeed,

$$\prod_{l=1}^d h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{2\bar{\beta}+d-2} \sum_{l=1}^d \frac{1}{\bar{\beta}_l}} = (\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}}.$$

the upper bound condition is therefore

$$(\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}} \leq (\frac{1}{\log T})^{\frac{3d}{d-2}},$$

which is true if and only if $\log T \leq T^{\frac{d-2}{3(2\bar{\beta}+d-2)}}$. Now we observe that $\frac{d-2}{3(2\bar{\beta}+d-2)} > 0$ for $\bar{\beta} > 1 - \frac{d}{2}$, which is always true since $d \geq 3$. In particular we can write $\frac{d-2}{3(2\bar{\beta}+d-2)} =: \gamma \in (0, 1)$ and, given that eventually for T going to ∞ it is $\log T \leq T^\gamma$, we have $\prod_{l=1}^d h_l(T) \leq (\frac{1}{\log T})^{\frac{3d}{d-2}}$. In the same way it is

$$(\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}} \geq \frac{(\log T)^{2d}}{T^{\frac{d}{3}}}.$$

For the same reasoning as here above it is true if $(\frac{1}{3} - \frac{1}{2\bar{\beta}+d-2})\frac{1}{2} =: \gamma$ is positive.

Being $\bar{\beta} > 1$ and $d \geq 3$, it turns out $\gamma > 0$, as we wanted.

Up to consider $\tilde{h}_l(T) := \frac{1}{\lfloor T^{\frac{\bar{\beta}}{\bar{\beta}_l(2\bar{\beta}+d-2)}} \rfloor}$ instead of $h_l(T)$, which is asymptotically equivalent and

which leads to the same convergence rate, we have that the rate optimal choice belongs to the set of candidate bandwidths \mathcal{H}_T proposed in (10).

Having now $h(T) \in \mathcal{H}_T$, for how we have found the rate optimal choice in Proposition 2, the $\inf_{h \in \mathcal{H}_T} (B(h) + V(h))$ is clearly realized by it and so the bound stated in Theorem 1 is actually (see also Corollary 1)

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 (\frac{1}{T})^{\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2 (\log T)^2},$$

as we wanted. \square

We have showed Theorem 1 using, as main tool, the bound on $\mathbb{E}[A(h)]$ stated in Proposition 5. Its proof, as we will see in the next section, relies on the use of Berbee's coupling method as in Viennet [31] and on a version of Talagrand inequality given in Klein and Rio [14].

6.3 Proof of Proposition 5

Proof. To find a bound for $\mathbb{E}[A(h)]$, for each $h \in \mathcal{H}_T$ we want to use Talagrand inequality, stated on random variables which are independent. Therefore, we start introducing some blocks which are mutually independent. We do it through the use of Berbee's coupling method as done in Viennet [31], Proposition 5.1 and its proof p. 484.

We assume that $T = 2p_T q_T$, with p_T integer and q_T a real to be chosen. We split the initial process $X = (X_t)_{t \in [0, T]}$ in $2p_T$ processes of a length q_T : for each $j \in \{1, \dots, p_T\}$ we consider $X^{j,1} := (X_t)_{t \in [2(j-1)q_T, (2j-1)q_T]}$ and $X^{j,2} := (X_t)_{t \in [(2j-1)q_T, 2jq_T]}$. Then, there exist a process $(X_t^*)_{t \in [0, T]}$ satisfying the following properties:

1. For $j \in \{1, \dots, p_T\}$ the processes $X^{j,1}$ and $X^{*,j,1} := (X_t^*)_{t \in [2(j-1)q_T, (2j-1)q_T]}$ have the same distribution and so have the processes $X^{j,2}$ and $X^{*,j,2} := (X_t^*)_{t \in [(2j-1)q_T, 2jq_T]}$.
2. For $j \in \{1, \dots, p_T\}$, $\mathbb{P}(X^{j,1} \neq X^{*,j,1}) \leq \beta_X(q_T)$ and $\mathbb{P}(X^{j,2} \neq X^{*,j,2}) \leq \beta_X(q_T)$, where β_X is the β -mixing coefficient of the process X as in (4).
3. For each $k \in \{1, 2\}$, $X^{*,1,k}, \dots, X^{*,p_T,k}$ are independent.

We denote by $\hat{\mu}_h^*$ the estimator computed using X_t^* instead of X_t and we write $\hat{\mu}_h^* = \frac{1}{2}(\hat{\mu}_h^{*(1)} + \hat{\mu}_h^{*(2)})$ to separate the part coming from $X^{*,1}$ (super-index (1)) and those coming from $X^{*,2}$ (super-index (2)), having $\hat{\mu}_h^{*(1)} := \frac{1}{p_T q_T} \sum_{j=1}^{p_T} \int_{2(j-1)q_T}^{(2j-1)q_T} \mathbb{K}_h(X_u^* - x) du$.

In a natural way we define moreover $\hat{\mu}_{h,\eta}^* := \mathbb{K}_\eta * \hat{\mu}_h^*$, which can be written again as $\frac{1}{2}(\hat{\mu}_{h,\eta}^{*(1)} + \hat{\mu}_{h,\eta}^{*(2)})$, to separate the contribution of $X^{*,1}$ and $X^{*,2}$.

With this background we can evaluate $\mathbb{E}[A(h)]$. We recall that, as defined in (12),

$$A(h) := \sup_{\eta \in \mathcal{H}_T} (\|\hat{\mu}_{h,\eta} - \hat{\mu}_\eta\|_A^2 - V(\eta))_+.$$

Now we can see $\hat{\mu}_{h,\eta} - \hat{\mu}_\eta$ as sum of different terms which we deal with singularly:

$$\hat{\mu}_{h,\eta} - \hat{\mu}_\eta := (\hat{\mu}_{h,\eta} - \hat{\mu}_{h,\eta}^*) + (\hat{\mu}_{h,\eta}^* - \mu_{h,\eta}) + (\mu_{h,\eta} - \mu_\eta) + (\mu_\eta - \hat{\mu}_\eta^*) + (\hat{\mu}_\eta^* - \hat{\mu}_\eta).$$

As a consequence of the triangular inequality and of the definition of $A(h)$ the following estimation holds true:

$$\begin{aligned} A(h) &\leq \sup_{\eta \in \mathcal{H}_T} [\|\hat{\mu}_{h,\eta} - \hat{\mu}_{h,\eta}^*\|_A^2 + (\|\hat{\mu}_{h,\eta}^* - \mu_{h,\eta}\|_A^2 - \frac{V(\eta)}{2})_+ + \|\mu_{h,\eta} - \mu_\eta\|_A^2 + (\|\mu_\eta - \hat{\mu}_\eta^*\|_A^2 - \frac{V(\eta)}{2})_+ + \|\hat{\mu}_\eta^* - \hat{\mu}_\eta\|_A^2] = \\ &= \sup_{\eta \in \mathcal{H}_T} [\sum_{j=1}^5 I_j^{h,\eta}] \end{aligned}$$

