Estimation of the invariant density for discretely observed diffusion processes: impact of the sampling and of the asynchronicity.

Chiara Amorino* Arnaud Gloter †

March 2, 2022

Abstract

We aim at estimating in a non-parametric way the density π of the stationary distribution of a d-dimensional stochastic differential equation $(X_t)_{t\in[0,T]}$, for $d \geq 2$, from the discrete observations of a finite sample $X_{t_0}, \dots , X_{t_n}$ with $0 = t_0 < t_1 < \dots < t_n =: T_n$. We propose a kernel density estimation and we study its convergence rates for the pointwise estimation of the invariant density under anisotropic Holder smoothness constraints. First of all, we find some conditions on the discretization step that ensures it is possible to recover the same rates as when the continuous trajectory of the process was available. As proven in the recent work [5], such rates are optimal and new in the context of density estimator. Then we deal with the case where such a condition on the discretization step is not satisfied, which we refer to as intermediate regime. In this new regime we identify the convergence rate for the estimation of the invariant density over anisotropic Holder classes, which is the same convergence rate as for the estimation of a probability density belonging to an anisotropic Holder class, associated to n iid random variables $X_1, ..., X_n$. After that we focus on the asynchronous case, in which each component can be observed in different moments. Even if the asynchrony of the observations implies some difficulties, we are able to overcome them by considering some combinatorics and by proving some sharper bounds on the variance which allow us to lighten the condition needed on the discretization step.

Non-parametric estimation, stationary measure, discrete observation, convergence rate, ergodic diffusion, anisotropic density estimation, asynchronous framework.

1 Introduction

In this paper we aim at estimating the invariant density belonging to an anisotropic Holder class starting from the discrete observation of the d-dimensional process

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

for $d \geq 2$; with $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $W = (W_t, t \geq 0)$ a d-dimensional Brownian motion.

The model presented in (1) is interesting from a theoretical point of view and because of its applications in many fields. For example, it has application in biology [64] and epidemiology [12] as well as in physics [61] and mechanics [48]. Some other classical examples are neurology [41], mathematical finance [43] and economics [18].

Because of the importance of the model, statistical inference for stochastic differential equations has been widely investigated in a lot of different context: disposing of continuous or discrete observations; on a fixed time interval or on long time intervals; in the parametric or in the non parametric frameworks.

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

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

Hence, there have been a big amount of papers on the topic. Among them, we quote Comte et al [26], Dalalyan and Reiss [27], Genon-Catalot [33], Gobet et al [36], Hoffmann [40], Kessler [44], Larédo [50], Marie and Rosier [54] and Yoshida [70].
The detailed study on stochastic differential equations leads the way for statistical inference for more complicated models such as stochastic partial differential equations ([3], [24]), diffusions with mixed effects ([63], [28]), SDEs driven by Lévy processes ([56]), diffusions with jumps ([10], [11], [55], [66]), Hawkes processes ([13], [30], [9]) and diffusions with discontinuous coefficients ([57], [51], [52]).

In this paper we focus on the non parametric estimation of the invariant density starting from the discrete observations of the stochastic differential equation in (1). In particular, we will consider at the beginning the case where the data is synchronously available and we will study, then, the case where the observations are given asynchronously. The asynchronous case is extremely important for the applications; see for example Chapters 3 and 4 of [23] and [2] for respectively autoregressive and log-Gaussian statistical approaches, whilst [19] proposes a recurrent neural network approaches. All of these models are applied on asynchronous financial market closing prices. It relies on the fact that any two assets rarely trade at the same instant. The treatment of non-synchronous trading effects dates back to Fisher [31]. For several years researchers focused mainly on the effects that stale quotes have on daily closing prices. Campbell et al. (Chapter 3 of [22]) provides a survey of this literature. Some other examples in which the authors deal with asynchronicity are [17] and [62]. In [39] the authors consider the problem of estimating the covariance of two diffusion processes when they are observed at discrete times in a non-synchronous manner.

In this context, we aim at proposing a kernel density estimator based on the discrete observations of (1) and we aim at finding the convergence rates of estimation for the stationary measure π associated to it, assuming that it belongs to an anisotropic Holder class. As the smoothness properties of elements of a function space may depend on the chosen direction of $\mathbb{R}^d$, the notion of anisotropy plays an important role. We will present some conditions that the discretization step needs to satisfy in order to recover the same (optimal) convergence rate achievable when a continuous trajectory of the process X was available and we will discuss the convergence rates we find in the intermediate regime, which is when the discretization step is not small enough and so the discretization error is not negligible.

With regard to the literature already existing about the estimation of the invariant measure, it is important to say that it is a problem already widely faced in many different frameworks by many authors. See for example [16], [20], [29], [1], [58] and [71]. The reason why such a problem results very attractive is the huge amount of physical applications and numerical methods connected to it, such as the Markov Chain Monte Carlo method. For example, the analysis of the invariant distributions is used to to analyze the stability of stochastic differential systems in [38] and [16]. In [49] and [60], instead, it is possible to find an approximation algorithm for the computation of the invariant density. The non-parametric estimation of the invariant density can also be used in order to estimate the drift coefficient in a non-parametric way (see [54] and [65]).
As a consequence, kernel estimators are widely employed as powerful tools: in [16] and [21] some kernel estimators are used to estimate the marginal density of a continuous time process. They are used also in more complicated model, such as in jump-diffusion framework (see [7], [8] and [53]).

Some references in a context closer to ours are [27], [68] and [5]. In all three papers kernel density estimators have been used for the study of the convergence rate for the estimation of the invariant density associated to a reversible diffusion process with unit diffusion part (in the first two works) or to the same stochastic differential equation as in (1) (in the third one). They are all based on the continuous observation of the process considered.
In particular in [5] the invariant density π has been estimated by means of the kernel estimator $\hat{\pi}_{h,T}$ assuming to have the continuous record of the process X solution to (1) up to time T and the following upper bound for the mean squared error has been showed, for $d \geq 3$:

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

where Σ is a class of coefficients for which the stationarity density has some prescribed regularity, $\beta_1 \leq \beta_2 \leq ... \leq \beta_d$ and $\bar{\beta}_3$ is the harmonic mean over the smoothness after having removed the two smallest. In particular, it is $\frac{1}{\bar{\beta}_3} := \frac{1}{d-2}\sum_{l\geq 3}\frac{1}{\beta_l}$ and so it clearly follows that $\bar{\beta}_3 \geq \bar{\beta}$. It has also been proven that the convergence rates here above are optimal.

In this paper, we propose to estimate the invariant density π starting from the discrete observations of the process X by means of the kernel estimator $\hat{\pi}_{h,n}$, which is the discretized version of $\hat{\pi}_{h,T}$ (see Section 4.1). Then, we prove an upper bound on the variance which is composed of two terms. The first is the same as when the continuous trajectory of the process was available, while the second is the discretization error. Depending on the fact that the second is negligible with respect to the first or the vice-versa, we get a condition on the discretization step to recover the continuous convergence rates or the new convergence rate in the intermediate regime. In particular, for $d \geq 3$, we show the following:

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] \lesssim \begin{cases} (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}} & \text{if } \beta_2 < \beta_3 \text{ and } \Delta_n \leq h_1^* h_2^* \sum_{j=1}^d |\log h_j^*|, \\ (\frac{1}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}} & \text{for } \beta_2 = \beta_3 \text{ and } \Delta_n \leq h_1^* h_2^*, \end{cases}$$

where Σ and $\bar{\beta}_3$ are as in [5], Δ_n is the discretization step and $h^*(T_n) = (h_1^*(T_n), ..., h_d^*(T_n))$ is the rate optimal choice for the bandwidth h (see Theorem 1 and the discussion below for details about the dependence of h^* in T_n).
On the other side, when the condition above on the discretization step are not respected, we obtain

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

We remark that the convergence rate in the intermediate regime is the same as for the estimation of a probability density belonging to an anisotropic Holder class, associated to n iid random variables $X_1, ..., X_n$. An analogous result is showed also in the bi-dimensional case.

After that, we focus on the asynchronous frameworks, in which each component can be observed in a different moment. Such a context implies many difficulties, as in this way even the choice of the estimator to propose appears challenging. The idea is to introduce d functions $\varphi_{n,l} : [0,T] \to \mathbb{R}$ such that

$$\varphi_{n,l}(t) = \sup \left\{ t_i^l \,|\, t_i^l \leq t \right\},$$

where $(t_i^l)_i$ are the instants of time in which the component X^l is observed, for $l \in \{1, ..., d\}$. In this way it is possible to write the sums as integrals and to propose an estimator which is the natural adaptation of $\hat{\pi}_{h,T}$, the one considered in the continuous framework in [5]. Moreover, the non-synchronicity involves some other challenges in the computation of the upper bound on the variance, one above all the fact that the observation times are not ordered, which entails some issues in the bound of the transition density. We are able to overcome such issues by considering some combinatorics and by proving some technical sharp bounds which allows us to show that, in the asynchronous context, the condition $\Delta_n \leq h_1^* h_2^*$ is enough to recover the variance obtained in the continuous case, for $\beta_2 < \beta_3$.

From both our findings here above and the results in [5] it appears clearly that the first two components do not have the same weight as all the others. Hence, we decide to deal with a different framework, in which the first two components are observed continuously and all the others are observed in a discrete way. In this context we are able to remove, under a non restrictive condition on Δ, the discretization error in the bound on the variance and to recover the same convergence rates as when the continuous trajectory of the process was available. We also show the bias term, which generally does not provide any condition on the discretization step and it is easier to deal with, surprisingly gives in this case a stronger constraint, which derives from the asynchronicity of the problem. However, it is enough to ask $\Delta_n \leq (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$ to recover the same convergence rate as when the continuous trajectory of the process is available.

The outline of the paper is the following. In Section 2 we introduce the model and we list the assumptions we will need in the sequel, while in Section 3 we recall the results in the case where the continuous trajectory of the process X is available. In Section 4 we consider the synchronous framework. We propose the kernel estimator and we state the upper bounds for the variance which will result in conditions on the discretization step to obtain the continuous regime and in the convergence rates in the intermediate regime. Section 5 is devoted to the statement of our results in the asynchronous framework. In Section 6 we prove the results stated in Section 4 while Section 7 is devoted to the proof of results under asynchronicity.

2 Model Assumptions

We aim at proposing a non-parametric estimator for the invariant density associated to a d-dimensional diffusion process X. In the sequel, we will recall what happens when a continuous record of the process $X^T = \{X_t, 0 \le t \le T\}$ up to time T is available. After that, we will be working in a high frequency setting and we will wonder which conditions on the discretization step will ensure the achievement of the same results as in the continuous case. The diffusion is a strong solution of the following stochastic differential equation:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s, \quad t \in [0, T], \tag{2}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $W = (W_t, t \ge 0)$ is a d-dimensional Brownian motion. The initial condition X_0 and W are independent. We denote $\tilde{a} := a \cdot a^T$.

We denote with $|.|$ and $< ., . >$ respectively the Euclidian norm and the scalar product in $\mathbb{R}^d$, and for a matrix in $\mathbb{R}^d \otimes \mathbb{R}^d$ we denote its operator norm by $|.|$.

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

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

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

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

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

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

Under the assumptions A1 - A2 the process X admits a unique invariant distribution μ and the ergodic theorem holds. We suppose that the invariant probability measure μ of X is absolutely continuous with respect to the Lebesgue measure and from now on we will denote its density as π: $d\mu = \pi dx$.

We want to estimate the invariant density π belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ defined below.

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

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

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

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

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

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

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

- *a and b satisfy A1 with the constants $(a_{min}, a_0, a_1, b_0, b_1)$,*
- *b satisfies A2 with the constants $(\tilde{C}, \tilde{\rho})$,*

- *the density $\pi_{(a,b)}$ of the invariant measure associated to the stochastic differential equation* (2) *belongs to $\mathcal{H}_d(\beta, \mathcal{L})$.*

We aim at estimating the invariant density π starting from discrete observations of the process X. In particular, we want to find some conditions that the discretization step has to satisfy in order to recover the same convergence rates we had when a continuous record of the process was available. Moreover, one may wonder which are the convergence rates in intermediate regime, i.e. when the discretization step goes to zero but the associated error is not negligible. In order to answer to these questions we recall what happens when the whole trajectory of the process X is available, as detailed discussed in [5]. This is the purpose of next section.

3 Continuous observations

When a continuous trajectory of X is available, it is natural to estimate the invariant density $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ by means of a kernel estimator. We therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \infty, \quad \text{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^l dx = 0, \tag{3}$$

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

$$\hat{\pi}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m})du. \tag{4}$$

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

For $d \geq 3$, from Theorem 1 of [5] we have the following convergence rate for the kernel estimator proposed in (4) and for the optimal bandwidth given below, in (6):

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

where Σ is the set defined in Definition 2 and $\bar{\beta}_3$ is such that

$$\frac{1}{\bar{\beta}_3} := \frac{1}{d-2} \sum_{j=3}^d \frac{1}{\beta_j}.$$

Moreover, from Theorems 3 and 4 of [5] we know they are optimal.

Regarding the bi-dimensional case we know, from Theorem 2 in [5] that the following holds true

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

and the convergence rate here above is optimal (see Theorem 5 of [5]).

We now suppose that the continuous record of the process, up to time T, is no longer available. In its place, we dispose of the discretization of the process at the instants $0 = t_0 \leq t_1 \leq ... \leq t_n = T$, given by $X_{t_0}, ..., X_{t_n}$. The first goal of Section 4 is to find some conditions the discretization step has to satisfy in order to recover the same convergence rates as in this section.

4 Discrete observations, synchronous framework

In this section we suppose that we observe a finite sample $X_{t_0}, ..., X_{t_n}$, with $0 = t_0 \leq t_1 \leq ... \leq t_n =: T_n$. The process X is solution of the stochastic differential equation (2). Every observation time point depends also on n but, in order to simplify the notation, we suppress this index. We assume the discretization scheme to be uniform which means that, for any $i \in \{0, ..., n-1\}$, it is $t_{i+1} - t_i =: \Delta_n$. We will be working in a high-frequency setting i.e. the discretization step $\Delta_n \to 0$ for $n \to \infty$. We assume moreover that $T_n = n\Delta_n \to \infty$ for $n \to \infty$ and that $\Delta_n > n^{-k}$ for some $k \in (0, 1)$.

4.1 Construction estimator

As in Section 3, we propose to estimate the invariant density $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ associated to the process X, solution to (2). To do that, we propose a kernel estimator which is the discretized version of the one introduced in (4). For $x = (x_1, ..., x_d) \in \mathbb{R}^d$, we define

$$\begin{aligned} \hat{\pi}_{h,n}(x) &:= \frac{1}{n\Delta_n} \frac{1}{\prod_{l=1}^d h_l} \sum_{i=0}^{n-1} \prod_{l=1}^d K(\frac{x_l - X_{t_i}^l}{h_l})(t_{i+1} - t_i) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{K}_h(x - X_{t_i}), \end{aligned} \tag{5}$$

with K a kernel function as in (3).

In this context we have two objectives:

1. Find some conditions on Δ_n to get the same convergence rates we had when a continuous record of the process was available.
2. Find the convergence rates in the intermediate regime (Δ_n tends to zero but the discretization error is not negligible).

To achieve them, we have to study the behaviour of the mean squared error $\mathbb{E}[|\hat{\pi}_{n,h}(x) - \pi(x)|^2]$ when $d \geq 3$ and when $d = 2$, and we have deal with two asymptotic regimes. The idea mainly consists in some upper bounds for the variance of the estimator. The difference, with respect to the continuous case, is that now we get an extra term which derives from the discretization.
If the new discretization term is negligible compared to the others, then the convergence rates are the same they were in the continuous case; otherwise we will find new convergence rates.

4.2 Main results, synchronous framework

The asymptotic behaviour of the estimator proposed in (5) is based on the bias-variance decomposition. To find the convergence rates it achieves we need a bound on the variance, as stated in the following propositions. We recall that the estimator performs differently depending on the dimension d. In this paper we provide the main results for $d \geq 2$.

