Adaptive estimation of the stationary density of a stochastic differential equation driven by a fractional Brownian motion

Karine Bertin, Nicolas Klutchnikoff, Fabien Panloup, Maylis Varvenne

▶ **To cite this version:**

Karine Bertin, Nicolas Klutchnikoff, Fabien Panloup, Maylis Varvenne. Adaptive estimation of the stationary density of a stochastic differential equation driven by a fractional Brownian motion. Statistical Inference for Stochastic Processes, 2020, 23 (2), pp.271-300. 10.1007/s11203-020-09218-0 . hal-02503902

HAL Id: hal-02503902

https://univ-angers.hal.science/hal-02503902

Submitted on 10 Mar 2020

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 of the stationary density of a stochastic differential equation driven by a fractional Brownian motion

Karine Bertin[*], Nicolas Klutchnikoff[†], Fabien Panloup[‡], Maylis Varvenne[§]

March 10, 2020

Abstract

We build and study a data-driven procedure for the estimation of the stationary density f of an additive fractional SDE. To this end, we also prove some new concentrations bounds for discrete observations of such dynamics in stationary regime.

Keywords: Fractional Brownian motion; Non-parametric Inference; Stochastic Differential Equation; Stationary density; Rate of convergence; Adaptive density estimation.

AMS classification (2010): 60G22, 60H10, 62M09.

1 Introduction

We consider the $\mathbb{R}^d$-random process $X = (X_t : t \geq 0)$ governed by stochastic differential equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \sigma B_t^H, \quad t \geq 0, \tag{1}$$

where X_0 is the initial value of X, $b\colon \mathbb{R}^d \to \mathbb{R}^d$ is a continuous function, σ is a constant $d \times d$ matrix and $B^H = (B_t^H : t \geq 0)$ is a d-dimensional two-sided

[*]CIMFAV-INGEMAT, Universidad de Valparaiso, General Cruz 222, Valparaíso, Chile. E-mail: karine.bertin@uv.cl

[†]Université de Rennes, CNRS, IRMAR – UMR 6625, F-35000 Rennes, France, E-mail: nicolas.klutchnikoff@univ-rennes2.fr

[‡]Laboratoire angevin de recherche en mathématiques (LAREMA), Université d'Angers, CNRS, 49045 Angers Cedex 01, France, E-mail: fabien.panloup@univ-angers.fr

[§]Institut de Mathématiques de Toulouse (IMT), Université de Toulouse 1 Capitole, 2 Rue du Doyen-Gabriel-Marty, 31042 Toulouse, France. E-mail: maylis.varvenne@math.univ-toulouse.fr; maylis.varvenne@ut-capitole.fr

fractional Brownian motion with Hurst parameter $0 < H < 1$. Even in this non-markovian framework (if $H \neq 1/2$), the process X can be embedded into an infinite Markovian structure (see Hairer, 2005). This allows us to define, under usual assumptions on the coefficients b and σ, a *unique* invariant distribution of X which admits a density $f : \mathbb{R}^d \to \mathbb{R}$. In this paper we are interested in the non-parametric estimation of f based on the observation of X at n equally spaced sampling times $t_1 = \Delta_n, \ldots, t_n = n\Delta_n$ where Δ_n is a non-increasing positive sequence such that $n\Delta_n \to \infty$.

In the case of diffusion models driven by standard Brownian motion ($H = \frac{1}{2}$), the problem of non parametric estimation of the invariant density has been extensively studied, in both discrete and continuous time. In the continuous time framework, the process X is observed for all $0 \leq t \leq T$. Castellana and Leadbetter (1986) proved that, under some specific assumption on the joint density of (X_0, X_t), the parametric rate of convergence $T^{-1/2}$ can be reached. Among other, see also Bosq et al. (1997), Kutoyants (1998), Dalalyan (2001), Comte and Merlevède (2005) and Bosq (2012). Without this specific assumption, classical non-parametric rates of convergence of the form $T^{-s/(2s+1)}$ can be obtained (see Comte and Merlevède, 2002) where s is the smoothness parameter of the function $f : \mathbb{R} \to \mathbb{R}$. The case of discrete observations (which corresponds to our framework) has been studied in a univariate setting in Tribouley and Viennet (1998), Comte and Merlevède (2002) and Schmisser (2013) for integrated risk and in Bertin and Klutchnikoff (2017) for pointwise risk. In these papers, the rate of convergence is proved to depend only on $T = n\Delta_n$ and (adaptive) minimax rates of convergence are of the form $(n\Delta_n)^{-s/(2s+1)}$ (up a a logarithmic term) where s is the smoothness of the density function. See also Bertin et al. (2018) that consider integrated risk in a multivariate setting. When $H \neq 1/2$, nonparametric estimation methods for the model (1) have mainly focused on estimation of the drift term b on the continuous case, see e.g. Mishra and Prakasa Rao (2011) (where the authors study the consistency and the rate of convergence of a nonparametric estimator of the whole trend of the solution to a fractional SDE) and Comte and Marie (2018) (where the authors the consistency of some Nadaraya-Watson's-type estimators of the drift function in a fractional SDE). Note that these papers only consider the case $H > 1/2$ in the continuous case.

Our goal in this paper is to construct a data-driven procedure to estimate the stationary density f of X in the discrete case for both $H < 1/2$ and $H > 1/2$. To this aim, new concentration inequalities are obtained for the "stationary" process, following the strategy of Varvenne (2019). In this paper, the idea was to use a pathwise interpretation of the concentration phenomenon by studying the distance between a functional and its average as a sum of differences of "conditioned paths". Then, the result was obtained by making use of the contraction properties of the dynamics (under strong convexity assumptions). In our paper, the novelty with respect to this paper is to assume that Δ_n may depend on n but mostly, that the process is observed in its stationary regime (instead of starting from a given x like in Varvenne (2019)). Actually, if this

modification is easy to overcome in a Markovian setting, here, this is not the case since at time 0, the process has already a past. In other words, an invariant distribution of (1) is a probability on $\mathbb{R}^d \times \mathcal{W}$ where $\mathcal{W}$ is a functional space (see Section 2.1 for details). In short, proving concentration bounds in stationary regime requires to strongly modify the original proof given in Varvenne (2019) (see Section 6.1 for more detailed explanations).

These tools are used for two purposes. First we obtain rates of convergence for the pointwise risk of classical kernel estimators assuming that f belongs to a Hölder class with a known smoothness parameter $s = (s_1, \dots, s_d) \in (0, \infty)^d$. More precisely, choosing adequately a bandwidth that depends on s, we obtain the rate $\phi_n(s) = (n\Delta_n)^{-\beta_H \gamma(\mathbf{s})}$ where

$$\gamma(\mathbf{s}) = \frac{\bar{s}}{2\left(1 + \frac{1}{\min_j s_j}\right)\bar{s} + 2} \qquad \text{and} \qquad \beta_H = 2 - \max(2H, 1). \tag{2}$$

Here $\bar{s} = \left(\sum_{i=1}^d 1/s_i\right)^{-1}$ denotes a classical parameter in multivariate nonparametric estimation that can be viewed as the effective smoothness of f. Next, we propose a data-driven procedure based on the ideas developed by Goldenshluger and Lepski (see Goldenshluger and Lepski, 2011a, 2014, and references therein) to select the bandwidth. The concentration tools we develop in this paper allow us to prove an oracle-type inequality. This ensures that our data-driven procedure performs almost as well as the best estimator in a given family of estimators. As a direct consequence, our procedure is proved to be adaptive: assuming that f is Hölder with unknown smoothness s, it converges at the rate $\phi_n(s)$ up to a $\log(n\Delta_n)$ factor.