We start considering $I_5^{h,\eta}$. We define the set

$$\Omega^* := \{X_t = X_t^* \quad \forall t \in [0, T]\}.$$

As a consequence of the second property of the process X^* and of the β -mixing with exponential decay showed in Lemma 2 we get, recalling that $2p_T q_T = T$ (with q_T and p_T to be chosen),

$$\mathbb{P}(\Omega^{*c}) \leq 2p_T \beta_X(q_T) \leq c \frac{T}{q_T} e^{-\gamma q_T}. \quad (31)$$

From the definition of $\hat{\mu}_h^*$ and Jensen inequality it is

$$\begin{aligned} \|\hat{\mu}_\eta^* - \hat{\mu}_\eta\|_A^2 &= \int_A \left(\frac{1}{T} \int_0^T \mathbb{K}_\eta(X_t - x) - \mathbb{K}_\eta(X_t^* - x) dt \right)^2 dx 1_{\Omega^{*c}} \leq \\ &\leq c \int_A \frac{1}{T^2} T \left(\int_0^T \mathbb{K}_\eta^2(X_t - x) + \mathbb{K}_\eta^2(X_t^* - x) dt \right) dx 1_{\Omega^{*c}} \leq c \|\mathbb{K}_\eta\|_\infty^2 1_{\Omega^{*c}} \end{aligned} \quad (32)$$

By the definition (11) we get that, $\forall h \in \mathcal{H}_T$, $\|\mathbb{K}_h\|_\infty \leq (\prod_{l=1}^d h_l)^{-1}$.

We recall that, from how we have defined in (9) the set \mathcal{H}_T , we have that $\forall h \in \mathcal{H}_T$

$\prod_{l=1}^d h_l > \frac{(\log T)^{2d}}{T^{\frac{d}{3}}}$ and so

$$\|\mathbb{K}_\eta\|_\infty^2 < \frac{1}{(\prod_{l=1}^d \eta_l)^2} \leq \frac{T^{\frac{2d}{3}}}{(\log T)^{4d}}.$$

Replacing this bound in (32) it follows

$$\sup_{\eta \in \mathcal{H}_T} \|\hat{\mu}_\eta^* - \hat{\mu}_\eta\|_A^2 \leq c \frac{T^{\frac{2d}{3}}}{(\log T)^{4d}} 1_{\Omega^{*c}}.$$

We take its expectation and we use (31), getting a term which depends on q_T , a real to be chosen. From the arbitrariness of q_T we get a convergence to zero as fast as we want, for T going to ∞ . Indeed, taking $q_T := (\log T)^2$ yields $\forall h, \eta \in \mathcal{H}_T$,

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_5^{h,\eta}] = \mathbb{E}[\sup_{\eta \in \mathcal{H}_T} \|\hat{\mu}_\eta^* - \hat{\mu}_\eta\|_A^2] \leq c \frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma(\log T)^2}. \quad (33)$$

Regarding $\sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta}$, we estimate it through (33) and the following lemma, which will be proven in the appendix.

Lemma 3. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded, measurable function with support \mathcal{S} satisfying $\text{diam}(\mathcal{S}) \leq 2\sqrt{d}$; \tilde{A} a compact set in \mathbb{R}^d such that $A \subset \tilde{A}$ and $\tilde{A} = \{\zeta : d(\zeta, A) \leq 2\sqrt{d}\}$ and g a function in $L^2(\tilde{A})$. Then,*

$$\|f * g\|_A \leq \|f\|_{1, \mathbb{R}^d} \|g\|_{2, \tilde{A}},$$

where we have denoted as $\|\cdot\|_A$ the usual L^2 norm on A , $\|\cdot\|_{1, \mathbb{R}^d}$ the L^1 norm on \mathbb{R}^d and $\|\cdot\|_{2, \tilde{A}}$ the L^2 norm on \tilde{A} .

We recall that $\hat{\mu}_{h,\eta} = \mathbb{K}_\eta * \hat{\mu}_h$ and $\hat{\mu}_{h,\eta}^* = \mathbb{K}_\eta * \hat{\mu}_h^*$. Therefore, remarking that $\text{diam}(K) \leq 2$ and so by the definition of \mathbb{K}_η it is $\text{diam}(\mathbb{K}_\eta) \leq 2\sqrt{d}$, we can use Lemma 3, which yields

$$\sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta} = \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta * (\hat{\mu}_h - \hat{\mu}_h^*)\|_A^2 \leq \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta\|_{1, \mathbb{R}^d}^2 \|\hat{\mu}_h - \hat{\mu}_h^*\|_{\tilde{A}}^2.$$

Taking the expected value, using that $\|\mathbb{K}_\eta\|_{1, \mathbb{R}^d} \leq c \forall \eta \in \mathcal{H}_T$ and the equation (33), remarking that the dependence on the integration set considered is hidden in the constant c in which this time will appear $|\tilde{A}|$ instead of $|A|$ we get

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta}] \leq c \frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma(\log T)^2}. \quad (34)$$

We still use Lemma 3 to study $\sup_{\eta \in \mathcal{H}_T} I_3^{h,\eta}$, recalling that $\mu_{h,\eta} = \mathbb{K}_\eta * \mu_h$ and $\mu_\eta = \mathbb{K}_\eta * \mu_b$. It yields

$$\sup_{\eta \in \mathcal{H}_T} I_3^{h,\eta} = \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta * (\mu_h - \mu)\|_A^2 \leq \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta\|_{1, \mathbb{R}^d}^2 \|\mu_h - \mu\|_{\tilde{A}}^2 \leq c \|\mu_h - \mu\|_{\tilde{A}}^2 = cB(h). \quad (35)$$

We are left to study $I_2^{h,\eta}$ and $I_4^{h,\eta}$, for which we need the following lemma that will be showed right after the proof of this proposition.