Proposition 1. *Suppose that A1-A2 hold with some constant* $0 < a_{min} \leq a_0$ *and* $a_1 > 0$, $b_0 > 0, b_1 > 0, \tilde{C}_b, \tilde{\rho}_b$ *and that* $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, *for* $d \geq 3$. *Suppose moreover that* $\beta_1 = \beta_2 = ... = \beta_{k_0} < \beta_{k_0+1} \leq ... \leq \beta_d$, *for some* $k_0 \in \{1, ..., d\}$. *If* $\hat{\pi}_{h,n}$ *is the estimator proposed in* (5), *then there exist* $c > 0$ *and* $T_0 > 0$ *such that, for* $T_n \geq T_0$, *the following holds true.*

- *If* $k_0 = 1$ *and* $\beta_2 < \beta_3$ *or* $k_0 = 2$, *then*
$$Var(\hat{\pi}_{h,n}(x)) \leq \frac{c}{T_n} \frac{\sum_{j=1}^d |\log(h_j)|}{\prod_{l=3}^d h_l} + \frac{c}{T_n} \frac{\Delta_n}{\prod_{l=1}^d h_l}.$$
- *If* $k_0 \geq 3$, *then*
$$Var(\hat{\pi}_{h,n}(x)) \leq \frac{c}{T_n} \frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}} (\prod_{l \geq k_0+1} h_l)} + \frac{c}{T_n} \frac{\Delta_n}{\prod_{l=1}^d h_l}.$$
- *If otherwise* $k_0 = 1$ *and* $\beta_2 = \beta_3$, *then*
$$Var(\hat{\pi}_{h,T}(x)) \leq \frac{c}{T_n} \frac{1}{\prod_{l \geq 4} h_l \sqrt{h_2 h_3}} + \frac{c}{T_n} \frac{\Delta_n}{\prod_{l=1}^d h_l}.$$

Moreover, the constant c *is uniform over the set of coefficients* $(a, b) \in \Sigma$.

Proposition 1 leads us to the first main result of this section.

Theorem 1. *[Discretization term negligible]*
Suppose that A1-A2 hold and that $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, for $d \ge 3$, with $\beta_1 \le \beta_2 \le ... \le \beta_d$. Let $h^* = (h_1^*, ..., h_d^*)$ be the rate optimal choice for the bandwidth h as given in (6)*. Then, there exist $c > 0$ and $T_0 > 0$ such that, for $T_n \ge T_0$, the following hold true.*

- *If $\beta_2 < \beta_3$ and $\Delta_n \lesssim h_1^* h_2^* \sum_{j=1}^d |\log h_j^*|$, then*

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

- *If otherwise $\beta_2 = \beta_3$ and $\Delta_n \lesssim h_1^* h_2^*$, then*

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

Comparing the results here above with the ones included in Section 3 of [5] we deduce that, when the discretization step satisfies the constraint $\Delta_n \lesssim h_1^* h_2^* \sum_{j=1}^d |\log h_j^*|$ (or $\Delta_n \lesssim h_1^* h_2^*$ respectively) it is possible to recover the same optimal convergence rates we had when the trajectory of the process was observed continuously.
As seen in the proof of Theorem 1 of [5], for $\beta_2 < \beta_3$ the rate optimal choice for the bandwidth h is given by $h_j^* = (\frac{\log T_n}{T_n})^{a_j}$, while for $\beta = \beta_3$ it is $h_j^* = (\frac{1}{T_n})^{a_j}$ with

$$a_j = \frac{\bar{\beta}_3}{\beta_j(2\bar{\beta}_3 + d - 2)} \qquad \text{for any } j \in \{1, ..., d\}. \tag{6}$$

We remark that, according to [5], it is also possible to improve the choice of h_1^* and h_2^* in the case $\beta_2 < \beta_3$. However, in order to make the condition on the discretization step as weak as possible, it is convenient to choose $h_1^* h_2^*$ as large as possible, which leads us to the choice gathered in (6). Hence, replacing the optimal choice for the bandwidth as in (6) one can recover the same upper bound for the mean squared error as in Theorem 1 of [5] when the following conditions hold:

$$\Delta_n \lesssim (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}(\frac{1}{\beta_1} + \frac{1}{\beta_2})} \log T_n \qquad \text{for } \beta_2 < \beta_3,$$

$$\Delta_n \lesssim (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}(\frac{1}{\beta_1} + \frac{1}{\beta_2})} \qquad \text{for } \beta_2 = \beta_3.$$

When such conditions are not respected, instead, we get a different convergence rate. It is achieved by choosing the optimal bandwidth as $h_j(n) := (\frac{1}{n})^{\frac{\bar{\beta}}{\beta_j(2\bar{\beta}+d)}}$, where $\bar{\beta}$ is the harmonic mean over the d different smoothness:

$$\frac{1}{\bar{\beta}} = \sum_{j=1}^d \frac{1}{\beta_j}.$$

It leads us to the following result.

Theorem 2. *[Discretization term non-negligible]*
Suppose that A1-A2 hold and that $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ for $d \ge 3$, with $\beta_1 \le ... \le \beta_d$. Assume that the conditions on the discretization step gathered in Theorem 1 are not respected, and so one of the following holds

- $\Delta_n > (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}(\frac{1}{\beta_1} + \frac{1}{\beta_2})} \log T_n$ *and* $\beta_2 < \beta_3$.

- $\Delta_n > (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}(\frac{1}{\beta_1} + \frac{1}{\beta_2})}$ *and* $\beta_2 = \beta_3$.

Then, there exist $c > 0$ and $T_0 > 0$ such that, for $T_n \geq T_0$,

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

where $\bar{\beta}$ is the harmonic mean of the smoothness over the d direction, defined as

$$\frac{1}{\bar{\beta}} := \frac{1}{d}\sum_{j=1}^{d} \frac{1}{\beta_j}.$$

It is interesting to remark that it is also the convergence rate for the estimation of a probability density belonging to an Holder class, associated to n independent and identically distributed random variables $X_1, ..., X_n$.

For $d = 2$, analogous results hold. In particular, we have the following proposition.

Proposition 2. *Suppose that A1-A2 hold and that $d = 2$. If $\pi \in \mathcal{H}_2(\beta, \mathcal{L})$ and $\hat{\pi}_{h,n}$ is the estimator proposed in (5), then there exist $c > 0$ and $T_0 > 0$ such that, for $T_n \geq T_0$,*

$$Var(\hat{\pi}_{h,n}(x)) \leq \frac{c}{T_n}\sum_{j=1}^{d} |\log(h_j)| + \frac{1}{T_n}\frac{\Delta_n}{h_1 h_2},$$

where the constant c is uniform over the set of coefficients $(a, b) \in \Sigma$.

As before, it leads us to a condition on Δ_n that allows us to recover the continuous convergence rate gathered in Theorem 2 of [5], in the continuous case.

Theorem 3. *Suppose that A1-A2 hold and that $\pi \in \mathcal{H}_2(\beta, \mathcal{L})$. Let $h^* = (h_1^*, h_2^*)$ be the rate optimal choice (6) for the bandwidth h. Then, here exist $c > 0$ and $T_0 > 0$ such that, for $T_n \geq T_0$, the following hold true.*

- *If $\Delta_n \leq h_1^* h_2^* \sum_{j=1}^{2} |\log h_j^*| = (\frac{\log T_n}{T_n})^{\frac{1}{\bar{\beta}}} \log T_n$, then*

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] \leq c\frac{\log T_n}{T_n}$$

- *If otherwise $\Delta_n > (\frac{\log T_n}{T_n})^{\frac{1}{\bar{\beta}}} \log T_n$, then*

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

In this section we have found the convergence rates for the estimation of the invariant density starting from the discrete observation of the process X. Such observations are, in this section, all taken at the same instant. One may wonder if it is possible to recover the same results when the process X is observed asynchronously. The goal of next section is to answer to such a question. In particular, as in the main results of this section we have discovered that the first two components of the process do not have the same weight as the others, we may think it is possible to improve the result by observing continuously X^1 and X^2 and discretely and asynchronously all the others components.

5 Main results, asynchronous framework

In this section we assume $d \geq 3$ and we suppose that the components of the process X are observed asynchronously, i.e. in different instants. As we have seen that the first two components have a different weight compared to the others, we will consider two different cases. First of all we will assume that all the components are discretely observed in different moments and we will see that, up to require the discretization step to satisfy the condition $\Delta_n \leq (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$, it is possible to obtain the continuous convergence rate $(\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$. After that, we will assume that the first two components are continuously observed while the others $d-2$ are discretely and asynchronously observed, in order to lighten the condition on Δ_n to get the same bound on the variance.

5.1 Discrete asynchronous observations

We assume that we dispose of the discrete observations $X^l_{t^l_1}, ..., X^l_{t^l_n}$ for any $l \in \{1, ..., d\}$, with $0 \le t^l_1 \le ... \le t^l_n \le T_n$. We also define

$$\Delta_n := \sup_{l=1,\dots,d} \sup_{i=0,\dots,n-1} (t^l_{i+1} - t^l_i).$$

We remark it would have been possible to consider a different number of points on the different directions, having in particular n_l observations for the coordinate X^l. We have decided to take $n_1 = ... = n_d$ in order to lighten the notation.
Before we proceed with the statements of our results, we introduce some functions. First of all we observe that we have defined d partition of $[0, T_n]$ and so, for any $u \in [0, T_n]$, there exist some indexes $i_1, ..., i_d$ such that $u \in [t^l_{i_l}, t^l_{i_l+1})$, depending on the considered direction. We introduce then the following d functions $\varphi_{n,l} : [0, T_n] \to \mathbb{R}$ such that $\varphi_{n,l}(u) := t^l_{i_l}$, for any $l \in \{1, ..., d\}$. Thanks to these functions we can write the sums in the form of integrals. It leads us to the following estimator, which is the natural adaptation of the one in (4):

$$\begin{aligned}\hat{\pi}^a_{h,T_n}(x) &= \frac{1}{T_n \prod_{l=1}^d h_l} \int_0^{T_n} \prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(u)}}{h_l}) du \\ &=: \frac{1}{T_n} \int_0^{T_n} \prod_{l=1}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(u)}) du.\end{aligned}$$

Then, the following result holds true.

Proposition 3. *Suppose that A1-A2 hold with some constant $0 < a_{min} \le a_0$ and $a_1 > 0$, $b_0 > 0, b_1 > 0, \tilde{C}_b, \tilde{\rho}_b$ and that $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, for $d \ge 3$. Let $\beta_1 \le ... \le \beta_d$ and let $h^* = (h^*_1, ..., h^*_d)$ be the rate optimal choice for the bandwidth h given in (6). We suppose moreover that $\Delta_n \le \frac{1}{4} h^*_1 h^*_2$, then there exist $c > 0$ and $T_0 > 0$ such that, for $T_n \ge T_0$,*

$$Var(\hat{\pi}^a_{h,n}(x)) \le \frac{c}{T_n} \frac{\sum_{j=1}^d |\log(h_j)|}{\prod_{l=3}^d h_l}.$$

Moreover, the constant c is uniform over the set of coefficients $(a, b) \in \Sigma$.

Comparing the bound here above with the results gathered in Proposition 1 it appears clearly that asking the condition $\Delta_n \le \frac{1}{4} h^*_1 h^*_2$ is enough both in the synchronous and asynchronous frameworks to recover the same bound on the variance as in the continuous case, which is optimal for $\beta_2 < \beta_3$. We remark it is not worth extending Proposition 3 to remove the logarithm (which leads to the optimal convergence rate in the case $\beta_2 = \beta_3$), as the strongest condition derives in any case from the bias part. In particular, the following bound on the bias holds true.

Proposition 4. *Suppose that A1 holds and that a and b are $\mathcal{C}^3$ with bounded derivatives. If $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, with $\beta_1 \ge 1$, then there exists $c > 0$ such that for all $T_n > 0$, $0 < h_i < 1$,*

$$\left|\mathbb{E}[\hat{\pi}^a_{h,T_n}(x)] - \pi(x)\right| \le c \sum_{i=1}^d h_i^{\beta_i} + c\sqrt{\Delta_n}.$$

We remark that Proposition 4 still holds true if some of the components are observed continuously. This will be useful in next section.
Surprisingly the bias term, which is generally easier to deal with and does not generate any condition, in this setting provides a constraint on the discretization step stronger than the one on the variance. The condition derives from the fact that the different components are observed in different moments. From Propositions 3 and 4 next theorem easily follows.

Theorem 4. *Suppose that A1-A2 hold, that a and b are $\mathcal{C}^3$ with bounded derivatives. If $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, with $\beta_1 \ge 1$ and $\Delta_n \le (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3 + d - 2}}$, then there exist $\tilde{C}$, $\tilde{\rho}$, $c > 0$ and $T_0 > 0$ such that, for $T_n \ge T_0$,*

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

The theorem here above relies on the fact that the condition on the discretization step gathered in Proposition 3 is negligible compared to the one in Proposition 4. Indeed, after having replaced the value of the optimal bandwidth as in (6), we have that

$$h_1^* h_2^* = (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})}.$$

As $\beta_i \ge 1 \; \forall i$, it is $\frac{1}{\beta_1}+\frac{1}{\beta_2} \le 2$, from which we derive that

$$(\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})} > (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}.$$

Hence, if $\Delta_n \le (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$, then it is also $\Delta_n \le h_1^* h_2^*$.
We now wonder if it is possible to lighten the condition on the discretization step by observing continuously the first two components of the process. This is discussed in the next section.

5.2 What if two out of d components are continuously observed?

We suppose that the first two components of the process X are continuously observed, while we only dispose of discrete observations for the other $d-2$ components. In particular, we consider the asynchronous case, for which each of the last $d-2$ directions is observed in a different moment. Therefore, the continuous trajectories of X^1 and X^2 are available, as well as the discrete observations $X^l_{t^l_1}, ..., X^l_{t^l_n}$ for any $l \in \{3, ..., d\}$, with $0 \le t^l_1 \le ... \le t^l_n \le T_n$. As before, the discretization step is defined as

$$\Delta_n := \sup_{l=3,...,d} \; \sup_{i=0,...,n-1} (t^l_{i+1} - t^l_i).$$

Before we proceed with the statements of our results, we introduce the kernel estimator in this context. Using the notation introduced in previous section we have

$$\begin{aligned}\bar{\pi}_{h,T_n}(x) &= \frac{1}{T_n \prod_{l=1}^d h_l} \int_0^{T_n} \prod_{m=1,2} K(\frac{x_m - X^m_u}{h_m}) \prod_{l=3}^d K(\frac{x_l - X^l_{\varphi_{n,l}(u)}}{h_l}) du \\ &=: \frac{1}{T_n} \int_0^{T_n} \prod_{m=1,2} K_{h_m}(x_m - X^m_u) \prod_{l=3}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(u)}) du.\end{aligned}$$

What is unexpected is that, up to observe continuously X^1 and X^2, it is possible to recover the same upper bound on the variance as in Proposition 2 of [5], where all the components of the process X were continuously available. Indeed, as we will see in the proposition below, to dispose of the trajectories of the first two components of X and to ask a not restrictive condition on Δ_n is enough to make the condition on the discretization term of Proposition 1 disappear.

Proposition 5. *Suppose that A1-A2 hold and that $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$ with $\beta_1 \le ... \le \beta_d$. Let $h^* = (h^*_1, ..., h^*_d)$ be the rate optimal choice for the bandwidth h given in (6) and suppose that $\Delta_n \le \frac{1}{2}(\prod_{l\ge 3} h^*_l)^{\frac{2}{d-2}} = \frac{1}{2}(\frac{\log T_n}{T_n})^{\frac{2}{2\bar{\beta}_3+d-2}}$, then there exist $c > 0$ and $T_0 > 0$ such that, for $T_n \ge T_0$,*

$$Var(\bar{\pi}_{h^*,T_n}(x)) \le \frac{c}{T_n} \frac{\sum_{j=1}^d |\log(h^*_j)|}{\prod_{l=3}^d h^*_l}.$$

Moreover, the constant c is uniform over the set of coefficients $(a, b) \in \Sigma$.

One can see, comparing the proposition here above with Proposition 2 of [5] that, when $k_0 = 1, 2$, not having the continuous record of the last $(d-2)$ components does not interfere in the computations of the upper bound of the variance, up to requiring the mild condition on the discretization step requested here above. Therefore, in contrast with the results gathered in Proposition 1 and Theorem 1, even when $h^*_1 h^*_2 \sum_{j=1}^d |\log h^*_j| \lesssim \Delta_n$, the bound on the variance is the same it would have been with a continuous record of the whole process. Regarding the condition appearing here above, we remark it is $h^*_1 h^*_2 \sum_{j=1}^d |\log h^*_j| \lesssim (\prod_{l\ge 3} h^*_l)^{\frac{2}{d-2}}$ as it is equivalent to ask $(\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})} \log T_n \lesssim (\frac{\log T_n}{T_n})^{\frac{2}{2\bar{\beta}_3+d-2}}$, which holds true in a pure anisotropic context, for which $\bar{\beta}_3(\frac{1}{\beta_1} + \frac{1}{\beta_2})\frac{1}{2} > 1$. The smoothness are indeed ordered and so the harmonic mean over β_1 and β_2 is strictly smaller than the harmonic mean $\bar{\beta}_3$, computed over the other $d-2$ smoothness.

