



HAL
open science

Adaptive estimation for stochastic damping Hamiltonian systems under partial observation

Fabienne Comte, Clémentine Prieur, Adeline Samson

► **To cite this version:**

Fabienne Comte, Clémentine Prieur, Adeline Samson. Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. *Stochastic Processes and their Applications*, 2017, 127 (11), pp.3689 - 3718. 10.1016/j.spa.2017.03.011 . hal-01659337

HAL Id: hal-01659337

<https://hal.science/hal-01659337v1>

Submitted on 15 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE ESTIMATION FOR STOCHASTIC DAMPING HAMILTONIAN SYSTEMS UNDER PARTIAL OBSERVATION.

FABIENNE COMTE ⁽¹⁾, CLÉMENTINE PRIEUR ⁽²⁾, AND ADELIN SAMSON ⁽²⁾

ABSTRACT. The paper considers a process $Z_t = (X_t, Y_t)$ where X_t is the position of a particle and Y_t its velocity, driven by a hypoelliptic bi-dimensional stochastic differential equation. Under adequate conditions, the process is stationary and geometrically β -mixing. In this context, we propose an adaptive non-parametric kernel estimator of the stationary density p of Z , based on n discrete time observations with time step δ . Two observation schemes are considered: in the first one, Z is the observed process, in the second one, only X is measured. Estimators are proposed in both settings and upper risk bounds of the mean integrated squared error (MISE) are proved and discussed in each case, the second one being more difficult than the first one. We propose a data driven bandwidth selection procedure based on the Goldenshluger and Lepski [2011] method. In both cases of complete and partial observations, we can prove a bound on the MISE asserting the adaptivity of the estimator. In practice, we take advantage of a very recent improvement of the Goldenshluger and Lepski [2011] method provided by Lacour et al. [2016], which is computationally efficient and easy to calibrate. We obtain convincing simulation results in both observation contexts.

⁽¹⁾ UNIVERSITÉ PARIS DESCARTES, LABORATOIRE MAP5, UMR CNRS 8145, PARIS

⁽²⁾ LABORATOIRE JEAN KUNTZMANN, UMR CNRS 5224, GRENOBLE

Keywords: adaptive bandwidth selection, hypoelliptic diffusion, kernel density estimation, partial observations.

AMS Subject Classification: Primary 62M05; Secondary 62G07, 60H10, 35H10.

1. INTRODUCTION

In this paper, we focus on a hypoelliptic bi-dimensional stochastic differential equation, called *Langevin* or *hypoelliptic, damping Hamiltonian* system. It can be viewed as describing the dynamic of a particle: the first coordinate represents the position of the particle and the second its velocity. The position is defined as the (deterministic) integral of the velocity, but the dynamic of the velocity is modeled with a stochastic noise. Thus, the noise occurs only in the second component leading to a hypoelliptic stochastic differential equation: the diffusion coefficient matrix is singular. Such models appear in many domains such as random mechanics, finance, biology, ecology... We refer to Pokern et al. [2009] for some examples of such models arising in applications. There is thus a need for efficient estimation methods. But the singularity of the diffusion coefficient prevents to apply standard estimation methods that have been developed for (multi-dimensional) stochastic differential equations (see more details below). The aim of this paper is to propose a data-driven non-parametric estimator of the stationary density of the system.

Date: March 13, 2017.

More precisely, let us denote X_t the position of the particle, Y_t its velocity, and $Z_t := (X_t, Y_t)$. We assume that $(Z_t, t \geq 0)$ is governed by the following stochastic differential equation:

$$(1) \quad \begin{cases} dX_t = Y_t dt \\ dY_t = \sigma dW_t - (c(X_t, Y_t) Y_t + \nabla D(X_t)) dt. \end{cases}$$

with $c(x, y)$ the damping force, D the potential, $\sigma > 0$ the diffusion coefficient and $(W_t)_{t \geq 0}$ a standard Brownian motion. We assume that the damping force c and the potential D are regular enough to ensure the existence and uniqueness of a non explosive solution of (1).

Our objective is to estimate the stationary density p of Z_t . Except from specific choices of c and D , this stationary density has no closed expression. We thus consider the non-parametric framework.

In some applications, it is not possible to measure the two coordinates and only discrete time observations are available. We will proceed in two steps: First we consider the case of “complete observation” where we have at hand a sample $(X_{i\delta}, Y_{i\delta})_{1 \leq i \leq n}$ for a time step δ of observations. Second, we consider the “partial observation” case where only the sample $(X_{i\delta})_{1 \leq i \leq n}$ is available.

Estimation of such hypoelliptic systems has already been studied. In the parametric framework, we refer to Gloter [2006], Samson and Thieullen [2012] who propose contrasts to estimate parameters σ, c and those appearing in D . Pokern et al. [2009] propose estimation strategies for parameters appearing in D through a complex Gibbs algorithm in a Bayesian parametric framework.

Non-parametric estimation in stochastic differential equations (SDE) has been widely studied in the elliptic case, i.e. when a noise also drives the first coordinate and when the diffusion coefficient is not singular. For elliptic SDE, we can cite among others Comte et al. [2007] for a data-driven non-parametric estimation procedure of the drift and diffusion coefficient in a one-dimensional stochastic differential equation discretely observed with small step. The strategy has been extended in Comte et al. [2009] when only discrete integrated observations are available. The method is based on projection estimators minimizing a mean-square contrast with model selection. The problem here is different since estimation is bi-dimensional and focused on the stationary density. Extension of adaptive non-parametric drift estimation to multidimensional setting has been considered in Schmisser [2013a]. Especially adaptive projection estimation of the stationary density is developed in Schmisser [2013b], see also the references on the topic therein. However, the results in Schmisser [2013b] are first stated for one-dimensional mixing processes, and in a second step, applied to elliptic stochastic differential equation in order to estimate the drift when the observation step is fixed. We emphasize that the elliptic assumption is not easy to weaken. Thus the hypoelliptic case considered in Model (1) is not considered in the previous estimation methods for multi-dimensional SDE. The main element that explains why hypoellipticity is difficult to work with is the form of the transition density. The transition density of hypoelliptic diffusion explodes when the step size δ between two observations goes to 0 at a specific rate: $1/\delta^2$. This rate directly deteriorates the variance of the estimator and impacts the rate of convergence of the estimator.

The main references in non-parametric estimation of multi-dimensional hypoelliptic stochastic differential equations are Cattiaux et al. [2014a,b]. In Cattiaux et al. [2014b], a kernel estimator

of the stationary density p is proposed, in both the complete observation and the partial observation cases. The authors prove the consistency and the asymptotic normality of their estimators. But the problem of the automatic data-driven selection of the bandwidth is not studied.

Our study is related to adaptive density estimation for dependent variables, which was considered by Comte and Merlevède [2002, 2005], relying on useful tools and inequalities proved in Viennet [1997]; these authors study univariate estimation and use penalized projection estimators. In most cases, the penalty used in the procedure involves an unknown coefficient related to the dependence of the data. From this point of view, their proposal has been improved by Lerasle [2011], who studies an original penalization based on a resampling strategy. Note that different types of dependence have been studied by Gannaz and Wintenberger [2010]. Recently, Asin and Johannes [2016] investigated the topic, still for univariate projection estimators: using a specific assumption on the joint density of two observations, they can avoid the mixing coefficient in the penalty. Lastly, we mention that a strategy similar to the one presented here but in a pointwise univariate setting is presented in Bertin and Klutchnikoff [2016]: more precisely, they propose a kernel estimator with point-by-point selection of the bandwidth. Their procedure is more demanding than the global strategy studied and implemented in the present paper, but may be adapted to inhomogeneous functions; however, it involves a log-loss in the rate, which is known to be unavoidable in the independent setting.

In this paper, we want to provide a non-parametric estimator of p with a fully data-driven procedure of the bandwidth. We first study the integrated risk of the kernel estimator. We distinguish the case of an estimation of p on a compact subset of \mathbb{R}^2 or on the whole real plane, which give two different bounds. Then, under regularity assumptions on p , we provide rates of convergence of the estimators. Finally, we propose a non-asymptotic fully data-driven selection procedure of the bandwidth of the kernel estimator, the procedure is inspired by the methodology detailed in Goldenshluger and Lespki [2011]. The method has the decisive advantage of being anisotropic: the bandwidths selected in each direction are in general different, which is coherent with the possibly different regularities with respect to each variable. Finally, we prove oracle results of our final estimators. However, the implementation of the Goldenshluger and Lespki [2011] method is complex in the multi-dimensional setting: the convolution of estimators is computationally demanding and the influence of the constant is not intuitive, thus preventing for using automatic procedure such as the slope heuristic. We thus implement a simplified version of the method of Goldenshluger and Lespki [2011], that has been most recently developed by Lacour et al. [2016]. Their simplification is a decisive improvement for the implementability of the method in the multidimensional setting: it avoids the computation of convolution of estimators and allows to calibrate automatically the constant by slope heuristic. We illustrate the two estimators on two hypoelliptic systems: the Harmonic Oscillator and the Van Der Pol Oscillator.

The paper is organized as follows. We give in Section 2 assumptions on the system ensuring that the aim is meaningful. In particular, we take advantage of previous probabilistic studies of the system provided in Konakov et al. [2010] and Cattiaux et al. [2014b]. We study first the estimation of p from discrete but complete observations of Z in Section 3, and then the more realistic but difficult case of partial observations in Section 4. Simulation experiments illustrate the methodology in the two observation contexts in Section 5.

2. PROBABILISTIC USEFUL RESULTS

2.1. Assumptions on equation (1). In this section, we recall some results proved in Cattiaux et al. [2014b]. We first give some assumptions on the model.

In all the following, we assume that the potential D is lower bounded, smooth over \mathbb{R} , that D and ∇D have polynomial growth at infinity and that

$$(2) \quad 0 < d \leq \liminf_{|x| \rightarrow +\infty} \frac{x \cdot \nabla D(x)}{|x|} \leq +\infty,$$

the latter being often called the **drift condition**. We also assume that the damping coefficient $c(x, y)$ is smooth and bounded, and there exist $c, L > 0$ such that $c(x, y) \geq cId > 0$, $\forall (|x| > L, y \in \mathbb{R})$.

Hypoelliptic property of the system ensures that the distribution $P_t(z, \cdot)$ of the process Z_t starting from $Z_0 = z$ has a smooth density $p_t(z, \cdot)$. Some bounds are given in Appendix 6.1, especially the fact that transition density $p_t(z, \cdot)$ is explosive when $t \rightarrow 0$ with a specific (hypoelliptic) rate of explosion $1/t^2$ (bound (27)). Moreover, the drift condition (2) ensures that the process is positive recurrent with a unique invariant probability measure μ . This measure has a smooth density $\mu(dz) = p(z)dz$ [Cattiaux et al., 2014b], with

$$\mu(dx, dy) = \exp\left(-\frac{2c}{\sigma^2}H(x, y)\right) dx dy$$

where $H(x, y) = \frac{1}{2}y^2 + D(x)$ is the Hamiltonian (see [Roberts and Spanos, 2003]).

Under the drift condition, it is proved in Wu [2001] (see Theorem 3.1 therein) that the Markov chain $(Z_i)_{i \in \mathbb{N}}$, with initial condition $Z_0 \sim p(z)dz$ is exponentially β -mixing. We recall hereafter the meaning of this property as well as useful bounds and properties.

2.2. About β -mixing coefficients and processes. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let Γ be a random variable with values in a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. Let \mathcal{M} be a σ -algebra of \mathcal{A} . Let $\mathbb{P}_{\Gamma|\mathcal{M}}$ be a conditional distribution of Γ given \mathcal{M} , \mathbb{P}_{Γ} be the distribution of Γ , and let $\mathcal{B}(\mathbb{B})$ be the Borel σ -algebra on $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. Define now

$$\beta(\mathcal{M}, \sigma(\Gamma)) = \mathbb{E}\left(\sup_{A \in \mathcal{B}(\mathbb{B})} |\mathbb{P}_{\Gamma|\mathcal{M}}(A) - \mathbb{P}_{\Gamma}(A)|\right).$$

The coefficient $\beta(\mathcal{M}, \sigma(\Gamma))$ is the mixing coefficient introduced by Volkonskii and Rozanov [1960]. Let now $\mathbf{\Gamma} = (\Gamma_i)_{i \geq 1}$ be a strictly stationary sequence of real-valued random variables. For any $k \geq 0$, the coefficients $\beta_{\mathbf{\Gamma}, 1}(k)$ are defined by

$$(3) \quad \beta_{\mathbf{\Gamma}, 1}(k) = \beta(\sigma(\Gamma_1), \sigma(\Gamma_{1+k})),$$

Let $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$. The coefficients $\beta_{\mathbf{\Gamma}, \infty}(k)$ are defined by

$$\beta_{\mathbf{\Gamma}, \infty}(k) = \sup_{i \geq 1, l \geq 1} \sup \{\beta(\mathcal{M}_i, \sigma(\Gamma_{i_1}, \dots, \Gamma_{i_l})), i+k \leq i_1 < \dots < i_l\}.$$

We recall the coupling properties of these coefficients. Assume that Ω is rich enough, which means that there exists U uniformly distributed over $[0, 1]$ and independent of $\mathcal{M} \vee \sigma(\Gamma)$. There exists a $\mathcal{M} \vee \sigma(U) \vee \sigma(\Gamma)$ -measurable random variable Γ_1^* distributed as Γ and independent of \mathcal{M} such that

$$(4) \quad \beta(\mathcal{M}, \sigma(\Gamma)) = \mathbb{P}(\Gamma \neq \Gamma_1^*).$$

The equality in (4) is due to Berbee [1979], and is used in our proofs for adaptive estimators.