Lemma 4. *For $i = 1, 2$, there exist some positive constants c_1^*, c_2^*, c_3^* and a constant k_0^* such that, for any $\bar{k} \geq k_0^*$,*

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\|\mu_\eta - \hat{\mu}_\eta^{*(i)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq c_1^* T^{c_2^*} e^{-c_3^* (\log T)^2}. \quad (36)$$

Moreover there exist k_0 such that, for any $\bar{k} \geq k_0$,

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\|\mu_{h,\eta} - \hat{\mu}_{h,\eta}^{*(i)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq c_1^* T^{c_2^*} e^{-c_3^* (\log T)^2}. \quad (37)$$

Concerning $I_4^{h,\eta}$ observe it is $\mu_\eta - \hat{\mu}_\eta^* = \frac{1}{2}(2\mu_\eta - \hat{\mu}_\eta^{*(1)} - \hat{\mu}_\eta^{*(2)})$. Hence, from triangular inequality and the definition of positive part function, we get

$$I_4^{h,\eta} \leq c(\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 - \frac{V(\eta)}{2})_+ + c(\|\mu_\eta - \hat{\mu}_\eta^{*(2)}\|_A^2 - \frac{V(\eta)}{2})_+.$$

From (36), for a k in the definition of $V(\eta)$ big enough, for which we have $\frac{k}{2} > (k_0^* \vee k_0)$, we get

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_4^{h,\eta}] \leq c_1 T^{c_2} e^{-c_3(\log T)^2}. \quad (38)$$

In the same way, remarking that

$$I_2^{h,\eta} \leq c(\|\hat{\mu}_{h,\eta}^{*(1)} - \mu_{h,\eta}\|_A^2 - \frac{V(\eta)}{2})_+ + c(\|\hat{\mu}_{h,\eta}^{*(2)} - \mu_{h,\eta}\|_A^2 - \frac{V(\eta)}{2})_+$$

and using (37) it follows

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_2^{h,\eta}] \leq c_1 T^{c_2} e^{-c_3(\log T)^2}. \quad (39)$$

From (33), (34), (35), (38) and (39) we obtain, for any $h \in \mathcal{H}_T$,

$$\mathbb{E}[A(h)] \leq c \frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma(\log T)^2} + cB(h) + c_1 T^{c_2} e^{-c_3(\log T)^2} \leq cB(h) + c_1 e^{-c_2(\log T)^2},$$

as we wanted. \square

To conclude the proof of the adaptive procedure we need to show Lemma 4, which core is the use of the Talagrand inequality. First of all, we recall the following version of the Talagrand inequality, which has been stated as Lemma 2 in [8] and which is a straightforward consequence of the Talagrand inequality given in Klein and Rio [14].

Lemma 5. *Let T_1, \dots, T_n be independent random variables with values in some Polish space \mathcal{X} , \mathcal{R} a countable class of measurable functions from \mathcal{X} into $[-1, 1]^p$ and $v_p(r) := \frac{1}{p} \sum_{j=1}^p [r(T_j) - \mathbb{E}[r(T_j)]]$. Then,*

$$\mathbb{E}[(\sup_{r \in \mathcal{R}} |v_p(r)|^2 - 2H^2)_+] \leq c(\frac{v}{p} e^{-c \frac{pH^2}{v}} + \frac{M^2}{p^2} e^{-c \frac{pH}{M}}), \quad (40)$$

with c a universal constant and where

$$\sup_{r \in \mathcal{R}} \|r\|_\infty \leq M, \quad \mathbb{E}[\sup_{r \in \mathcal{R}} |v_p(r)|] \leq H, \quad \sup_{r \in \mathcal{R}} \frac{1}{p} \sum_{j=1}^p \text{Var}(r(T_j)) \leq v.$$

6.4 Proof of Lemma 4

Proof. Since the two cases $i = 1$ and $i = 2$ are similar, we study only one of them. We start proving (36), the proof of inequality (37) follows the same line. We first observe it is

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq \sum_{\eta \in \mathcal{H}_T} \mathbb{E}[(\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+]. \quad (41)$$

Our goal is now to find a bound for the right hand side of the inequality here above using the version of the Talagrand inequality gathered in Lemma 5. To do it, we need to introduce some notation.

We observe that $\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 = \sup_{r, \|r\|=1} < \mu_\eta - \hat{\mu}_\eta^{*(1)}, r >^2$, and the supremum can be considered over a countable dense set of function r such that $\|r\| = 1$; let us denote this set by $\mathcal{B}(1)$.

We define

$$T_j(z) := \frac{1}{qT} \int_{(2j-1)qT}^{2jqT} \mathbb{K}_\eta(X_t^{*,j,1} - z) dt; \quad r(T_j) := \int_A T_j(z) r(z) dz.$$

Thus

$$v_{p_T}(r) = \langle \hat{\mu}_\eta^{*(1)} - \mu_\eta, r \rangle = \frac{1}{p_T} \sum_{j=1}^{p_T} [r(T_j) - \mathbb{E}[r(T_j)]]$$

is a centered empirical process with independent variables

$$\psi_r(X^{*j,1}) := r(T_j) - \mathbb{E}[r(T_j)] = \int_A \left(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} [\mathbb{K}_\eta(X_t^{*j,1} - z) - \mathbb{E}[\mathbb{K}_\eta(X_t^{*j,1} - z)]] dt \right) r(z) dz,$$

to which we want to apply Talagrand inequality (40). Therefore, we have to compute M , H and v as defined in Lemma 5. We start by the calculation of M . For any $r \in \mathcal{B}(1)$ it is, using the definition of r and Cauchy - Schwartz inequality,

$$\left| \int_A T_j(z) r(z) dz \right| \leq \left(\int_A T_j^2(z) dz \right)^{\frac{1}{2}}.$$

Now from the definition of T and Jensen inequality it follows

$$\int_A T_j^2(z) dz \leq \int_A \frac{1}{q_T^2} q_T \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta^2(X_t^{*j,1} - z) dt dz \leq c \|K_\eta\|_\infty^2 |\mathcal{S}| \leq c \left(\prod_{l=1}^d \eta_l \right)^{-2} \left(\prod_{l=1}^d \eta_l \right) = c \left(\prod_{l=1}^d \eta_l \right)^{-1},$$

where we have also used that the support of K_η is on \mathcal{S} which size is $\prod_{l=1}^d \eta_l$. Hence,

$$\left| \int_A T_j(z) r(z) dz \right| \leq c \left(\prod_{l=1}^d \eta_l \right)^{-\frac{1}{2}} =: M. \quad (42)$$

Regarding the computation of H , from the definition of $v_{p_T}(r)$ and the fact that the random variables $\psi_r^{*j,1}$ are centered and independents it follows

$$\begin{aligned} \mathbb{E} \left[\sup_{r \in \mathcal{B}(1)} |v_{p_T}(r)|^2 \right] &= \mathbb{E}_b \left[\left\| \mu_\eta - \hat{\mu}_\eta^{*(1)} \right\|_A^2 \right] = \int_A \text{Var} \left(\frac{1}{p_T} \sum_{j=1}^{p_T} \frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} (\mathbb{K}_\eta(X_t^{*j,1} - z) - \mathbb{E}[\mathbb{K}_\eta(X_t^{*j,1} - z)]) dt \right) dz = \\ &= \int_A \frac{1}{p_T} \text{Var} \left(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z) dt \right) dz \leq c |A| \frac{1}{p_T} \frac{1}{q_T} \left(\prod_{l=1}^d \eta_l \right)^{\frac{2}{d}-1}, \end{aligned}$$

where in the last inequality we have used the estimation for the variance in the case $d \geq 3$ gathered in Proposition 1, considering that taking the Kernel function as f we have that its support \mathcal{S} is such that $|\mathcal{S}| \leq c \left(\prod_{l=1}^d \eta_l \right)$. It yields

$$\mathbb{E} \left[\sup_{r \in \mathcal{B}(1)} |v_{p_T}(r)|^2 \right] \leq \frac{c}{T} \left(\prod_{l=1}^d \eta_l \right)^{\frac{2}{d}-1} =: H^2 \quad (43)$$

In order to use Lemma 5, we are left to compute v .