Regarding the bias term, it is easy to see that the bound gathered in Proposition 4 still holds true when $\varphi_{n,l}(t) = t$ for $l = 1, 2$ and so with $\bar{\pi}_{h,T_n}(x)$ instead of $\hat{\pi}^a_{h,T_n}(x)$. It yields the following theorem

Theorem 5. *Suppose that A1-A2 hold, that a and b are $\mathcal{C}^3$ with bounded derivatives. If $\pi \in \mathcal{H}_d(\beta, \mathcal{L})$, with $\beta_1 \geq 1$, then there exist $\tilde{C}$, $\tilde{\rho}$, $c > 0$ and $T_0 > 0$ such that, for $T_n \geq T_0$,*

$$|\mathbb{E}[\bar{\pi}_{h,T_n}(x)] - \pi(x)| \leq c \sum_{i=1}^{d} h_i^{\beta_i} + c\sqrt{\Delta_n}.$$

If moreover $\Delta_n \leq (\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}}$, then

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

We remark that, in the theorem here above, the condition on the discretization step is given by the bias term, while the one deriving from the bound on the variance gathered in Proposition 5 is negligible, as $\bar{\beta}_3 \geq 1$ and so

$$(\frac{\log T_n}{T_n})^{\frac{2\bar{\beta}_3}{2\bar{\beta}_3+d-2}} \leq (\frac{\log T_n}{T_n})^{\frac{2}{2\bar{\beta}_3+d-2}}.$$

It implies that, even if the bound on the variance gathered in Proposition 5 allows us to lighten the conditions on the discretization step to recover the continuous variance, the final condition on the discretization step to recover the continuous convergence rate is the same as for X^1 and X^2 discretely observed. The reason why it happens is that the bias term, which is generally the easier term to deal with and does not provide any condition on the discretization step, is not at all negligible in the asynchronous context. It provides some motivations for the improvement of the bound gathered in Proposition 4, as future perspective.

The proofs of all the propositions stated in this section can be found in Section 7.

6 Proof main results, synchronous framework

This section is devoted to the proof of our main results in the case where all the components are observed at the same time.

6.1 Proof of Proposition 1

Proof. The proof of Proposition 1 heavily relies on the proof of the upper bound on the variance of (4), in the continuous case. Intuitively, the integrals in Proposition 2 of [5] will be now replaced by sums, that we will split in order to use some different bounds on each of them. The main tools are the exponential ergodicity of the process as gathered in Proposition 1 of [5] and a bound on the transition density as in Proposition 5.1 of [35]. From the definition of our estimator $\hat{\pi}_{h,n}$, using also the fact that we are considering a uniform discretization step, it follows

$$\begin{aligned}
Var(\hat{\pi}_{h,n}(x)) &= Var(\frac{1}{n\Delta_n} \sum_{j=0}^{n-1} \mathbb{K}_h(x - X_{t_j})\Delta_n) \\
&= \frac{\Delta_n^2}{T_n^2} \sum_{j=0}^{n-1} (n-j)\, Cov(\mathbb{K}_h(x - X_0), \mathbb{K}_h(x - X_{t_j})) \\
&=: \frac{\Delta_n^2}{T_n^2} (\sum_{j=0}^{j_{\delta_1}} + \sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} + \sum_{j=j_{\delta_2}+1}^{j_D} + \sum_{j=j_D+1}^{n-1})\,(n-j)\,k(t_j) \\
&=: I_1 + I_2 + I_3 + I_4,
\end{aligned}$$

having introduced $0 \leq j_{\delta_1} \leq j_{\delta_2} \leq j_D \leq n-1$ and set $\delta_1 = \Delta_n j_{\delta_1}$, $\delta_2 = \Delta_n j_{\delta_2}$, $D = \Delta_n j_D$ and

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

Recall also that $t_j = \Delta_n j$. The quantities j_{δ_1}, j_{δ_2}, j_D (and consequently δ_1, δ_2 and D) will be chosen later, in order to get an upper bound on the variance as sharp as possible. We will provide some bounds for I_1, and I_4 which do not depend on k_0, while we will bound differently I_2 and I_3 depending on whether or not k_0 is larger than 3. For j small we use Cauchy-Schwarz inequality, the stationarity of the process, the boundedness of π and the definition of the kernel function to obtain

$$|k(t_j)| \leq Var(\mathbb{K}_h(x - X_0))^{\frac{1}{2}} Var(\mathbb{K}_h(x - X_{t_j}))^{\frac{1}{2}} \leq \int_{\mathbb{R}^d} (\mathbb{K}_h(x - y))^2 \pi(y) dy \leq \frac{c}{\prod_{l=1}^d h_l}. \tag{7}$$

It follows

$$|I_1| \leq \frac{\Delta_n^2 n}{T_n^2} \sum_{j=0}^{j_{\delta_1}} \frac{c}{\prod_{l=1}^d h_l} = c \frac{\Delta_n^2 n}{T_n^2} \frac{1}{\prod_{l=1}^d h_l} (j_{\delta_1} + 1). \tag{8}$$

For $j \in [j_{\delta_1} + 1, j_{\delta_2}]$ we act differently depending on k_0 as done for $s \in [\delta_1, \delta_2)$ in Proposition 2 of [5]. For $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$ it provides (see Equation (15) in [5])

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

that, for $s = t_j$, becomes

$$|k(t_j)| \leq \frac{c}{\prod_{l \geq 3} h_l} \frac{1}{t_j}.$$

It yields

$$|I_2| \leq \frac{\Delta_n^2 n}{T_n^2} \sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} \frac{c}{\prod_{l \geq 3} h_l} \frac{1}{t_j} = c \frac{\Delta_n}{T_n} \frac{1}{\prod_{l \geq 3} h_l} \sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} \frac{1}{t_j}.$$

We recall that t_j can be seen as $\Delta_n j$. Therefore,

$$\sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} \frac{1}{t_j} \leq \frac{c}{\Delta_n} \log(\frac{j_{\delta_2}}{j_{\delta_1}}) = \frac{c}{\Delta_n} \log(\frac{\Delta_n j_{\delta_2}}{\Delta_n j_{\delta_1}}) = \frac{c}{\Delta_n} \log(\frac{\delta_2}{\delta_1}).$$

It follows

$$|I_2| \leq c \frac{1}{T_n} \frac{1}{\prod_{l \geq 3} h_l} \log(\frac{\delta_2}{\delta_1}). \tag{9}$$

When $k_0 \geq 3$ instead, acting as to get (17) in [5] and taking $s = t_j$, we obtain

$$|k(t_j)| \leq \frac{c}{\prod_{l \geq k_0+1} h_l} t_j^{-\frac{k_0}{2}}.$$

Therefore,

$$\begin{aligned}
|I_2| &\leq \frac{\Delta_n^2 n}{T_n^2} \sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} \frac{c}{\prod_{l \geq k_0+1} h_l} t_j^{-\frac{k_0}{2}} \\
&= c \frac{\Delta_n}{T_n} \frac{1}{\prod_{l \geq k_0+1} h_l} \sum_{j=j_{\delta_1}+1}^{j_{\delta_2}} \Delta_n^{-\frac{k_0}{2}} j^{-\frac{k_0}{2}} \\
&\leq c \frac{\Delta_n^{1-\frac{k_0}{2}}}{T_n} \frac{1}{\prod_{l \geq k_0+1} h_l} j_{\delta_1}^{1-\frac{k_0}{2}} \\
&= c \frac{\delta_1^{1-\frac{k_0}{2}}}{T_n} \frac{1}{\prod_{l \geq k_0+1} h_l}
\end{aligned} \tag{10}$$

where we have used that, as $k_0 \geq 3$, $1 - \frac{k_0}{2}$ is negative.
To conclude the analysis of I_2 we assume that $k_0 = 1$ and $\beta_2 = \beta_3$. In this case the estimation here above still holds but, as $1 - \frac{k_0}{2} = \frac{1}{2}$ is now positive, it provides

$$|I_2| \leq c \frac{\delta_2^{\frac{1}{2}}}{T_n} \frac{1}{\prod_{l\geq 2} h_l}. \tag{11}$$

We now deal with I_3. With the same bound on the covariance as in (20) of Proposition 2 in [5] we get in any case, but for $k_0 = 1$ and $\beta_2 = \beta_3$,

$$|k(s)| \leq c(s^{-\frac{d}{2}} + 1).$$

Therefore, taking $s = t_j$,

$$\begin{aligned}
|I_3| &\leq \frac{\Delta_n^2 n}{T_n^2} \sum_{j=j_{\delta_2}+1}^{j_D} c(t_j^{-\frac{d}{2}} + 1) \qquad (12)\\
&\leq c\frac{\Delta_n}{T_n}(\sum_{t_j \leq 1,\, j=j_{\delta_2}+1}^{j_D} \Delta_n^{-\frac{d}{2}} j^{-\frac{d}{2}} + \sum_{t_j>1,\, j=j_{\delta_2}+1}^{j_D} 1)\\
&\leq c\frac{\Delta_n}{T_n}(\Delta_n^{-\frac{d}{2}} j_{\delta_2+1}^{1-\frac{d}{2}} + j_D)\\
&= \frac{c}{T_n}(\delta_2^{1-\frac{d}{2}} + D).
\end{aligned}$$

For $k_0 = 1$ and $\beta_2 = \beta_3$, instead, (22) of [5] provides

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

It follows

$$|I_3| \leq \frac{c\Delta_n^2 n}{T_n^2} \frac{1}{\prod_{l\geq 4} h_l} \sum_{j=j_{\delta_2}+1}^{j_D} (t_j^{-\frac{3}{2}} + 1).$$

Acting as above we obtain

$$|I_3| \leq \frac{c}{T_n}(\frac{1}{\prod_{l\geq 4} h_l} \frac{1}{\delta_2^{\frac{1}{2}}} + D). \tag{13}$$

To conclude, we need to evaluate the case where $j \in [j_D + 1, n-1]$. In this interval we use the exponential ergodicity of the process, as in Proposition 1 of [5]. It follows

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

for c and ρ positive constant uniform over the set of coefficients $(a, b) \in \Sigma$. It implies

$$\begin{aligned}
|I_4| &\leq c\frac{\Delta_n}{T_n} \frac{1}{(\prod_{l=1}^d h_l)^2} \sum_{j=j_D+1}^{n-1} e^{-\rho \Delta_n j} \qquad (14)\\
&\leq c\frac{\Delta_n}{T_n} \frac{1}{(\prod_{l=1}^d h_l)^2} e^{-\rho\Delta_n(j_D+1)}\\
&\leq c\frac{\Delta_n}{T_n} \frac{1}{(\prod_{l=1}^d h_l)^2} e^{-\rho D}.
\end{aligned}$$

From (8), (9), (12) and (14) we obtain the following bound for the case $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$:

$$Var(\hat{\pi}_{h,n}(x)) \leq \frac{c}{T_n} \Big[\frac{1}{\prod_{l=1}^d h_l} (\delta_1 + \Delta_n) + \frac{1}{\prod_{l\geq 3} h_l} \log(\frac{\delta_2}{\delta_1}) + \delta_2^{1-\frac{d}{2}} + D + \frac{1}{(\prod_{l=1}^d h_l)^2} e^{-\rho D} \Big].$$

It is easy to see that, except for the term $\frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l}\Delta_n$, the bound is the same as in the continuous case (see (22) in [5]). Hence, also the optimal choice for the parameters δ_1, δ_2 and D should be the same as in the continuous case, for which $\delta_1 = h_1 h_2$, $\delta_2 := (\prod_{j\ge 3} h_j)^{\frac{2}{d-2}}$ and $D := [\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1)\wedge T]$. Recalling that j_{δ_1}, j_{δ_2} and j_D have to be some integers, we can not propose exactly the same choice as above but we can take

$$j_{\delta_1} := \lfloor \frac{h_1 h_2}{\Delta_n} \rfloor, \qquad j_{\delta_2} := \lfloor \frac{(\prod_{j\ge 3} h_j)^{\frac{2}{d-2}}}{\Delta_n} \rfloor, \qquad j_D := \lfloor \frac{[\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1)\wedge T]}{\Delta_n} \rfloor.$$

It yields

$$Var(\hat{\pi}_{h,n}(x)) \le \frac{c}{T_n}\frac{\sum_{l=1}^d |\log(h_l)|}{\prod_{l\ge 3} h_l} + \frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l}\Delta_n,$$

as we wanted.
When $k_0 \ge 3$, instead, we replace the bound gathered in (9) with the one in (10). It follows

$$Var(\hat{\pi}_{h,n}(x)) \le \frac{c}{T_n}\Big[\frac{1}{\prod_{l=1}^d h_l}(\delta_1 + \Delta_n) + \frac{1}{\prod_{l\ge k_0+1} h_l}\delta_1^{1-\frac{k_0}{2}} + \delta_2^{1-\frac{d}{2}} + D + \frac{1}{(\prod_{l=1}^d h_l)^2}e^{-\rho D}\Big].$$

Again, every term but $\frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l}\Delta_n$ was already present in the proof of Proposition 2 of [5] (see (23)) and so the best choice would be to take the parameters as before. With this purpose in mind we choose

$$j_{\delta_1} := \lfloor \frac{(\prod_{l=1}^{k_0} h_l)^{\frac{2}{k_0}}}{\Delta_n} \rfloor, \qquad j_{\delta_2} := \lfloor \frac{1}{\Delta_n} \rfloor, \qquad j_D := \lfloor \frac{[\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1)\wedge T]}{\Delta_n} \rfloor.$$

It follows

$$Var(\hat{\pi}_{h,n}(x)) \le \frac{c}{T_n}\frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}}(\prod_{l\ge k_0+1} h_l)} + \frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l}\Delta_n.$$

We are left to study the case where $k_0 = 1$ and $\beta_2 = \beta_3$. Here, from (8), (11), (13) and (14) we obtain

$$Var(\hat{\pi}_{h,n}(x)) \le \frac{c}{T_n}\Big[\frac{1}{\prod_{l=1}^d h_l}(\delta_1 + \Delta_n) + \frac{\delta_2^{\frac{1}{2}}}{\prod_{l\ge 2} h_l} + \frac{1}{\prod_{l\ge 4} h_l\, \delta_2^{\frac{1}{2}}} + D + \frac{1}{(\prod_{l=1}^d h_l)^2}e^{-\rho D}\Big].$$

We take

$$j_{\delta_1} := 1, \qquad j_{\delta_2} := \lfloor \frac{h_2 h_3}{\Delta_n} \rfloor, \qquad j_D := \lfloor \frac{[\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1)\wedge T]}{\Delta_n} \rfloor$$

to get

$$Var(\hat{\pi}_{h,n}(x)) \le \frac{c}{T_n}\frac{1}{\sqrt{h_2 h_3}\prod_{l\ge 4} h_l} + \frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l}\Delta_n.$$

All the constant are uniform over the set of coefficients $(a, b) \in \Sigma$. The proof of Proposition 1 is then complete. □

6.2 Proof of Theorem 1

Proof. We will act differently depending on k_0.