Moreover, several covariance inequalities can be given for β -mixing variables. We state the inequality that we apply to the β -mixing process $(Z_{i\delta})_{i \in \mathbb{N}}$. This result follows from the proof p.479 in Viennet [1997]. It asserts that there exist functions $g_{k\delta}(\cdot)$ such that, for any measurable bounded function ψ ,

$$(5) \quad |Cov(\psi(Z_0), \psi(Z_{k\delta}))| \leq 2 \int \psi^2(z) g_{k\delta}(z) d\mathbb{P}_{Z_0}(z)$$

with $\int g_{k\delta}(z) d\mathbb{P}_{Z_0}(z) = \beta(k\delta) = \beta_{Z, \delta, 1}(k)$, the β -mixing coefficient of $(Z_{k\delta})$.

Let us remark that the β -coefficients for covariance inequality are coefficients of type $\beta_{\mathbf{Z}, \delta, 1}$, while those used for coupling in the proofs are of type $\beta_{\mathbf{Z}, \delta, \infty}$. As $Z := (Z_{i\delta}; i \geq 1)$ is a Markov chain, by the Markov property and an elementary argument, for each $n \geq 1$, $\beta_{\mathbf{Z}, \delta, \infty}(k) = \beta_{\mathbf{Z}, \delta, 1}(k)$ (see e.g., Bradley et al. [2005]). They are indifferently denoted by $\beta(k\delta)$ in the following.

Finally, Cattiaux et al. [2014b] prove the geometric β -mixing property of the system: there exist a coefficient $\rho \in (0, 1)$ and a constant C such that for all $i \in \mathbb{N}$:

$$(6) \quad \beta(i\delta) \leq C\rho^{i\delta}.$$

3. ESTIMATION OF THE DENSITY p IN THE COMPLETE OBSERVATION CASE.

As already mentioned, our aim is to estimate the invariant density p non parametrically in two cases: from discrete observations of Z at discrete times $i\delta$, $i = 1, \dots, n$ with discretization step δ and from only discrete observations of the first component X .

In this section we define an estimator of the invariant density p in the complete case from observations $Z_{i\delta} = (X_{i\delta}, Y_{i\delta})$. We study the integrated quadratic risk, first on \mathbb{R}^2 and then on a compact set. Then, we study the rate of convergence of the estimator and propose an adaptive procedure to select the bandwidth. In the following, for $g \in \mathbb{L}^2(\mathbb{R}^2)$, $\|g\|_2 = \sqrt{\int g^2(x, y) dx dy}$ denotes the \mathbb{L}^2 -norm with respect to Lebesgue on \mathbb{R}^2 . Sometimes we omit the sub-index 2.

3.1. The estimator. The density p is estimated with a kernel estimator. Let K denote the kernel, and assume that K is a C^1 function such that its partial derivatives functions $\partial K/\partial x$ and $\partial K/\partial y$ are in $\mathbb{L}^2(\mathbb{R})$, $\int K(x, y) dx dy = 1$ and $\int K^2(x, y) dx dy < \infty$. For all $b = (b_1, b_2) > 0$, for all $(x, y) \in \mathbb{R}^2$, we denote

$$K_b(x, y) = \frac{1}{b_1 b_2} K\left(\frac{x}{b_1}, \frac{y}{b_2}\right).$$

When both coordinates are observed, we propose the following estimator of p for all $z = (x, y) \in \mathbb{R}^2$:

$$(7) \quad \tilde{p}_b(z) = \tilde{p}_b(x, y) := \frac{1}{n} \sum_{i=1}^n K_b(x - X_{i\delta}, y - Y_{i\delta}) = \frac{1}{n} \sum_{i=1}^n K_b(z - Z_{i\delta}).$$

3.2. Integrated quadratic risk. In this section we study the integrated risks of \tilde{p}_b computed on \mathbb{R}^2 or on a compact set. We introduce $p_b(z) = p_b(x, y) = K_b * p(x, y) = K_b * p(z)$, which is the function that is estimated without bias by $\tilde{p}_b(z)$, i.e. $\mathbb{E}(\tilde{p}_b(z)) = p_b(z)$. Following the proof of Viennet [1997] for the variance bound, we can obtain the Proposition hereafter.

Proposition 1. *Consider the estimator given by (7), under the drift condition (2). Then we have*

$$(8) \quad \mathbb{E} \left[\int (\tilde{p}_b(z) - p(z))^2 dz \right] \leq \int (p_b(z) - p(z))^2 dz + \frac{4\|K\|^2}{nb_1b_2} \sum_{i=0}^{n-1} \beta(i\delta).$$

where $\beta(\cdot)$ denotes the β -mixing coefficient defined in Section 2.

The proof of this result, like most other results, is given in Section 6.

On the other hand, restricting the estimation of p on a compact A , we can provide more specific computations using the coefficient ρ of the β -mixing property (6) and obtain the following bound.

Proposition 2. *Let A be a compact subset of \mathbb{R}^2 . Consider the estimator given by (7). Then there exists a constant $C > 0$ such that*

$$(9) \quad \mathbb{E} \left[\int_A (\tilde{p}_b(z) - p(z))^2 dz \right] \leq \int_A (p_b(z) - p(z))^2 dz + C \left(\frac{1}{nb_1b_2} + \frac{1}{n} \sum_{i=1}^{n-1} \left[\left(1 + \frac{1}{i^2\delta^2}\right) \wedge \frac{\rho^{i\delta}}{b_1b_2} \right] \right).$$

Let us emphasize that the term $1/(i^2\delta^2)$ is a direct consequence of the hypoellipticity of model (1) and of the bound on the transition density $p_t(z, \cdot)$ (equation (27) in Appendix 6.1). This term is thus explosive when $\delta \rightarrow 0$. We enhance in the remark below, the difference between the variance bounds stated in Propositions 1 and 2.

Remark 1. Consider first the last term in the bound given by (8), and assume that $n\delta^2 \geq 1$. We get that $\sum_{i=0}^n \beta(i\delta) \leq C/(1 - \rho^\delta)$. For any $0 < \delta < 1$, we get $\sum_{i=0}^n \beta(i\delta) \leq C/(\delta\rho|\log(\rho)|)$. As a consequence, the variance bound in (8), for $0 < \delta < 1$, is

$$(10) \quad \frac{4\|K\|^2}{nb_1b_2} \sum_{i=0}^n \beta(i\delta) \leq \frac{4\|K\|^2}{\rho|\log(\rho)|} \frac{C}{nb_1b_2\delta}.$$

Now, consider the variance bound stated in (9). Choose i_n an integer larger than $1/\delta^2$, $i_n = \lceil 1/\delta^2 \rceil + 1$. Then we have

$$\begin{aligned} \sum_{i=1}^{n-1} \left[\left(1 + \frac{1}{i^2\delta^2}\right) \wedge \frac{\rho^{i\delta}}{b_1b_2} \right] &= \sum_{i=1}^{i_n-1} \left(1 + \frac{1}{i^2\delta^2}\right) + \sum_{i=i_n}^{n-1} \frac{\rho^{i\delta}}{b_1b_2} \\ &\leq i_n + \frac{1}{\delta^2} \sum_{i \geq 1} \frac{1}{i^2} + \frac{1}{b_1b_2} \frac{\rho^{i_n\delta}}{1 - \rho^\delta} \\ &\leq \frac{3}{\delta^2} + \frac{1}{b_1b_2} \frac{\rho^{1/\delta}}{1 - \rho^\delta} \end{aligned}$$

Here, for $0 < \delta < 1$, we have $\rho^{1/\delta}/(1 - \rho^\delta) \leq 1/(\rho \log^2(\rho))$. Thus

$$(11) \quad \frac{1}{nb_1b_2} + \frac{1}{n} \sum_{i=1}^{n-1} \left[\left(1 + \frac{1}{i^2\delta^2}\right) \wedge \frac{\rho^{i\delta}}{b_1b_2} \right] \leq \left(1 + \frac{1}{\rho(\log(\rho))^2}\right) \frac{1}{nb_1b_2} + \frac{3}{n\delta^2}.$$

As a consequence, we will have to compare orders $O(1/(nb_1b_2\delta))$ and $O(1/(nb_1b_2) + 1/(n\delta^2))$, the term $1/(n\delta^2)$ being due to the hypoellipticity.

3.3. Rates of convergence. To evaluate the rate of convergence of the estimator \tilde{p}_b , we need to set regularity assumptions on p , and to make adequate choices of the bandwidths b_1 and b_2 .

This is why in this section, we consider the particular case of the Hölder's or Nikol'ski spaces for the density p , depending on the type of risk which is considered.

Definition 1. We say that p is Hölder $\mathcal{H}(a, L)$ with index $a = (a_1, a_2)$ around (x, y) that is we assume that p admits derivatives with respect to x (resp. y) up to order $\lfloor a_1 \rfloor$ (resp. $\lfloor a_2 \rfloor$) in a neighborhood of (x, y) , with

$$\begin{aligned} \left| \frac{\partial^{\lfloor a_1 \rfloor} p}{(\partial x)^{\lfloor a_1 \rfloor}}(x + \delta_x, y) - \frac{\partial^{\lfloor a_1 \rfloor} p}{(\partial x)^{\lfloor a_1 \rfloor}}(x, y) \right|^2 dx &\leq L |\delta_x|^{2(a_1 - \lfloor a_1 \rfloor)}, \\ \left| \frac{\partial^{\lfloor a_2 \rfloor} p}{(\partial y)^{\lfloor a_2 \rfloor}}(x, y + \delta_y) - \frac{\partial^{\lfloor a_2 \rfloor} p}{(\partial y)^{\lfloor a_2 \rfloor}}(x, y) \right|^2 dx &\leq L |\delta_y|^{2(a_2 - \lfloor a_2 \rfloor)}. \end{aligned}$$

Definition 2. We say that p is Nikol'ski $\mathcal{N}(\gamma, L)$ with index $\gamma = (\gamma_1, \gamma_2)$, that is we assume that p admits derivatives with respect to x (resp. y) up to order $\lfloor \gamma_1 \rfloor$ (resp. $\lfloor \gamma_2 \rfloor$) on \mathbb{R}^2 , with

$$\begin{aligned} \int \left| \frac{\partial^{\lfloor \gamma_1 \rfloor} p}{(\partial x)^{\lfloor \gamma_1 \rfloor}}(x + \delta_x, y) - \frac{\partial^{\lfloor \gamma_1 \rfloor} p}{(\partial x)^{\lfloor \gamma_1 \rfloor}}(x, y) \right|^2 dx &\leq L |\delta_x|^{2(\gamma_1 - \lfloor \gamma_1 \rfloor)}, \\ \int \left| \frac{\partial^{\lfloor \gamma_2 \rfloor} p}{(\partial y)^{\lfloor \gamma_2 \rfloor}}(x, y + \delta_y) - \frac{\partial^{\lfloor \gamma_2 \rfloor} p}{(\partial y)^{\lfloor \gamma_2 \rfloor}}(x, y) \right|^2 dx &\leq L |\delta_y|^{2(\gamma_2 - \lfloor \gamma_2 \rfloor)}. \end{aligned}$$

Remark 2. If c and σ^2 are constants and if the potential D satisfies the conditions stated in Section 2, then the above regularity assumptions are fulfilled, see Cattiaux et al. [2014a,b].

Following the strategy given in Tsybakov [2009], we need to specify the order of the kernel K and recall the following definition.

Definition 3. A kernel K is of order $\ell = (\ell_1, \ell_2)$ if for $i \in \{1, 2\}$, $1 \leq k \leq \ell_i - 1$, $\int z_i^k K(z) dz = 0$.

Examples of kernels of arbitrary order are given in Tsybakov [2009], Kerkycharian et al. [2001]. We can now compute the order of the integrated squared bias term $\int (p_b(z) - p(z))^2 dz$.

Proposition 3. We assume that $p(x, y)$ is in $\mathcal{H}(a, L)$ for any $(x, y) \in S$ where S is a subset of \mathbb{R}^2 , and K is a kernel with order $\ell = (\ell_1, \ell_2)$ such that $K(x, y) = K_1(x)K_2(y)$ with K_1 a kernel of order $\ell_1 \geq \lfloor a_1 \rfloor$ and K_2 a kernel of order $\ell_2 \geq \lfloor a_2 \rfloor$ and $\int |x|^{\ell_1} K_1(x) dx < +\infty$, $\int |y|^{\ell_2} K_2(y) dy < +\infty$. Then,

$$(12) \quad \iint_S (p_b(z) - p(z))^2 dz \leq C \left(b_1^{2a_1} + b_2^{2a_2} \right).$$

The same order holds for integrated bias on Nikol'ski space $\mathcal{N}((a_1, a_2), L)$, see Tsybakov [2009].

We can now minimize the quadratic risk of the estimator $\tilde{p}(z)$ in the complete observations case.

• Estimation on \mathbb{R}^2 .

Plugging in (8) the variance bound stated in (10) and the bias bound given in Proposition 3, we obtain the following order for the integrated quadratic risk. For $0 < \delta < 1$, we have

$$\mathbb{E} \left[\int (\tilde{p}_b(z) - p(z))^2 dz \right] \leq C \left\{ \left(b_1^{2a_1} + b_2^{2a_2} \right) + \frac{4 \|K\|^2}{\rho |\log(\rho)|} \frac{1}{nb_1 b_2 \delta} \right\}.$$

Then we need to calculate the trade-off between the two terms $b_1^{2a_1} + b_2^{2a_2}$ and $\frac{1}{n\delta b_1 b_2}$. We can choose the following (anisotropic) bandwidths $b_j = b_{j,opt}$

$$(13) \quad b_{j,opt} \propto (n\delta)^{-\bar{a}/(2a_j(\bar{a}+1))} \quad \text{where} \quad 1/\bar{a} = 1/2(1/a_1 + 1/a_2).$$

Thus we obtain the rate

$$(14) \quad \mathbb{E} \left[\int (\tilde{p}_b(z) - p(z))^2 dz \right]^2 \leq O((n\delta)^{-\bar{a}/(\bar{a}+1)})$$

• Estimation on the compact A .