We observe it is

$$\frac{1}{p_T} \sum_{j=1}^{p_T} \text{Var} \left(\int_A \frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z) dt r(z) dz \right) = \frac{1}{p_T} \sum_{j=1}^{p_T} \text{Var} \left(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} (\mathbb{K}_\eta * r)(X_t^{*j,1}) dt \right).$$

We want to prove a tight upper bound for the variance of the integral functional $\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1}) dt$ of the diffusion X^* , where we have denoted $f_\eta := \mathbb{K}_\eta * r$.

Following the proof we have given of Proposition 1 we have,

$$\begin{aligned} \text{Var} \left(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1}) dt \right) &\leq \frac{c}{q_T^2} \int_0^{q_T} (q_T - u) \mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})] du = \\ &= \frac{c}{q_T^2} \left(\int_0^D (q_T - u) \mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})] du + \int_D^{q_T} (q_T - u) \mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})] du \right), \end{aligned}$$

where we have introduced $f_{\eta,c}(x) := f_\eta(x) - \mu(f_\eta)$ and D a quantity to be chosen in order to balance the contribution of the two integrals here above. We denote moreover as P_t^* the transition

semigroup of the process X^* . Concerning the integral between 0 and D , we act like we did in (15) and we use Cauchy - Schwartz inequality and the fact that P_t^* is a contraction. It follows

$$\begin{aligned} \frac{c}{q_T^2} \int_0^D (q_T - u) \mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})] du &\leq \frac{c}{q_T} \left| \int_0^D \langle P_t^* f_\eta, f_\eta \rangle_\mu dt \right| \leq \\ &\leq \frac{c}{q_T} \int_0^D (\|P_t^* f_\eta\|_\mu^2 \|f_\eta\|_\mu^2)^{\frac{1}{2}} dt \leq \frac{c}{q_T} \int_0^D \|f_\eta\|_\mu^2 dt \leq \frac{cD}{q_T}, \end{aligned} \quad (44)$$

where in the last inequality we have used the fact that $\|\mu\|_\infty \leq c$, Young inequality and the definition of the Kernel function and of r in order to say

$$\|f_\eta\|_\mu^2 = \|\mathbb{K}_\eta * r\|_\mu^2 \leq c \|\mathbb{K}_\eta * r\|_{2,\mathbb{R}^d}^2 \leq c \|\mathbb{K}_\eta\|_{1,\mathbb{R}^d}^2 \|r\|_{2,\mathbb{R}^d}^2 \leq c. \quad (45)$$

Regarding the integral between D and q_T , we act like we did in (19) using the exponential ergodicity gathered in Lemma 2 to get

$$\frac{c}{q_T^2} \int_D^{q_T} (q_T - u) \mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})] du \leq \frac{c}{q_T} \left| \int_D^{q_T} \langle P_t^* f_{\eta,c}, f_{\eta,c} \rangle_\mu dt \right| \leq \frac{c}{q_T} \int_D^{q_T} e^{-\rho t} \|f_{\eta,c}\|_\infty^2 dt. \quad (46)$$

We now recall that $f_{\eta,c}(x) = f_\eta(x) - \mu(f_\eta)$ with $\mu(f_\eta) = \int_{\mathbb{R}^d} f_\eta(x) \mu(x) dx$. Hence, from Cauchy-Schwartz inequality and (45), we get $|\mu(f_\eta)| \leq c$ and therefore

$$|f_{\eta,c}(x)| \leq |f_\eta(x)| + c.$$

To estimate the infinity norm of $f_{\eta,c}$ we remark it is, $\forall y \in \mathbb{R}^d$,

$$|f_\eta(y)| = |\mathbb{K}_\eta * r(y)| = \left| \int_A \mathbb{K}_\eta(y-z) r(z) dz \right| \leq \left(\int_A \mathbb{K}_\eta^2(y-z) dz \right)^{\frac{1}{2}} \left(\int_A r^2(z) dz \right)^{\frac{1}{2}} = \left(\int_S \mathbb{K}_\eta^2(y-z) dz \right)^{\frac{1}{2}},$$

where we have used Cauchy - Schwartz inequality and the fact that $r \in \mathcal{B}(1)$ and so its 2 - norm is equal to 1 by definition. Moreover, the Kernel function is different from 0 only on its support \mathcal{S} , which size is $\prod_{l=1}^d \eta_l$. Hence, recalling also that the infinite norm of \mathbb{K}_η is upper bounded by $c(\prod_{l=1}^d \eta_l)^{-1}$, we get

$$|\mathbb{K}_\eta * r(y)| \leq c(\|\mathbb{K}_\eta\|_\infty^2 |\mathcal{S}|)^{\frac{1}{2}} \leq \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}}.$$

It follows

$$\|f_{\eta,c}\|_\infty \leq \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}} + c \leq \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}}, \quad (47)$$

given that the second term is negligible compared to the first. Replacing (47) in (46), using also (44), we obtain

$$\text{Var}\left(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1}) dt\right) \leq \frac{c}{q_T} \left(D + \frac{e^{-\rho D}}{\prod_{l=1}^d \eta_l}\right).$$

We look for a D for which the first and the second term of the right hand side of the inequality here above have the same magnitude. Therefore, we choose $D := [\max(-\frac{1}{\rho} \log(\prod_{l=1}^d \eta_l), 1)] \wedge q_T$.