The paper is organized as follows. We first present the model and the new concentration inequalities in Section 2. We introduce the statistical framework in Section 3. Section 4 is devoted to the description of our estimation procedures and their theoretical properties are stated in Section 5. The proofs are postponed to Section 6 (for the concentration inequalities) and 7 (for the properties of statistical procedures).

2 Model and Probabilistic background

2.1 Model

We recall that in the non-Markovian setting given by (1), the well definition of "the" invariant distribution of the process X requires the embedding of the dynamics into an infinite-dimensional Markovian structure. More precisely, the Markovian process above the dynamics, called *Stochastic Dynamical System* (SDS) can be realized as a map on the space $\mathbb{R}^d \times \mathcal{W}$ where $\mathcal{W}$ denotes an appropriate space of Hölder functions from $(-\infty, 0]$ to $\mathbb{R}^d$, equipped with the Wiener measure. This construction is strongly based on the Mandelbrot Van-

Ness representation of the fBm:

$$\forall t \in \mathbb{R}, \quad B_t^H = c_H \int_{\mathbb{R}} (t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \mathrm{d}W_s, \tag{3}$$

where $(W_t)_{t\in\mathbb{R}}$ is a two-sided d-dimensional Brownian motion and $c_H > 0$. We denote by $(\mathcal{Q}_t(x,w))_{t\geq 0,(x,w)\in\mathbb{R}^d\times\mathcal{W}}$ the related semi-group (for details on regularity properties of the SDS, see Hairer (2005)).

For this type of dynamics, an *initial condition* is given by a couple (X_0, W^-), where $W^- = (W_t)_{t\leq 0}$ and $X_0 \in \mathbb{R}^d$. In other words, an *initial condition* is a distribution μ on $\mathbb{R}^d \times \mathcal{W}$ such that the projection on the second coordinate is $\mathbf{P}_{W^-}$.

Then, an invariant distribution ν for $(\mathcal{Q}_t(x,w))_{t\geq 0,(x,w)\in\mathbb{R}^d\times\mathcal{W}}$ is an initial condition which is such that the distribution $\mathbf{P}_{X^\nu}$ of the process $(X_t^\nu)_{t\geq 0}$ built with this initial condition is invariant by a time-shift. We say that the invariant distribution is unique if $\mathbf{P}_{X^\nu}$ is unique. Finally, if the invariant distribution exists, we will denote by $\bar{\nu}$, its first marginal: $\bar{\nu}(dx) = \int_{\mathcal{W}} \nu(dx, dw)$. Such a distribution (on $\mathbb{R}^d$) will be usually called "marginal invariant distribution". We will denote by f the density of $\bar{\nu}$ with respect to the Lebesgue measure on $\mathbb{R}^d$ (denoted by λ_d in the sequel) when exists. In Proposition 1 below, we recall some sufficient conditions which ensure existence, uniqueness of the invariant distribution and absolute continuity of $\bar{\nu}$ with respect to the Lebesgue measure. To this end, let us first state the assumptions used throughout our paper:

($\mathbf{H_1}$) *(stability)* The function $b : \mathbb{R}^d \to \mathbb{R}$ is continuous and there exists a constant $\alpha > 0$ such that: For every $x, y \in \mathbb{R}^d$, we have

$$\langle b(x) - b(y),\, x - y\rangle \leq -\alpha |x-y|^2$$

($\mathbf{H_2}$) *(strong regularity)* For every $x, y \in \mathbb{R}^d$,

$$|b(x) - b(y)| \leq L|x-y|.$$

($\mathbf{H_3}$) *(nondegeneracy)* The matrix σ is invertible.

Proposition 1. *Assume* ($\mathbf{H_1}$) *and* ($\mathbf{H_2}$). *Then, existence holds for the invariant distribution ν. If* ($\mathbf{H_3}$) *is also fulfilled, then uniqueness holds for ν (unique in the sense of Hairer, 2005). Furthermore, if b is $\mathcal{C}^1$, then the marginal invariant distribution $\bar{\nu}$ admits a density f with respect to λ_d.*

Existence and uniqueness are consequences of Hairer (2005). For the existence of density for $\bar{\nu}$, we rely on the one hand, on (Hairer, 2005, Theorems 1.2, 1.3), which state that (X_t^x) converges in total variation distance towards $\bar{\nu}$ and, on the other hand, to the fact that, under ($\mathbf{H_3}$), the distribution of (X_t^x) has a density with respect to the Lebesgue measure for any $t > 0$ (see *e.g.* (Besalú et al., 2016, Theorem 1.2) or (Baudoin and Hairer, 2007, Theorem 4.3) when $H > 1/2$). The combination of these two properties implies that $\bar{\nu}$ is absolutely continuous *w.r.t.* to the Lebesgue measure (or equivalently that the density exists).

Remark 1. $\triangleright$ *For the existence of the invariant distribution, Assumption* $(\mathbf{H_1})$ *could be alleviated. More precisely, the contraction assumption may be only assumed out of a compact set. However, we chose to recall the result only under this assumption since, in the sequel, we will need* $(\mathbf{H_1})$ *to obtain concentration properties.*

$\triangleright$ *In this paper, we do not discuss about the smoothness of* f*. This problem is out of the scope of the paper. However, we can expect that the smoothness of* f *strongly relies on the one of* b*. For instance, in the setting of gradient diffusions* $dx_t = -\nabla U(x_t) + \sigma_0 dW_t$ *(where* σ_0 *is a positive number) , it is well-known that the density is given by* $f(x) = \frac{1}{Z_{U,\sigma_0}} \exp\left(-\frac{U(x)}{2\sigma_0^2}\right)$*. This involves that in this particular case, when* b *is of class* $\mathcal{C}^{r+\alpha}$ *(*$\alpha \in (0,1]$*), then* f *is of class* $\mathcal{C}^{r+1+\alpha}$*. We conjecture that this property is still true in our setting.*

2.2 Concentration inequalities for stationary solution

Let $n \in \mathbb{N}^*$. We denote by d_n the following L^1–distance:

$$\forall (x,y) \in (\mathbb{R}^d)^n \times (\mathbb{R}^d)^n, \quad d_n(x,y) := \sum_{k=1}^{n} |x_i - y_i|, \tag{4}$$

where $|\,.\,|$ stands for the Euclidean norm on $\mathbb{R}^d$. For a given $d \times d$-matrix A with real entries, we also denote by $\|\,.\,\|$ a given matrix-norm, subordinated to the Euclidean norm.