In that framework, we consider that δ is small, otherwise, we apply the previous result. It follows now by plugging (11) and the bias bound of Proposition 3 in (9). The integrated quadratic risk on A has the order

$$O \left(b_1^{2a_1} + b_2^{2a_2} + \left(1 + \frac{1}{\rho(\log(\rho))^2} \right) \frac{1}{nb_1 b_2} + \frac{3}{n\delta^2} \right).$$

Choosing now the following bandwidths $b_j = b_{j,opt}^*$

$$(15) \quad b_{j,opt}^* \propto n^{-\bar{a}/(2a_j(\bar{a}+1))},$$

we get

$$\mathbb{E} \left[\int_A (\tilde{p}_b(z) - p(z))^2 dz \right]^2 \leq O \left(n^{-\bar{a}/(\bar{a}+1)} + \frac{1}{n\delta^2} \right).$$

For small δ , the first term in the rate is smaller than the order obtained in (14). We study conditions ensuring that the last term $1/(n\delta^2)$ is negligible w.r.t. the first one. Let us assume $\delta = n^{-\omega}$ with $0 < \omega < \frac{1}{2(\bar{a}+1)}$ (note that this condition implies $0 < \omega < 1/2$ and then $1/n\delta^2 \rightarrow 0$). Then the rate of convergence is improved with respect to (14):

$$(16) \quad \mathbb{E} \left[\int_A (\tilde{p}_b(z) - p(z))^2 dz \right]^2 = O(n^{-\bar{a}/(\bar{a}+1)}).$$

Note that the condition $n\delta^2 \rightarrow \infty$ with which the rate (14) is improved is surprisingly the opposite than the standard condition $n\delta^2 \rightarrow 0$ required in the parametric context to obtain the asymptotic normality of the parametric estimator [see Samson and Thiellens, 2012, for example].

Note also that the rate in (16) is the optimal rate in the minimax sense, for density estimation when n i.i.d. observations are available, see Goldenshluger and Lepski [2011].

In all cases, we emphasize that the two selected bandwidths are different, and this anisotropy property is important in our setting: regularity in each direction can be different. The bandwidth selection procedure has to be able to provide such different choices for b_1 and b_2 .

3.4. Adaptive procedure. It is clear from the previous section that the proposed bandwidth choice depends on the regularity of the function p , which is unknown. This is why we study a data driven bandwidth selection device. To select b adequately, we propose the following method, inspired from Goldenshluger and Lepski [2011]. Let us define for all $z = (x, y)$

$$\tilde{p}_{b,b'}(z) = \tilde{p}_{b,b'}(x, y) = K_{b'} * \tilde{p}_b(x, y) = \frac{1}{n} \sum_{i=1}^n K_{b'} * K_b(x - X_{i\delta}, y - Y_{i\delta})$$

where $u * v(z) = \int u(z-t)v(t)dt$. Note that $\tilde{p}_{b,b'} = \tilde{p}_{b',b}$. Let $\mathcal{B}_n = \{(b_{1,k}, b_{2,\ell}), k, \ell = 1, \dots, B_n\}$ be a discrete set of bandwidths among which we want to make a choice. Then let

$$A(b) = \sup_{b' \in \mathcal{B}_n} (\|\tilde{p}_{b,b'} - \tilde{p}_{b'}\|^2 - V(b'))_+ \quad \text{with} \quad V(b) = \kappa \frac{\|K\|_1^2 \|K\|^2}{nb_1 b_2} \sum_{i=0}^{n-1} \beta(i\delta),$$

where κ is a numerical constant to be calibrated by simulations. Heuristically, $A(b)$ is an estimate of the squared bias and $V(b)$ of the variance bound. Thus, the selection is made by setting

$$\tilde{b} = \arg \min_{b \in \mathcal{B}_n} (A(b) + V(b)).$$

Remark. We can also choose $V(b) = \kappa/(1 - \rho^\delta) \|K\|_1^2 \|K\|^2 / (nb_1 b_2)$. In any case, we face the problem which is standard for non independent data, of having an unknown quantity ρ in the penalty. There are suggestions of block empirical variance estimation for such terms in Lerasle [2011], Lerasle [2012].

Under conditions on \mathcal{B}_n , we can prove the following result for the integrated risk on \mathbb{R}^2 .

Theorem 1. *Assume that \mathcal{B}_n is such that*

$$(17) \quad \sum_{b \in \mathcal{B}_n} e^{-\lambda/\sqrt{b}}/\sqrt{b} \leq C(\lambda) < +\infty, \quad \forall \lambda > 0,$$

with $1/b = 1/(b_1 b_2) \leq n$, and

$$(18) \quad \frac{\log(n)}{\delta} e^{-\lambda\sqrt{n}\delta/\log(n)} \text{card}(\mathcal{B}_n) \leq C(\lambda, \delta) < +\infty, \quad \forall \lambda > 0.$$

Then we have

$$\mathbb{E}(\|\tilde{p}_{\tilde{b}} - p\|^2) \leq C \inf_{b \in \mathcal{B}_n} (\|p - p_b\|^2 + V(b)) + C' \frac{\log(n)}{n\delta}$$

for C and C' constants.

Clearly, the bound stated in Theorem 1 shows that the estimator leads to an automatic trade-off between the bias $\|p - p_b\|^2$ and the variance $V(b)$, up to a multiplicative constant C . The last term $C' \log(n)/(n\delta)$ is indeed negligible. Moreover this procedure allows to choose different bandwidths for the two directions x and y .

Let us give an example of set \mathcal{B}_n of bandwidths fitting our conditions. Choose $b_{i,k} = 1/k$ for $i = 1, 2$, and $B_n = \lfloor \sqrt{n} \rfloor$. Then $1/b = 1/(b_{1,k} b_{2,\ell}) \leq n$ and $\text{card}(\mathcal{B}_n) \leq n$, so that for any fixed δ and $\lambda > 0$, $\log(n)\delta^{-1} e^{-\lambda\sqrt{n}\delta/\log(n)} \text{card}(\mathcal{B}_n) \leq C(\lambda, \delta) < +\infty$ and condition (18) is fulfilled. Moreover as for $k, \ell \geq 1$, $e^{-\lambda\sqrt{k\ell}} \leq \exp(-\lambda\sqrt{k}/2) \exp(-\lambda\sqrt{\ell}/2)$, we get

$$\sum_{k=1}^{B_n} \sum_{\ell=1}^{B_n} \sqrt{k\ell} e^{-\lambda\sqrt{k\ell}} \leq \left(\sum_{k=1}^{B_n} \sqrt{k} e^{-\lambda\sqrt{k}/2} \right)^2 < +\infty$$

Thus condition (17) is satisfied. The choice $b_{i,k} = 2^{-k}$, for $i = 1, 2$ and $B_n = \log(n)/(2 \log 2)$ would suit also. It is important to note that these sets contain the optimal bandwidth.

4. ESTIMATION OF THE DENSITY p IN THE PARTIAL OBSERVATION CASE.

For the partially observed case, where only the position process X is available at discrete time $(i\delta)_{i=1,\dots,n}$, we approximate the velocity $Y_{i\delta}$ by the increment

$$\bar{Y}_{i\delta} = \frac{X_{(i+1)\delta} - X_{i\delta}}{\delta}$$

and define the 2-dimensional kernel estimator by:

$$(19) \quad \hat{p}_b(x, y) := \frac{1}{n} \sum_{i=1}^n K_b \left(x - X_{i\delta}, y - \frac{X_{(i+1)\delta} - X_{i\delta}}{\delta} \right).$$

We present the integrated risk of \hat{p}_b and an adaptive procedure to select automatically the bandwidth b .

4.1. Integrated quadratic risk. Let us choose $K(z) = K(x, y) = K_1(x)K_2(y)$ and study the integrated risk of the new estimator in this more difficult setting.

Proposition 4. *Consider, in the partial observation case, the estimator given by (19), under the drift condition (2). Then there exists a constant $C > 0$ such that*

$$(20) \quad \mathbb{E} \left[\int (\hat{p}_b(z) - p(z))^2 dz \right] \leq \int (p_b(z) - p(z))^2 dz + \frac{4\|K\|^2}{nb_1b_2} \sum_{i=0}^{n-1} \beta(i\delta) + \frac{C\delta}{b_1b_2^3}.$$

First, note that integrating on a compact set A would allow to replace

$$\frac{4\|K\|^2}{nb_1b_2} \sum_{i=0}^{n-1} \beta(i\delta) \quad \text{by} \quad C \left(\frac{1}{nb_1b_2} + \frac{1}{n} \sum_{i=1}^{n-1} \left[\left(1 + \frac{1}{i^2\delta^2}\right) \wedge \frac{\rho^{i\delta}}{b_1b_2} \right] \right)$$

as previously.

Second, let us discuss the rate implied by the bound (20) under the condition that $p \in \mathcal{N}((a_1, a_2), L)$, and $\beta(i\delta) \leq \rho^{i\delta}$. In this case, the order of the bound of the integrated quadratic risk becomes, for $0 < \delta < 1$,

$$\mathcal{O} \left(b_1^{2a_1} + b_2^{2a_2} + \frac{1}{nb_1b_2\delta} + \frac{\delta}{b_1b_2^3} \right).$$

Assume that b_1 and b_2 are chosen as in the complete data case i.e. without the last error term. In that case, $b_{j,opt} \asymp (n\delta)^{-\bar{a}/[2a_j(\bar{a}+1)]}$. Assume that $\delta = n^{-\omega}$, $0 < \omega < 1$. Then the rate is

$$\mathcal{O} \left(n^{-(1-\omega)\bar{a}/(\bar{a}+1)} + n^{-\omega} n^{(1-\omega) \left[\frac{\bar{a}}{2a_1(\bar{a}+1)} + \frac{3\bar{a}}{2a_2(\bar{a}+1)} \right]} \right).$$

The estimator is consistent if the power of the second term is negative i.e. if $\omega > \omega_0(a_1, a_2)$ with

$$\omega_0(a_1, a_2) = \frac{1 + \frac{\bar{a}}{a_2}}{2 + \bar{a} + \frac{\bar{a}}{a_2}}.$$

For instance in the isotropic case, $a_1 = a_2 = a$, $\omega_0(a, a) = 2/(3+a) \in (0, 2/3)$ for $a \in (0, +\infty)$. Therefore, the estimator is convergent if $\omega > 2/3$. If $a > 1$, $\omega > 1/2$ is sufficient.

Now we can wonder when the rate of the second term is less than the first term rate, i.e. conditions on ω ensuring that

$$-\omega + (1-\omega) \frac{\bar{a}}{\bar{a}+1} \left[\frac{1}{2a_1} + \frac{3}{2a_2} \right] \leq -(1-\omega) \frac{\bar{a}}{\bar{a}+1}.$$

This yields $\omega > \omega_1(a_1, a_2)$ with

$$\omega_1(a_1, a_2) = \frac{1 + \frac{\bar{a}}{a_2(1+\bar{a})}}{2 + \frac{\bar{a}}{a_2(1+\bar{a})}}.$$

For instance in the isotropic case, $a_1 = a_2 = a$, $\omega_1(a, a) = (2 + a)/(3 + 2a) \in (1/2, 2/3)$ for $a \in (0, +\infty)$. Therefore, the additional term does not degrade the optimal rate if $\omega \geq 2/3$.

4.2. Adaptive procedure for partial observations. We extend the Goldenshluger and Lepski [2011] procedure to the present case as follows. Let us define, in the same spirit as in the complete observation case, for all $z = (x, y)$

$$\hat{p}_{b,b'}(z) = \hat{p}_{b,b'}(x, y) = K_{b'} * \hat{p}_b(x, y) = \frac{1}{n} \sum_{i=1}^n K_{b'} * K_b(x - X_{i\delta}, y - \bar{Y}_{i\delta}).$$

Now let

$$A^{(P)}(b) = \sup_{b' \in \mathcal{B}_n} \left(\|\hat{p}_{b,b'} - \hat{p}_{b'}\|^2 - V^{(P)}(b') \right)_+$$

with

$$V^{(P)}(b) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{nb_1 b_2} \sum_{i=0}^{n-1} \beta(i\delta) + \kappa_2 \frac{\delta}{b_1 b_2^3} := V_1^{(P)}(b) + V_2^{(P)}(b),$$

where κ_1 is a numerical constant to be calibrated by simulations and κ_2 depends on K and σ^2 . The selection is made by setting

$$(21) \quad \hat{b} = \arg \min_{b \in \mathcal{B}_n} \left(A^{(P)}(b) + V^{(P)}(b) \right).$$

Then we can prove

Theorem 2. *Assume that \mathcal{B}_n satisfies conditions (17) and (18). Then we have*

$$\mathbb{E}(\|\hat{p}_{\hat{b}} - p\|^2) \leq C \inf_{b \in \mathcal{B}_n} \left(\|p - p_b\|^2 + V^{(P)}(b) \right) + C' \frac{\log(n)}{n\delta}$$

for two constants C and C' .

The result states that the compromise between all terms involved in the risk bound is automatically performed with this strategy. As for the complete observation case, this procedure allows to select anisotropic bandwidths.

5. SIMULATIONS

The two procedures of estimation (complete and partial observation cases) are implemented in R and evaluated on two examples: the Harmonic and the Van Der Pol Oscillators. We first explain the implementation of the estimators and then give the numerical results obtained for the two models.

5.1. Implementation of the estimators. The implementation of the Goldenshluger and Lepski [2011] is not efficient numerically in the multi-dimensional setting. Indeed, the convolution estimator $\hat{p}_{b,b'}$ has to be computed for a double collection of bandwidths $((b_1, b_2), (b'_1, b'_2))$ which is numerically demanding. Moreover, the penalty $V(b)$ appears twice in the procedure, the two constants κ_1, κ_2 are then difficult to calibrate. Especially, the slope heuristic can not be used in that case.

In practice, we therefore propose to implement a method, inspired from Goldenshluger and Lespki [2011], as rewritten most recently by Lacour et al. [2016]. The method is proposed in their paper for an independent sample and we naturally extend the implementation to the dependent context.