Replacing such a value we get, if $q_T > -\frac{1}{\rho} \log(\prod_{l=1}^d \eta_l)$,

$$\frac{1}{p_T} \sum_{j=1}^{p_T} \text{Var}\left(\int_A \frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z) dt r(z) dz\right) \leq \frac{c}{q_T} \left(1 + \log\left(\frac{1}{|\prod_{l=1}^d \eta_l|}\right)\right)$$

Otherwise, if $q_T \leq -\frac{1}{\rho} \log(\prod_{l=1}^d \eta_l)$, by the definition of D we have $D = q_T$. We still have the contribution of $\frac{c}{q_T} D$ which is in this case less than $\frac{c}{q_T} (\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))$ and, moreover, the contribution of the integral between D and q_T is now null since we have $D = q_T$. Hence we have

$$v := \frac{c}{q_T} \left(1 + \log\left(\frac{1}{|\prod_{l=1}^d \eta_l|}\right)\right). \quad (48)$$

We use Lemma 5 on the right hand side of (41), recalling that $\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 = \sup_{r \in \mathcal{B}(1)} < \mu_\eta - \hat{\mu}_\eta^{*(1)}, r >^2 = \sup_{r \in \mathcal{B}(1)} |v_{p_T}(r)|^2$; with M , H and v as found in (42), (43) and (48). It follows

$$\begin{aligned} & \mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq \\ & \leq c \sum_{\eta \in \mathcal{H}_T} \frac{(1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}{p_T q_T} e^{-c \frac{p_T (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1}}{q_T (1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}} + \frac{(\prod_{l=1}^d \eta_l)^{-1}}{p_T^2} e^{-c \frac{p_T \frac{1}{\sqrt{T}} (\prod_{l=1}^d \eta_l)^{\frac{1}{d}-\frac{1}{2}}}{(\prod_{l=1}^d \eta_l)^{-\frac{1}{2}}}}. \end{aligned}$$

We recall that $2p_T q_T = T$, where q_T is chosen above equation (33) as $(\log T)^2$; we can therefore upper bound the right hand side of the equation here above with

$$\begin{aligned} & c \sum_{\eta \in \mathcal{H}_T} \frac{(1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}{T} e^{-\frac{c}{(\prod_{l=1}^d \eta_l)^{1-\frac{2}{d}} (1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}} + \frac{(\log T)^4}{(\prod_{l=1}^d \eta_l) T^2} e^{-c \frac{\sqrt{T}}{(\log T)^2} (\prod_{l=1}^d \eta_l)^{\frac{1}{d}}} \leq \\ & \leq (\frac{\log T}{T} e^{-c(\log T)^2} + \frac{T^{\frac{d}{3}-2}}{(\log T)^{2d-2}} e^{-cT^{\frac{1}{6}}}) |\mathcal{H}_T|, \end{aligned}$$

where in the last inequality we have used that, by the definition (9) we have given of \mathcal{H}_T , $\forall h \in \mathcal{H}_T$ we have $\frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d h_l \leq (\frac{1}{\log T})^{\frac{3d}{d-2}}$ and so $(\prod_{l=1}^d \eta_l)^{1-\frac{2}{d}} (1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|})) \leq c \frac{1}{(\log T)^2}$ and $\frac{\sqrt{T}}{(\log T)^2} (\prod_{l=1}^d \eta_l)^{\frac{1}{d}} \geq cT^{\frac{1}{6}}$; we have therefore upper bounded each element of the sum with a quantity which does not depend on η .

We have moreover assumed that $|\mathcal{H}_T|$ has polynomial growth in T ; and so there is a constant $c > 0$ such that

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_A^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq (\frac{\log T}{T} e^{-c(\log T)^2} + \frac{T^{\frac{d}{3}-2}}{(\log T)^{2d-2}} e^{-cT^{\frac{1}{6}}}) T^c;$$

inequality (36) follows.

As a consequence of Lemma 3 and the definition of Kernel function we moreover have

$$\|\mu_{h,\eta} - \hat{\mu}_{h,\eta}^{*(1)}\|_A^2 = \|\mathbb{K}_h * (\mu_\eta - \hat{\mu}_\eta^{*(1)})\|_A^2 \leq \|\mathbb{K}_h\|_{1,\mathbb{R}^d}^2 \|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_{2,\tilde{A}}^2 \leq c \|\mu_\eta - \hat{\mu}_\eta^{*(1)}\|_{2,\tilde{A}}^2.$$

Using that and (36) we have just shown we obtain (37). □

A Appendix

A.1 Proof of Lemma 1

Proof. Lemma 1 relies on the first point of Theorem 1.1 in [4], which needs the following assumptions on the coefficients $\tilde{a} = a \cdot a^T$, b , γ and on the jumps.

($H^{\tilde{a}}$) There are $c_1 > 0$ and $\beta \in (0, 1)$ such that for all $x, y \in \mathbb{R}^d$, $|\tilde{a}(x) - \tilde{a}(y)| \leq c_1 |x - y|^\beta$ and, for some $c_2 \geq 1$, $c_2^{-1} \mathbb{I}_{d \times d} \leq \tilde{a}(x) \leq c_2 \mathbb{I}_{d \times d}$.

(H^k) The function $k(x, z) := \frac{1}{|Det(\gamma(x))|} |z|^{d+\alpha} |F(\gamma(x)^{-1} z)|$ is bounded, measurable and, if $\alpha = 1$, for any $0 < r < R < \infty$ it is

$$\int_{r \leq |z| \leq R} z k(x, z) |z|^{-d-1} dz = 0. \quad (49)$$

(H^b) The function b belongs to the Kato class \mathbb{K}^2 which is, as defined in [4],

$$\mathbb{K}^\lambda := \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} \text{ satisfies } \limsup_{\delta \rightarrow 0} \int_0^\delta \int_{\mathbb{R}^d} |f(x \pm y)| \eta_{\lambda, \lambda-1}(s, y) dy ds = 0 \right\},$$

where we have denoted $f(x \pm y)$ as an abbreviation for $f(x + y) + f(x - y)$ and

$$\eta_{\lambda, \gamma}(t, x) := t^{\frac{\gamma}{2}} (|x| + t^{\frac{1}{2}})^{-d-\lambda}.$$

We remark that, in [4], the authors assumed $d \geq 2$. After inspection of the proof it is possible to see that the result can be extended to the case $d = 1$: it was stated for $d \geq 2$ for the convenience in describing the Kato class function in K_2 (for the drift).

Through this paper we have assumed that Assumptions A1 - A3 hold.

From A1 it follows $H^{\tilde{a}}$ since we have asked the boundedness of \tilde{a} and, in the case in which x and y are such that $|x - y| > 1$ we have that $\exists c$ such that $|\tilde{a}(x) - \tilde{a}(y)| \leq |\tilde{a}(x)| + |\tilde{a}(y)| \leq 2c \leq 2c|x - y|^\beta$, for each $\beta \in (0, 1)$. When $|x - y| \leq 1$, instead, we have $|\tilde{a}(x) - \tilde{a}(y)| \leq L|x - y| = L|x - y|^{1-\beta}|x - y|^\beta \leq L|x - y|^\beta$, as consequence of the Lipschitz continuity.