Theorem 1. *Let* $H \in (0,1)$*. Assume* $(\mathbf{H_1})$*,* $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$ *and denote by* $(X_t)_{t\geq 0}$ *the stationary solution associated to* (1)*. Let* $n \in \mathbb{N}^*$ *and* $\Delta_n > 0$ *such that* $n\Delta_n \geq 1$*. Let* d_n *be the metric defined by* (4)*. Then, there exists some positive constant* $\mathfrak{C} = C(H, L, \alpha, |b(0_{\mathbb{R}^d})|, \|\sigma\|)$ *such that for all Lipschitz function* $F : \left((\mathbb{R}^d)^n, d_n\right) \to (\mathbb{R}, |\cdot|)$ *and for all* $\lambda > 0$*,*

$$\mathbf{E}\left[\exp\left(\lambda(F_X - \mathbf{E}[F_X])\right)\right] \leqslant \exp\left(\mathfrak{C}\|F\|_{\mathrm{Lip}}^2 \lambda^2 n^{a_H} \Delta_n^{-\beta_H}\right).$$

where $a_H := \max\{2H, 1\}$*,* $\beta_H := \min\{1, 2-2H\}$ *and* $F_X := F(X_{t_1}, \ldots, X_{t_n})$*. Moreover, we deduce from the previous inequality that*

$$\mathbf{P}\left(F_X - \mathbf{E}[F_X] > r\right) \leqslant \exp\left(\frac{-r^2}{4\mathfrak{C} n^{a_H} \Delta_n^{-\beta_H} \|F\|_{\mathrm{Lip}}^2}\right).$$

Corollary 1. *Let the assumptions of Theorem 1 be in force. Let* $F_X := \frac{1}{n}\sum_{k=1}^{n} g(X_{t_k})$ *where* $g : (\mathbb{R}^d, |\cdot|) \to (\mathbb{R}, |\cdot|)$ *is a given Lipschitz function. We have* $\|F\|_{\mathrm{Lip}}^2 = n^{-2}\|g\|_{\mathrm{Lip}}^2$ *and then there exists some positive constant* $C = C(H, L, \alpha, |b(0_{\mathbb{R}^d})|, \|\sigma\|)$ *such that*

$$\mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n}\left(g(X_{t_k}) - \mathbf{E}[g(X_{t_k})]\right) > r\right) \leqslant \exp\left(\frac{-r^2 n^{\beta_H} \Delta_n^{\beta_H}}{4\mathfrak{C}\|g\|_{\mathrm{Lip}}^2}\right) \tag{5}$$

since $2 - a_H = \beta_H$.

Remark 2. $\triangleright$ *In Theorem 1, assumption* $(\mathbf{H_3})$ *ensures the uniqueness of the stationary solution* $(X_t)_{t\geq 0}$ *but the concentration result remains true for every stationary solution to* (1) *when* $(\mathbf{H_3})$ *does not hold.*
$\triangleright$ *In the above results, the constant* $\mathfrak{C} = C(H, L, \alpha, |b(0_{\mathbb{R}^d})|, \|\sigma\|)$ *can be chosen in such a way that* $(L, \alpha, b_0, q) \mapsto C(H, L, \alpha, b_0, q)$ *is bounded on every compact set of* $\mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+ \times \mathbb{R}_+$ *(see Remarks 6 and 8 for more details).*
$\triangleright$ *In Corollary 1, we remark that concentration bounds can be deduced from* (5) *if we impose at least that* $\lim_{n\to+\infty} n\Delta_n = +\infty$, *i.e. that* $\lim_{n\to+\infty} t_n = +\infty$.

3 Adaptive framework

Let us recall that $\bar{\nu}$ denotes the marginal invariant distribution and that f denotes its density *w.r.t.* the Lebesgue measure λ_d, which is assumed to exist in whole the paper (see Proposition 1 for conditions of existence). To measure the accuracy of an estimator $\tilde{f} = \tilde{f}(\cdot, X_{t_1}, \ldots, X_{t_n})$ of f, we define the pointwise risk

$$R_n(\tilde{f}, f) = \left(\mathbf{E}\left|\tilde{f}(x_0) - f(x_0)\right|^p\right)^{1/p}$$

where $p \geq 1$ is fixed. Let $\mathbb{F}$ be a subset of $\mathcal{C}(\mathbb{R}^d, \mathbb{R})$. In what follows, we will consider specific Hölder classes. The maximal risk of $\tilde{f}_n$ over $\mathbb{F}$ is defined by:

$$R_n(\tilde{f}, \mathbb{F}) = \sup_{f\in\mathbb{F}} R_n(\tilde{f}, f).$$

We say that an estimator $\tilde{f}$ converges at the rate of convergence $\phi_n(\mathbb{F})$ over $\mathbb{F}$ if

$$\limsup_{n\to\infty} \phi_n^{-1}(\mathbb{F}) R_n(\tilde{f}, \mathbb{F}) < \infty \tag{6}$$

Note that such estimator may depend on the class $\mathbb{F}$. Moreover (6) ensures a specific behavior of the estimator $\tilde{f}$ over $\mathbb{F}$ but the same estimator can perform poorly over another functional space. The problem of adaptive estimation consists in finding a single estimation procedure with a good behavior over a scale of functional classes. More precisely, given a family $\{\mathbb{F}_\lambda : \lambda \in \Lambda\}$ of subsets of $\mathcal{C}(\mathbb{R}^d, \mathbb{R})$, the goal is to construct f_n^* such that $R_n(f_n^*, \mathbb{F}_\lambda)$ is asymptotically bounded, up to a small multiplicative factor (for example a constant or a logarithmic term), by $\phi_n(\mathbb{F}_\lambda)$ for any $\lambda \in \Lambda$. One of the main tools to prove that an estimation procedure is adaptive over a scale of functional classes is to prove an oracle-type inequality that guarantees that this procedure performs almost as well as the best estimator in a rich family of estimators. Ideally, we would like to have an inequality of the following form:

$$R_n(f^*, f) \leq \inf_{\eta\in H} R_n(\hat{f}_\eta, f), \tag{7}$$

where $\{\hat{f}_\eta \colon \eta \in H\}$ is a family of estimators satisfying: for any $\lambda \in \Lambda$, there exists $\eta(\lambda)$ such that $\hat{f}_{\eta(\lambda)}$ converges at the rate $\phi_n(\mathbb{F}_\lambda)$ over the class $\mathbb{F}_\lambda$. In

general, obtaining such an inequality is not possible. However in many situations, (7) can be relaxed and a weaker inequality of the following type can be proved:

$$R_n(f^*, f) \leq \Upsilon_1 \inf_{\eta \in H} R_n^*(f, \eta) + \Upsilon_2 \delta(n), \tag{8}$$

where Υ_1 and Υ_2 are two positive constants, $R_n^*(f, \eta)$ is an appropriate quantity to be determined that can be viewed as a tight upper bound on $R_n(\hat{f}_\eta, f)$ and $\delta(n)$ is a reminder term. Inequalities of the form (8) are called *oracle-type* inequalities.

4 Estimation procedure

To estimate f we construct a procedure defined through classical kernel density estimators. It is well known that the accuracy of these estimators is mainly determined by the *bandwidth* vector. Thus, obtaining a *data-driven* choice of this parameter is the central problem in our model. In this section, after introducing a family of kernel density estimators, we define a selection procedure based on the ideas developed in Goldenshluger and Lepski (2011b).

4.1 Kernel density estimators

In this paper a function $K : \mathbb{R} \to \mathbb{R}$ is called a kernel if the support of K is included into $[-1, 1]$, K is a Lipschitz function with Lipschitz constant $\mathfrak{L}_K > 0$ and K satisfies $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$. Following Rosenblatt et al. (1956) and Parzen (1962), we consider kernel density estimators $\hat{f}_h$ defined, for $h = (h_1, \ldots, h_d) \in (0, +\infty)^d$, by:

$$\hat{f}_h(x_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(x_0 - X_{t_i}), \tag{9}$$

where, for $(u_1, \ldots, u_d) \in \mathbb{R}^d$, we define $K_h(u_1, \ldots, u_d) = \prod_{i=1}^d h_i^{-1} K(h_i^{-1} u_i)$. We say that a kernel K is of order $M \in \mathbb{N}$ if for any $1 \leq \ell \leq M$, we have $\int_{\mathbb{R}} K(x) x^\ell dx = 0$. In the following paragraph, we propose a data-driven procedure to select the bandwidth h in the finite subset $\mathcal{H}$ of $(0, 1)^d$.