More precisely, let us define a discrete set of admissible bandwidths

$$\mathcal{B}_n = \{(b_{1,k}, b_{2,\ell}), k, \ell = 1, \dots, B_n, b_{1,k} \geq b_{1,\min}, b_{2,\ell} \geq b_{2,\min}\}$$

among which we want to make a choice. We propose to select the bandwidth with the following criterion:

$$(22) \quad \tilde{b} = \arg \min_{b \in \mathcal{B}_n} (\|\tilde{p}_b - \tilde{p}_{b_{\min}}\|^2 + V(b)).$$

Heuristically, $\|\tilde{p}_b - \tilde{p}_{b_{\min}}\|^2$ is an estimate of the squared bias and replaces the term $A(b) = \sup_{b' \in \mathcal{B}_n} (\|\tilde{p}_{b,b'} - \tilde{p}_{b'}\|^2 - V(b'))_+$. The main advantage is that we avoid the computation of the convolved $\tilde{p}_{b,b'}$ and the constant κ appears only once, and in a standard manner. This allows to calibrate it with the slope heuristic, as explained below. The final estimator is $\tilde{p}_{\tilde{b}}$.

A Gaussian kernel is chosen for function K . The collection of estimators $(\tilde{p}_b)_{b \in \mathcal{B}_n}$ is computed for the following collection of bandwidths: $\mathcal{B}_n = \{(b_1, b_2) : b_1 \in \{1/\sqrt{4n}, 2/\sqrt{4n}, \dots, 30/\sqrt{4n}\}, b_2 \in \{1/\sqrt{4n}, 2/\sqrt{4n}, \dots, 30/\sqrt{4n}\}\}$. Note that $b_{\min} = (1/\sqrt{4n}, 1/\sqrt{4n})$. Then the automatic selection of the best bandwidth with the criterion (22) is implemented: the norm $\|\tilde{p}_b - \tilde{p}_{b_{\min}}\|^2$ is computed by discretization of the integral. Then we approximate $V(b)$. The β -mixing coefficients $\beta(\delta)$ are generally unknown but bounded by $C\rho^\delta$ (equation (6)). If $0 < \rho < 1$, we have $\sum_{i=1}^n \beta(i\delta) \leq C/(\delta\rho|\log \rho|)$. As ρ is unknown, we let the constant κ absorb the term in ρ and approximate the penalty $V(b)$ by

$$\tilde{V}(b) = \kappa \frac{1}{n\delta b_1 b_2}.$$

The constant κ can then be calibrated numerically by the slope heuristic [Arlot and Massart, 2009, Lacour et al., 2016] and we find $\kappa = 0.1$. By plugging \tilde{b} into (7) we obtain $\tilde{p}_{\tilde{b}}$ which is our estimator for the complete observation case.

For the partial observation case, we do not implement neither the Goldenshluger and Lespki [2011] procedure (21) but also extend the Lacour et al. [2016]'s procedure to this case. The bandwidth \hat{b} is selected by the following criterion

$$(23) \quad \hat{b} = \arg \min_{b \in \mathcal{B}_n} (\|\hat{p}_b - \hat{p}_{b_{\min}}\|^2 + \hat{V}^{(P)}(b)), \quad \text{with} \quad \hat{V}^{(P)} = \kappa_1 \frac{1}{n\delta b_1 b_2} + \kappa_2 \frac{\delta}{b_1 b_2^3},$$

with $\kappa_1 = 0.1$ (same than for the complete observation case) and $\kappa_2 = 0.001$ (obtained numerically after a set of simulation experiments to calibrate it). The second term is almost always neglected. By plugging \hat{b} into (19) we obtain $\hat{p}_{\hat{b}}$ which is our estimator for the partial observation case.

5.2. Numerical results for the Harmonic Oscillator. We first present the Harmonic Oscillator. The process (X_t, Y_t) is solution of the process

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= -(\alpha X_t + \gamma Y_t) dt + \sigma dW_t \end{aligned}$$

Complete observations	0.11	0.05	0.02
$n\delta = 400$	$n = 500, \delta = 0.8$	$n = 1000, \delta = 0.4$	$n = 2000, \delta = 0.2$
Complete observations	0.090	0.013	0.011
Partial observations	0.302	0.038	0.014
$n\delta = 200$	$n = 500, \delta = 0.4$	$n = 1000, \delta = 0.2$	$n = 2000, \delta = 0.1$
Complete observations	0.023	0.018	0.016
Partial observations	0.048	0.021	0.018
$n\delta = 200$	$n = 250, \delta = 0.8$	$n = 500, \delta = 0.4$	$n = 1000, \delta = 0.2$
Complete observations	0.187	0.023	0.018
Partial observations	0.400	0.049	0.021

TABLE 1. Simulation study, Harmonic Oscillator. MISE of the estimation of the stationary density $p(x, y)$ using the estimators $\tilde{p}_{\tilde{b}}$ (complete observations) and $\hat{p}_{\hat{b}}$ (partial observations) from 100 simulated trajectories, with different values of n and δ .

with $\alpha > 0, \gamma > 0$. In the following, we choose $\alpha = 4, \gamma = 0.5, \sigma = 0.5$. The potential is then $D(x) = \alpha/2x^2$. The stationary distribution is Gaussian, with mean zero and explicit diagonal variance matrix:

$$p(x, y) = \frac{\gamma\sqrt{\alpha}}{\pi\sigma^2} \exp\left(-\frac{2\gamma}{2\sigma^2}y^2 - \frac{2\gamma\alpha}{2\sigma^2}x^2\right)$$

with diagonal variances equal to $1/16$ and $1/4$, respectively in our case.

The two estimators $\tilde{p}_{\tilde{b}}$ and $\hat{p}_{\hat{b}}$ are applied to 100 trajectories of the process (X_t, Y_t) . Several designs are compared: $n\delta = 400$ with $(n = 500, \delta = 0.8)$, $(n = 1000, \delta = 0.4)$ and $(n = 2000, \delta = 0.2)$; $n\delta = 200$ with $(n = 500, \delta = 0.4)$, $(n = 1000, \delta = 0.2)$, $(n = 2000, \delta = 0.1)$ and finally $n\delta = 200$ with $(n = 250, \delta = 0.8)$, $(n = 500, \delta = 0.4)$, $(n = 1000, \delta = 0.2)$. For each n, δ , each of the 100 trajectories has been simulated with a Euler scheme with a step size equal to $\delta/10$.

An example of the collection of estimators $(\tilde{p}_b)_{b \in \mathcal{B}_n}$ and the associated final estimator $\tilde{p}_{\tilde{b}}$, obtained for a trajectory with $n = 2000, \delta = 0.2$, is presented in the top line of Figure 1. The selected bandwidth is anisotropic and equal to $\tilde{b} = (8/\sqrt{4n}, 17/\sqrt{4n})$. On the same trajectory, assuming that only X is observed (partial observations), we obtain the collection of estimators $(\hat{p}_b)_{b \in \mathcal{B}_n}$ and the associated final estimator $\hat{p}_{\hat{b}}$ presented in the bottom line of Figure 1. The selected bandwidth is $\hat{b} = (9/\sqrt{4n}, 19/\sqrt{4n})$. The densities are well estimated in both cases. The partial observations impacts slightly the estimation of the marginal in y .

The influence of n, δ is then studied by comparing the Mean Integrated Squared Error (MISE) of the two estimators. Results are presented in Table 1. As expected, the MISE are always smaller in the complete observation case compared to the partial one. When n increases, the MISE decreases. The influence of δ is more complex to analyse and depends on the evolution of $n\delta$: when δ decreases and n increases (this corresponds to a fixed $n\delta$), the MISE decreases. When δ is fixed and n increases (this correspond to a increased $n\delta$), the MISE decreases. When $n\delta$ is fixed and δ decreases, the MISE decreases. But when n is fixed and δ decreases, the MISE does not always decrease.

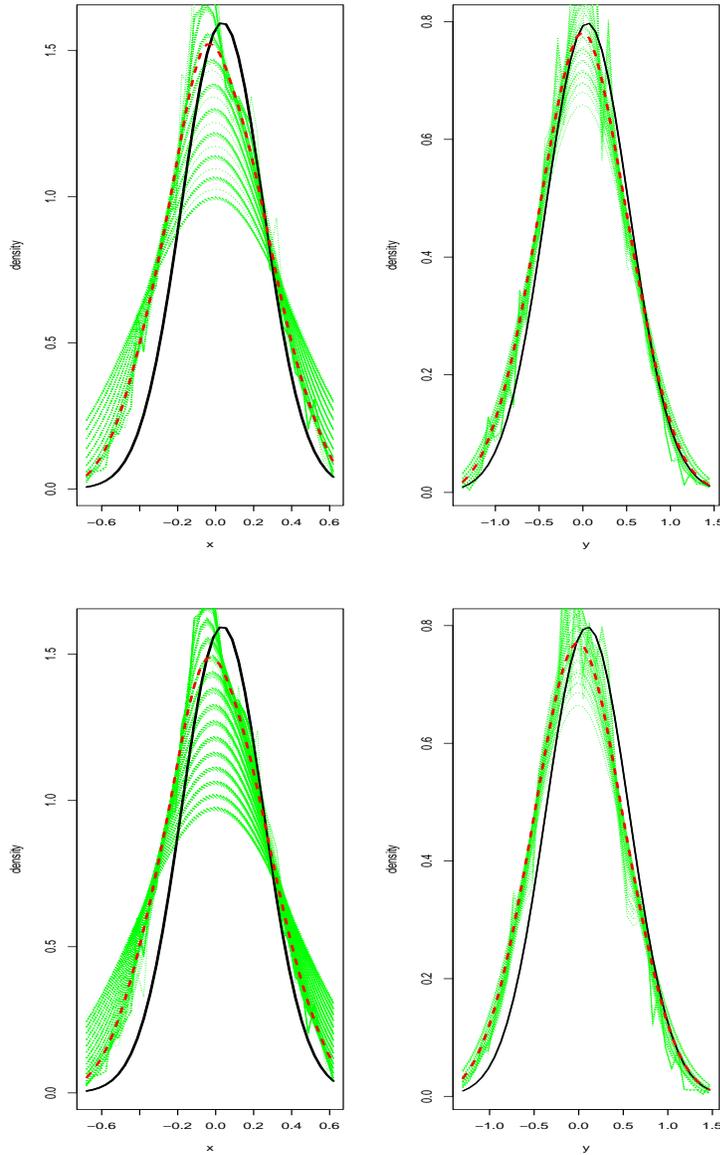


FIGURE 1. Harmonic Oscillator. Top line, complete observation case: marginals in x (left) and in y (right) of the collection of estimators $(\tilde{p}_b)_{b \in \mathcal{B}_n}$ obtained for a sample $(X_{i\delta}, Y_{i\delta})_{i=1, \dots, n}$ with $n = 2000, \delta = 0.2$ (green dotted lines), final estimator $\tilde{p}_{\hat{b}}$ (red dashed line) and true density p (black plain line). Bottom line, partial observation case: marginals in x (left) and in y (right) of the collection of estimators $(\hat{p}_b)_{b \in \mathcal{B}_n}$ obtained for the same sample $(X_{i\delta}, Y_{i\delta})_{i=1, \dots, n}$ (green dotted lines), final estimator $\hat{p}_{\hat{b}}$ (red dashed line) and true density p (black plain line).

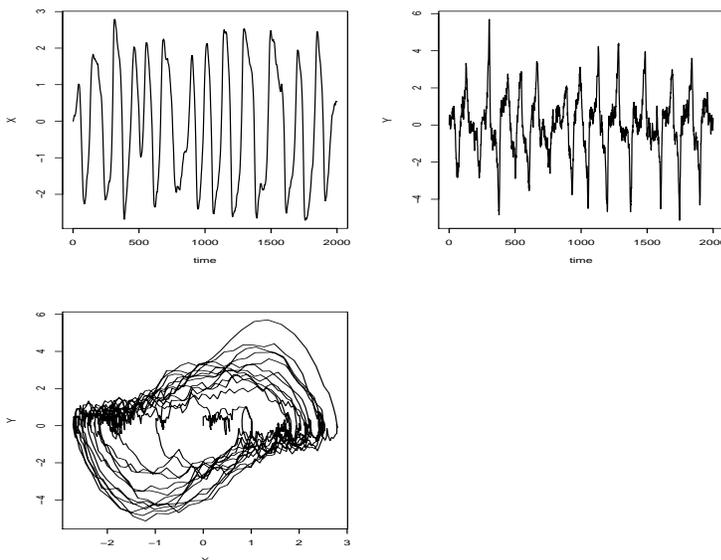


FIGURE 2. Van der Pol oscillator. Sample $(X_{i\delta})_{i=0,\dots,n}$ (top left), $(Y_{i\delta})_{i=0,\dots,n}$ (top right) and state phase (bottom) for $\delta = 0.05$ and $n = 2000$.

5.3. Numerical results for the Van Der Pol Oscillator. We consider the Van der Pol oscillator defined by

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= -((c_1 X_t^2 - c_2) Y_t + \omega_0^2 X_t) dt + \sigma dW_t \end{aligned}$$

with $\sigma, c_1, c_2, \omega_0^2 > 0$. In the following, we choose $\sigma = c_1 = c_2 = \omega_0^2 = 1$. The potential is then $D(x) = \omega_0^2/2x^2$. The invariant density has no explicit expression ($c(x)$ is not constant). To compare the estimators $\tilde{p}_{\tilde{b}}$ and $\hat{p}_{\hat{b}}$ to the invariant density p , we approximate p by solving its corresponding Fokker-Planck equation:

$$(24) \quad \frac{1}{2} \frac{\partial^2 p(x, y)}{\partial y^2} - y \frac{\partial p(x, y)}{\partial x} + c(x) p(x, y) + (c(x) y + \nabla D(x)) \frac{\partial p(x, y)}{\partial y} = 0$$

The partial differential equation (24) is approximated using a finite difference scheme [Kumar and Narayanan, 2006].

A sample $(X_{i\delta}, Y_{i\delta})_{i=0,\dots,n}$ is obtained using the Euler scheme with $\delta = 0.05$ and $n = 2000$. The trajectory and its state phase is presented in Figure 2.