Regarding H^k , we have that k is a bounded function as a consequence of second and third points of A3. Indeed,

$$|k(x, z)| = \frac{1}{|Det(\gamma(x))|} |z|^{d+\alpha} |F(\gamma(x)^{-1}z)| \leq \frac{1}{|Det(\gamma(x))|} |z|^{d+\alpha} \frac{|\gamma(x)|^{d+\alpha}}{|z|^{d+\alpha}} \leq \frac{1}{c_2} \gamma_{max}^{d+\alpha} < \infty.$$

Observing moreover that (49) holds true on the basis of the fourth point of A3, H^k clearly follows. We are left to show that $b \in \mathbb{K}^2$. We start considering the case $d \geq 2$. Noticing that (cf also Remark 2.6 in [4])

$$\int_0^\delta \eta_{\lambda, \lambda-1}(s, y) ds = \int_0^\delta \frac{s^{\frac{(\lambda-1)}{2}}}{(|y| + s^{\frac{1}{2}})^{d+\lambda}} ds \leq c \frac{(|y|^2 \wedge \delta)^{\frac{(1+\lambda)}{2}}}{|y|^{d+\lambda}} = \frac{1}{|y|^{d-1}} (1 \wedge \frac{\delta}{|y|^2})^{\frac{(1+\lambda)}{2}}, \quad (50)$$

it is enough to show that

$$M_b^2(\delta) := \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |b(x+y)| \frac{1}{|y|^{d-1}} (1 \wedge \frac{\delta}{|y|^2})^{\frac{3}{2}} dy \longrightarrow 0 \quad \text{as } \delta \rightarrow 0$$

to get $b \in \mathbb{K}^2$.

From Assumption A1 we know b is upper bounded by a constant. Therefore we have

$$\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |b(x+y)| \frac{1}{|y|^{d-1}} (1 \wedge \frac{\delta}{|y|^2})^{\frac{3}{2}} dy \leq c \int_{\{y: |y| \leq \sqrt{\delta}\}} \frac{1}{|y|^{d-1}} dy + c \int_{\{y: |y| > \sqrt{\delta}\}} \frac{1}{|y|^{d-1}} (\frac{\delta}{|y|^2})^{\frac{3}{2}} dy. \quad (51)$$

We move to polar coordinates system for $d > 1$ (otherwise it is immediate), getting the right hand side of the equation (51) here above is upper bounded by

$$\int_0^{\sqrt{\delta}} c \rho d\rho + c \delta^{\frac{3}{2}} \int_{\sqrt{\delta}}^\infty \rho^{-3} d\rho = c\sqrt{\delta} + c\delta^{\frac{3}{2}} \frac{1}{2\delta},$$

which clearly goes to 0 for $\delta \rightarrow 0$. It yields that H^b holds true.

It entails we can use Theorem 1.1 of [4]; Lemma 1 follows. □

A.2 Proof of Lemma 2

Proof. The exponential ergodicity and the exponential β -mixing of the process X are showed in Proposition 3.8 and the second point of Theorem 2.2 of [21].

To use them we have to show that Assumptions 1, 2 and 3* stated in [21] hold.

Assumption 1 of [21] is a regularity condition which corresponds to our assumption A1.

We want to show that point (b) of Assumption 2 of [21] holds, which is the following:

(b) There exists a constant $\Delta > 0$ such that X_Δ admits a density $p_\Delta(x, y)$ with respect to the Lebesgue measure on \mathbb{R}^d for every $x \in \mathbb{R}^d$, and $(x, y) \mapsto p_\Delta(x, y)$ is bounded in $y \in \mathbb{R}^d$ and in $x \in K$ for every compact $K \subset \mathbb{R}^d$. Moreover, for every $x \in \mathbb{R}^d$ and every open ball $U \subset \mathbb{R}^d$ there exists a point $z = z(x, U) \in \text{supp}(F)$ such that $\gamma(x) \cdot z \in U$.

We observe that the existence of a bounded density has already been proven in Lemma 1. Moreover, from second and third points of A3, we know that $\text{supp}(F) = \mathbb{R}^d$ and that γ is an invertible matrix. Hence, for every $x \in \mathbb{R}^d$ and every open ball $U \subset \mathbb{R}^d$ there exists a point $z = z(x, U) \in \mathbb{R}^d$ such that $\gamma(x) \cdot z \in U$.

To conclude, we have to prove that Assumption 3* holds and so we have to show the existence of a

Lyapunov function. We therefore want to provide a function f^* which satisfies the drift condition $Af^* \leq -c_1 f^* + c_2$ for $c_1 > 0$ and $c_2 > 0$. A denotes the generator of the diffusion, which is the sum of the continuous and discrete part

$$A_c f(x) := \frac{1}{2} \sum_{i,j=1}^d \tilde{a}_{i,j}(x) \partial_{i,j}^2 f(x) + \sum_{i=1}^d b_i(x) \partial_i f(x) \quad \text{and}$$

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

for every function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $f \in \mathcal{C}^2(\mathbb{R}^d)$.

From the fifth point of condition A3 we know there exists $\epsilon_0 > 0$ such that $\int_{\mathbb{R}^d} |z|^2 e^{\epsilon_0 |z|} F(z) dz \leq c_0$. For $\epsilon \leq \min(\frac{\epsilon_0}{\|\gamma\|_\infty}, \epsilon_0)$ and for x such that $|x| \geq 1$ we define $f^*(x) := e^{\epsilon|x|}$, for $|x| < 1$ we take it such that $f^*(x)$ is C^∞ on \mathbb{R}^d . We have built $f^*(x)$ as a piecewise function to avoid any regularity problem in 0. We observe that, when x is such that $|x| \geq 1$, it is $\partial_i f^*(x) = \epsilon e^{\epsilon|x|} \frac{x_i}{|x|}$ and

$$\partial_{i,j}^2 f^*(x) = \epsilon \frac{x_i x_j}{|x|^2} e^{\epsilon|x|} (\epsilon - \frac{1}{|x|}) + \epsilon e^{\epsilon|x|} \frac{1}{|x|} 1_{j=i}. \quad (52)$$

We therefore have, using also the drift condition gathered in assumption A2, $\forall x : |x| > \bar{\rho} \wedge 2$,

$$|A_c f^*(x)| \leq \frac{1}{2} \epsilon e^{\epsilon|x|} (\epsilon + \frac{2}{|x|}) \sum_{i,j=1}^d |\tilde{a}_{i,j}(x)| + \epsilon e^{\epsilon|x|} \frac{1}{|x|} \langle x, b(x) \rangle \leq$$

$$\leq c \epsilon e^{\epsilon|x|} (\epsilon + \frac{1}{|x|}) \sum_{i,j=1}^d |\tilde{a}_{i,j}(x)| - \tilde{C} \epsilon e^{\epsilon|x|}. \quad (53)$$