• For $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$, when $\Delta_n \le h_1^* h_2^* \sum_{j=1}^d |\log h_j^*| = (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})}\log T_n$ the result is a straightforward consequence of the bias variance decomposition and of the bound on the variance gathered in Proposition 1. The rate optimal choice of the bandwidth h^* as in (6) provides

$$\begin{aligned} Var(\hat{\pi}_{h,n}(x)) &\le \frac{c}{T_n}\frac{\sum_{l=1}^d |\log(h_l^*)|}{\prod_{l\ge 3} h_l^*} + \frac{c}{T_n}\frac{1}{\prod_{l=1}^d h_l^*}\Delta_n \\ &\le \frac{c}{T_n}\frac{\sum_{l=1}^d |\log(h_l^*)|}{\prod_{l\ge 3} h_l^*}. \end{aligned}$$

Hence,

$$\sup_{(a,b)\in\Sigma} \mathbb{E}[|\hat{\pi}_{h,n}(x)-\pi(x)|^2] \le c\sum_{j=1}^{d} {h_j^*}^{2\beta_j} + \frac{c}{T_n}\frac{\sum_{l=1}^{d}|\log(h_l^*)|}{\prod_{l\ge 3} h_l^*}$$
$$= (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{\beta_l(2\bar{\beta}_3+d-2)}}$$

• For $k_0 \ge 3$, when $\Delta_n \le (h_1^* h_2^*) = (h_1^*)^2 = (\frac{1}{T_n})^{\frac{2\bar{\beta}_3}{\beta_1(2\bar{\beta}_3+d-2)}}$, we get the continuous convergence rate by the bias-variance decomposition and the second point of Proposition 1, choosing $h_1^* = ... = h_{k_0}^*$ and $h_l^*(T_n) : (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{\beta_l(2\bar{\beta}_3+d-2)}}$ for any $l \in \{1, ..., d\}$.
The reasoning is the same for $k_0 = 1$ and $\beta_2 = \beta_3$, recalling that we no longer have $h_1^* = h_2^*$ (as $\beta_1 < \beta_2$) and so the condition on the discretization step becomes $\Delta_n \le h_1^* h_2^* =: (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})}$. □

6.3 Proof of Theorem 2

Proof. Even if the final result is the same for any k_0, the proof of Theorem 2 is substantially different depending on the ordering of the smoothness and so on the value of k_0.
• We start assuming $k_0 = 1$ and $\beta_2 < \beta_3$ or $k_0 = 2$. When $\Delta_n > (\frac{\log T_n}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})} \log T_n$, the form of the rate optimal bandwidth is no longer as in the continuous case. We observe that $T_n = n\Delta_n$ and

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = \frac{d}{\bar{\beta}} - \frac{d-2}{\bar{\beta}_3},$$

and so the condition here above is equivalent to ask

$$\Delta_n > (\log T_n)(\frac{1}{n})^{\frac{\alpha}{1+\alpha}}, \tag{15}$$

with

$$\alpha := \frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2}) = \frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{d}{\bar{\beta}} - \frac{d-2}{\bar{\beta}_3}) = \frac{\bar{\beta}_3 d - (d-2)\bar{\beta}}{\bar{\beta}(2\bar{\beta}_3+d-2)}. \tag{16}$$

From the bias variance decomposition together with Proposition 1 we now obtain

$$\mathbb{E}[|\hat{\pi}_{h,n}(x)-\pi(x)|^2] \le c\sum_{j=1}^{d} h_j^{2\beta_j} + \frac{c}{T_n}\frac{\sum_{l=1}^{d}|\log(h_l)|}{\prod_{l\ge 3} h_l} + \frac{c}{T_n}\frac{1}{\prod_{l=1}^{d} h_l}\Delta_n.$$

We look for the rate optimal choice of the bandwidth by choosing $a_1, \dots, a_d$ such that $h_l = (\frac{1}{n})^{a_l}$.
We get the following bound

$$\mathbb{E}[|\hat{\pi}_{h,n}(x)-\pi(x)|^2] \le c\sum_{j=1}^{d} (\frac{1}{n})^{2\beta_j a_j} + \frac{c}{n\Delta_n}\frac{\sum_{l=1}^{d} a_l \log n}{\prod_{l\ge 3}(\frac{1}{n})^{a_l}} + \frac{c}{n}\frac{1}{\prod_{l=1}^{d}(\frac{1}{n})^{a_l}} \tag{17}$$
$$\le c\sum_{j=1}^{d} (\frac{1}{n})^{2\beta_j a_j} + \frac{c}{n}\frac{1}{\log(n\Delta_n)(\frac{1}{n})^{\frac{\alpha}{1+\alpha}}}\frac{\log n}{\prod_{l\ge 3}(\frac{1}{n})^{a_l}} + \frac{c}{n}\frac{1}{\prod_{l=1}^{d}(\frac{1}{n})^{a_l}},$$

having used the condition on Δ_n gathered in (15). We recall we have assumed $n\Delta_n \to \infty$ for $n \to \infty$ and that $\Delta_n > n^{-k}$ for some $k \in (0, 1)$. It follows that $\log(n\Delta_n) < (1-k)\log n$, which implies $\frac{\log n}{\log(n\Delta_n)} \le c$. Then, we observe that the balance between the three terms in (17) is achieved for $h_l(n) = (\frac{1}{n})^{\frac{\bar{\beta}}{\beta_j(2\bar{\beta}+d)}}$. In this way $\prod_{l\ge 3} h_l = (\frac{1}{n})^{\frac{\bar{\beta}}{2\bar{\beta}+d}\frac{d-2}{\bar{\beta}_3}}$, which implies in particular that the second term in the right hand side of (17) is upper bounded by

$$(\frac{1}{n})^{\frac{1}{1+\alpha} - \frac{\bar{\beta}}{2\bar{\beta}+d}\frac{d-2}{\bar{\beta}_3}}$$
$$= (\frac{1}{n})^{\frac{\bar{\beta}(2\bar{\beta}_3+d-2)}{\bar{\beta}_3(2\bar{\beta}+d)} - \frac{\bar{\beta}(d-2)}{\bar{\beta}_3(2\bar{\beta}+d)}} \tag{18}$$
$$= (\frac{1}{n})^{\frac{2\bar{\beta}}{2\bar{\beta}+d}},$$

which is clearly the size of the other terms as well, after having replaced the rate optimal choice for $h_l(n)$.

• We now consider the case where $\Delta_n > (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})}$ and $\beta_2 = \beta_3$. We start assuming that $k_0 = 1$. We can write

$$\Delta_n > (\frac{1}{n})^{\frac{\alpha}{1+\alpha}},$$

with α as in (16). From the bound on the variance gathered in Proposition 1 and the bias-variance decomposition easily follows the wanted result, acting as above.

When $k_0 \geq 3$ it is $\beta_1 = \beta_2$ and so $\Delta_n > (\frac{1}{T_n})^{\frac{\bar{\beta}_3}{2\bar{\beta}_3+d-2}(\frac{1}{\beta_1}+\frac{1}{\beta_2})} = (\frac{1}{T_n})^{\frac{2\bar{\beta}_3}{\beta_1(2\bar{\beta}_3+d-2)}}$. As $T_n = n\Delta_n$, the previous condition is equivalent to ask $\Delta_n > (\frac{1}{n})^{\frac{\alpha}{1+\alpha}}$, with

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

We underline that α is exactly the same as in (16), having now $\beta_1 = \beta_2$. The previous remark, together with the second point of Proposition 1 leads to the following bound for the mean squared error, for $k_0 \geq 3$:

$$\begin{aligned}\mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] &\leq c\sum_{j=1}^{d} h_j^{2\beta_j} + \frac{c}{T_n}\frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}} \prod_{l\geq k_0+1} h_l} + \frac{c}{T_n}\frac{\Delta_n}{\prod_{l=1}^{d} h_l} \\ &\leq c\sum_{j=1}^{d} h_j^{2\beta_j} + \frac{c}{n(\frac{1}{n})^{\frac{\alpha}{1+\alpha}}}\frac{1}{(\prod_{l=1}^{k_0} h_l)^{1-\frac{2}{k_0}} \prod_{l\geq k_0+1} h_l} + \frac{c}{n}\frac{1}{\prod_{l=1}^{d} h_l}.\end{aligned}$$

As before, we choose the rate optimal bandwidth as $h_l(n) := (\frac{1}{n})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d)}}$. As $\beta_1 = ... = \beta_{k_0}$ it follows in particular that $h_1(n) = ... = h_{k_0}(n)$. Replacing the value of $h_l(n)$ in the bound of the mean squared error we get

$$\mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] \leq c(\frac{1}{n})^{\frac{2\bar{\beta}}{2\bar{\beta}+d}} + c(\frac{1}{n})^{1-\frac{\alpha}{1+\alpha}} n^{\frac{\bar{\beta}}{2\bar{\beta}+d}(\frac{k_0-2}{\beta_1})} n^{\frac{\bar{\beta}}{2\bar{\beta}+d}(\frac{d-k_0}{\bar{\beta}_k})} + c(\frac{1}{n})^{1-\frac{\bar{\beta}}{2\bar{\beta}+d}(\frac{d}{\bar{\beta}})},$$

recalling that $\bar{\beta}_k$ is the mean smoothness over $\beta_{k_0+1}, \dots, \beta_d$ and it is such that $\frac{1}{\bar{\beta}_k} = \frac{1}{d-k_0}\sum_{l\geq k_0+1}\frac{1}{\beta_l}$. We remark that

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

Then, using also (18), we have that the exponent of $\frac{1}{n}$ in the second term here above is

$$\frac{1}{1+\alpha} - \frac{\bar{\beta}}{2\bar{\beta}+d}(\frac{d-2}{\bar{\beta}_3}) = \frac{2\bar{\beta}}{2\bar{\beta}+d}.$$

Remarking that the constant c does not depend on $(a, b) \in \Sigma$, it follows

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

as we wanted. □

6.4 Proof of Proposition 2

Proof. The proof follows the procedure to bound the variance for $d = 2$ proposed in the continuous case (see Theorem 2 of [5]). We split the sum in three terms:

$$\begin{aligned}Var(\hat{\pi}_{h,n}(x)) &= Var(\frac{1}{n\Delta_n}\sum_{j=0}^{n-1}\mathbb{K}_h(x - X_{t_j})\Delta_n) \\ &= \frac{\Delta_n^2}{T_n^2}(\sum_{j=0}^{j_\delta} + \sum_{j=j_\delta+1}^{j_D} + \sum_{j=j_D+1}^{n-1})\,(n-j)\,k(t_j) \qquad (19) \\ &=: \tilde{I}_1 + \tilde{I}_2 + \tilde{I}_3.\end{aligned}$$

Now we act on $\tilde{I}_1$ and $\tilde{I}_3$ as we did on I_1 and I_4 (defined in the proof of Proposition 1, respectively. It provides

$$|\tilde{I}_1| \le \frac{c}{T_n}\frac{1}{h_1 h_2}(\delta + \Delta_n), \tag{20}$$

$$|\tilde{I}_3| \le \frac{c}{T_n}\frac{1}{(h_1 h_2)^2}e^{-\rho D}. \tag{21}$$

Regarding $\tilde{I}_2$, we act here as we did on I_3 in Proposition 1. Recalling that here $d = 2$ we get

$$|k(t_j)| \le c(t_j^{-\frac{d}{2}} + 1) \le c(\frac{1}{t_j} + 1). \tag{22}$$

We need to consider separately what happens when t_j is larger or smaller than 1.

$$\begin{aligned}
|\tilde{I}_2| &\le c\frac{\Delta_n}{T_n}\sum_{j=j_\delta+1}^{j_D}(\frac{1}{t_j}+1) \\
&\le c\frac{\Delta_n}{T_n}(\sum_{t_j\le 1,\, j=j_\delta+1}^{j_D}\frac{1}{t_j} + \sum_{t_j>1,\, j=j_\delta+1}^{j_D} 1) \\
&\le \frac{c}{T_n}(|\log D| + |\log \delta| + D).
\end{aligned}$$

Putting all the pieces together, one can see that the choice $j_\delta := \lfloor \frac{h_1 h_2}{\Delta_n} \rfloor$ and $j_D := \lfloor \frac{[\max(-\frac{2}{\rho}\log(h_1 h_2), 1) \wedge T]}{\Delta_n} \rfloor$ leads to the wanted result. □

6.5 Proof of Theorem 3

Proof. The scheme we follow to prove Theorem 3 is the one provided in the proof of Theorem 1. We start considering the case where

$$\Delta_n \le h_1^* h_2^* \sum_{j=1}^{2} |\log h_j^*| = (\frac{\log T_n}{T_n})^{(\frac{1}{2\beta_1}+\frac{1}{2\beta_2})}\log T_n = (\frac{\log T_n}{T_n})^{\frac{1}{\bar{\beta}}}\log T_n,$$

where we have used that, from the proof of Theorem 2 of [5], $h_l^*(T_n) = (\frac{\log T_n}{T_n})^{a_l}$ with $a_l \ge \frac{1}{2\beta_l}$ and that, for $d = 2$, $\frac{1}{2}(\frac{1}{\beta_1} + \frac{1}{\beta_2}) = \frac{1}{\bar{\beta}}$.

From the bias variance decomposition together with Proposition 2, taking the rate optimal choice $h_l^*(T_n)$ as above directly follows

$$\mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] \le c\frac{\log T_n}{T_n} + c\frac{\log(\log T_n)}{T_n} = c\frac{\log T_n}{T_n}.$$

If $\Delta_n > (\frac{\log T_n}{T_n})^{\frac{1}{\bar{\beta}}}\log T_n$, instead, it is also $\Delta_n > (\frac{1}{n})^{\frac{1}{\bar{\beta}+1}}(\log(n\Delta_n))$. Using the bias variance decomposition and Proposition 2 we obtain

$$\begin{aligned}
\mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] &\le c(h_1^{2\beta_1} + h_2^{2\beta_2}) + \frac{c}{T_n}\sum_{j=1}^{2}|\log h_j| + \frac{c}{T_n}\frac{\Delta_n}{h_1 h_2} \\
&\le c(h_1^{2\beta_1} + h_2^{2\beta_2}) + c(\frac{1}{n})^{1-\frac{1}{\bar{\beta}+1}}\frac{\sum_{j=1}^{2}|\log h_j|}{\log(n\Delta_n)} + \frac{c}{n h_1 h_2}.
\end{aligned}$$

We choose the rate optimal bandwidth $h_l(n) := (\frac{1}{n})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d)}}$, for $l = 1, 2$. We have already discussed the behaviour of $n\Delta_n$, saying in particular that $\frac{\log n}{\log n\Delta_n} \le c$ in the proof of Theorem 2. It yields

$$\begin{aligned}
\mathbb{E}[|\hat{\pi}_{h,n}(x) - \pi(x)|^2] &\le c(\frac{1}{n})^{\frac{2\bar{\beta}}{2\bar{\beta}+d}} + c(\frac{1}{n})^{1-\frac{1}{\bar{\beta}+1}} + c(\frac{1}{n})^{1-\frac{\bar{\beta}}{2\bar{\beta}+d}(\frac{1}{\beta_1}+\frac{1}{\beta_2})} \\
&\le c(\frac{1}{n})^{\frac{2\bar{\beta}}{2\bar{\beta}+d}} + c(\frac{1}{n})^{\frac{\bar{\beta}}{\bar{\beta}+1}} + (\frac{1}{n})^{\frac{2\bar{\beta}+d-2}{2\bar{\beta}+d}}
\end{aligned}$$

which is what we wanted, as $d = 2$. □

7 Proof main results, asynchronous framework

In Section 5 we assume to observe the components of the process X in different moments. In particular, we assume that all the components are observed asynchronously in Section 5.1 and we show that the condition $\Delta_n \le h_1^* h_2^*$ is enough to obtain the same variance as when the continuous record of the process was available. In Section 5.2, instead, we suppose that the trajectory of the first two components is observed continuously and that ensures the possibility to get the continuous variance with a weaker condition on the discretization step. As the proof of the results is easier when the first two components are observed in the same moment, we start by proving the results gathered in Proposition 5, assuming that the first two components are continuously observed.