The collections of estimators $(\tilde{p}_{\tilde{b}})_{\tilde{b} \in \mathcal{B}_n}$ (complete observations) and $(\hat{p}_{\hat{b}})_{\hat{b} \in \mathcal{B}_n}$ (partial observations) are presented in Figure 3. In the complete observation case, the two marginals are well estimated. The influence of the partial observation is rather limited. Same results are obtained for different values of n and δ (not shown).

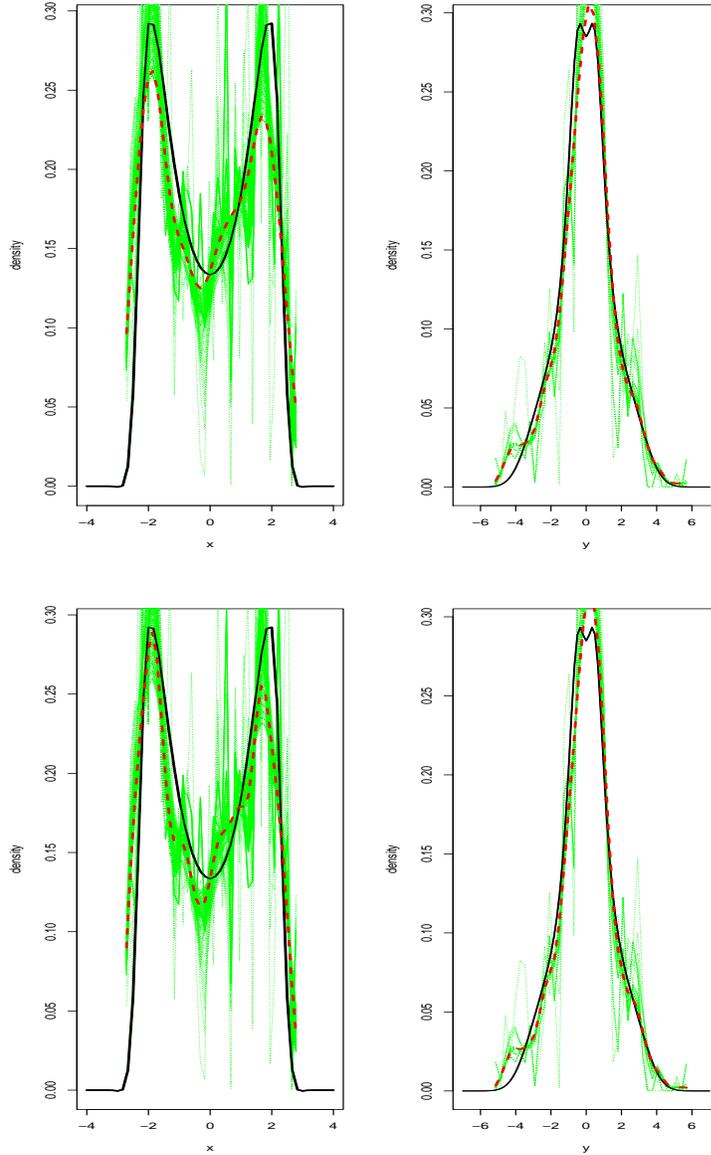


FIGURE 3. Van der Pol Oscillator. Top line, complete observation case: marginals in x (left) and in y (right) of the collection of estimators $(\tilde{p}_b)_{b \in \mathcal{B}_n}$ obtained for a sample $(X_{i\delta}, Y_{i\delta})_{i=1, \dots, n}$ with $n = 2000, \delta = 0.05$ (green dotted lines), final estimator $\hat{p}_{\hat{b}}$ (red dashed line) and true density p (black plain line). Bottom line, partial observation case: marginals in x (left) and in y (right) of the collection of estimators $(\hat{p}_b)_{b \in \mathcal{B}_n}$ obtained for the same sample $(X_{i\delta}, Y_{i\delta})_{i=1, \dots, n}$ (green dotted lines), final estimator $\hat{p}_{\hat{b}}$ (red dashed line) and true density p (black plain line).

ACKNOWLEDGEMENTS

Adeline Samson was partially supported by the LabExPERSYVAL-Lab (ANR-11-LABX-0025-01). Clémentine PRIEUR was partially supported by the MATH AMSUD project 16-MATH-03 SIDRE. The authors thank Angie Pineda from Universidad Central Venezuela in Caracas for sharing the numerical codes of the PDE resolution.

REFERENCES

- S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10:245–279, 2009.
- N. Asin and J. Johannes. Adaptive non-parametric estimation in the presence of dependence. *Preprint arXiv:1602.00531*, 2016.
- H. Berbee. Random walks with stationary increments and renewal theory. *Mathematical Centre Tracts 112. Amsterdam: Mathematisch Centrum*, 1979.
- K. Bertin and N. Klutchnikoff. Pointwise adaptive estimation of the marginal density of a weakly dependent process. *Preprint arXiv:1604.00039*, 2016.
- Richard C Bradley et al. Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2(2):107–144, 2005.
- P. Cattiaux, J.R. León, and C. Prieur. Estimation for stochastic damping hamiltonian systems under partial observation. ii. drift term. *ALEA*, 11:359–384, 2014a.
- P. Cattiaux, J.R. León, and C. Prieur. Estimation for stochastic damping hamiltonian systems under partial observation. i. invariant density. *Stochastic Process. Appl.*, 124:1236–1260, 2014b.
- F. Comte and F. Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238, 2002. ISSN 1292-8100. New directions in time series analysis (Luminy, 2001).
- F. Comte and F. Merlevède. Super optimal rates for nonparametric density estimation via projection estimators. *Stochastic Process. Appl.*, 115(5):797–826, 2005. ISSN 0304-4149.
- F. Comte, V. Genon-Catalot, and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007. ISSN 1350-7265. doi: 10.3150/07-BEJ5173. URL <http://dx.doi.org/10.3150/07-BEJ5173>.
- F. Comte, V. Genon-Catalot, and Y. Rozenholc. Nonparametric adaptive estimation for integrated diffusions. *Stochastic Process. Appl.*, 119:811–834, 2009.
- I. Gannaz and O. Wintenberger. Adaptive density estimation under weak dependence. *ESAIM Probab. Stat.*, 14:151–172, 2010. ISSN 1292-8100.
- V. Genon-Catalot, T. Jeantheau, and C. Laredo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6:1051–1079, 2000.
- A. Gloter. Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Statist.*, 33(1):83–104, 2006.
- A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39:1608–1632, 2011.
- G. Kerkycharian, O. Lepski, and D. Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121:137–170, 2001.
- T. Klein and E. Rio. Empirical processes and concentration around the mean. *The Annals of Probability*, 33(3):1060–1077, 2005.

- V. Konakov, S. Menozzi, and S. Molchanov. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Ann. Institut Henri Poincaré*, 46:908–923, 2010.
- P. Kumar and S. Narayanan. Solution of fokker-planck equation by finite element and finite difference methods for nonlinear systems. *Sadhana*, 31:445–461, 2006.
- C. Lacour, P. Massart, and V. Rivoirard. Estimator selection: a new method with applications to kernel density estimation. Arxiv Preprint arXiv:1607.05091v1, 2016. URL <https://arxiv.org/pdf/1607.05091v1.pdf>.
- M. Lerasle. Optimal model selection for density estimation of stationary data under various mixing conditions. *Ann. Statist.*, 39(4):1852–1877, 2011. ISSN 0090-5364. doi: 10.1214/11-AOS888. URL <http://dx.doi.org/10.1214/11-AOS888>.
- M. Lerasle. Optimal model selection in density estimation. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(3):884–908, 2012. ISSN 0246-0203. doi: 10.1214/11-AIHP425. URL <http://dx.doi.org/10.1214/11-AIHP425>.
- Y. Pokern, A.M. Stuart, and P. Wiberg. Parameter estimation for partially observed hypoelliptic diffusions. *J. Roy. Stat. Soc. B*, 71(1):49–73, 2009.
- J. B. Roberts and P. D. Spanos. *Random vibration and statistical linearization*. Dover Publications, Inc., Mineola, NY, 2003. ISBN 0-486-43240-8. Revised reprint of the 1990 original [Wiley, Chichester; MR1076193].
- A. Samson and M. Thieullen. Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 2012.
- E. Schmitter. Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics*, 47(1):61–84, 2013a. ISSN 0233-1888. doi: 10.1080/02331888.2011.591931. URL <http://dx.doi.org/10.1080/02331888.2011.591931>.
- E. Schmitter. Nonparametric estimation of the derivatives of the stationary density for stationary processes. *ESAIM Probab. Stat.*, 17:33–69, 2013b. ISSN 1292-8100. doi: 10.1051/ps/2011102. URL <http://dx.doi.org/10.1051/ps/2011102>.
- A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL <http://dx.doi.org/10.1007/b13794>. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields*, 107:467–492, 1997.
- V. Volkonskii and Y. Rozanov. Some limit theorems for random functions. i. *Theor. Probab. Appl.*, 4:178–197, 1960.
- L. Wu. Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stochastic Processes and their Applications*, 91:205–238, 2001.

6. PROOFS.

In the following, for $p = 1, 2$, $\|g\|_p = (\int g^p(x, y) dx dy)^{1/p}$ denotes the \mathbb{L}^p norm with respect to Lebesgue on \mathbb{R}^2 . For $p = 2$, we often omit the sub-index 2.

6.1. Useful results on the transition density. We recall the following result used in the sequel (see Theorem 2.1 in Konakov et al. [2010], Corollary 2.12 in Cattiaux et al. [2014b], Proposition 2.14 in Cattiaux et al. [2014b]).

Proposition 5. *For the system (1) under the drift condition (2), for all z and all bounded, open neighborhood U of z , the density $p_t(z, \cdot)$ can be bounded: $\forall t \in (0, T), \forall z' \in U$*

$$(25) \quad p_t(z, z') \leq C' \frac{1}{t^2} \exp \left(-C \left[\frac{|y - y'|^2}{4t} + \frac{3 \left| x' - x - \frac{t(y+y')}{2} \right|^2}{t^3} \right] \right) + c(U) e^{-\frac{c'(U)}{t}}.$$

Moreover, for all $t \geq 0$ and all pair (z, z') ,

$$(26) \quad p_{t_0+t}(z, z') \leq \sup_w p_{t_0}(w, z') < +\infty.$$

Thanks to the results in Proposition 5, we can show that for any compacts A', A'' , if $T > 1$, then for any $t \in (0, 1]$

$$\sup_{z \in A', z' \in A''} p_t(z, z') \leq C(A', A'') (t^{-2} + 1).$$

One can indeed choose in Proposition 5 a common open neighborhood U of all $z \in A'$ compact, satisfying moreover $A'' \subset U$. Furthermore, using (26), we get

$$p_{1+t}(z, z') \leq \sup_w p_1(w, z'),$$

and thus for any $t \in (0, \infty)$,

$$(27) \quad \sup_{z \in A', z' \in A''} p_t(z, z') \leq C(A', A'') (t^{-2} + 1).$$

Note that this means that the transition density is explosive when t goes to 0.

6.2. Proof of Proposition 1. We have from the standard bias variance decomposition

$$\mathbb{E} \left(\int (\tilde{p}_b(z) - p(z))^2 dz \right) = \int [p_b(z) - p(z)]^2 dz + \mathbb{E} \int [\tilde{p}_b(z) - p_b(z)]^2 dz.$$

Let us denote $T_2 = \mathbb{E} [\tilde{p}_b(z) - p_b(z)]^2$. Recall that there exists a function $g_{k\delta}$ such that $\int g_{k\delta}(z) d\mathbb{P}_{Z_0}(z) = \beta(k\delta)$ (see (5)). Thus, we have

$$\begin{aligned} \int T_2(z) dz &= \int \mathbb{E} [(\tilde{p}_b(z) - p_b(z))^2] = \int \text{Var} \left(\frac{1}{n} \sum_{i=1}^n K_b(x - X_{i\delta}, y - Y_{i\delta}) \right) dx dy \\ &\leq \frac{4}{n} \int \left[\int \left(\frac{1}{b_1 b_2} K \left(\frac{x-u}{b_1}, \frac{y-v}{b_2} \right) \right)^2 \left(\sum_{k=0}^{n-1} g_{k\delta}(u, v) \right) d\mathbb{P}_{(X,Y)}(u, v) \right] dx dy \\ &= \frac{4}{n} \int \left(\int \frac{1}{(b_1 b_2)^2} K^2 \left(\frac{x-u}{b_1}, \frac{y-v}{b_2} \right) dx dy \right) \left(\sum_{k=0}^{n-1} g_{k\delta}(u, v) \right) d\mathbb{P}_{(X,Y)}(u, v) \\ &= \frac{4}{n b_1 b_2} \left(\int K^2(s, t) ds dt \right) \int \left(\sum_{k=0}^{n-1} g_{k\delta}(u, v) \right) d\mathbb{P}_{(X,Y)}(u, v) = \frac{4 \|K\|^2}{n b_1 b_2} \sum_{k=0}^{n-1} \beta(k\delta). \end{aligned}$$

Gathering the terms gives the announced result. \square

6.3. Proof of Proposition 2. We have from the standard bias variance decomposition

$$\mathbb{E} \left(\int_A (\tilde{p}_b(z) - p(z))^2 dz \right) = \int_A [p_b(z) - p(z)]^2 dz + \mathbb{E} \int_A [\tilde{p}_b(z) - p_b(z)]^2 dz.$$

Let us denote $U_{i,n}(z) = K_b(z - Z_{i\delta})$. Using the definition of $T_2(z)$ from the proof of Proposition 1, we have:

$$\begin{aligned} \int_A T_2(z) dz &= \int_A \mathbb{E} \left(\frac{1}{n^2} \left(\sum_{i=1}^n (K_b(z - Z_{i\delta}) - \mathbb{E}K_b(z - Z_{i\delta})) \right)^2 \right) dz \\ &= \int_A \left(\frac{1}{n^2} \sum_{i=1}^n \text{Var}(K_b(z - Z_{i\delta})) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{Cov}(U_{i,n}(z), U_{j,n}(z)) \right) dz \\ &= \int_A \left(\frac{1}{n} \text{Var}(U_{0,n}(z)) + \frac{2}{n^2} \sum_{i=1}^{n-1} (n-i) \text{Cov}(U_{0,n}(z), U_{i,n}(z)) \right) dz \end{aligned}$$