Concerning the discrete part of the generator, from intermediate value theorem we have

$$|A_d f^*(x)| \leq \int_{\mathbb{R}^d} \int_0^1 (\gamma(x) \cdot z)^T \cdot H^2 f_{(x+s\gamma(x) \cdot z)} \cdot (\gamma(x) \cdot z) ds F(z) dz,$$

where $H^2 f_{(x+s\gamma(x) \cdot z)}$ denotes the hessian matrix of the function f computed in the point $x+s\gamma(x) \cdot z$. We now split the integral in the right hand side here above, to act differently depending on whether $|z|$ is more or less than $\frac{|x|}{2\|\gamma\|_\infty}$. We therefore get

$$|A_d f^*(x)| \leq \int_{z: |z| \leq \frac{|x|}{2\|\gamma\|_\infty}} \int_0^1 (\gamma(x) \cdot z)^T \cdot H^2 f_{(x+s\gamma(x) \cdot z)} \cdot (\gamma(x) \cdot z) ds F(z) dz +$$

$$+ \int_{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}} \int_0^1 (\gamma(x) \cdot z)^T \cdot H^2 f_{(x+s\gamma(x) \cdot z)} \cdot (\gamma(x) \cdot z) ds F(z) dz =: I_1 + I_2.$$

Concerning I_1 , we first of all observe that, as $|z| \leq \frac{|x|}{2\|\gamma\|_\infty}$ and $|x| > 2$, it is $|x + s\gamma(x) \cdot z| > 1$. From (52) it follows

$$I_1 \leq c \int_{z: |z| \leq \frac{|x|}{2\|\gamma\|_\infty}} \int_0^1 |z|^2 \|\gamma\|_\infty^2 \epsilon e^{\epsilon|x+s\gamma(x) \cdot z|} (\epsilon + \frac{1}{|x+s\gamma(x) \cdot z|}) F(z) dz ds \leq$$

$$\leq c \epsilon e^{\epsilon|x|} \int_{z: |z| \leq \frac{|x|}{2\|\gamma\|_\infty}} |z|^2 e^{\epsilon|z|\|\gamma\|_\infty} (\epsilon + \frac{1}{|x|-|z|\|\gamma\|_\infty}) F(z) dz \leq c \epsilon e^{\epsilon|x|} (\epsilon + \frac{1}{|x|}), \quad (54)$$

where in the last inequality we have used the fifth point of A3 and the fact that on the integral we are considering it is $|z| \leq \frac{|x|}{2\|\gamma\|_\infty}$ and so $\frac{1}{|x|-|z|\|\gamma\|_\infty} \leq \frac{2}{|x|}$.

We now study the term I_2 , which is $I_{2,1} + I_{2,2} :=$

$$\int_{\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}, |x+s\gamma(x) \cdot z| \leq 1\}} \int_0^1 (\gamma(x) \cdot z)^T \cdot H^2 f_{(x+s\gamma(x) \cdot z)} \cdot (\gamma(x) \cdot z) ds dz +$$

$$+ \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}, |x + sz\gamma(x)| > 1\right\}} \int_0^1 (\gamma(x) \cdot z)^T \cdot H^2 f_{(x+sz\gamma(x) \cdot z)} \cdot (\gamma(x) \cdot z) ds dz.$$

On $I_{2,1}$ we can upper bound the hessian matrix with c and so we get

$$\begin{aligned} I_{2,1} &\leq c \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}, |x + sz\gamma(x)| \leq 1\right\}} |z|^2 \|\gamma\|_\infty^2 F(z) dz \\ &\leq c \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}\right\}} |z|^2 e^{\epsilon|z|} e^{-\epsilon|z|} F(z) dz \leq c e^{-\epsilon \frac{|x|}{2\|\gamma\|_\infty}} \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}\right\}} |z|^2 e^{\epsilon|z|} F(z) dz \leq c e^{-\epsilon \frac{|x|}{2\|\gamma\|_\infty}}, \end{aligned} \quad (55)$$

where we have used that $|z| > \frac{|x|}{2\|\gamma\|_\infty}$; after that we have enlarged the domain of integration and used the fifth point of Assumption A3 to upper bound the integral with a constant. On $I_{2,2}$ we still use (52), getting

$$\begin{aligned} I_{2,2} &\leq c \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}, |x + sz\gamma(x)| > 1\right\}} \int_0^1 |z|^2 \|\gamma\|_\infty^2 \epsilon e^{\epsilon|x+sz\gamma(x)|} \left(\epsilon + \frac{1}{|x + sz\gamma(x)|}\right) F(z) dz ds \leq \\ &\leq c(\epsilon + 1) \epsilon e^{\epsilon|x|} \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}, |x + sz\gamma(x)| > 1\right\}} |z|^2 e^{\epsilon|z|} \|\gamma\|_\infty F(z) dz. \end{aligned} \quad (56)$$

We define $J(x) := \int_{\left\{z: |z| > \frac{|x|}{2\|\gamma\|_\infty}\right\}} |z|^2 e^{2\epsilon|z|} \|\gamma\|_\infty F(z) dz$. From (54) (55) and (56), observing that the domain of integration of the integral defined as J contains the one in (56), we get

$$|A_d f^*(x)| \leq c e^{\epsilon|x|} (\epsilon^2 + \frac{\epsilon}{|x|} + e^{-c\epsilon|x|} + \epsilon(\epsilon + 1)J(x)). \quad (57)$$

From (53) and (57) we get, using also the boundedness of a and J which follows from A1 and the fifth point of A3,

$$|A f^*(x)| \leq \epsilon e^{\epsilon|x|} (c\epsilon - \tilde{C}) + c e^{\epsilon|x|} \left(\frac{\epsilon}{|x|} + e^{-c\epsilon|x|} + \epsilon J(x)\right). \quad (58)$$

Since $\epsilon > 0$ can be chosen small; $c\epsilon$ is less than \tilde{C} and so the first term here above turns out being negative.

Moreover, the second term on the right hand side of (58) is $o(f^*)$. Indeed, $c\epsilon(\frac{1}{|x|} + e^{-c\epsilon|x|})$ clearly goes to 0 for $|x| \rightarrow \infty$ and the domain of integration of the integral defined as $J(x)$ is the set of z such that $|z| > \frac{|x|}{2\|\gamma\|_\infty}$. Therefore, for $|x| \rightarrow \infty$ the contribution of the integral becomes null.

It follows $A f^* \leq -c_1 f^* + o(f^*)$, as we wanted.