7.1 Proof of Proposition 5

Proof. In analogy to the previous proofs, we introduce

$$k(t,s) := Cov(\prod_{m=1,2} K(\frac{x_m - X_t^m}{h_m})\prod_{l=3}^d K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l}), \prod_{m=1,2} K(\frac{x_m - X_s^m}{h_m})\prod_{l=3}^d K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})),$$

such that

$$\begin{aligned} Var(\bar{\pi}_{h^*,T_n}(x)) &= \frac{2}{T_n^2}\int_0^{T_n}\int_0^t k(t,s)1_{s<t}dsdt \\ &= \frac{2}{T_n^2}\int_0^{t}\int_0^{T_n} k(t,s)1_{s<t}\big(1_{|t-s|\le h_1^*h_2^*\sum_{j=1}^d|\log h_j^*|} \\ &+ 1_{h_1^*h_2^*\sum_{j=1}^d|\log h_j^*|\le|t-s|\le(\prod_{j\ge3}h_j^*)^{\frac{2}{d-2}}} + 1_{(\prod_{j\ge3}h_j^*)^{\frac{2}{d-2}}\le|t-s|\le D} + 1_{D\le|t-s|\le T_n}\big)dsdt \\ &= \sum_{j=1}^4 I_j, \end{aligned}$$

with h^* the rate optimal choice of the bandwidth as in the proof of Theorem 1 of [5], given by (6). We start considering I_1. Here, acting as in Proposition 1 in order to get (7), we obtain

$$|k(t,s)| \le \frac{c}{\prod_{l=1}^d h_l^*}.$$

Therefore, after the change of variable $t \to t' := t - s$, we have

$$\begin{aligned} I_1 &\le \frac{1}{T_n^2}\int_0^{T_n}\int_0^{h_1^*h_2^*\sum_{j=1}^d|\log h_j^*|}\frac{c}{\prod_{l=1}^d h_l^*}dt'ds \\ &\le \frac{c}{T_n}\frac{h_1^*h_2^*\sum_{j=1}^d|\log h_j^*|}{\prod_{l=1}^d h_l^*} = \frac{c}{T_n}\frac{\sum_{j=1}^d|\log h_j^*|}{\prod_{l=3}^d h_l^*}. \end{aligned} \tag{23}$$

We now study I_2, which is the most complicated term we have to deal with. Intuitively, we would like to bound the variance as in the interval $[\delta_1, \delta_2)$ in the proof of Proposition 2 of [5]. It relies on a bound for the transition density (as in Proposition 5.1 of [35] or Lemma 1 of [5]). In order to use it we need to know the ordering between the quantities s, t, $\varphi_{n,3}(s)$, ... , $\varphi_{n,d}(s)$, $\varphi_{n,3}(t)$, ... , $\varphi_{n,d}(t)$. We know that $s < t$, $\varphi_{n,l}(s) \le s$ and $\varphi_{n,l}(t) \le t$ for any $l \in \{3, ..., d\}$. Moreover, it is possible to have $\varphi_{n,l}(t) \le s < t$, but it implies $\varphi_{n,l}(t) = \varphi_{n,l}(s)$. Then, we can consider a permutation w_3, ... , w_d of $\varphi_{n,3}(s)$, ... , $\varphi_{n,d}(s)$ which is such that $w_3 \le ... \le w_d$. In particular, we denote by $w_3 \le ... \le w_d$ a reordering of $\varphi_{n,l}(t)$ and σ an element of the permutation group on $\{3, ..., d\}$ such that $w_i = \varphi_{n,\sigma(i)}(t)$ for all $i \in \{3, ..., d\}$. In the same way we introduce the permutation $\tilde{w}_3$, ... , $\tilde{w}_d$ of $\varphi_{n,3}(t)$, ... , $\varphi_{n,d}(t)$ which is properly ordered, i.e. $\tilde{w}_3 \le ... \le \tilde{w}_d$. In particular, we introduce $\tilde{\sigma}$ which is an element of the permutation group such that $\tilde{w}_i = \varphi_{n,\tilde{\sigma}(i)}(s)$ for all $i \in \{3, ..., d\}$. In order to admit the possibility that $\varphi_{n,l}(t) \le s$, and then $\varphi_{n,l}(t) = \varphi_{n,l}(s)$ for some index l, we say that $\tilde{w}_j \in \{w_3, ..., w_d\}$ for $j \le h$ and $\tilde{w}_{h+1} \ge s$. It follows that

$$w_3 \le ... \le w_d \le s \le \tilde{w}_{h+1} \le ... \le \tilde{w}_d \le t.$$

Hence we can write, for $l \leq h$, $\tilde{w}_l = w_{\tau(l)}$ for some $\tau(l) \in \{3, ..., d\}$. We now introduce the following vectors, which represent the positions in the instants previously discussed. At the instant w_j we have the vector y^j, for $j \in \{3, ..., d\}$. The vector z is instead connected to the time s, while $\tilde{z}$ is connected to t. At the time $\tilde{w}_j$ we have the vectors $\tilde{y}^j$, for $j \in \{h+1, ..., d\}$. We remark that y^j_l is the l-component of the vector y^j, which gives the position at the instant w_j. We observe that, as $\tilde{w}_l = w_{\tau(l)}$ for $l \leq h$ and so we can write, for any $l \leq h$, $\tilde{y}^l_{\sigma(\tilde{l})} = y^{\tau(l)}_{\sigma(\tau(l))}$. We have

$$|k(t,s)| \leq |\tilde{k}(t,s)| +$$
$$|\mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X^m_t}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l})]||\mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X^m_s}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})]|, \quad (24)$$

where

$$\tilde{k}(t,s) := \mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X^m_t}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l}) \prod_{m=1,2} K(\frac{x_m - X^m_s}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})]. \quad (25)$$

We aim at proving that $|\tilde{k}(t,s)| \leq \frac{c}{(t-s)\prod_{l=3}^{d} h_l}$. We write the expectation in (25) using the law of the random vector $(X_{w_3}, X_{w_4} \dots, X_{w_d}, X_s, X_{\tilde{w}_{h+1}}, \dots, X_{\tilde{w}_d}, X_t)$. For simplicity, we assume that all the instants appearing in this vector are different $w_3 < w_4 < \dots < w_d < s < \tilde{w}_{h+1} < \dots < \tilde{w}_d < t$ and in turn the law of this vector admits a density as product of the transition density of the process X. If we are not in the situation where all the instants are distinct, it is possible to slightly move some values of $w_3, \dots, w_d, \tilde{w}_h, \dots, \tilde{w}_d$ in order to get different instants, and then conclude by a density argument in order to get the upper bound on $|\tilde{k}(s,t)|$. With these considerations, we can write

$$\begin{aligned}
|\tilde{k}(t,s)| \leq & \int_{\mathbb{R}^d} \prod_{m=1}^{2} |K_{h^*_m}(x_m - z_m)| \int_{\mathbb{R}^{d(d-2)}} |\prod_{l=3}^{d} K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})| \int_{\mathbb{R}^d} \prod_{m=1}^{2} |K_{h^*_m}(x_m - \tilde{z}_m)| \\
& \times \prod_{l=3}^{h} |K_{h^*_{\sigma(\tau(l))}}(x_{\sigma(\tau(l))} - y^{\tau(l)}_{\sigma(\tau(l))})| \int_{\mathbb{R}^{d(d-h)}} \prod_{l=h+1}^{d} |K_{h_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})| \\
& \times p_{w_4 - w_3}(y^3, y^4) p_{w_5 - w_4}(y^4, y^5) \times ... \times p_{w_d - w_{d-1}}(y^{d-1}, y^d) p_{s - w_d}(y^d, z) \quad (26) \\
& \times p_{\tilde{w}_{h+1} - s}(z, \tilde{y}^{h+1}) \times ... \times p_{\tilde{w}_d - \tilde{w}_{d-1}}(\tilde{y}^{d-1}, \tilde{y}^d) p_{t - \tilde{w}_d}(\tilde{y}^d, \tilde{z}) \pi(y^3) dz dy^3 ... dy^d d\tilde{z} d\tilde{y}^{h+1} ... d\tilde{y}^d.
\end{aligned}$$

We bound

$$\begin{aligned}
& |\prod_{l=3}^{h} K_{h^*_{\sigma(\tau(l))}}(x_{\sigma(\tau(l))} - y^{\tau(l)}_{\sigma(\tau(l))}) \prod_{l=h+1}^{d} K_{h_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})| \\
& = |\prod_{l=3}^{h} K_{h^*_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)}) \prod_{l=h+1}^{d} K_{h_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})| \\
& \leq \frac{c}{\prod_{l \geq 3} h^*_{\tilde{\sigma}(l)}} = \frac{c}{\prod_{l \geq 3} h^*_l},
\end{aligned}$$

it follows

$$\begin{aligned}
|\tilde{k}(t,s)| \leq & \frac{1}{\prod_{l \geq 3} h^*_l} \int_{\mathbb{R}^d} \prod_{m=1}^{2} K_{h^*_m}(x_m - z_m) \prod_{l=3}^{d} K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)}) \int_{\mathbb{R}^d} \prod_{m=1}^{2} K_{h^*_m}(x_m - \tilde{z}_m) \\
& \times \int_{\mathbb{R}^{d(d-2)}} p_{w_4 - w_3}(y^3, y^4) p_{w_5 - w_4}(y^4, y^5) \times ... \times p_{w_d - w_{d-1}}(y^{d-1}, y^d) p_{s - w_d}(y^d, z) \quad (27) \\
& \times p_{t-s}(y^d, \tilde{z}) \pi(y^3) dz dy^3 ... dy^d d\tilde{z}.
\end{aligned}$$

Using Gaussian upper bounds on the transition density (Proposition 5.1 in [35]), we have

$$p_{t-s}(y^d, \tilde{z}) \leq \frac{c}{(t-s)} q_{t-s}(\tilde{z}_3, ..., \tilde{z}_d | \tilde{z}_1, \tilde{z}_2, y^d),$$

with

$$q_{t-s}(\tilde{z}_3, ..., \tilde{z}_d|\tilde{z}_1, \tilde{z}_2, y^d) = e^{-\lambda_0 \frac{(\tilde{z}_1 - y_1^d)^2}{t-s}} e^{-\lambda_0 \frac{(\tilde{z}_2 - y_2^d)^2}{t-s}} \frac{1}{\sqrt{t-s}} e^{-\lambda_0 \frac{(\tilde{z}_3 - y_3^d)^2}{t-s}} \times ... \times \frac{1}{\sqrt{t-s}} e^{-\lambda_0 \frac{(\tilde{z}_d - y_d^d)^2}{t-s}}.$$

We observe that

$$\sup_{t-s\in(0,1)} \sup_{y^d, \tilde{z}_1, \tilde{z}_2 \in \mathbb{R}^{d+2}} \int_{\mathbb{R}^{d-2}} q_{t-s}(\tilde{z}_3, ..., \tilde{z}_d|\tilde{z}_1, \tilde{z}_2, y^d) d\tilde{z}_3 ... d\tilde{z}_d < c, \tag{28}$$

remarking that $t - s \in (0, 1)$ as

$$0 \le t - s \le (\prod_{j\ge3} h_j^*)^{\frac{2}{d-2}} < 1.$$

Moreover, we easily bound

$$\int_{\mathbb{R}^2} \prod_{m=1}^{2} K_{h_m^*}(x_m - \tilde{z}_m) d\tilde{z}_1 d\tilde{z}_2 < c.$$

Replacing everything in (27) we obtain

$$\begin{aligned} |\tilde{k}(t, s)| \le{}& \frac{c}{t-s} \frac{1}{\prod_{l\ge3} h_l^*} \int_{\mathbb{R}^d} \prod_{m=1}^{2} K_{h_m^*}(x_m - z_m) \prod_{l=3}^{d} K_{h_{\sigma(l)}^*}(x_{\sigma(l)} - y_{\sigma(l)}^l) \\ &\times \int_{\mathbb{R}^{d(d-2)}} \prod_{l=4}^{d} p_{w_l - w_{l-1}}(y^{l-1}, y^l) p_{s-w_d}(y^d, z) \pi(y^3) dz dy^3 ... dy^d. \end{aligned} \tag{29}$$

We aim at showing that the integral here above is upper bounded by a constant. To do that we use the first point of Lemma 2, which is stated and proven at the end of this section. We remark that in our case $r = d-1$ and, in particular, $u_1, ..., u_{r-1}, u_r$ are in this case $y^3, ..., y^d, z$ while $q_1 = q_2 = r$ and, for $l \ge 3$, $q_l = \sigma^{-1}(l)$. We obtain

$$|\tilde{k}(t, s)| \le \frac{c}{t-s} \frac{1}{\prod_{l\ge3} h_l^*}.$$

Using again the first point of Lemma 2, there exists a constant c such that

$$|\mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X_u^m}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X_{\varphi_{n,l}(u)}^l}{h_l})]| \le c, \quad \forall u$$

and it implies recalling (24) that

$$|k(t, s)| \le \frac{c}{t-s} \frac{1}{\prod_{l\ge3} h_l^*} + c.$$

In the equation above we have also used that, thanks to the first point of Lemma 2, there exists a constant $c > 0$ such that

$$|\mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X_t^m}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X_{\varphi_{n,l}(t)}^l}{h_l})]||\mathbb{E}[\prod_{m=1,2} K(\frac{x_m - X_s^m}{h_m}) \prod_{l=3}^{d} K(\frac{x_l - X_{\varphi_{n,l}(s)}^l}{h_l})]| \le c.$$

It yields

$$I_2 \le \frac{c}{T_n^2} \int_0^t \int_0^{T_n} (\frac{1}{\prod_{l\ge3} h_l^*} \frac{1}{t-s} + 1) 1_{h_1^* h_2^* \sum_{j=1}^d |\log h_j^*| \le |t-s| \le (\prod_{j\ge3} h_j^*)^{\frac{2}{d-2}}} ds dt.$$

We apply the change of variable $t - s =: t'$. It follows

$$\begin{aligned} |I_2| &\le \frac{c}{T_n} \frac{1}{\prod_{l\ge3} h_l^*} \int_{h_1^* h_2^* \sum_{j=1}^d |\log h_j^*|}^{(\prod_{j\ge3} h_j^*)^{\frac{2}{d-2}}} \frac{1}{t'} dt' \qquad (30) \\ &\le \frac{c}{T_n} \frac{\sum_{j=1}^2 |\log h_j^*|}{\prod_{l\ge3} h_l^*} \qquad (31) \\ &\le \frac{c}{T_n} \frac{\sum_{j=1}^2 |\log h_j^*|}{\prod_{l\ge3} h_l^*}. \qquad (32) \end{aligned}$$

Regarding I_3, it is

$$I_3 := \frac{1}{T_n^2}\int_0^{T_n}\int_0^{T_n} k(t,s)1_{(\prod_{j\geq 3} h_j^*)^{\frac{2}{d-2}}\leq|t-s|\leq D}dtds.$$

We can write $\tilde{k}(t,s)$ as in (26). The only difference is that now s and t are distant to each other, which implies that $\varphi_{n,l}(s) < \varphi_{n,l}(t)$ for any $l \in \{3,...,d\}$ and so, in particular, the ordering of the quantities previously introduced is the following:

$$w_3 \leq ... \leq w_d \leq s \leq \tilde{w}_3 \leq ... \leq \tilde{w}_d \leq t.$$

This holds true because

$$|t-s| \geq (\prod_{j\geq 3} h_j^*)^{\frac{2}{d-2}} \qquad \text{and } \Delta_n \leq (\prod_{j\geq 3} h_j^*)^{\frac{2}{d-2}}.$$

Hence, we have

$$|\tilde{k}(t,s)| \leq \int_{\mathbb{R}^d}\prod_{m=1}^2 |K_{h_m^*}(x_m - z_m)|\int_{\mathbb{R}^{d(d-2)}}\prod_{l=3}^d |K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})|\int_{\mathbb{R}^d}\prod_{m=1}^2 |K_{h_m^*}(x_m - \tilde{z}_m)|$$
$$\times \int_{\mathbb{R}^{d(d-2)}}\prod_{l=3}^d |K_{h^*_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})| p_{w_4-w_3}(y^3,y^4)p_{w_5-w_4}(y^4,y^5)\times ... \times p_{w_d-w_{d-1}}(y^{d-1},y^d)p_{s-w_d}(y^d,z)$$
$$\times p_{\tilde{w}_3-s}(z,\tilde{y}^3)\times ... \times p_{\tilde{w}_d-\tilde{w}_{d-1}}(\tilde{y}^{d-1},\tilde{y}^d)p_{t-\tilde{w}_d}(\tilde{y}^d,\tilde{z})\pi(y^3)dzdy^3...dy^d d\tilde{z}d\tilde{y}^3...d\tilde{y}^d.$$

We remark that the largest interval of time above is $\tilde{w}_3 - s$. We use on it the rough estimation

$$p_{\tilde{w}_3-s}(z,\tilde{y}^3) \leq \frac{c}{(\tilde{w}_3-s)^{\frac{d}{2}}}\prod_{l=1}^d e^{-\lambda_0\frac{(z_l-\tilde{y}^3_l)^2}{\tilde{w}_3-s}} \leq \frac{c}{(\tilde{w}_3-s)^{\frac{d}{2}}}.$$

It follows

$$|\tilde{k}(t,s)| \leq \frac{c}{(\tilde{w}_3-s)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\prod_{m=1}^2 |K_{h_m^*}(x_m - z_m)|\int_{\mathbb{R}^{d(d-2)}}\prod_{l=3}^d |K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})|\int_{\mathbb{R}^d}\prod_{m=1}^2 |K_{h_m^*}(x_m - \tilde{z}_m)|$$
$$\times \int_{\mathbb{R}^{d(d-2)}}\prod_{l=3}^d |K_{h^*_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})|\prod_{l=4}^d p_{w_l-w_{l-1}}(y^{l-1},y^l)p_{s-w_d}(y^d,z)$$
$$\times \prod_{l=4}^d p_{\tilde{w}_l-\tilde{w}_{l-1}}(\tilde{y}^{l-1},\tilde{y}^l)p_{t-\tilde{w}_d}(\tilde{y}^d,\tilde{z})\pi(y^3)dzdy^3...dy^d d\tilde{z}d\tilde{y}^3...d\tilde{y}^d.$$

We apply twice the first point of Lemma 2 (having on each integral $r = (d-1)$ as we are considering the integrals in $y^3,...,y^d,z$ and in $\tilde{y}^3,...,\tilde{y}^d,\tilde{z}$). It implies

$$|k(t,s)| \leq \frac{c}{(\tilde{w}_3-s)^{\frac{d}{2}}} + c.$$

We now observe that it is

$$|t-s| \leq |t-\tilde{w}_3| + |\tilde{w}_3 - s| \leq \Delta_n + |\tilde{w}_3-s| \leq \frac{1}{2}(\prod_{l\geq 3} h_l^*)^{\frac{2}{d-2}} + |\tilde{w}_3 - s|. \tag{33}$$