First we compute $\text{Var}(U_{0,n}(z))$.

$$\begin{aligned} \text{Var}(U_{0,n}(z)) &= \text{Var} \left(\frac{1}{b_1 b_2} K \left(\frac{x - X_0}{b_1}, \frac{y - Y_0}{b_2} \right) \right) \\ &\leq \frac{1}{b_1^2 b_2^2} \int \int K^2 \left(\frac{x-s}{b_1}, \frac{y-t}{b_2} \right) p(s, t) ds dt \\ &\leq \frac{1}{b_1 b_2} \int \int K^2(u, v) p(x - ub_1, y - vb_2) dudv. \end{aligned}$$

Thus

$$\int_A \text{Var}(U_{0,n}(z)) dz \leq \frac{\|K\|^2}{b_1 b_2}.$$

Then we compute the covariance terms.

$$\begin{aligned} |\text{Cov}(U_{0,n}(z), U_{i,n}(z))| &= \frac{1}{b_1^2 b_2^2} \left| \text{Cov} \left(K \left(\frac{x - X_0}{b_1}, \frac{y - Y_0}{b_2} \right), K \left(\frac{x - X_{i\delta}}{b_1}, \frac{y - Y_{i\delta}}{b_2} \right) \right) \right| \\ &\leq \frac{1}{b_1^2 b_2^2} \left(\mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) K \left(\frac{z - Z_{i\delta}}{b} \right) \right| + \left(\mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) \right| \right)^2 \right) \end{aligned}$$

We have to compute $\mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) K \left(\frac{z - Z_{i\delta}}{b} \right) \right|$:

$$\begin{aligned} &\mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) K \left(\frac{z - Z_{i\delta}}{b} \right) \right| \\ &= \int \int \left| K \left(\frac{x-s}{b_1}, \frac{y-t}{b_2} \right) K \left(\frac{x-s'}{b_1}, \frac{y-t'}{b_2} \right) \right| p_{i\delta}((s, t), (s', t')) p(s, t) ds dt ds' dt' \\ &= b_1^2 b_2^2 \int \int |K(u, v) K(u', v')| \times \\ &\quad p_{i\delta}((x - ub_1, y - vb_2), (x - u'b_1, y - v'b_2)) p(x - ub_1, y - vb_2) dudvdu'dv' \end{aligned}$$

We apply bound (27) to $t = i\delta$ and $z = (x - ub_1, y - vb_2)$, $z' = (x - u'b_1, y - v'b_2)$ which is explosive when $\delta \rightarrow 0$. The support A_K of the kernel K is assumed to be compact and A is

compact. Then

$$\begin{aligned} & \mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) K \left(\frac{z - Z_{i\delta}}{b} \right) \right| \\ & \leq C_{A, A_K} b_1^2 b_2^2 \left(\frac{1}{(i\delta)^2} + 1 \right) \int \int |K(u, v) K(u', v')| p(x - ub_1, y - vb_2) dudvdu' dv'. \end{aligned}$$

Thus

$$\int_A \mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) K \left(\frac{z - Z_{i\delta}}{b} \right) \right| dz \leq C_{A, A_K} b_1^2 b_2^2 \left(\frac{1}{(i\delta)^2} + 1 \right) \|K\|_1^2$$

We have also, denoting ℓ_{A_K} the length of A_K :

$$\begin{aligned} \int_A \left(\mathbb{E} \left| K \left(\frac{z - Z_0}{b} \right) \right| \right)^2 dz &= b_1^2 b_2^2 \int_A \left(\int_{A_K} |K(u, v)| p(x - ub_1, y - vb_2) dudv \right)^2 dx dy \\ &\leq b_1^2 b_2^2 \ell_{A_K} \int_A \int_{A_K} K^2(u, v) p^2(x - ub_1, y - vb_2) dudv dx dy \\ &\leq b_1^2 b_2^2 \ell_{A_K} \|K\|_2^2 \|p\|_{2, A+A_K}^2 \end{aligned}$$

Finally

$$\int_A Cov(U_{0,n}(z), U_{i,n}(z)) dz \leq C \left(\frac{1}{(i\delta)^2} + 1 \right)$$

where C is a constant depending on C_{A, A_K} , ℓ_{A_K} , $\|K\|_2^2$, $\|K\|_1^2$, $\|p\|_{2, A+A_K}^2$.

We can also bound $|Cov(U_{0,n}(z), U_{i,n}(z))|$ following Viennet [1997] and the functions $g_{k\delta}$ introduced in (5). Thus

$$\begin{aligned} \int_A |Cov(U_{0,n}(z), U_{i,n}(z))| dz &\leq \frac{4}{n} \int \left[\int \left(\frac{1}{b_1 b_2} K \left(\frac{x-u}{b_1}, \frac{y-v}{b_2} \right) \right)^2 g_{i\delta}(u, v) d\mathbb{P}_{Z_0}(u, v) \right] dx dy \\ &= \frac{4}{n} \int \left(\int \frac{1}{(b_1 b_2)^2} K^2 \left(\frac{x-u}{b_1}, \frac{y-v}{b_2} \right) dx dy \right) g_{i\delta}(u, v) d\mathbb{P}_{Z_0}(u, v) \\ &\leq \frac{4}{n b_1 b_2} \int K^2(s, t) ds dt \int g_{i\delta}(u, v) d\mathbb{P}_{Z_0}(u, v) \\ \int_A |Cov(U_{0,n}(z), U_{i,n}(z))| dz &\leq \frac{4\|K\|^2}{n b_1 b_2} \beta(i\delta) \leq C \frac{4\|K\|^2}{n b_1 b_2} \rho^{i\delta}. \end{aligned}$$

Finally, we can bound $\int_A T_2(z) dz$:

$$\int_A T_2(z) dz \leq \frac{C}{n b_1 b_2} + \frac{C}{n} \sum_{i=1}^{n-1} \left(\left(\frac{1}{(i\delta)^2} + 1 \right) \wedge \frac{\rho^{i\delta}}{b_1 b_2} \right).$$

By gathering the bounds, we get the result. \square

6.4. Proof of Theorem 1.

• Preliminary results.

To prove Theorem 1, we use Berbee's coupling method as in Viennet (1997), Proposition 5.1

and its proof p.484. Let $Z_{i\delta} = (X_{i\delta}, Y_{i\delta})$. We assume that $n = 2p_n q_n$ with integers p_n and q_n to be chosen. Then there exist random variables $Z_{i\delta}^*$, $i = 1, \dots, n$ satisfying the following properties:

- For $\ell = 1, \dots, p_n$, the random vectors $\vec{U}_{\ell,1} = (Z_{[2(\ell-1)q_n+1]\delta}, \dots, Z_{(2\ell-1)q_n\delta})'$ and $\vec{U}_{\ell,1}^* = (Z_{[2(\ell-1)q_n+1]\delta}^*, \dots, Z_{(2\ell-1)q_n\delta}^*)'$ have the same distribution, and so have the vectors $\vec{U}_{\ell,2} = (Z_{[(2\ell-1)q_n+1]\delta}, \dots, Z_{2\ell q_n\delta})'$ and $\vec{U}_{\ell,2}^* = (Z_{[(2\ell-1)q_n+1]\delta}^*, \dots, Z_{2\ell q_n\delta}^*)'$.
- For $\ell = 1, \dots, p_n$, $\mathbb{P}(\vec{U}_{\ell,1} \neq \vec{U}_{\ell,1}^*) \leq \beta_Z(q_n\delta)$ and $\mathbb{P}(\vec{U}_{\ell,2} \neq \vec{U}_{\ell,2}^*) \leq \beta_Z(q_n\delta)$.
- For each $\epsilon \in \{1, 2\}$, the random vectors $\vec{U}_{1,\epsilon}^*, \dots, \vec{U}_{p_n,\epsilon}^*$ are independent.

We denote by \tilde{p}_b^* the estimator computed with the $Z_{i\delta}^*$ instead of the $Z_{i\delta}$ and write $\tilde{p}_b^* = (\tilde{p}_b^{*(1)} + \tilde{p}_b^{*(2)})/2$ and $\tilde{p}_{b,b'}^* = (\tilde{p}_{b,b'}^{*(1)} + \tilde{p}_{b,b'}^{*(2)})/2$ to separate the part with $Z_{i\delta}^*$ with odd i (super-index (1)) or even i (super-index (2)).

For $Z_{i\delta} = (X_{i\delta}, Y_{i\delta})$, let us define the set $\Omega^* = \{Z_{i\delta} = Z_{i\delta}^*, i = 1, \dots, n\}$. We have

$$(28) \quad \mathbb{P}(\Omega^{*c}) \leq 2p_n\beta_Z(q_n\delta) \leq n\beta_Z(q_n\delta).$$

Now writing that

$$(29) \quad \begin{aligned} \|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2 &= \int \left(\frac{1}{n} \sum_{i=1}^n K_{b'}(z - Z_{i\delta}) - K_{b'}(z - Z_{i\delta}^*) \right)^2 dz \mathbf{1}_{\Omega^{*c}} \\ &\leq \frac{2}{n} \sum_{i=1}^n \int (K_{b'}^2(z - Z_{i\delta}) + K_{b'}^2(z - Z_{i\delta}^*)) dz \mathbf{1}_{\Omega^{*c}} \\ &\leq \frac{4}{b'} \int K^2(w) dw \mathbf{1}_{\Omega^{*c}}, \end{aligned}$$

this implies, for $1/b \leq n$, $\forall b \in \mathcal{B}_n$,

$$\mathbb{E}(\sup_{b'} \|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2) \leq 4n\|K\|^2 \mathbb{P}(\Omega^{*c}) \leq 4\|K\|^2 n^2 \beta_Z(q_n\delta).$$

Using $\beta_Z(q_n\delta) \leq \rho^{q_n\delta}$, we choose

$$(30) \quad q_n = 3 \left\lceil \frac{\log(n)}{\delta |\log(\rho)|} \right\rceil,$$

and we get

$$(31) \quad \mathbb{E}(\sup_{b'} \|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2) \leq \frac{4\|K\|^2}{n}.$$

Using Young's Inequality $\|u * v\|_r \leq \|u\|_p \|v\|_q$, $(1/p) + (1/q) = (1/r) + 1$ with $r = 2$, $p = 1$ and $q = 2$, we get for any $b, b' \in \mathcal{B}_n$,

$$\|\tilde{p}_{b,b'} - \tilde{p}_{b,b'}^*\|^2 = \|K_{b'} * (\tilde{p}_b - \tilde{p}_b^*)\|^2 \leq \|K_{b'}\|_1^2 \|\tilde{p}_b - \tilde{p}_b^*\|^2 = \|K\|_1^2 \|\tilde{p}_b - \tilde{p}_b^*\|^2.$$

Thus

$$(32) \quad \mathbb{E}(\sup_{b'} \|\tilde{p}_{b,b'} - \tilde{p}_{b,b'}^*\|^2) \leq \frac{4\|K\|_1^2 \|K\|^2}{n}.$$

Moreover, we can prove by using Talagrand Inequality (see (37)),

Lemma 1. For $i = 1, 2$, and q_n chosen in (30), there exist a constant C^* and a constant κ_0^* such that, for any $\kappa \geq \kappa_0^*$:

$$(33) \quad \mathbb{E} \left[\sup_{b' \in \mathcal{B}_n} \left(\|\tilde{p}_{b'}^{*(i)} - p_{b'}\|^2 - \kappa \sum_{i=0}^n \beta(i\delta) \frac{\|K_{b'}\|^2}{n} \right)_+ \right] \leq C^* \frac{\log(n)}{n\delta}.$$

Moreover there exist a constant C and a constant κ_0 such that, for any $\kappa \geq \kappa_0$:

$$(34) \quad \mathbb{E} \left[\sup_{b' \in \mathcal{B}_n} \left(\|\tilde{p}_{b,b'}^{*(i)} - p_{b,b'}\|^2 - \kappa \sum_{i=0}^n \beta(i\delta) \frac{\|K\|_1^2 \|K_{b'}\|^2}{n} \right)_+ \right] \leq C \frac{\log(n)}{n\delta}.$$

• **Proof of Theorem 1.**

The definition of \tilde{b} implies that $A(\tilde{b}) + V(\tilde{b}) \leq A(b) + V(b)$ for any $b \in \mathcal{B}_n$. Thus, for any b in \mathcal{B}_n , recalling that $p_b = K_b * p$, $p_{b,b'} = K_b * K_{b'} * p$, we have the decomposition

$$\begin{aligned} \|\tilde{p}_{\tilde{b}} - p\|^2 &\leq 3(\|\tilde{p}_{\tilde{b}} - \tilde{p}_{b,\tilde{b}}\|^2 + \|\tilde{p}_{b,\tilde{b}} - \tilde{p}_b\|^2 + \|\tilde{p}_b - p\|^2) \\ &\leq 3(A(b) + V(\tilde{b})) + 3(A(\tilde{b}) + V(b)) + 3\|\tilde{p}_b - p\|^2 \\ &\leq 6(A(b) + V(b)) + 3\|\tilde{p}_b - p\|^2. \end{aligned}$$

Therefore

$$(35) \quad \mathbb{E}[\|\tilde{p}_{\tilde{b}} - p\|^2] \leq 3\mathbb{E}[\|\tilde{p}_b - p\|^2] + 6V(b) + 6\mathbb{E}[A(b)].$$