4.2 Bandwidth selection

Our procedure depends on a hyperparameter $\mathfrak{K} > 0$. We refer the reader to Section 5 for detailed comments on the impact of the choice of this parameter. For any bandwidth vector $h = (h_1, \ldots, h_d) \in (0, 1)^d$, we define:

$$V_h = \min(h_1, \ldots, h_d) \prod_{i=1}^{d} h_i \quad \text{and} \quad \varphi_n(h) = \left(\frac{4d\mathfrak{L}_K^2}{V_h^2 (n\Delta_n)^{\beta_H}} \right)^{1/2}, \tag{10}$$

where β_H is defined by (2). We consider the following subset of $(0,1)^d$

$$\mathcal{H} = \left\{ (e^{-l_1}, \ldots, e^{-l_d}) : l_i = 0, \ldots, \left\lfloor \frac{\beta_H}{2} \log(n\Delta_n) \right\rfloor \right\} \cap \widetilde{\mathcal{H}}$$

where $\lfloor \cdot \rfloor$ denotes the integer part and

$$\widetilde{\mathcal{H}} = \left\{ h \in (0,1)^d : V_h \geq \left(\frac{1}{n\Delta_n} \right)^{\beta_H/2} \right\}.$$

Without loss of generality, we assume that $(n\Delta_n)^{\beta_H} \geq e$, which implies in particular that $\mathcal{H}$ is not empty.

Following Goldenshluger and Lepski (2011b), we define for h and $\mathfrak{h}$ in $\mathcal{H}$ the following quantities

$$M_n(h,\mathfrak{h}) = M_n(\mathfrak{h}) + M_n(\mathfrak{h} \vee h) \quad \text{with} \quad M_n(h) = \varphi_n(h)\sqrt{\hat{\kappa} p |\log V_h|} \tag{11}$$

and

$$B(h,x_0) = \max_{\mathfrak{h} \in \mathcal{H}} \left\{ \left| \hat{f}_{\mathfrak{h} \vee h}(x_0) - \hat{f}_{\mathfrak{h}}(x_0) \right| - M_n(h,\mathfrak{h}) \right\}_+ .$$

Here $\{y\}_+ = \max(0,y)$ denotes the nonnegative part of $y \in \mathbb{R}$ and $\mathfrak{h} \vee h$ denotes the component-wise maximum of the bandwidth h and $\mathfrak{h}$. Our procedure consists of selecting a bandwidth $\hat{h}(x_0)$ such that

$$\hat{h}(x_0) = \arg\min_{h \in \mathcal{H}} \left(B(h,x_0) + M_n(h) \right). \tag{12}$$

The final estimator of $f(x_0)$ is then defined as the plugin estimator:

$$\hat{f}(x_0) = \hat{f}_{\hat{h}(x_0)}(x_0).$$

This selection rule follows the principles and the ideas developed by Goldenshluger and Lepski. The quantity $M_n(h)$, called a *majorant*, is a penalized version of the standard deviation of the estimator $\hat{f}_h$ while the quantity $B(h,x_0)$ is, in some sense, closed to its bias term. Finding tight *majorants* is the key point of the method since $\hat{h}(x_0)$ is chosen in (12) in order to realize an empirical trade-off between these two quantities.

5 Results

We first recall the definition of Hölder balls $\Sigma_d(\boldsymbol{s}, \mathbf{L})$. For two d-tuples of positive reals $\boldsymbol{s} = (s_1, \ldots, s_d)$ and $\mathbf{L} = (L_1, \ldots, L_d)$,

$$\Sigma_d(\boldsymbol{s}, \mathbf{L}) = \Big\{ f : \mathbb{R}^d \to \mathbb{R} \text{ s.t. } \forall 1 \leq i \leq d \quad \left\| \frac{\partial^m f}{\partial x_i^m} \right\|_\infty \leq L_i, \quad m = 0, \ldots, \lfloor s_i \rfloor$$
$$\text{and for all } t \in \mathbb{R} \quad \left\| \frac{\partial^{\lfloor s_i \rfloor} f}{\partial x_i^{\lfloor s_i \rfloor}}(\cdot + te_i) - \frac{\partial^{\lfloor s_i \rfloor} f}{\partial x_i^{\lfloor s_i \rfloor}}(\cdot) \right\|_\infty \leq L_i |t|^{s_i - \lfloor s_i \rfloor} \Big\}$$

where for any i, $\lfloor s_i \rfloor = \max\{l \in \mathbb{N} : l < s_i\}$ and e_i is the vector where all coordinates are null except the i-th one which is equal to 1.

5.1 Properties of the kernel estimators

The two following propositions give upper-bounds of the bias and the stochastic term of the estimator $\hat{f}_h$.

Proposition 2. *Let* $\mathbf{s} = (s_1, \ldots, s_d) \in (0, +\infty)^d$ *and* $\mathbf{L} = (L_1, \ldots, L_d) \in (0, +\infty)^d$. *Assume that* $f \in \Sigma_d(\boldsymbol{s}, \mathbf{L})$ *and assume that* K *is a kernel of order greater than* $\max_i \lfloor s_i \rfloor$. *Under* $(\mathbf{H_1})$, $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$, *we have for all* $h \in \mathcal{H}$

$$\left|\mathbf{E}\hat{f}_h(x_0) - f(x_0)\right| \leq \sum_{i=1}^{d} \frac{L_i h_i^{s_i}}{\lfloor s_i \rfloor !} \int_{\mathbb{R}} \left|v^{s_i} K(v)\right| dv \tag{13}$$

and

$$E_h(x_0) = \max_{\mathfrak{h} \in \mathcal{H}} \left|\mathbf{E}\hat{f}_{h \vee \mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right| \leq 2 \sum_{i=1}^{d} \frac{L_i h_i^{s_i}}{\lfloor s_i \rfloor !} \int_{\mathbb{R}} \left|v^{s_i} K(v)\right| dv. \tag{14}$$

Proposition 3. *Under* $(\mathbf{H_1})$, $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$, *we have for all* $h \in \mathcal{H}$

$$\left(\mathbf{E}\left|\hat{f}_h(x_0) - \mathbf{E}\hat{f}_h(x_0)\right|^p\right)^{1/p} \leq \left(p\Gamma\left(\frac{p+1}{2}\right)\right)^{1/p} \mathfrak{C}^{1/2} \varphi_n(h). \tag{15}$$

Remark 3. ▷ *Note that the control of the bias term obtained in Proposition 2 is the same as those obtained for the problem of density estimation in an i.i.d. context. The control of the stochastic term, see Proposition 3, relies on the concentration inequality obtained in Corollary 1. The right hand side of* (15) *depends on the additional assumptions made on our model through the constant* $\mathfrak{C}$.

▷ *This result is valid for a large class of functional (only a Lipschitz condition is required) and under weak assumptions on the process. For* $H = 1/2$, *the concentration inequality is optimal, see Saussereau (2012) and Djellout et al. (2004). However, under strongest assumptions—for example on the joint distribution of* (X_0, X_t), *Bernstein-type inequalities can be used to derive a better upper bound on the stochastic term of order* $(nh_1 \ldots h_d)^{-1}$, *see Bertin and Klutchnikoff (2017) for* $d = 1$.

Using the above propositions we derive, over any Hölder balls $\Sigma_d(\boldsymbol{s}, \mathbf{L})$, the rate of convergence achieved by a kernel estimator defined in (9) with a specific choice of bandwidth that depend on the smoothness parameter $\boldsymbol{s}$.