It follows

$$|\tilde{w}_3 - s| \geq |t-s| - \frac{1}{2}(\prod_{l\geq 3} h_l^*)^{\frac{2}{d-2}} \geq \frac{1}{2}(\prod_{l\geq 3} h_l^*)^{\frac{2}{d-2}}.$$

Moreover,

$$|\tilde{w}_3 - s| \leq |t-s| \leq D.$$

From the change of coordinates $s \rightarrow s' := t-s$ we obtain

$$\begin{aligned} I_3 &\leq \frac{c}{T_n^2}\int_0^{T_n}\int_{\frac{1}{2}(\prod_{l\geq 3} h_l^*)^{\frac{2}{d-2}}}^D (\frac{c}{s'^{\frac{d}{2}}} + c)ds'dt \\ &\leq \frac{c}{T_n}((\prod_{l\geq 3} h_l^*)^{\frac{2}{d-2}(1-\frac{d}{2})} + D) \\ &= \frac{c}{T_n}((\prod_{l\geq 3} h_l^*)^{-1} + D), \end{aligned} \tag{34}$$

which is the order we wanted.
We are left to study I_4, where $D \le |t-s| \le T_n$. Here we want to use the fact that the process is exponential β-mixing. To do that, we use the definition of the covariance. We introduce the notation $K_{h_j^*}(t) := K_{h_j^*}(x - X_t)$. Then we need to study, up to reorder the components,

$$Cov(K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d), K_{h_1^*}(t)K_{h_2^*}(t)K_{h_3^*}(\tilde{w}_3)...K_{h_d^*}(\tilde{w}_d)),$$

where $w_3 \le ... \le w_d \le s \le \tilde{w}_3 \le ... \le \tilde{w}_d \le t$, $D \le |t-s| \le T$. We define

$$g(X_{\tilde{w}_3}) := \mathbb{E}[K_{h_1^*}(t)K_{h_2^*}(t)K_{h_3^*}(\tilde{w}_3)...K_{h_d^*}(\tilde{w}_d)|X_{\tilde{w}_3}].$$

It follows we can write the covariance as

$$\begin{aligned}
&\mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)K_{h_1^*}(t)K_{h_2^*}(t)K_{h_3^*}(\tilde{w}_3)...K_{h_d^*}(\tilde{w}_d)]+\\
&-\mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)]\mathbb{E}[K_{h_1^*}(t)K_{h_2^*}(t)K_{h_3^*}(\tilde{w}_3)...K_{h_d^*}(\tilde{w}_d)]\\
&=\mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)g(X_{\tilde{w}_3})] - \mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)]\mathbb{E}[g(X_{\tilde{w}_3})]\\
&=\mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)(g(X_{\tilde{w}_3})-\pi(g))]\\
&=\mathbb{E}[K_{h_1^*}(s)K_{h_2^*}(s)K_{h_3^*}(w_3)...K_{h_d^*}(w_d)(P_{\tilde{w}_3-s}g(X_s)-\pi(g))].
\end{aligned}$$

From the β-mixing gathered in Proposition 1 of [5] we easily obtain

$$\|P_{\tilde{w}_3-s}g(X_s)-\pi(g)\|_{L^1} \le ce^{-\rho(\tilde{w}_3-s)}\|g\|_\infty.$$

Therefore,

$$\begin{aligned}
|k(t,s)| &\le \prod_{l=1}^d \left\|K_{h_l^*}\right\|_\infty \|P_{\tilde{w}_3-s}g(X_s)-\pi(g)\|_{L^1}\\
&\le \frac{c}{\prod_{l=1}^d h_l^*}e^{-\rho(\tilde{w}_3-s)}\|g\|_\infty\\
&\le \frac{c}{(\prod_{l=1}^d h_l^*)^2}e^{-\rho(\tilde{w}_3-s)},
\end{aligned}$$

where we have also used that, from the definition of g, it is $\|g\|_\infty \le \frac{c}{\prod_{l=1}^d h_l^*}$. Moreover, acting as in (33) and remarking that $(\prod_{l\ge 3} h_l^*)^{\frac{2}{d-2}} \le D$, we easily get

$$|\tilde{w}_3 - s| \ge |t-s| - \Delta_n.$$

With the change of variable $s \to s' := t-s$ we obtain

$$\begin{aligned}
I_4 &\le \frac{c}{(\prod_{l=1}^d h_l^*)^2}\frac{1}{T_n^2}\int_0^{T_n}\int_D^{T_n} e^{-\rho s'}e^{\rho\Delta_n}dtds' \qquad (35)\\
&\le \frac{c}{T_n(\prod_{l=1}^d h_l^*)^2}e^{-\rho D}
\end{aligned}$$

Putting all the pieces together, using in particular (23), (30), (34) and (35), it yields

$$Var(\bar{\pi}_{h^*,T_n}(x)) \le \frac{c}{T_n}\frac{\sum_{j=1}^d|\log h_j^*|}{\prod_{l=3}^d h_l^*} + \frac{c}{T_n}\frac{\sum_{j=1}^d|\log h_j^*|}{\prod_{l=3}^d h_l^*} + \frac{c}{T_n}\frac{1}{\prod_{l=3}^d h_l^*} + \frac{D}{T_n} + \frac{c}{T_n(\prod_{l=1}^d h_l^*)^2}e^{-\rho D}.$$

By choosing $D := [\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1) \wedge T_n]$ we obtain the wanted result. □

7.2 Proof of Proposition 3

Proof. The proof of Proposition 3 follows the route of Proposition 5. We introduce now

$$k(t,s) := Cov(\prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l}), \prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})),$$

such that

$$Var(\hat{\pi}^a_{h^*,T_n}(x)) = \frac{2}{T_n^2}\int_0^{T_n}\int_0^t k(t,s)1_{s<t}dsdt.$$

We write

$$\begin{aligned}Var(\hat{\pi}^a_{h^*,T_n}(x)) &= \frac{2}{T_n^2}\int_0^{T_n}\int_0^t k(t,s)1_{s<t}\big(1_{|t-s|\le h_1^*h_2^*}\\ &+ 1_{h_1^*h_2^*\le|t-s|\le(\prod_{j\ge3}h_j^*)^{\frac{2}{d-2}}} + 1_{(\prod_{j\ge3}h_j^*)^{\frac{2}{d-2}}\le|t-s|\le D} + 1_{D\le|t-s|\le T_n}\big)dsdt\\ &= \sum_{j=1}^4 \tilde{I}_j,\end{aligned}$$

with h^* the rate optimal choice of the bandwidth given by (6). Regarding $\tilde{I}_1$, exactly as in (23) we have

$$\tilde{I}_1 \le \frac{c}{T_n}\frac{\sum_{j=1}^d |\log h_j^*|}{\prod_{l=3}^d h_l^*}.$$

In order to bound $\tilde{I}_2$ we order the quantities $\varphi_{n,1}(s)$, ... , $\varphi_{n,d}(s)$, $\varphi_{n,1}(t)$, ... , $\varphi_{n,d}(t)$ in a way similar to the proof of Proposition 5 above, introducing w_1, ... , w_d, $\tilde{w}_1$, ... , $\tilde{w}_d$. In a similar way as before, σ and $\tilde{\sigma}$ are elements of the permutation group such that $w_i = \varphi_{n,\sigma(i)}(s)$ and $\tilde{w}_i = \varphi_{n,\tilde{\sigma}(i)}(t)$ for all $i \in \{1,...,d\}$. Moreover, as $\Delta_n \le \frac{1}{4}h_1^*h_2^*$ and $|t-s| > h_1^*h_2^*$, we have

$$w_1 \le ... \le w_d \le s \le \tilde{w}_1 \le ... \le \tilde{w}_d \le t.$$

It means that, in the notation introduced in the proof of Proposition 5, h is equal to 0. We also introduce the vectors, which represent the positions in the instants w_j and $\tilde{w}_j$. At the instant w_j we have the vector y^j, for $j \in \{1,...,d\}$, while the time $\tilde{w}_j$ are associated to the vectors $\tilde{y}^j$, for $j \in \{1,...,d\}$. As before, we observe we can write

$$\begin{aligned}|k(t,s)| &\le |\tilde{k}(t,s)| + |\mathbb{E}[\prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l})]||\mathbb{E}[\prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})]|\\ &\le |\tilde{k}(t,s)| + c,\end{aligned}$$

where

$$\tilde{k}(t,s) := \mathbb{E}[\prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(t)}}{h_l})\prod_{l=1}^d K(\frac{x_l - X^l_{\varphi_{n,l}(s)}}{h_l})].$$

Hence, we have

$$\begin{aligned}|\tilde{k}(t,s)| \le \int_{\mathbb{R}^{d^2}}\prod_{l=1}^d |K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})|\int_{\mathbb{R}^{d^2}}\prod_{l=1}^d |K_{h_{\tilde{\sigma}(l)}}(x_{\tilde{\sigma}(l)} - \tilde{y}^l_{\tilde{\sigma}(l)})|p_{w_2-w_1}(y^1,y^2)p_{w_3-w_2}(y^2,y^3) \quad (36)\\ \times ... \times p_{w_d-w_{d-1}}(y^{d-1},y^d)p_{\tilde{w}_1-w_d}(y^d,\tilde{y}^1)\times ... \times p_{\tilde{w}_d-\tilde{w}_{d-1}}(\tilde{y}^{d-1},\tilde{y}^d)\pi(y^1)dy^1dy^2...dy^d d\tilde{y}^1...d\tilde{y}^d.\end{aligned}$$

As it is important to remove the contribution of the two smallest bandwidth, we need to reorder the components to $\tilde{y}$. To do that we introduce $\tilde{\sigma}^{-1}$, the inverse of the permutation $\tilde{\sigma}$, and we write

$$\begin{aligned}|\tilde{k}(t,s)| \le \int_{\mathbb{R}^{d^2}}\prod_{l=1}^d |K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})|\int_{\mathbb{R}^{d^2}}\prod_{l=1}^d |K_{h_l}(x_l - \tilde{y}_l^{\tilde{\sigma}^{-1}(l)})|p_{w_2-w_1}(y^1,y^2)p_{w_3-w_2}(y^2,y^3)\\ \times ... \times p_{w_d-w_{d-1}}(y^{d-1},y^d)p_{\tilde{w}_1-w_d}(y^d,\tilde{y}^1)\times ... \times p_{\tilde{w}_d-\tilde{w}_{d-1}}(\tilde{y}^{d-1},\tilde{y}^d)\pi(y^1)dy^1dy^2...dy^d d\tilde{y}^1...d\tilde{y}^d.\end{aligned}$$

We use the upper bound $\prod_{l=3}^d \left|K_{h_l^*}(x_l - \tilde{y}_l^{\tilde{\sigma}^{-1}(l)})\right| \leq C/\prod_{l=3}^d h_l^*$ and we integrate with respect to the variables $\tilde{y}_l^{\sigma^{-1}(l)}$, $l = 3, \ldots, d$ to get

$$|\tilde{k}(t,s)| \leq \frac{C}{\prod_{l=3}^d h_l^*} \int_{\mathbb{R}^{d^2}} \prod_{l=1}^d \left|K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})\right| \int_{\mathbb{R}^{2d}} \left|K_{h_1^*}(x_1 - \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}) K_{h_2^*}(x_2 - \tilde{y}_2^{\tilde{\sigma}^{-1}(2)})\right|$$
$$\times p_{w_2 - w_1}(y^1, y^2) p_{w_3 - w_2}(y^2, y^3) \times \ldots \times p_{w_d - w_{d-1}}(y^{d-1}, y^d)$$
$$\times p_{\tilde{w}_{i_\star} - w_d}(y^d, \tilde{y}^{i_\star}) p_{\tilde{w}_{i^\star} - \tilde{w}_{i_\star}}(\tilde{y}^{i_\star}, \tilde{y}^{i^\star}) \pi(y^1) dy^1 dy^2 \ldots dy^d d\tilde{y}^{\tilde{\sigma}^{-1}(1)} d\tilde{y}^{\tilde{\sigma}^{-1}(2)}, \quad (37)$$

where $i_\star = \min(\tilde{\sigma}^{-1}(1), \tilde{\sigma}^{-1}(2))$ and $i^\star = \max(\tilde{\sigma}^{-1}(1), \tilde{\sigma}^{-1}(2))$. Using a Gaussian bound on the transition density

$$p_{\tilde{w}_{i_\star} - w_d}(y^d, \tilde{y}^{i_\star}) p_{\tilde{w}_{i^\star} - \tilde{w}_{i_\star}}(\tilde{y}^{i_\star}, \tilde{y}^{i^\star})$$
$$\leq \frac{C}{\tilde{w}_{i_\star} - w_d} q\left((\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 1}, (\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 2} \mid y^d, \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}, \tilde{y}_2^{\tilde{\sigma}^{-1}(2)}\right) \quad (38)$$

where

$$q\left((\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 1}, (\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 2} \mid y^d, \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}, \tilde{y}_2^{\tilde{\sigma}^{-1}(2)}\right) =$$
$$\sqrt{w_{i_\star} - w_d} \prod_{\substack{j=1 \\ j \neq \tilde{\sigma}(i_\star)}}^d \frac{e^{-c\frac{(y_j^{i_\star} - y_j^d)^2}{\tilde{w}_{i_\star} - w_d}}}{\sqrt{\tilde{w}_{i_\star} - w_d}} \prod_{j=1}^d \frac{e^{-c\frac{(y_j^{i^\star} - y_j^{i^\star})^2}{\tilde{w}_{i^\star} - \tilde{w}_{i_\star}}}}{\sqrt{\tilde{w}_{i^\star} - \tilde{w}_{i^\star}}}.$$

We now prove that

$$\sup_{(y^d, \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}, \tilde{y}_2^{\tilde{\sigma}^{-1}(2)}) \in \mathbb{R}^{d+2}} \int_{\mathbb{R}^{2(d-1)}} q\left((\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 1}, (\tilde{y}_j^{\tilde{\sigma}^{-1}(1)})_{j \neq 2} \mid y^d, \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}, \tilde{y}_2^{\tilde{\sigma}^{-1}(2)}\right)$$
$$\prod_{j=2}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(1)}) \prod_{\substack{j=1 \\ j \neq 2}}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(2)}) \leq C. \quad (39)$$

To prove (39), assume, in order the simplify the notations, that $i_\star = \tilde{\sigma}^{-1}(1)$ and $i^\star = \tilde{\sigma}^{-2}(2)$, as the other case can be proved symmetrically. Then, the integral in the left hand side of (39) is

$$\int_{\mathbb{R}^{2(d-1)}} \sqrt{w_{\tilde{\sigma}^{-1}(1)} - w_d} \prod_{j=2}^d \frac{e^{-c\frac{(y_j^{\tilde{\sigma}^{-1}(1)} - y_j^d)^2}{\tilde{w}_{\tilde{\sigma}^{-1}(1)} - w_d}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(1)} - w_d}} \prod_{j=1}^d \frac{e^{-c\frac{(y_j^{\tilde{\sigma}^{-1}(2)} - y_j^{\tilde{\sigma}^{-1}(1)})^2}{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}} \prod_{j=2}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(1)}) \prod_{\substack{j=1 \\ j \neq 2}}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(2)}).$$

Integrating with respect to the measures $\prod_{j=3}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(1)}) \prod_{j=3}^d d(\tilde{y}_j^{\tilde{\sigma}^{-1}(2)})$, we get that the last integral is upper bounded by

$$\int_{\mathbb{R}^2} \sqrt{w_{\tilde{\sigma}^{-1}(1)} - w_d} \frac{e^{-c\frac{(y_2^{\tilde{\sigma}^{-1}(1)} - y_2^d)^2}{\tilde{w}_{\tilde{\sigma}^{-1}(1)} - w_d}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(1)} - w_d}} \prod_{j=1}^2 \frac{e^{-c\frac{(y_j^{\tilde{\sigma}^{-1}(2)} - y_j^{\tilde{\sigma}^{-1}(1)})^2}{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}} d\tilde{y}_2^{\tilde{\sigma}^{-1}(1)} d\tilde{y}_1^{\tilde{\sigma}^{-1}(2)}.$$

Then, integrating with respect to $d\tilde{y}_2^{\tilde{\sigma}^{-1}(1)}$, the convolution of Gaussian kernels yields to the following upper bound for the LHS of (39),

$$\int_{\mathbb{R}} \sqrt{w_{\tilde{\sigma}^{-1}(1)} - w_d} \frac{e^{-c\frac{(y_2^{\tilde{\sigma}^{-1}(2)} - y_2^d)^2}{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - w_d}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - w_d}} \frac{e^{-c\frac{(y_1^{\tilde{\sigma}^{-1}(2)} - y_1^{\tilde{\sigma}^{-1}(1)})^2}{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}}}{\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - \tilde{w}_{\tilde{\sigma}^{-1}(1)}}} d\tilde{y}_1^{\tilde{\sigma}^{-1}(2)}.$$