We get the drift condition holding true on the proposed function f^* and so the Assumption 3* does. Therefore, using Proposition 3.8 in [21] we know that there exist a finite-valued function h on \mathbb{R}^d and a positive constant ρ for which

$$\|P_t(x, \cdot) - \mu(\cdot)\|_{TV} \leq h(x) e^{-\rho t}.$$

Proposition 3.8 in [21] follows from Theorem 5.1 and Theorem 6.1 in [23], the latter providing a specific choice $h(x) = 1 + f^*(x)$. Hence

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

We get for free that $\mu(f^*) < \infty$. Indeed, as Assumptions 1,2 and 3* in [21] hold true, we can also use Theorem 2.2 therein. Its second point provides $\mu(f^*) \leq \frac{c_2}{c_1}$. Hence, for any function f such that $\mu(f) = 0$, it is

$$\|P_t(x, f)\|_{TV} \leq (1 + f^*(x)) e^{-\rho t} \|f\|_\infty.$$

We get

$$\|P_t f\|_{L^1(\mu)} = \int_{\mathbb{R}^d} |P_t f(x)| \mu(x) dx \leq \|f\|_\infty e^{-\rho t} \int_{\mathbb{R}^d} |1 + f^*(x)| \mu(x) dx \leq c \|f\|_\infty e^{-\rho t},$$

as we wanted. □

A.3 Proof of Lemma 3

Proof. We observe that, from the definition of convolution and Cauchy-Schwartz inequality it is

$$\begin{aligned} |f * g(x)| &= \int_{\mathbb{R}^d} |f(x-y)| |g(y)| dy = \int_{\mathbb{R}^d} |f(x-y)|^{\frac{1}{2}} |g(y)| |f(x-y)|^{\frac{1}{2}} dy \leq \\ &\leq c \left(\int_{\mathbb{R}^d} |f(x-y)| |g(y)|^2 dy \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} |f(x-y)| dy \right)^{\frac{1}{2}} = c \left(\int_{\mathbb{R}^d} |f(x-y)| |g(y)|^2 dy \right)^{\frac{1}{2}} \|f\|_{1, \mathbb{R}^d}^{\frac{1}{2}}. \end{aligned}$$

Therefore the following bound on the L^2 norm on A holds true:

$$\|f * g\|_A^2 = \int_A |f * g(x)|^2 dx \leq c \|f\|_{1, \mathbb{R}^d} \int_A \left(\int_{\mathbb{R}^d} |f(x-y)| |g(y)|^2 dy \right) dx. \quad (59)$$

We now use the fact that \mathcal{S} , the support of f , satisfies $\text{diam}(\mathcal{S}) < 2\sqrt{d}$ and so, since $x \in A$, it follows $\int_{\tilde{A}^c} |f(x-y)| |g(y)|^2 dy = 0$ for \tilde{A} compact set of \mathbb{R}^d such that $\tilde{A} := \left\{ \zeta \in \mathbb{R}^d : d(\zeta, A) \leq 2\sqrt{d} \right\}$. Using moreover Fubini's theorem, the right hand side of (59) becomes

$$c \|f\|_{1, \mathbb{R}^d} \int_A \left(\int_{\tilde{A}} |f(x-y)| |g(y)|^2 dy \right) dx = c \|f\|_{1, \mathbb{R}^d} \int_{\tilde{A}} |g(y)|^2 \left(\int_A |f(x-y)| dx \right) dy \leq c \|f\|_{1, \mathbb{R}^d}^2 \|g\|_{2, \tilde{A}}^2,$$

where in the last estimation we have enlarged the integration domain of f to \mathbb{R}^d . The estimation here above joint with (59) gives us

$$\|f * g\|_A^2 \leq c \|f\|_{1, \mathbb{R}^d}^2 \|g\|_{2, \tilde{A}}^2.$$

□

References

- [1] Applebaum, David. Lévy processes and stochastic calculus. Cambridge university press, 2009.
- [2] 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.
- [3] Bates, D.S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark. The Review of Financial Studies, 9(1), 69-107.
- [4] 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.
- [5] Comte, F., Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. In Annales de l'IHP Probabilités et statistiques (Vol. 49, No. 2, pp. 569-609).
- [6] Comte, F., Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. Stochastic processes and their applications, 115(5), 797-826.
- [7] Comte, F., Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM: Probability and Statistics, 6, 211-238.
- [8] Comte, F., Prieur, C., Samson, A. (2017). Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. Stochastic processes and their applications, 127(11), 3689-3718.
- [9] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probab. Theory Relat. Fields, 137(1), 2547.
- [10] Ditlevsen, S. and Greenwood, P. (2013). The MorrisLecar neuron model embeds a leaky integrate-and-fire model. Journal of Mathematical Biology 67 239-259.
- [11] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media.

- [12] Funke, B., Schmisser, E. (2018). Adaptive nonparametric drift estimation of an integrated jump diffusion process. *ESAIM: Probability and Statistics*, 22, 236-260.
- [13] Goldenshluger, A., Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3), 1608-1632.
- [14] Klein, T., Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3), 1060-1077.
- [15] Kou, S.G. (2002). A Jump-Diffusion Model for Option Pricing. *Management Science*, 48, 1086-1101.
- [16] Kusuoka, S., Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probability Theory and Related Fields*, 116(4), 457-484.
- [17] Kutoyants, Y. A. (2013). Statistical inference for ergodic diffusion processes. Springer Science and Business Media.
- [18] Lacour, C., Massart, P., Rivoirard, V. (2017). Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2), 298-335.
- [19] Lepski, O. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *The Annals of Statistics*, 41(2), 1005-1034.
- [20] Lepski, O. V., Levit, B. Y. (1999). Adaptive non-parametric estimation of smooth multivariate functions.
- [21] Masuda, H. (2007). Ergodicity and exponential beta - mixing bounds for multidimensional diffusions with jumps. *Stochastic processes and their applications*, 117(1), 35-56.
- [22] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3, 125-144
- [23] Meyn, S. P., Tweedie, R. L. (1993). Stability of Markovian processes III: FosterLyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3), 518-548.
- [24] Nikolskii, S. M. Approximation of Functions of Several Variables and Embedding Theorems (Springer, Berlin, 1975).
- [25] Schmisser, E. (2013). Nonparametric estimation of the derivatives of the stationary density for stationary processes. *ESAIM: Probability and Statistics*, 17, 33-69.
- [26] Schmisser, E. (2019). Non parametric estimation of the diffusion coefficients of a diffusion with jumps. *Stochastic Processes and their Applications*, 129(12), 5364-5405.
- [27] Stramer, O., Tweedie, R. L. (1997). Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients. *Statistica Sinica*, 577-593.
- [28] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *The Annals of Statistics*, 46(6B), 3451-3480.
- [29] Tsybakov, A. B. (2008). Introduction to nonparametric estimation. Springer Science & Business Media.
- [30] Veretennikov, A. Y. (1988). Bounds for the mixing rate in the theory of stochastic equations. *Theory of Probability and Its Applications*, 32(2), 273-281. ISO 690
- [31] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probability theory and related fields*, 107(4), 467-492.