Theorem 2. *Let* $\boldsymbol{s} = (s_1, \ldots, s_d) \in (0, +\infty)^d$ *and* $\mathbf{L} = (L_1, \ldots, L_d) \in (0, +\infty)^d$. *Assume that* $f \in \Sigma_d(\boldsymbol{s}, \mathbf{L})$ *and assume that* K *is a kernel of order greater than* $\max_i \lfloor s_i \rfloor$. *Under* $(\mathbf{H_1})$, $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$, *the estimator* $\hat{f}_{h(\boldsymbol{s})}$ *defined through the bandwidth* $h(\boldsymbol{s}) = (h_1(\boldsymbol{s}), \ldots, h_d(\boldsymbol{s}))$ *where for any* $i = 1, \ldots, d$

$$h_i(\boldsymbol{s}) = \left(\frac{1}{(n\Delta_n)^{\beta_H}}\right)^{\frac{\gamma(\boldsymbol{s})}{s_i}} \quad \textit{with} \quad \gamma(\boldsymbol{s}) = \frac{\bar{s}}{2\left(1 + \frac{1}{\min_j s_j}\right)\bar{s} + 2} \in (0, 1/2)$$

and satisfies

$$R_n(\hat{f}_{h(\boldsymbol{s})}, f) \leq (\Lambda_1 + \Lambda_2) \left(n\Delta_n\right)^{-\beta_H \gamma(\boldsymbol{s})},$$

where

$$\Lambda_1 = \sum_{i=1}^{d} \frac{L_i}{\lfloor s_i \rfloor !} \int_{\mathbb{R}} \left|v^{s_i} K(v)\right| dv, \quad \Lambda_2 = 2\sqrt{d\mathfrak{C}} \left(p\Gamma\left(\frac{p+1}{2}\right)\right)^{1/p} \mathfrak{L}_K.$$

To our best knowledge few papers deal with nonparametric rate of convergence in our model. Only Comte and Marie (2018) have considered the estimation of the trend function b based on continuous observations when $d = 1$ and $H > 1/2$. They obtain the same rates of convergence only in the case $s = 1$ assuming a Lipschitz condition on the function b.

5.2 Properties of the data-driven procedure

The estimator $\hat{f}$, defined in Section 4.2 using the hyperparameter $\mathfrak{K}$ and the family of bandwidths $\mathcal{H}$ satisfies the following oracle inequality.

Theorem 3. *Under* $(\mathbf{H_1})$, $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$, *if* $\mathfrak{K} > \mathfrak{C}$, *we have:*

$$R_n(\hat{f}, f) \leq \min_{h \in \mathcal{H}} \left\{ R_n(\hat{f}_h, f) + 4M_n(h) + 3E_h(x_0) \right\} + C_0 (n\Delta_n)^{-\beta_H/2} \tag{16}$$

with

$$C_0 = 6\Lambda_2 \left(\frac{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}} - 1\right)}}{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}} - 1\right)} - 1} \right)^{d/p}$$

Remark 4. ▷ *The validity of this result depends on the hyperparameter* $\mathfrak{K}$ *of our procedure. Using Remark 2, for a given value of* $\mathfrak{K}$, *Theorem 3 can be applied for a wide class of models, as soon as the constant* $\mathfrak{C}$, *which depends on* H, L, α, $b(0_{\mathbb{R}^d})$ *and* $\|\sigma\|$ *is less than* $\mathfrak{K}$.

The oracle inequality allows us to obtain upper bound for the rates of convergence over Hölder balls $\Sigma_d(\boldsymbol{s}, \mathbf{L})$ in an adaptive framework.

Theorem 4. *Let* M *be a nonnegative integer and assume that* K *is a kernel of order greater than* M. *Set* $\mathbf{s} \in (0, M+1]^d$ *and* $\mathbf{L} \in (0, +\infty)^d$. *Asume that* $f \in \Sigma_d(\boldsymbol{s}, \mathbf{L})$, *then under* $(\mathbf{H_1})$, $(\mathbf{H_2})$ *and* $(\mathbf{H_3})$ *we have:*

$$R_n(\hat{f}, f) \leq C_1 \left((n\Delta_n)^{-\beta_H} \log\left((n\Delta_n)^{\beta_H} \right) \right)^{\gamma(\mathbf{s})} \tag{17}$$

where

$$C_1 = 8\sqrt{\mathfrak{K} d(d + 3/2)p}\,\mathfrak{L}_K + 7e^{M+1}\Lambda_1 + \Lambda_2 + C_0.$$

Remark 5. ▷ *This result ensures that the estimator* $\hat{f}$ *achieves the rate of convergence obtained in Theorem 2 up to a logarithmic factor. Such behavior is well-known for pointwise adaptive estimation, see Lepskiĭ (1990); Tsybakov*

(1998); Klutchnikoff (2014) among others.
▷ If we let the hyperparameter $\mathfrak{K}$ depend on n (e.g. $\mathfrak{K} = \log(n)$) then the procedure is also asymptotically adaptive with respect to the values of L, α, $b(0_{\mathbb{R}^d})$ and $\|\sigma\|$. In the case $\mathfrak{K} = \log(n)$, the rate of convergence in (17) is multiplied by $(\log n)^{1/2}$.

6 On concentration inequalities for fractional SDEs in stationary regime

6.1 Sketch of proof of Theorem 1

We denote by $(\Omega, \mathcal{F}, \mathbf{P})$ the probability space on which the fBm is defined. Let $(\mathcal{F}_t)_{t\geqslant 0}$ be the natural filtration associated to the two-sided Brownian motion $(W_t)_{t\in\mathbb{R}}$ induced by the Mandelbrot-Van Ness representation (see (3)). As in Varvenne (2019), let us first introduce the following decomposition. For all $k \in \mathbb{N}$, set

$$M_k := \mathbf{E}[F_X \mid \mathcal{F}_k] \tag{18}$$

where we recall that $F_X = F(X_{t_1}, \dots, X_{t_n})$ and $(X_t)_{t\geq 0}$ is the stationary solution of (1). Then, we have:

$$F_X - \mathbf{E}[F_X] = M_{\lceil t_n \rceil} = \sum_{k=1}^{\lceil t_n \rceil} M_k - M_{k-1} + \ \mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X].$$

Our strategy of proof is decomposed as follows, we show that :

(1) for all $1 \leq k \leq \lceil t_n \rceil$, there exists $u_k^{(n)} > 0$ (deterministic) such that for all $\lambda > 0$,

$$\mathbf{E}\left[\exp\left(\lambda(M_k - M_{k-1})\right) \mid \mathcal{F}_{k-1}\right] \leqslant \exp(\lambda^2 u_k^{(n)}) \quad a.s.$$

and then

$$\mathbf{E}\left[\exp\left(\lambda \sum_{k=1}^{\lceil t_n \rceil}(M_k - M_{k-1})\right) \;\middle|\; \mathcal{F}_0\right] \leqslant \exp\left(\lambda^2 \sum_{k=1}^{\lceil t_n \rceil} u_k^{(n)}\right) \quad a.s.$$

(2) there exists $u_0^{(n)} > 0$ (deterministic) such that for all $\lambda > 0$,

$$\mathbf{E}\left[\exp\left(\lambda(\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X])\right)\right] \leqslant \exp(\lambda^2 u_0^{(n)}) \quad a.s.$$

From (1) and (2), we finally get

$$\mathbf{E}\left[\exp\left(\lambda(F_X - \mathbf{E}[F_X])\right)\right] \leqslant \exp\left(\lambda^2 \sum_{k=0}^{\lceil t_n \rceil} u_k^{(n)}\right).$$

We are thus reduced to study conditional exponential moments in (1) and (2). The related results are given in Proposition 4 and 7 (and Theorem 1 easily follows).