To study $A(b)$, we note that for any b, b'

$$\begin{aligned} \|\tilde{p}_{b'} - \tilde{p}_{b,b'}\|^2 &\leq \frac{7}{4} \left(\|\tilde{p}_{b'}^{*(1)} - p_{b'}\|^2 + \|\tilde{p}_{b'}^{*(2)} - p_{b'}\|^2 + \|\tilde{p}_{b,b'}^{*(1)} - p_{b,b'}\|^2 + \|\tilde{p}_{b,b'}^{*(2)} - p_{b,b'}\|^2 \right) \\ &\quad + 7(\|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2 + \|\tilde{p}_{b,b'} - \tilde{p}_{b,b'}^*\|^2 + \|p_{b'} - p_{b,b'}\|^2). \end{aligned}$$

This implies that

$$\begin{aligned} A(b) &\leq \frac{7}{4} \sup_{b'} \left(\|\tilde{p}_{b'}^{*(1)} - p_{b'}\|^2 - \frac{V(b')}{7} \right)_+ + \frac{7}{4} \sup_{b'} \left(\|\tilde{p}_{b'}^{*(2)} - p_{b'}\|^2 - \frac{V(b')}{7} \right)_+ \\ &\quad + \frac{7}{4} \sup_{b'} \left(\|\tilde{p}_{b,b'}^{*(1)} - p_{b,b'}\|^2 - \frac{V(b')}{7} \right)_+ + \frac{7}{4} \sup_{b'} \left(\|\tilde{p}_{b,b'}^{*(2)} - p_{b,b'}\|^2 - \frac{V(b')}{7} \right)_+ \\ &\quad + 7 \sup_{b'} \|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2 + 7 \sup_{b'} \|\tilde{p}_{b,b'} - \tilde{p}_{b,b'}^*\|^2 + 7 \sup_{b'} \|p_{b'} - p_{b,b'}\|^2 \end{aligned}$$

Using Young's Inequality again, we get for any $b, b' \in \mathcal{B}_n$,

$$(36) \quad \|p_{b'} - p_{b,b'}\|^2 = \|K_{b'} * (p - p_b)\|^2 \leq \|K_{b'}\|_1^2 \|p - p_b\|^2 = \|K\|_1^2 \|p - p_b\|^2.$$

Now using (31), (32), and inequalities (33) and (34) in Lemma 1, implies

$$\mathbb{E}(A(b)) \leq 7\|K\|_1^2 \|p - p_b\|^2 + 7C \vee C' \frac{\log(n)}{n\delta}$$

and plugging this in (35), we get

$$\mathbb{E}(\|\tilde{p}_{\tilde{b}} - p\|^2) \leq (3 + 42\|K\|_1^2) \|p_b - p\|^2 + \left(\frac{12}{\kappa\|K\|_1} + 6 \right) V(b) + 336 \frac{\|K\|^2 (\|K\|_1^2 \vee 1)}{n} + \frac{42C \vee C' \log n}{n\delta}$$

where we use the bound of Proposition 1 which implies that $\mathbb{E}(\|\tilde{p}_b - p\|^2) \leq \|p_b - p\|^2 + (4/\kappa)(V(b)/\|K\|_2^2)$. This ends the proof of Theorem 1. \square

6.5. Proof of Lemma 1.

First, we recall the following version of Talagrand inequality.

Lemma 2. *Let T_1, \dots, T_n be independent random variables and $\nu_n(r) = (1/n) \sum_{j=1}^n [r(T_j) - \mathbb{E}(r(T_j))]$, for r belonging to a countable class \mathcal{R} of measurable functions. Then, for $\epsilon > 0$,*

$$(37) \quad \mathbb{E}[\sup_{r \in \mathcal{R}} |\nu_n(r)|^2 - (1 + 2\epsilon)H^2]_+ \leq C \left(\frac{v}{n} e^{-K_1 \epsilon \frac{nH^2}{v}} + \frac{M^2}{n^2 C^2(\epsilon)} e^{-K_2 C(\epsilon) \sqrt{\epsilon} \frac{nH}{M}} \right)$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\epsilon) = \sqrt{1 + \epsilon} - 1$ and C a universal constant and where

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

Inequality (37) is a straightforward consequence of the Talagrand inequality given in Klein and Rio [2005]. Moreover, standard density arguments allow to apply it to the unit ball of spaces.

We prove Inequality (33), and Inequality (34) follows the same line. We first write that

$$\begin{aligned} & \mathbb{E} \left[\sup_{b'} \left(\|\tilde{p}_{b'} - p_{b'}\|^2 - \kappa B_\delta \frac{\|K_{b'}\|^2}{n} \right)_+ \right] \\ & \leq 3\mathbb{E} \left[\sup_{b'} \|\tilde{p}_{b'} - \tilde{p}_{b'}^*\|^2 \right] + 3\mathbb{E} \left[\sup_{b'} \left(\|\tilde{p}_{b'}^{*(1)} - p_{b'}\|^2 - \kappa B_\delta \frac{\|K_{b'}\|^2}{n} \right)_+ \right] \\ & \quad + 3\mathbb{E} \left[\sup_{b'} \left(\|\tilde{p}_{b'}^{*(2)} - p_{b'}\|^2 - \kappa B_\delta \frac{\|K_{b'}\|^2}{n} \right)_+ \right] \end{aligned}$$

The first expectation is bounded by using (31), and the two others are similar, so we study only one of them.

$$(38) \quad \mathbb{E} \left[\sup_{b'} \left(\|\tilde{p}_{b'}^{*(2)} - p_{b'}\|^2 - \kappa B_\delta \frac{\|K_{b'}\|^2}{n} \right)_+ \right] \leq \sum_{b \in \mathcal{B}_n} \mathbb{E} \left[\left(\|\tilde{p}_b^{*(2)} - p_b\|^2 - \kappa B_\delta \frac{\|K_b\|^2}{n} \right)_+ \right].$$

Next we note that $\|\tilde{p}_b^{*(2)} - p_b\|^2 = \sup_{t, \|t\|=1} \langle \tilde{p}_b^{*(2)} - p_b, t \rangle^2$, and the supremum can be considered over a countable dense set of functions t such that $\|t\| = 1$; let us denote this set by $\mathcal{B}(1)$. Thus,

$$\nu_n(t) := \langle \tilde{p}_b^{*(2)} - p_b, t \rangle = \frac{1}{p_n} \sum_{\ell=1}^{p_n} \frac{1}{q_n} \sum_{j=1}^{q_n} \int \left[K_b(z - Z_{[2(\ell-1)q_n+j]\delta}^*) - \mathbb{E}[K_b(z - Z_{[2(\ell-1)q_n+j]\delta}^*)] \right] t(z) dz$$

is a centered empirical process with independent variables

$$\psi_t(\vec{U}_{\ell,1}^*) = \frac{1}{q_n} \sum_{j=1}^{q_n} \int \left[K_b(z - Z_{[2(\ell-1)q_n+j]\delta}^*) - \mathbb{E}[K_b(z - Z_{[2(\ell-1)q_n+j]\delta}^*)] \right] t(z) dz,$$

to which we apply Talagrand Inequality. We compute H^2, v and M as defined in Lemma 2. As $|\nu_n(t)| \leq \|\tilde{p}_b^{*(2)} - p_b\| \|t\|$, we have

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in \mathcal{B}(1)} (\nu_n(t))^2\right] &\leq \mathbb{E}[\|\tilde{p}_b^{*(2)} - p_b\|^2] = \frac{1}{p_n} \int \text{Var} \left(\frac{1}{q_n} \sum_{j=1}^{q_n} [K_b(z - Z_{j\delta}^*) - \mathbb{E}[K_b(z - Z_{j\delta}^*)]] \right) dz \\ &\leq \frac{4\|K\|^2}{p_n q_n b_1 b_2} (1 + \sum_{i=1}^{q_n-1} \rho^{i\delta}) \leq \frac{8\|K\|^2}{n b_1 b_2} (1 + \sum_{i=1}^{n-1} \rho^{i\delta}). \end{aligned}$$

Therefore $H^2 = 8\|K\|^2 / (n b_1 b_2) (1 + \sum_{i=1}^{n-1} \rho^{i\delta})$.

For v , we have to bound for $t \in \mathcal{B}(1)$, the term $\text{Var} \left(\psi_t(\vec{U}_{\ell,1}^*) \right)$. We apply the variance inequality in Viennet [1997] which yields

$$\text{Var} \left(\psi_t(\vec{U}_{\ell,1}^*) \right) \leq \frac{4}{q_n} \int [K_b * t(z)]^2 \theta(z) p(z) dz$$

where $\mathbb{E}(\theta^q(Z)) \leq q C_{q+1}$ with $C_r = \sum_{\ell \in \mathbb{N}} (\ell + 1)^{r-2} \beta_{\ell\delta}$ (see Lemma 4.2 in Viennet [1997]). By Schwarz Inequality, we get

$$\begin{aligned} \text{Var} \left(\psi_t(\vec{U}_{\ell,1}^*) \right) &\leq \frac{4}{q_n} \left(\int [K_b * t(z)]^4 p(z) dz \right)^{1/2} \left(\int \theta^2(z) p(z) dz \right)^{1/2} \\ &\leq \frac{8}{q_n} \|p\|_{\infty}^{1/2} \|K_b * t\|_4^2 \left(\sum_{\ell \in \mathbb{N}} (\ell + 1) \beta_{\ell\delta} \right)^{1/2}. \end{aligned}$$

Now, we apply Young's Inequality $\|u * v\|_r \leq \|u\|_p \|v\|_q$, $(1/p) + (1/q) = (1/r) + 1$ with $r = 4$ and $q = 2$ and thus $p = 4/3$. We get

$$\text{Var} \left(\psi_t(\vec{U}_{\ell,1}^*) \right) \leq \frac{8}{q_n} \|p\|_{\infty}^{1/2} \|K_b\|_{4/3}^2 \|t\|_2^2 \left(\sum_{\ell \in \mathbb{N}} (\ell + 1) \beta_{\ell\delta} \right)^{1/2}.$$

Lastly,

$$\|K_b\|_{4/3}^2 = \left(\int \left(\frac{1}{b} K\left(\frac{z}{b}\right) \right)^{4/3} dz \right)^{3/2} = b^{-1/2} \|K\|_{4/3}^2.$$

So we get $v = 8\|p\|_{\infty}^{1/2} \|K\|_{4/3}^2 C_3^{1/2} / (q_n \sqrt{b}) = \mathbb{C}_v / (\delta q_n \sqrt{b})$ where $\mathbb{C}_v = 8\|p\|_{\infty}^{1/2} \|K\|_{4/3}^2 C_3^{1/2} \delta$ and C_3 is of order $1/\delta^2$ explaining the normalization.

Lastly as $|\int K_b(z - U)t(z)dz| \leq \sqrt{\int K_b^2(z - U)dz \int t^2(z)dz}$, we get $M = \|K\|/\sqrt{b}$. Thus we apply Talagrand Inequality and get

$$\begin{aligned} \mathbb{E} \left[\left(\|\tilde{p}_b^{*(2)} - p_b\|^2 - \frac{1}{7} \frac{\kappa \|K\|_1^2}{8} H^2 \right)_+ \right] &\leq \frac{C_1}{p_n} \left\{ \frac{\mathbb{C}_v}{\delta q_n \sqrt{b}} \exp \left(- \frac{8C_2 \|K\|^2}{\mathbb{C}_v} \frac{\delta}{1 - \rho^\delta} \frac{1}{\sqrt{b}} \right) \right. \\ &\quad \left. + \frac{\|K\|^2}{p_n b} \exp \left(-C_3 \sqrt{p_n} / (q_n (1 - \rho^\delta))^{1/2} \right) \right\}. \end{aligned}$$

Note that $\exp(-C_3\sqrt{p_n}/(q_n(1-\rho^\delta))^{1/2})$ is of order $\exp(-C\sqrt{n}\delta/\log(n))$ when replacing q_n by its value as given by (30). We get the result under conditions (17) and (18) and as $1/b \leq n$. \square

6.6. Proof of Proposition 4. The difference with the complete observation case is the term

$$T_3 = \mathbb{E} \left[\int \left(\frac{1}{nb_1b_2} \sum_{i=1}^n K\left(\frac{z - \bar{Z}_{i\delta}}{b}\right) - K\left(\frac{z - Z_{i\delta}}{b}\right) \right)^2 dz \right]$$

where $\bar{Z}_{i\delta} = \left(X_{i\delta}, \frac{X_{(i+1)\delta} - X_{i\delta}}{\delta}\right)$. We have

$$\begin{aligned} & \mathbb{E} \left[\int \left(\frac{1}{nb_1b_2} \sum_{i=1}^n K\left(\frac{z - \bar{Z}_{i\delta}}{b}\right) - K\left(\frac{z - Z_{i\delta}}{b}\right) \right)^2 dz \right] \\ & \leq \frac{1}{n^2b_1^2b_2^2} \int \mathbb{E} \left[\sum_{i=1}^n \left(K\left(\frac{z - \bar{Z}_{i\delta}}{b}\right) - K\left(\frac{z - Z_{i\delta}}{b}\right) \right)^2 \right] dz \\ & \leq \frac{1}{n^2b_1^2b_2^2} \int n^2 \mathbb{E} \left[\left(K\left(\frac{x - X_0}{b_1}, \frac{y - \bar{Y}_0}{b_2}\right) - K\left(\frac{x - X_0}{b_1}, \frac{y - Y_0}{b_2}\right) \right)^2 \right] dz \end{aligned}$$