Using that $\sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(1)} - w_d} \leq \sqrt{\tilde{w}_{\tilde{\sigma}^{-1}(2)} - w_d}$ and that the first exponential inside the integral above is smaller than 1, we deduce that (39) holds true. Using (37)–(39) we deduce

$$|\tilde{k}(t,s)| \leq \frac{C}{\prod_{l=3}^d h_l^*} \frac{1}{\tilde{w}_{i_\star} - w_d} \int_{\mathbb{R}^{d^2}} \prod_{l=1}^d \left|K_{h^*_{\sigma(l)}}(x_{\sigma(l)} - y^l_{\sigma(l)})\right| \int_{\mathbb{R}^2} \left|K_{h_1^*}(x_1 - \tilde{y}_1^{\tilde{\sigma}^{-1}(1)}) K_{h_2^*}(x_2 - \tilde{y}_2^{\tilde{\sigma}^{-1}(2)})\right|$$
$$\times p_{w_2 - w_1}(y^1, y^2) p_{w_3 - w_2}(y^2, y^3) \times \ldots \times p_{w_d - w_{d-1}}(y^{d-1}, y^d) \pi(y^1) dy^1 dy^2 \ldots dy^d d\tilde{y}_1^{\tilde{\sigma}^{-1}(1)} d\tilde{y}_2^{\tilde{\sigma}^{-1}(2)}.$$

Using that π is bounded, and the second point of Lemma 2 with $q_i = \tilde{\sigma}^{-1}(i)$ for $i \in \{1, \dots, d\}$ we deduce

$$|\tilde{k}(t,s)| \le \frac{c}{\tilde{w}_{i_\star} - w_d} \frac{1}{\prod_{l \ge 3} h_l^*} \le \frac{c}{\tilde{w}_1 - w_d} \frac{1}{\prod_{l \ge 3} h_l^*}$$

which implies

$$|k(t,s)| \le \frac{c}{\tilde{w}_1 - w_d} \frac{1}{\prod_{l \ge 3} h_l^*} + c.$$

In order to bound I_2 we also observe that

$$\begin{aligned} |t-s| &\le |t - \tilde{w}_1| + |\tilde{w}_1 - w_d| + |w_d - s| \\ &\le |\tilde{w}_1 - w_d| + 2\Delta_n \\ &\le |\tilde{w}_1 - w_d| + \frac{1}{2} h_1^* h_2^*. \end{aligned}$$

It implies

$$\frac{1}{|\tilde{w}_1 - w_d|} \le \frac{1}{|t-s| - \frac{1}{2} h_1^* h_2^*}.$$

As a consequence

$$|\tilde{I}_2| \le \frac{c}{T_n^2} \int_0^{T_n} \int_0^t \frac{1}{|t-s| - \frac{1}{2} h_1^* h_2^*} \frac{1}{\prod_{l \ge 3} h_l^*} 1_{\frac{1}{2} h_1^* h_2^* \le |t-s| \le \frac{3}{2} (\prod_{l \ge 3} h_l^*)^{\frac{2}{d-2}}} dsdt.$$

By the change of variable $t - s =: t'$, we obtain

$$\begin{aligned} |\tilde{I}_2| &\le \frac{c}{T_n} \int_{\frac{1}{2} h_1^* h_2^*}^{\frac{3}{2} (\prod_{l \ge 3} h_l^*)^{\frac{2}{d-2}}} \frac{c}{t' - \frac{1}{2} h_1^* h_2^*} \frac{1}{\prod_{l \ge 3} h_l^*} dt' \\ &\le \frac{c}{T_n} \frac{\sum_{j=1}^2 |\log h_j^*|}{\prod_{l \ge 3} h_l^*}. \end{aligned}$$

Regarding $\tilde{I}_3$, it is

$$\tilde{I}_3 := \frac{1}{T_n^2} \int_0^{T_n} \int_0^t k(t,s) 1_{(\prod_{j \ge 3} h_j^*)^{\frac{2}{d-2}} \le |t-s| \le D} dsdt.$$

We can write $\tilde{k}(t,s)$ as in (36). We use the rough estimation

$$p_{\tilde{w}_1 - w_d}(y^d, \tilde{y}^1) \le \frac{c}{(\tilde{w}_1 - w_d)^{\frac{d}{2}}}.$$

We replace it in (36) and we recall we have already proven that the everything else is upper bound by a constant. It implies

$$|k(t,s)| \le \frac{c}{(\tilde{w}_1 - w_d)^{\frac{d}{2}}} + c. \tag{40}$$

Acting as before it is easy to see that

$$|\tilde{w}_1 - w_d| \ge |t-s| - 2\Delta_n \ge |t-s| - \frac{1}{2} h_1^* h_2^* \ge \frac{1}{2} |t-s|.$$

From the change of coordinates $t' := t - s$ we obtain

$$\begin{aligned} \tilde{I}_3 &\le \frac{c}{T_n^2} \int_0^{T_n} \int_{\frac{1}{2} (\prod_{l \ge 3} h_l^*)^{\frac{2}{d-2}}}^{D} \left(\frac{c}{t'^{\frac{d}{2}}} + c\right) dt' ds \\ &\le \frac{c}{T_n} \left(\left(\prod_{l \ge 3} h_l^*\right)^{\frac{2}{d-2}(1 - \frac{d}{2})} + D \right) \\ &= \frac{c}{T_n} \left(\left(\prod_{l \ge 3} h_l^*\right)^{-1} + D \right), \end{aligned}$$

which is the order we wanted.
We are left to study $\tilde{I}_4$, which the case where $D \le |t-s| \le T_n$. Here we act as in order to bound I_4, in the proof of Proposition 5. To do that, we introduce

$$g(X_{\tilde{w}_1}) := \mathbb{E}[K_{h_1^*}(\tilde{w}_1)...K_{h_d^*}(\tilde{w}_d)|X_{\tilde{w}_1}].$$

It follows we can write the covariance as

$$\begin{aligned}&\mathbb{E}[K_{h_1^*}(w_1)...K_{h_d^*}(w_d)g(X_{\tilde{w}_1})] - \mathbb{E}[K_{h_1^*}(w_1)...K_{h_d^*}(w_d)]\mathbb{E}[g(X_{\tilde{w}_1})]\\ &= \mathbb{E}[K_{h_1^*}(w_1)...K_{h_d^*}(w_d)(g(X_{\tilde{w}_1}) - \pi(g))]\\ &= \mathbb{E}[K_{h_1^*}(w_1)...K_{h_d^*}(w_d)(P_{\tilde{w}_1 - w_d}g(X_{w_d}) - \pi(g))].\end{aligned}$$

From the β-mixing gathered in Proposition 1 of [5] we get Therefore,

$$\begin{aligned}|k(t,s)| &\le \prod_{l=1}^d \left\|K_{h_l^*}\right\|_\infty \left\|P_{\tilde{w}_1 - w_d}g(X_{w_d}) - \pi(g)\right\|_{L^1}\\ &\le \frac{c}{\prod_{l=1}^d h_l^*} e^{-\rho(\tilde{w}_1 - w_d)} \left\|g\right\|_\infty \\ &\le \frac{c}{(\prod_{l=1}^d h_l^*)^2} e^{-\rho(\tilde{w}_1 - w_d)}.\end{aligned} \tag{41}$$

As before, it is clearly $|\tilde{w}_1 - w_d| \ge |t-s| - 2\Delta_n$. Hence,

$$\begin{aligned}\tilde{I}_4 &\le \frac{c}{(\prod_{l=1}^d h_l^*)^2} \frac{1}{T_n^2} \int_0^{T_n} \int_D^{T_n} e^{-\rho s'} e^{\rho \Delta_n} dt ds'\\ &\le \frac{c}{T_n(\prod_{l=1}^d h_l^*)^2} e^{-\rho D}\end{aligned}$$

Putting all the pieces together, it yields

$$Var(\hat{\pi}^a_{h^*,T_n}(x)) \le \frac{c}{T_n} \frac{\sum_{j=1}^d |\log h_j^*|}{\prod_{l=3}^d h_l^*} + \frac{c}{T_n} \frac{\sum_{j=1}^d |\log h_j^*|}{\prod_{l=3}^d h_l^*} + \frac{c}{T_n}\frac{1}{\prod_{l=3}^d h_l^*} + \frac{D}{T_n} + \frac{c}{T_n(\prod_{l=1}^d h_l^*)^2} e^{-\rho D}.$$

By choosing $D := [\max(-\frac{2}{\rho}\log(\prod_{j=1}^d h_j), 1) \wedge T_n]$ we obtain the wanted result. □

7.3 Proof of Proposition 4

Proof. From the expression of $\hat{\pi}^a_{h,T_n}(x)$ given in Section 4 we have

$$\mathbb{E}[\hat{\pi}^a_{h,T_n}(x)] = \frac{1}{T_n} \int_0^{T_n} \mathbb{E}\left[\prod_{l=1}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(t)})\right] dt. \tag{42}$$

Hence, we focus on $\mathbb{E}\left[\prod_{l=1}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(t)})\right]$ for $t \in [0, T_n]$. We denote by $w_1 \le w_2 \le \dots \le w_d$ a reordering of $(\varphi_{n,l}(t))_{l=1,\dots,d}$, and let σ an element of the permutation group such that $w_i = \varphi_{n,\sigma(i)}(t)$ for all $i \in \{1,\dots,d\}$. With this notations, $\mathbb{E}\left[\prod_{l=1}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(t)})\right] = \mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right]$ and we now show the following control

$$\left|\mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right] - \int_{\mathbb{R}^d} \pi(y) \prod_{i=1}^d K_{h_i}(x_i - z_i) dz_1 \dots dz_d\right| \le c\sqrt{\Delta_n}, \tag{43}$$

for some constant c independent of $(w_i)_i$ and $(h_i)_i$.

In the proof of (43) we can assume by a density argument that $w_i < w_{i+1}$ for $i = 1, \dots, d-1$. In this case, we write

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right] &= \int_{\mathbb{R}^{d^2}} \pi(y^1) \prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - y^i_{\sigma(i)}) \prod_{i=1}^{d-1} p_{w_{i+1}-w_i}(y^i, y^{i+1}) dy^1 \dots dy^d\\ &= \int_{\mathbb{R}^d} \left(\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - y^i_{\sigma(i)})\right) \xi_{(w_i)_i,\pi}(y^1_{\sigma(1)}, \dots, y^d_{\sigma(d)}) dy^1_{\sigma(1)} dy^2_{\sigma(2)} \cdots y^d_{\sigma(d)},\end{aligned} \tag{44}$$

where for any function ϕ we have set

$$\xi_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \int_{\mathbb{R}^{d(d-1)}} \phi(y^1)\prod_{i=1}^{d-1} p_{w_{i+1}-w_i}(y^i,y^{i+1})d\widehat{y}^1\dots d\widehat{y}^d, \tag{45}$$

with $\widehat{y}^i = (y^i_j)_{j\in\{1,\dots,d\}\setminus\{\sigma(i)\}}$. Applying Lemma 1 below with $\phi=\pi$, and (44), we deduce

$$\mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right]$$
$$= \int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - y^i_{\sigma(i)})\right)\pi(y_1^{\sigma^{-1}(1)},\dots,y_d^{\sigma^{-1}(d)})dy^1_{\sigma(1)}\dots dy^d_{\sigma(d)} + O(\sqrt{\Delta_n}).$$

Changing the notation $z_i = y_i^{\sigma^{-1}(i)}$, which is such that $y^i_{\sigma(i)} = z_{\sigma(i)}$, we get

$$\mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right] = \int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - z_{\sigma(i)})\right)\pi(z_1,\dots,z_d)dz_{\sigma(1)}\dots dz_{\sigma(d)} + O(\sqrt{\Delta_n})$$
$$= \int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_i}(x_i - z_i)\right)\pi(z_1,\dots,z_d)dz_1\dots z_d + O(\sqrt{\Delta_n}).$$

This implies (43). Then, recalling (42), and $\mathbb{E}\left[\prod_{l=1}^d K_{h_l}(x_l - X^l_{\varphi_{n,l}(t)})\right] = \mathbb{E}\left[\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)} - X^{\sigma(i)}_{w_i})\right]$ we deduce that

$$\left|\mathbb{E}[\hat{\pi}^a_{h,T_n}(x)] - \int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_i}(x_i - z_i)\right)\pi(z_1,\dots,z_d)dz_1\dots z_d\right| \le c\sqrt{\Delta_n}.$$

The proposition follows from the following upper bound on the bias of the synchronous case (see [8])

$$\left|\int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_i}(x_i - z_i)\right)\pi(z_1,\dots,z_d)dz_1\dots z_d - \pi(x_1,\dots,x_d)\right| \le c\sum_{i=1}^d h_i^{\beta_i}.$$

□

Lemma 1. *Suppose that A1 holds and that a and b are $\mathcal{C}^3$ with bounded derivatives. Let $\phi:\mathbb{R}^d\to\mathbb{R}$ be a bounded $\mathcal{C}^1$ function with bounded derivatives and let us denote*

$$d_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \xi_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) - \phi(y_1^{\sigma^{-1}(1)},\dots,y_d^{\sigma^{-1}(d)}).$$

Then, there exists some constant C such that we have

$$\int_{\mathbb{R}^d}\left(\prod_{i=1}^d \left|K_{h_{\sigma(i)}}(x_{\sigma(i)} - y^i_{\sigma(i)})\right|\right)\left|d_{(w_i)_i,\pi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})\right|dy^1_{\sigma(1)}\dots dy^d_{\sigma(d)} \le C\sqrt{\Delta_n}.$$

The constant C is independent of $(h_i)_i$.

Proof. Let us denote by $g_\alpha(z)$ the density of a centred Gaussian variable with covariance matrix α. We recall the approximation of the diffusion transition density by the Gaussian kernel given in [37]. Specifying $s=1/N$ and $T=1$ in the notations of the statement of Theorem 3 [37], we have for all $s\le 1$,

$$\left|p_s(z,z') - g_{\tilde{a}(z)s}(z-z')\right| \le C\sqrt{s}g_{\lambda_0 \mathrm{Id}s}(z-z'), \tag{46}$$

where $\lambda_0>0$ and $C>0$ are some constant and $\tilde{a} = a\cdot a^T$. This leads us to introduce a Gaussian approximation of (45)

$$\xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \int_{\mathbb{R}^{d(d-1)}}\phi(y^1)\prod_{i=1}^{d-1} g_{(w_{i+1}-w_i)\tilde{a}(y_i)}(y^{i+1}-y^i)d\widehat{y}^1\dots d\widehat{y}^d. \tag{47}$$

With this notation, we split $d_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})$ as $\sum_{l=1}^2 d^{(l)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})$, with

$$d^{(1)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \xi_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) - \xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}), \tag{48}$$

$$d^{(2)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) - \phi(y_1^{\sigma^{-1}(1)},\dots,y_d^{\sigma^{-1}(d)}). \tag{49}$$

The lemma is a consequence of the following upper bound for $l\in\{1,2\}$,

$$\int_{\mathbb{R}^d}\left(\prod_{i=1}^d \left|K_{h_{\sigma(i)}}(x_{\sigma(i)}-y^i_{\sigma(i)})\right|\right)\left|d^{(l)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})\right| dy^1_{\sigma(1)}\dots dy^d_{\sigma(d)} \le C\sqrt{\Delta_n}. \tag{50}$$

• We first prove (50) with $l=1$. Comparing (45) with (47), we can write

$$d^{(1)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \int_{\mathbb{R}^{d(d-1)}} \phi(y^1)\sum_{k=1}^{d-1}\prod_{1\le i<k} p_{w_{i+1}-w_i}(y^i,y^{i+1})$$
$$\times\left[p_{w_{k+1}-w_k}(y^k,y^{k+1}) - g_{(w_{k+1}-w_k)\tilde{a}(y^k)}(y^{k+1}-y^k)\right]\prod_{k<i\le d-1} g_{(w_{i+1}-w_i)\tilde{a}(y^i)}(y^{i+1}-y^i)d\widehat{y}^1\dots d\widehat{y}^d.$$

Using (46) and a Gaussian upper bound of the transition density, we deduce

$$\begin{aligned}\left|d^{(1)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})\right| \le C \sup_{i=1,\dots,d-1}\sqrt{w_{i+1}-w_i}\times\\ \int_{\mathbb{R}^{d(d-1)}}\left|\phi(y^1)\right|\prod_{1\le i\le d-1} g_{(w_{i+1}-w_i)\lambda_0 Id}(y^{i+1}-y^i)d\widehat{y}^1\dots d\widehat{y}^d,\\ \le C\sqrt{\Delta_n}\int_{\mathbb{R}^{d(d-1)}}\prod_{1\le i\le d-1} g_{(w_{i+1}-w_i)\lambda_0 Id}(y^{i+1}-y^i)d\widehat{y}^1\dots d\widehat{y}^d,\end{aligned}$$

for some constant $\lambda_0>0$. It yields,

$$\int_{\mathbb{R}^d}\left(\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)}-y^i_{\sigma(i)})\right)\left|d^{(1)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})\right| dy^1_{\sigma(1)}\dots dy^d_{\sigma(d)}$$
$$\le C\sqrt{\Delta_n}\int_{\mathbb{R}^{d^2}}\prod_{i=1}^d K_{h_{\sigma(i)}}(x_{\sigma(i)}-y^i_{\sigma(i)})\prod_{1\le i\le d-1} g_{(w_{i+1}-w_i)\lambda_0 Id}(y^{i+1}-y^i)dy^1\dots dy^d. \tag{51}$$

From Lemma 2 2) with $r=d$, $u^i=y^i$ for $i=1\dots,d$, and $q_i=\sigma^{-1}(i)$ we deduce that the last integral is upper bounded by some constant, and in turn that (50) holds true for $l=1$.