In order to provide such exponential bounds, we rely on the following key lemma (see Lemma 1.5 in Chapter 1 of Rigollet and Hütter (2017)):

Lemma 1. *Let Z be a **centered** random variable on $\mathbb{R}$ such that there exists $\zeta > 0$ such that for all $p \geq 2$,*

$$\mathbf{E}[|Z|^p] \leq \zeta^{\frac{p}{2}} p\Gamma\left(\frac{p}{2}\right).$$

Then, for all $\lambda > 0$, we have

$$\mathbf{E}[\exp(\lambda Z)] \leq \exp(2\lambda^2 \zeta).$$

6.2 Part 1: sum of martingale increments

In this subsection, our purpose is to prove the following result :

Proposition 4. *Assume $(\mathbf{H_1})$ and $(\mathbf{H_2})$. Let $H \in (0,1)$. There exists $K = K(H, L, \alpha, \|\sigma\|) > 0$ such that for all $\lambda > 0$,*

$$\mathbf{E}\left[\exp\left(\lambda\left(\sum_{k=1}^{\lceil t_n \rceil} M_k - M_{k-1}\right)\right) \;\middle|\; \mathcal{F}_0\right] \leqslant \exp\left(K\lambda^2 \|F\|_{\mathrm{Lip}}^2 \sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2\right) \quad a.s.$$

where

$$\psi_{n,k} := \left\lfloor \frac{k}{\Delta_n} \right\rfloor - \left\lceil \frac{k-1}{\Delta_n} \right\rceil + \sum_{i=\lceil \frac{k}{\Delta_n} \rceil}^{n} u_i^{H-3/2},$$

$u_i := t_i - k + 1 = i\Delta_n - k + 1$ and M_k is defined by (18) .
Moreover, there exists $c = c(H) > 0$ such that

$$\sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2 \leq c \begin{cases} n\Delta_n^{-1} & \text{if} \quad H \in (0, 1/2) \\ n^{2H}\Delta_n^{2H-2} & \text{if} \quad H \in (1/2, 1). \end{cases}$$

Remark 6. *Following carefully the constants in the proof of this proposition, one easily checks that $(L, \alpha, s) \mapsto K(H, L, \alpha, s)$ is bounded on every compact set of $\mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+$.*

Through equation (1) and the fact that b is Lipschitz continuous, for all $t \geqslant 0$, Y_t can be seen as a functional of the time t, the initial condition X_0 and the Brownian motion $(W_s)_{s \leq t}$. Denote by $\Phi : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{C}(\mathbb{R}, \mathbb{R}^d) \to \mathbb{R}^d$ this functional, we then have

$$\forall t \geqslant 0, \quad X_t := \Phi_t(X_0, (W_s)_{s \leq t}).$$

Now, let $k \geqslant 1$, we have

$$
\begin{aligned}
&|M_k - M_{k-1}| \\
&= |\mathbf{E}[F_X|\mathcal{F}_k] - \mathbf{E}[F_X|\mathcal{F}_{k-1}]| \\
&\leqslant \|F\|_{\mathrm{Lip}} \int_\Omega \sum_{t_i \geq k-1} \left|\Phi_{t_i}\left(X_0, (W_s)_{s\leq k} \sqcup \tilde{w}_{[k,t_i]}\right) - \Phi_{t_i}\left(X_0, (W_s)_{s\leq k-1} \sqcup \tilde{w}_{[k-1,t_i]}\right)\right| \mathbf{P}_W(\mathrm{d}\tilde{w}).
\end{aligned}
$$

Let us introduce now some notations. First, for all $t \geqslant k-1$ set $u := t-k+1$, then for all $u \geqslant 0$, we define

$$
Y_u := \begin{cases} \Phi_{u+k-1}\left(X_0, (W_s)_{s\leq k} \sqcup (\tilde{w}_s)_{s\in[k,u+k-1]}\right) & \text{if } u \geqslant 1 \\ \Phi_{u+k-1}\left(X_0, (W_s)_{s\leq u+k-1}\right) & \text{otherwise,} \end{cases}
$$

and

$$
\tilde{Y}_u := \Phi_{u+k-1}\left(X_0, (W_s)_{s\leq k-1} \sqcup (\tilde{w}_s)_{s\in[k-1,u+k-1]}\right).
$$

By using equation (1), we then have

$$
Y_u - \tilde{Y}_u = \int_0^u b(Y_s) - b(\tilde{Y}_s)\mathrm{d}s + c_H\sigma \int_0^{1\wedge u} (u-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s. \quad (19)
$$

where we have set $(W_s^{(k)})_{s\geqslant 0} := (W_{s+k-1} - W_{k-1})_{s\geqslant 0}$ which is a Brownian motion independent from $\mathcal{F}_{k-1}$ and $(\tilde{w}_s^{(k)})_{s\geqslant 0} := (\tilde{w}_{s+k-1} - \tilde{w}_{k-1})_{s\geqslant 0}$.

Finally, we have the following inequality for all $k \geq 1$:

$$
|M_k - M_{k-1}| \leq \|F\|_{\mathrm{Lip}} \int_\Omega \sum_{u_i \geq 0} \left|Y_{u_i} - \tilde{Y}_{u_i}\right| \mathbf{P}_W(\mathrm{d}\tilde{w}) \quad (20)
$$

where $u_i := t_i - k + 1 = i\Delta_n - k + 1$.

In the next section, we proceed to a control of the quantity $|Y_u - \tilde{Y}_u|$.

6.2.1 Control lemma

Lemma 2. *We have the two following inequalities:*

(*i*) *for all $u \in [0,1]$, there exists $K = K(H, L, \alpha, \|\sigma\|) > 0$ such that,*

$$
|Y_u - \tilde{Y}_u| \leqslant K \sup_{v\in[0,2]} \left|G_v^{(k)}(W - \tilde{w})\right|
$$

where

$$
G_v^{(k)}(W - \tilde{w}) := \int_0^{1\wedge v} (v-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s,
$$

(ii) for all $u \geq 1$, there exists $K = K(H, L, \alpha, \|\sigma\|) > 0$ such that,

$$|Y_u - \tilde{Y}_u| \leqslant K\ u^{H-\frac{3}{2}} \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|$$

Proof. $\triangleright$ First case: let $u \in [0,1]$.
By the triangle inequality and assumption $(\mathbf{H_2})$, we have in (19)

$$|Y_u - \tilde{Y}_u| \leqslant L \int_0^u |Y_s - \tilde{Y}_s| \mathrm{d}s + c_H \|\sigma\| \left| \int_0^{1\wedge u} (u-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right|$$

Then, from Gronwall's lemma, we deduce the following

$$\begin{aligned}
&|Y_u - \tilde{Y}_u| \\
&\leqslant c_H \|\sigma\| \left| \int_0^{1\wedge u} (u-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right| \\
&\qquad\qquad + c_H \|\sigma\| L \int_0^u \left| \int_0^{1\wedge v} (v-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right| e^{L(u-v)} \mathrm{d}v \\
&\leqslant c_H \|\sigma\| \sup_{v\in[0,2]} \left| \int_0^{1\wedge v} (v-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right| \left(1 + \left[-e^{L(u-v)}\right]_0^u\right) \\
&\leqslant c_H \|\sigma\| e^{2L} \sup_{v\in[0,2]} \left| \int_0^{1\wedge v} (v-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right|.
\end{aligned}$$

and Lemma 2 is shown for $u \in [0,1]$.