Then, we have

$$\begin{aligned} & K\left(\frac{x - X_0}{b_1}, \frac{y - \bar{Y}_0}{b_2}\right) - K\left(\frac{x - X_0}{b_1}, \frac{y - Y_0}{b_2}\right) \\ & = K_1\left(\frac{x - X_0}{b_1}\right) \frac{1}{b_2\delta} \int_0^\delta K_2'\left(\frac{y - Y_0 - M_v}{b_2}\right) (Y_v - Y_0) dv \end{aligned}$$

where $M_v = \frac{1}{\delta} \int_0^v (Y_s - Y_0) ds$. Using the compact support A_K of the kernel K , we have $|\partial_y K(a, b)| \leq C \mathbf{1}_{|a| \leq A_K} \mathbf{1}_{|b| \leq A_K}$. This yields to

$$\begin{aligned} & \mathbb{E} \left[\int \left(\frac{1}{nb_1b_2} \sum_{i=1}^n K\left(\frac{z - \bar{Z}_{i\delta}}{b}\right) - K\left(\frac{z - Z_{i\delta}}{b}\right) \right)^2 dz \right] \\ & \leq \frac{1}{b_1b_2^4\delta^2} \int K_1^2(x) dx \int \mathbb{E} \left[\delta \int_0^\delta |Y_v - Y_0|^2 (K_2')^2 \left(\frac{y - Y_0 - M_v}{b_2}\right) dv \right] dy \\ & \leq \frac{1}{b_1b_2^3\delta} \int K_1^2(x) dx \int (K_2')^2(y) dy \mathbb{E} \left[\int_0^\delta |Y_v - Y_0|^2 dv \right]. \end{aligned}$$

Using the definition of the model, we have

$$Y_v - Y_0 = \sigma(W_v - W_0) - \int_0^v (cY_t - \nabla D(X_t)) dt$$

Thus, there exists a integer $k \geq 0$

$$\mathbb{E}_{Z_0} |Y_v - Y_0|^2 \leq 2\sigma^2 v + 2cv \int_0^v \mathbb{E}_{Z_0} |Z_t|^k dt \leq C(\sigma^2 v + v^2)$$

Finally, we have

$$\begin{aligned} T_3 &\leq \frac{C}{\delta b_1 b_2^3} \int_0^\delta (\sigma^2 v + v^2) dv \\ &= \frac{C}{b_1 b_2^3} \delta (1 + \delta). \end{aligned}$$

Now adding the bound of T_3 to the bounds of Proposition 1 gives the result of proposition 4. \square

6.7. Proof of Theorem 2.

• **Preliminaries.** The preliminaries are analogous to the proof of Theorem 1. Let $\Gamma_{i\delta} = (\bar{Z}_{i\delta}, Y_{i\delta})$, with $\bar{Z}_{i\delta} = (X_{i\delta}, \frac{X_{(i+1)\delta} - X_{i\delta}}{\delta})$. Then the following Lemma holds.

Lemma 3. *The process $(\Gamma_{i\delta})_{i \geq 1}$ is β -mixing with $\beta_{\Gamma_\delta}(k) \leq \beta_Z((k-1)\delta)$.*

Proof of Lemma 3. We follow the proof of Proposition 3.2 in Genon-Catalot et al. [2000]. Since $(\Gamma_{i\delta})_{i \geq 1}$ is strictly stationary Markov, $\beta_{\Gamma_\delta}(k) = \beta(\sigma(\Gamma_{1\delta}), \sigma(\Gamma_{(k+1)\delta}))$. Hence $\beta_{\Gamma_\delta}(k) \leq \beta_Z((k-1)\delta)$. \square

Now, as in the proof of Theorem 1, we use Berbee's coupling method as in Viennet (1997), Proposition 5.1 and its proof p.484. We assume that $n = 2p_n q_n$ for integers p_n and q_n to be chosen. Then there exist random variables $\Gamma_{i\delta}^*$, $i = 1, \dots, n$ satisfying the following properties (we keep the same notation \vec{U} than in the proof of Theorem 1 even if they are not the same):

- For $\ell = 1, \dots, p_n$, the random vectors $\vec{U}_{\ell,1} = (\Gamma_{[(2\ell-1)q_n+1]\delta}, \dots, \Gamma_{[(2\ell-1)q_n]\delta})'$ and $\vec{U}_{\ell,1}^* = (\Gamma_{[(2\ell-1)q_n+1]\delta}^*, \dots, \Gamma_{[(2\ell-1)q_n]\delta}^*)'$ have the same distribution, and so have the vectors $\vec{U}_{\ell,2} = (\Gamma_{[(2\ell-1)q_n+1]\delta}, \dots, \Gamma_{[(2\ell)q_n]\delta})'$ and $\vec{U}_{\ell,2}^* = (\Gamma_{[(2\ell-1)q_n+1]\delta}^*, \dots, \Gamma_{[(2\ell)q_n]\delta}^*)'$.
- For $\ell = 1, \dots, p_n$, $\mathbb{P}(\vec{U}_{\ell,1} \neq \vec{U}_{\ell,1}^*) \leq \beta_Z((q_n-1)\delta)$ and $\mathbb{P}(\vec{U}_{\ell,2} \neq \vec{U}_{\ell,2}^*) \leq \beta_Z((q_n-1)\delta)$.
- For each $\epsilon \in \{1, 2\}$, the random vectors $\vec{U}_{1,\epsilon}^*, \dots, \vec{U}_{p_n,\epsilon}^*$ are independent.

We denote $\bar{Z}_{i\delta}^*$ the two first coordinates of $\Gamma_{i\delta}^*$. We denote by \hat{p}_b^* and $\hat{p}_{b,b'}^*$ the estimators computed with the $\bar{Z}_{i\delta}^*$ instead of the $\bar{Z}_{i\delta}$ and write $\hat{p}_b^* = (\hat{p}_b^{*(1)} + \hat{p}_b^{*(2)})/2$ to separate the part with $\bar{Z}_{i\delta}^*$ with odd i (super-index (1)) or even i (super-index (2)).

Let us define the set $\bar{\Omega}^* = \{\Gamma_{i\delta} = \Gamma_{i\delta}^*, i = 1, \dots, n\}$. We have

$$(39) \quad \mathbb{P}(\bar{\Omega}^{*c}) \leq 2p_n \beta_Z((q_n-1)\delta) \leq n \beta_Z((q_n-1)\delta).$$

Now using the analogous of (29), we obtain, for $1/b \leq n$, $\forall b \in \mathcal{B}_n$,

$$\mathbb{E}(\sup_{b'} \|\hat{p}_{b'} - \hat{p}_{b'}^*\|^2) \leq 4n \|K\|^2 \mathbb{P}(\bar{\Omega}^{*c}) \leq 4 \|K\|^2 n^2 \beta_Z((q_n-1)\delta).$$

Young Inequality also gives

$$\mathbb{E}(\sup_{b'} \|\hat{p}_{b,b'} - \hat{p}_{b,b'}^*\|^2) \leq 4 \|K\|^2 \|K\|_1^2 n^2 \beta_Z((q_n-1)\delta).$$

Using $\beta_Z((q_n-1)\delta) \leq \rho^{(q_n-1)\delta}$, we choose

$$(40) \quad q_n - 1 = 3 \left\lceil \frac{\log(n)}{\delta |\log(\rho)|} \right\rceil,$$

and we get

$$(41) \quad \mathbb{E}(\sup_{b'} \|\hat{p}_{b'} - \hat{p}_{b'}^*\|^2) \leq \frac{4\|K\|^2}{n}, \quad \mathbb{E}(\sup_{b'} \|\hat{p}_{b,b'} - \hat{p}_{b,b'}^*\|^2) \leq \frac{4\|K\|^2\|K\|_1^2}{n}.$$

Lastly, the following result holds.

Lemma 4. *For $i = 1, 2$, there exists a constant $\kappa_{1,0}$ such that for all $\kappa_1 \geq \kappa_{1,0}$,*

$$\mathbb{E} \left[\sup_{b'} \left(\|\hat{p}_{b'}^{*(i)} - \mathbb{E}(\hat{p}_{b'}^{*(i)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ \right] \leq C \frac{\log(n)}{n\delta},$$

and for all $b \in \mathcal{B}_n$:

$$\mathbb{E} \left[\sup_{b'} \left(\|\hat{p}_{b,b'}^{*(i)} - \mathbb{E}(\hat{p}_{b,b'}^{*(i)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ \right] \leq C \frac{\log(n)}{n\delta},$$

where C is a positive constant.

The proof of Lemma 4 follows a line similar to the proof of Lemma 1 and is omitted.

• **Proof of Theorem 2.** Now we start the proof of the theorem. The definition of \hat{b} implies that $A^{(P)}(\hat{b}) + V^{(P)}(\hat{b}) \leq A^{(P)}(b) + V^{(P)}(b)$ for any $b \in \mathcal{B}_n$. Thus, for any b in \mathcal{B}_n , recalling that $p_b = K_b * p$, $p_{b,b'} = K_b * K_{b'} * p$, we have the decomposition

$$\begin{aligned} \|\hat{p}_{\hat{b}} - p\|^2 &\leq 3(\|\hat{p}_{\hat{b}} - \hat{p}_{\hat{b},\hat{b}}\|^2 + \|\hat{p}_{\hat{b},\hat{b}} - \hat{p}_{\hat{b}}\|^2 + \|\hat{p}_{\hat{b}} - p\|^2) \\ &\leq 3(A^{(P)}(b) + V^{(P)}(\hat{b})) + 3(A^{(P)}(\hat{b}) + V^{(P)}(b)) + 3\|\hat{p}_{\hat{b}} - p\|^2 \\ &\leq 6(A^{(P)}(b) + V^{(P)}(b)) + 3\|\hat{p}_{\hat{b}} - p\|^2. \end{aligned}$$

Therefore

$$\mathbb{E}[\|\hat{p}_{\hat{b}} - p\|^2] \leq 3\mathbb{E}[\|\hat{p}_{\hat{b}} - p\|^2] + 6V^{(P)}(b) + 6\mathbb{E}[A^{(P)}(b)].$$

Let us study $A^{(P)}(b)$. For any b, b'

$$\begin{aligned} \|\hat{p}_{b'} - \hat{p}_{b,b'}\|^2 &\leq \frac{9}{4} \left(\|\hat{p}_{b'}^{*(1)} - \mathbb{E}(\hat{p}_{b'}^{*(1)})\|^2 + \|\hat{p}_{b'}^{*(2)} - \mathbb{E}(\hat{p}_{b'}^{*(2)})\|^2 \right) \\ &\quad + \|\hat{p}_{b,b'}^{*(1)} - \mathbb{E}(\hat{p}_{b,b'}^{*(1)})\|^2 + \|\hat{p}_{b,b'}^{*(2)} - p_{b,b'}\|^2 \\ &\quad + 9(\|\hat{p}_{b'} - \hat{p}_{b'}^*\|^2 + \|\hat{p}_{b,b'} - \hat{p}_{b,b'}^*\|^2 + \|p_{b'} - p_{b,b'}\|^2) \\ &\quad + 9(\|\mathbb{E}(\hat{p}_{b,b}) - p_{b,b'}\|^2 + \|\mathbb{E}(\hat{p}_{b'}) - p_{b'}\|^2) \end{aligned}$$

and thus

$$\begin{aligned} A^{(P)}(b) &\leq \frac{9}{4} \sup_{b'} \left(\|\hat{p}_{b'}^{*(1)} - \mathbb{E}(\hat{p}_{b'}^{*(1)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ + \frac{9}{4} \sup_{b'} \left(\|\hat{p}_{b'}^{*(2)} - \mathbb{E}(\hat{p}_{b'}^{*(2)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ \\ &\quad + \frac{9}{4} \sup_{b'} \left(\|\hat{p}_{b,b'}^{*(1)} - \mathbb{E}(\hat{p}_{b,b'}^{*(1)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ + \frac{9}{4} \sup_{b'} \left(\|\hat{p}_{b,b'}^{*(2)} - \mathbb{E}(\hat{p}_{b,b'}^{*(2)})\|^2 - \frac{V_1^{(P)}(b')}{9} \right)_+ \\ &\quad + 9 \sup_{b'} \|\hat{p}_{b'} - \hat{p}_{b'}^*\|^2 + 9 \sup_{b'} \|\hat{p}_{b,b'} - \hat{p}_{b,b'}^*\|^2 + 9 \sup_{b'} \|p_{b'} - p_{b,b'}\|^2 \\ &\quad + 9 \sup_{b'} \left(\|\mathbb{E}(\hat{p}_{b,b}) - p_{b,b'}\|^2 - V_2^{(P)}(b')/9 \right)_+ + 9 \sup_{b'} \left(\|\mathbb{E}(\hat{p}_{b'}) - p_{b'}\|^2 - V_2^{(P)}(b')/9 \right)_+ \end{aligned}$$

It follows from the proof of Proposition 4 and Schwarz inequality that, for κ_2 large enough, the last two terms are null.

We already noticed, using Young's inequality, that for any $b, b' \in \mathcal{B}_n$,

$$\|p_{b'} - p_{b,b'}\|^2 = \|K_{b'} * (p - p_b)\|^2 \leq \|K_{b'}\|_1^2 \|p - p_b\|^2 = \|K\|_1^2 \|p - p_b\|^2.$$

This, together with Inequality (41) and Lemma 4, implies that

$$\mathbb{E}(\|\hat{p}_b - p\|^2) \leq (54\|K\|_1^2)\|p_b - p\|^2 + \left(\frac{12}{\kappa\|K\|_1} + 6\right)V^{(P)}(b) + 432\frac{\|K\|^2(\|K\|_1^2 \vee 1)}{n} + \frac{54C \log n}{n\delta}$$

where we use the result of Proposition 4. \square

Fabienne COMTE, UFR DE MATHÉMATIQUES ET INFORMATIQUE UNIVERSITÉ PARIS DESCARTES LABORATOIRE MAP5, UMR 8145, 45 RUE DES SAINTS-PÈRES, 75270 PARIS CEDEX 06, FRANCE.

E-mail address: fabienne.comte@parisdescartes.fr

Clémentine PRIEUR, UNIVERSITÉ GRENOBLE ALPES, LABORATOIRE JEAN KUNTZMANN, CNRS UMR 5224, INRIA PROJECT/TEAM AIRSEA, F-38000 GRENOBLE, FRANCE.

E-mail address: clementine.prieur@imag.fr

Adeline SAMSON, UNIVERSITÉ GRENOBLE ALPES, LABORATOIRE JEAN KUNTZMANN, CNRS UMR 5224, SVH-TEAM, F-38000 GRENOBLE, FRANCE.

E-mail address: adeline.leclercq-samson@imag.fr