• We now prove (50) with $l=2$. In the integral defined by the right hand side of (47) we make a change of variables, replacing the variables $(\widehat{y}^1,\dots,\widehat{y}^d)=(y^j_l)_{j\in\{1,\dots,d\},l\in\{1,\dots,d\}\setminus\{\sigma(j)\}}$ by new integration variables $(z^j_l)_{1\le l\le d,2\le j\le d}$ defined in the following way. For $j=d$, we define $z^d=(z^d_1,\dots,z^d_d)$ through the change of variable

$$y^d_l\to z^d_l := \frac{y^d_l-y^{d-1}_l}{\sqrt{w_d-w_{d-1}}} \text{ for } l\in\{1,\dots,d\}\setminus\{\sigma(d)\},\quad y^{d-1}_{\sigma(d)}\to z^d_{\sigma(d)} := \frac{y^d_{\sigma(d)}-y^{d-1}_{\sigma(d)}}{\sqrt{w_d-w_{d-1}}},$$

and more generally for $2\le j\le d$ we define z^j through the formulae

$$y^j_l\to z^j_l := \frac{y^j_l-y^{j-1}_l}{\sqrt{w_j-w_{j-1}}} \text{ for } l\in\{1,\dots,d\}\setminus\{\sigma(d),\sigma(d-1),\dots,\sigma(j)\},$$
$$y^{j-1}_l\to z^j_j := \frac{y^j_l-y^{j-1}_l}{\sqrt{w_j-w_{j-1}}},\ \text{ for } l\in\{\sigma(d),\sigma(d-1),\dots,\sigma(j)\}.$$

From these definitions we have

$$d\widehat{y}^1,\dots,d\widehat{y}^d = \prod_{\substack{j=1,\dots,d\\ l\in\{1,\dots,d\}\setminus\{\sigma(j)\}}} dy^j_l = dz^2\dots dz^d\prod_{j=2}^d(\sqrt{w_j-w_{j-1}})^{d/2}.$$

Moreover by construction $z^j = \frac{y^j - y^{j-1}}{\sqrt{w_j - w_{j-1}}}$ for all $j \in \{2,\dots,d\}$. We deduce that (47) can be written after change of variables as

$$\xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \int_{\mathbb{R}^{d(d-1)}} \phi(\widetilde{y}^1) \prod_{j=1}^{d-1} g_{\tilde{a}(\widetilde{y}^j)}(z^{j+1}) dz^2 \dots dz^d,$$

where $\widetilde{y}^j = \widetilde{y}^j(z^2,\dots,z^d; y^1_{\sigma(1)},\dots,y^d_{\sigma(d)})$ is a notation for the expression of $(y^1,\dots,y^d)$ as a function of the new variables $z^2,\dots,z^d$. They are given by the explicit expression

$$\widetilde{y}^j_l = \begin{cases} y_l^{\sigma^{-1}(l)} + \sum_{u=0}^{j-\sigma^{-1}(l)-1} \sqrt{w_{u+\sigma^{-1}(l)+1} - w_{u+\sigma^{-1}(l)}}\, z_l^{u+1+\sigma^{-1}(l)}, & \text{if } j > \sigma^{-1}(l), \\ y_l^{\sigma^{-1}(l)}, \quad \text{if } j = \sigma^{-1}(l), \\ y_l^{\sigma^{-1}(l)} - \sum_{u=0}^{\sigma^{-1}(l)-j-1} \sqrt{w_{u+j+1} - w_{u+j}}\, z_l^{u+1+j}, & \text{if } j < \sigma^{-1}(l). \end{cases}$$

This leads us to introduce the following notation which stresses the dependence upon the time intervals $w_{u+1} - w_u$. We set for $s_1,\dots,s_{d-1} > 0$,

$$\widehat{y}^j_l(s_1,\dots,s_{d-1}) = \begin{cases} y_l^{\sigma^{-1}(l)} + \sum_{u=0}^{j-\sigma^{-1}(l)-1} s_{u+\sigma^{-1}(l)} z_l^{u+1+\sigma^{-1}(l)}, & \text{if } j > \sigma^{-1}(l), \\ y_l^{\sigma^{-1}(l)}, \quad \text{if } j = \sigma^{-1}(l), \\ y_l^{\sigma^{-1}(l)} - \sum_{u=0}^{\sigma^{-1}(l)-j-1} s_{u+j} z_l^{u+1+j}, & \text{if } j < \sigma^{-1}(l), \end{cases} \tag{52}$$

and we define

$$\widehat{\xi^{\mathbf{G}}}(s_1,\dots,s_{d-1}) = \int_{\mathbb{R}^{d(d-1)}} \phi(\widehat{y}^1(s_1,\dots,s_{d-1})) \prod_{j=1}^{d-1} g_{\tilde{a}(\widehat{y}^j(s_1,\dots,s_{d-1}))}(z^{j+1}) dz^2 \dots dz^d. \tag{53}$$

With these notations, $\widetilde{y}^j_l = \widehat{y}^j_l(\sqrt{w_2 - w_1},\dots,\sqrt{w_d - w_{d-1}})$ for all $1 \le j,l \le d$, and

$$\xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) = \widehat{\xi^{\mathbf{G}}}(\sqrt{w_2 - w_1},\dots,\sqrt{w_d - w_{d-1}}). \tag{54}$$

For $(s_1,\dots,s_{d-1}) = (0,\dots,0)$ these quantities have simpler expressions. Let us denote $y^\star = (y_1^{\sigma^{-1}(1)},\dots,y_1^{\sigma^{-1}(d)})$, and remark that from (52), we have $\widehat{y}^j(0,\dots,0) = y^\star$, for all $1 \le j \le d$. It follows

$$\widehat{\xi^{\mathbf{G}}}(0,\dots,0) = \int_{\mathbb{R}^{d(d-1)}} \phi(y^\star) \prod_{j=1}^{d-1} g_{\tilde{a}(y^\star)}(z^{j+1}) dz^2 \dots dz^d = \phi(y^\star) \tag{55}$$

where we have used that $y^\star$ does not depend on the integration variable and that the Gaussian kernel integrates to one. We deduce from (54)–(55),

$$\begin{aligned} \left| \xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) - \phi(y^\star) \right| &= \left| \widehat{\xi^{\mathbf{G}}}(\sqrt{w_2 - w_1},\dots,\sqrt{w_d - w_{d-1}}) - \widehat{\xi^{\mathbf{G}}}(0,\dots,0) \right| \\ &\le \sqrt{\Delta_n} \sum_{j=2}^{d} \sup_{0 \le s_1,\dots,s_{d-1} \le \sqrt{\Delta_n}} \left| \frac{\partial}{\partial s_j} \widehat{\xi^{\mathbf{G}}}(s_1,\dots,s_{d-1}) \right| \end{aligned} \tag{56}$$

where we used $\sqrt{w_j - w_{j-1}} \le \sqrt{\Delta_n}$ for all $2 \le j \le d$. From the definition (52), we have $\left| \frac{\partial \widehat{y}^i_l}{\partial s_j} \right| \le c \sum_{u=2}^d \|z^u\|$. Using that ϕ and $\tilde{a}$ are $\mathcal{C}^1$ functions, bounded with bounded derivative, and that $\tilde{a} \ge a^2_{\min} Id$, we deduce from (53) that

$$\left| \frac{\partial \widehat{\xi^{\mathbf{G}}}(s_2,\dots,s_d)}{\partial s_j} \right| \le C \int_{\mathbb{R}^{d(d-1)}} \sum_{u=2}^{d} (1 + \|z^u\|^3) \prod_{j=1}^{d-1} g_{\tilde{a}(\widehat{y}^j(s_1,\dots,s_{d-1}))}(z^{j+1}) dz^2 \dots dz^d.$$

Used that $\tilde{a}$ is a bounded function under Assumption A1, we deduce that the last integral is bounded independently of $s_1,\dots,s_{d-1}$, and thus $\left| \sup_{s_1,\dots,s_{d-1}} \frac{\partial \widehat{\xi^{\mathbf{G}}}(s_2,\dots,s_d)}{\partial s_j} \right| \le C$. In turn, (56) implies,

$$\left| \xi^{\mathbf{G}}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) - \phi(y^\star) \right| \le C\sqrt{\Delta_n},$$

for a constant C independent of $(w_i)_i$ and $y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}$. Recalling (49) and the notation $y^\star = (y_1^{\sigma^{-1}(1)},\dots,y_1^{\sigma^{-1}(d)})$, this is the upper bound $\left| d^{(2)}_{(w_i)_i,\phi}(y^1_{\sigma(1)},\dots,y^d_{\sigma(d)}) \right| \le C\sqrt{\Delta_n}$, and we deduce (50) with $l = 2$ by integration. □

Lemma 2. *1) Let $r \geq 1$ be an integer and $q_1, \ldots, q_r \in \{1, \ldots, d\}$. Then, we have*

$$\int_{\mathbb{R}^{dr}} \prod_{i=1}^{d} |K_{h_i}(x_i - u_i^{q_i})| \prod_{j=1}^{r-1} p_{w_{j+1}-w_j}(u^j, u^{j+1}) du^1 \ldots du^r \leq C, \tag{57}$$

for some constant C independent of $(h_i)_i$ and $(w_j)_j$.

2) The same upper bound holds true if we replace the transition densities $p_{w_{j+1}-w_j}$ by Gaussian kernels $g_{(w_{j+1}-w_j)\lambda_0 Id}$.

Proof. We only prove the first point as the proof of the second is similar. Using the Gaussian upper bound on the transition density (e.g. see Proposition 5.1 in [35])

$$p_{w_{j+1}-w_j}(u^j, u^{j+1}) \leq C \frac{1}{(w_{j+1} - w_j)^{d/2}} e^{-\lambda_0 \frac{|u^{j+1}-u^j|^2}{w_{j+1}-w_j}},$$

we deduce that the left hand side of (57) is smaller than

$$C \int_{\mathbb{R}^{dr}} \prod_{i=1}^{d} \left[|K_{h_i}(x_i - u_i^{q_i})| \prod_{j=1}^{r-1} \frac{1}{(w_{j+1} - w_j)^{1/2}} e^{-\lambda_0 \frac{(u_i^{j+1}-u_i^j)^2}{w_{j+1}-w_j}} \right] \prod_{i=1}^{d} \left(\prod_{j=1}^{r} du_i^j \right)$$

which is equal to

$$C \prod_{i=1}^{d} \left(\int_{\mathbb{R}^r} |K_{h_i}(x_i - u_i^{q_i})| \prod_{j=1}^{r-1} \frac{1}{(w_{j+1} - w_j)^{1/2}} e^{-\lambda_0 \frac{(u_i^{j+1}-u_i^j)^2}{w_{j+1}-w_j}} du_i^1 \ldots du_i^r \right). \tag{58}$$

It is sufficient to show that for all $i \in \{1, \ldots, d\}$ the integrals in the product (58) are smaller than some constant independent of $(h_i)_i$ and $(w_j)_j$. We successively integrate in $u_i^r, u_i^{r-1} \ldots, u_i^{q_i+1}$, using the change of variables $u_i^j \to z_i^j := \frac{u_i^j - u_i^{j-1}}{\sqrt{w_{j+1}-w_j}}$ for $j = r, r-1, \ldots, q_i + 1$. Next, we successively integrate in $u_i^1, \ldots, u_i^{q_i-1}$, using the change of variable $u_i^j \to z_i^j := \frac{u_i^{j+1} - u_i^j}{\sqrt{w_{j+1}-w_j}}$. We deduce that the integrals appearing in (58) are upper bounded by $C \int_{\mathbb{R}} K_{h_i}(x_i - u_i^{q_i}) du_i^{q_i} \leq C$. This proves the lemma. □

7.4 Proof of Theorem 4

Proof. The proof is a straightforward consequence of the bias-variance decomposition, Proposition 3 and Proposition 4. It is also based on the discussion on the conditions on the discretization step located after the statement of Theorem 4. □

7.5 Proof of Theorem 5

Proof. The proof is a straightforward consequence of the bias-variance decomposition, Proposition 5 and Proposition 4, with $\varphi_{n,l}(t) = t$ for $l = 1, 2$. Again, it is also based on the discussion on the conditions on the discretization step located after the statement of Theorem 5. □

References

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

[2] Ait-Sahalia, Y., Fan, J. and Xiu, D. (2010). High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data. Journal of the American Statistical Association 105 1504-1517.

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

[4] Amorino, C. (2020). Rate of estimation for the stationary distribution of jump-processes over anisotropic Holder classes. Electronic Journal of Statistics, to appear.

[5] Amorino, C., & Gloter, A. (2021). Minimax rate of estimation for invariant densities associated to continuous stochastic differential equations over anisotropic Holder classes. arXiv preprint arXiv:2110.02774.

[6] Amorino, C., & Gloter, A. (2021). Optimal convergence rate in intermediary regime for the invariant density of diffusions discretely observed.

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

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

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

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

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

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

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

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

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

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

[17] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics, 162(2), 149-169.

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

[19] Binkowski, M., Marti, G. and Donnat, P. (2017). Autoregressive Convolutional Neural Networks for Asynchronous Time Series.

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

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

[22] Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997). The Econometrics of Financial Markets. Princeton University Press.

[23] Cheang, C. W. (2018). Three Essays in Financial Econometrics: Fractional Cointegration, Nonlinearities and Asynchronicities, PhD thesis, University of Southampton.

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

[25] Clément, E. (2021). Hellinger distance in approximating L é vy driven SDEs and application to asymptotic equivalence of statistical experiments. arXiv preprint arXiv:2103.09648.

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

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

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

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

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

[31] Fisher, L. (1966). Some new stock-market indexes. Journal of Business 39, 191–225.

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

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

[34] Gloter, A., & Gobet, E. (2008). LAMN property for hidden processes: the case of integrated diffusions. In Annales de l'IHP Probabilités et statistiques (Vol. 44, No. 1, pp. 104-128).

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

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

[37] Gobet, E., Labart, C. Sharp estimates for the convergence of the density of the Euler scheme in small time. Elect. Comm. in Probab., 1,p 352-363, 2008

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

[39] Hayashi, T., & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. Bernoulli, 11(2), 359-379.

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

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

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

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

[44] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, 24(2), 211-229.

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

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

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

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

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

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

[51] Lejay, A. and Pigato, P. (2018) Statistical estimation of the Oscillating Brownian Motion. Bernoulli.

[52] Lejay, A. and Pigato, P. (2020) Maximum likelihood drift estimation for a threshold diffusion. Scandinavian Journal of Statistics, 47(3):609–637.

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

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

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

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

[57] Mazzonetto, S., & Pigato, P. (2020). Drift estimation of the threshold Ornstein-Uhlenbeck process from continuous and discrete observations. arXiv preprint arXiv:2008.12653.

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

[59] Nualart, D. (2006). The Malliavin calculus and related topics (Vol. 1995). Berlin: Springer.

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

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

[62] Peluso, S., Corsi, F., & Mira, A. (2014). A Bayesian high-frequency estimator of the multivariate covariance of noisy and asynchronous returns. Journal of Financial Econometrics, 13(3), 665-697.

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

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

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

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

[67] Strasser, H. (1985). Mathematical theory of statistics, volume 7 of de Gruyter Studies in Mathematics.

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

[69] Tsybakov, A. B. (2009). Introduction to Nonparametric Estimation. Springer series in statistics, I-XII.

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

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