$\triangleright$ Second case: let $u \geqslant 1$.
First, if $u \in [1,2]$, we have $|Y_u - \tilde{Y}_u| \leqslant K_H \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|$ by the first part of this proof combined with the following inequality :

$$\begin{aligned}
&\left| \int_0^{1\wedge v} (v-s)^{H-\frac{1}{2}} \mathrm{d}(W^{(k)} - \tilde{w}^{(k)})_s \right| \\
&= \left| (v - 1\wedge v)^{H-\frac{1}{2}} (W_{1\wedge v}^{(k)} - \tilde{w}_{1\wedge v}^{(k)}) + (H-1/2) \int_0^{1\wedge v} (v-s)^{H-\frac{3}{2}} (W_s^{(k)} - \tilde{w}_s^{(k)}) \mathrm{d}s \right| \\
&\leqslant K_H \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|.
\end{aligned}$$

Now, let us treat the case $u \geq 2$. In the following inequalities, we use assumption $(\mathbf{H_1})$ on the function b and the elementary Young inequality $\langle a, b\rangle \leqslant$

$\frac{1}{2}\left(\varepsilon|a|^2+\frac{1}{\varepsilon}|b|^2\right)$ with $\varepsilon=2\alpha$.

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}u}|Y_u-\tilde{Y}_u|^2\\
&=2\langle Y_u-\tilde{X}_u, b(Y_u)-b(\tilde{Y}_u)\rangle + c_H(2H-1)\langle Y_u-\tilde{Y}_u,\ \sigma\int_0^1 (u-s)^{H-\frac{3}{2}}\mathrm{d}(W^{(k)}-\tilde{w}^{(k)})_s\rangle\\
&\leqslant -2\alpha|Y_u-\tilde{Y}_u|^2+\alpha|Y_u-\tilde{Y}_u|^2+\frac{c_H^2(2H-1)^2\|\sigma\|^2}{4\alpha}\left|\int_0^1 (u-s)^{H-\frac{3}{2}}\mathrm{d}(W^{(k)}-\tilde{w}^{(k)})_s\right|^2\\
&\leqslant -\alpha|Y_u-\tilde{Y}_u|^2+\frac{c_H^2(2H-1)^2\|\sigma\|^2}{4\alpha}\left|\int_0^1 (u-s)^{H-\frac{3}{2}}\mathrm{d}(W^{(k)}-\tilde{w}^{(k)})_s\right|^2.
\end{aligned}
$$

We then apply Gronwall's lemma to obtain

$$
|Y_u-\tilde{Y}_u|^2\leqslant e^{-\alpha(u-2)}|Y_2-\tilde{Y}_2|^2+\alpha_H\int_2^u e^{-\alpha(u-v)}\left|\int_0^1 (v-s)^{H-\frac{3}{2}}\mathrm{d}(W^{(k)}-\tilde{w}^{(k)})_s\right|^2\mathrm{d}v \tag{21}
$$

with $\alpha_H:=\frac{c_H^2(2H-1)^2\|\sigma\|^2}{4\alpha}$.
Now, we set $\varphi_k(v):=\int_0^1(v-s)^{H-\frac{3}{2}}\mathrm{d}(W^{(k)}-\tilde{w}^{(k)})_s$ and we apply an integration by parts to φ_k taking into account that $W_0^{(k)}=\tilde{w}_0^{(k)}=0$:

$$
\varphi_k(v)=(v-1)^{H-\frac{3}{2}}(W_1^{(k)}-\tilde{w}_1^{(k)})-(3/2-H)\int_0^1(v-s)^{H-\frac{5}{2}}(W_s^{(k)}-\tilde{w}_s^{(k)})\mathrm{d}s.
$$

And then

$$
\begin{aligned}
&\int_2^u e^{-\alpha(u-v)}\left|\varphi_k(v)\right|^2\mathrm{d}v\\
&\leqslant 2|W_1^{(k)}-\tilde{w}_1^{(k)}|^2\int_2^u(v-1)^{2H-3}e^{-\alpha(u-v)}\mathrm{d}v\\
&\qquad+2(3/2-H)^2\int_2^u e^{-\alpha(u-v)}\left|\int_0^1(v-s)^{H-\frac{5}{2}}(W_s^{(k)}-\tilde{w}_s^{(k)})\mathrm{d}s\right|^2\mathrm{d}v\\
&\leqslant 2|W_1^{(k)}-\tilde{w}_1^{(k)}|^2\int_2^u(v-1)^{2H-3}e^{-\alpha(u-v)}\mathrm{d}v\\
&\qquad+2(3/2-H)^2\left|\int_0^1(W_s^{(k)}-\tilde{w}_s^{(k)})\mathrm{d}s\right|^2\int_2^u(v-1)^{2H-5}e^{-\alpha(u-v)}\mathrm{d}v
\end{aligned}
\tag{22}
$$

Lemma 3. *Let $\alpha,\beta>0$. Then, for all $u\geqslant 2$,*

$$
e^{-\alpha u}\int_2^u(v-1)^{-\beta}e^{\alpha v}\mathrm{d}v\leqslant\max\left(e^{-\alpha(u-2)}(u-1),\ (u-1)^{-\beta+1}\right).
$$

In the right hand side of (22), we apply an integration by parts on the first

term and then we use Lemma 3:

$$\begin{aligned}
&\int_2^u (v-1)^{2H-3} e^{-\alpha(u-v)} \mathrm{d}v \\
&= \frac{1}{\alpha}\left((u-1)^{2H-3} - e^{-\alpha(u-2)} + (3-2H)e^{-\alpha u}\int_2^u (v-1)^{2H-4} e^{\alpha v}\mathrm{d}v\right) \\
&\leqslant \frac{1}{\alpha}\left((u-1)^{2H-3} - e^{-\alpha(u-2)} + (3-2H)\max\left(e^{-\alpha(u-2)}(u-1),\ (u-1)^{2H-3}\right)\right) \\
&\leqslant C'_H (u-1)^{2H-3}
\end{aligned}$$

where $C'_H > 0$ is some constant. Finally, by using Lemma 3 also on the second term in (22), we finally get the existence of a constant $C''_H > 0$ such that:

$$\int_2^u e^{-\alpha(u-v)} \left|\varphi_k(v)\right|^2 \mathrm{d}v \leqslant C''_H (u-1)^{2H-3} \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|^2. \tag{23}$$

Now, putting inequality (23) into (21) and taking the square root, we finally get:

$$|Y_u - \tilde{Y}_u| \leqslant e^{-\frac{\alpha}{2}(u-2)}|Y_2 - \tilde{Y}_2| + C_H^{(3)}(u-1)^{H-\frac{3}{2}} \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|. \tag{24}$$

On the one hand, we can note that $e^{-\frac{\alpha}{2}(u-2)} \leqslant C' u^{H-\frac{3}{2}}$ for all $u \geqslant 2$. On the other hand, we have $|Y_2 - \tilde{Y}_2| \leqslant K_H \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}|$.

These two facts combined with (24) conclude the proof.

□

6.2.2 Conditional exponential moments of the martingale increments

Proposition 5. *Assume* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$. *Let* $H \in (0,1)$. *There exists* $K = K(H, L, \alpha, \|\sigma\|) > 0$ *and* $\zeta > 0$ *such that for all* $k \in \mathbb{N}^*$ *and for all* $p \geqslant 2$,

$$\mathbf{E}[|M_k - M_{k-1}|^p | \mathcal{F}_{k-1}] \leqslant C^p \|F\|_{\mathrm{Lip}}^p \psi_{n,k}^p \zeta^{p/2} p\Gamma\left(\frac{p}{2}\right) \quad a.s.$$

where

$$\psi_{n,k} := \left\lfloor \frac{k}{\Delta_n} \right\rfloor - \left\lceil \frac{k-1}{\Delta_n} \right\rceil + \sum_{i=\lceil \frac{k}{\Delta_n} \rceil}^{n} u_i^{H-3/2},$$

$u_i := t_i - k + 1 = i\Delta_n - k + 1$ *and* M_k *is defined by* (18).

Proof. Let $k \in \mathbb{N}^*$ and $p \geq 2$. By combining inequality (20) with the technical lemma 2, we immediately get that there exists $K = K(H, L, \alpha, \|\sigma\|) > 0$ such that:

$$\begin{aligned}
&|M_k - M_{k-1}|^p \\
&\leq K^p \|F\|_{\mathrm{Lip}}^p \psi_{n,k}^p \left(\int_\Omega \sup_{v\in[0,2]} \left|G_v^{(k)}(W - \tilde{w})\right| + \sup_{s\in[0,1]} |W_s^{(k)} - \tilde{w}_s^{(k)}| |\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p
\end{aligned}$$

The end of the proof consists in proving that

$$\mathbf{E}\left[\left(\int_\Omega \sup_{v\in[0,2]}\left|G_v^{(k)}(W-\tilde{w})\right| + \sup_{s\in[0,1]}|W_s^{(k)}-\tilde{w}_s^{(k)}|\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p \;\middle|\; \mathcal{F}_{k-1}\right] \leq \zeta^{p/2}p\Gamma\left(\frac{p}{2}\right).$$

We have

$$\begin{aligned}
&\mathbf{E}\left[|M_k - M_{k-1}|^p|\mathcal{F}_{k-1}\right] \\
&\leq 2^{p-1}\mathbf{E}\left[\left(\int_\Omega \sup_{v\in[0,2]}\left|G_v^{(k)}(W-\tilde{w})\right|\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p\;\middle|\;\mathcal{F}_{k-1}\right] \\
&\qquad + 2^{p-1}\mathbf{E}\left[\left(\int_\Omega \sup_{s\in[0,1]}|W_s^{(k)}-\tilde{w}_s^{(k)}|\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p\;\middle|\;\mathcal{F}_{k-1}\right] \\
&\leq 2^{p-1}\mathbf{E}\left[\left(\int_\Omega \sup_{v\in[0,2]}\left|G_v^{(k)}(W-\tilde{w})\right|\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p\right] \\
&\qquad + 2^{p-1}\mathbf{E}\left[\left(\int_\Omega \sup_{s\in[0,1]}|W_s^{(k)}-\tilde{w}_s^{(k)}|\mathbf{P}_W(\mathrm{d}\tilde{w})\right)^p\right] \qquad (25)
\end{aligned}$$

where the last inequality is obtained by using that $W^{(k)} = (W_{s+k-1} - W_{k-1})_{s\geq 0}$ is independent from $\mathcal{F}_{k-1}$. Now, if we denote by $\mathcal{F}^{(k)}$ the natural filtration associated to $W^{(k)}$, then the right hand side terms of (25) are just expectations of conditional expectations with respect to $\mathcal{F}_1^{(k)}$, so we finally get

$$\begin{aligned}
\mathbf{E}\left[|M_k - M_{k-1}|^p|\mathcal{F}_{k-1}\right] &\leq 2^{p-1}\mathbf{E}\left[\sup_{v\in[0,2]}\left|G_v^{(k)}(W-\tilde{W})\right|^p\right] \\
&\qquad + 2^{p-1}\mathbf{E}\left[\sup_{s\in[0,1]}|W_s^{(k)}-\tilde{W}_s^{(k)}|^p\right].
\end{aligned}$$

Since $W^{(k)}$ and $\tilde{W}^{(k)}$ are i.i.d. and have the same law as $W^{(1)}$, we can replace $W^{(k)} - \tilde{W}^{(k)}$ by $\sqrt{2}W^{(1)}$, which gives

$$\begin{aligned}
&\mathbf{E}\left[|M_k - M_{k-1}|^p|\mathcal{F}_{k-1}\right] \\
&\leq 2^{p-1}\sqrt{2}\mathbf{E}\left[\sup_{v\in[0,2]}\left|G_v^{(1)}(W)\right|^p\right] + 2^{p-1}\sqrt{2}\mathbf{E}\left[\sup_{s\in[0,1]}|W_s^{(1)}|^p\right].
\end{aligned}$$

To conclude the proof, we only have to prove that $\sup_{v\in[0,2]}\left|G_v^{(1)}(W)\right|$ and $\sup_{s\in[0,1]}|W_s^{(1)}|$ are sub-Gaussian. The proof of this result follows the lines of Varvenne (2019) Appendices A and B and we leave it to the patient reader. □

With Lemma 1 in hand, we finally get :

Proposition 6. *Assume* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$*. Let* $H \in (0,1)$*. Let* $k \in \mathbb{N}^*$*. There exists* $K = K(H, L, \alpha, \|\sigma\|) > 0$ *(independent of* k*) such that for all* $\lambda > 0$*,*

$$\mathbf{E}[\exp(\lambda(M_k - M_{k-1})) \mid \mathcal{F}_{k-1}] \leqslant \exp\left(K\lambda^2 \|F\|_{\mathrm{Lip}}^2 \psi_{n,k}^2\right) \quad a.s.$$

where

$$\psi_{n,k} := \left\lfloor \frac{k}{\Delta_n} \right\rfloor - \left\lceil \frac{k-1}{\Delta_n} \right\rceil + \sum_{i=\lceil \frac{k}{\Delta_n} \rceil}^{n} u_i^{H-3/2},$$

$u_i := t_i - k + 1 = i\Delta_n - k + 1$ *and* M_k *is defined by* (18)*.*

6.2.3 Proof of Proposition 4

Proof. The inequality on the conditional Laplace transform of $\sum_{k=1}^{\lceil t_n \rceil} M_k - M_{k-1}$ easily follows from Proposition 6. We thus have to prove the bound on $\sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2$.
Let us begin by the estimation of $\psi_{n,k}$ for all $k \in \{1, \dots, \lceil t_n \rceil\}$. First, we easily get that

$$\left\lfloor \frac{k}{\Delta_n} \right\rfloor - \left\lceil \frac{k-1}{\Delta_n} \right\rceil \leq \Delta_n^{-1}. \tag{26}$$

Secondly, let us consider the second part of $\psi_{n,k}$, we have

$$\begin{aligned}
&\sum_{i=\lceil \frac{k}{\Delta_n} \rceil}^{n} (i\Delta_n - k + 1)^{H-3/2} \\
&\leq 1 + \int_{\lceil k\Delta_n^{-1} \rceil}^{n} (t\Delta_n - k + 1)^{H-3/2} \mathrm{d}t \\
&= 1 + \Delta_n^{H-3/2} \left[\frac{(t - (k-1)\Delta_n^{-1})^{H-1/2}}{H - 1/2} \right]_{\lceil k\Delta_n^{-1} \rceil}^{n} \\
&= 1 + \frac{\Delta_n^{H-3/2}}{H - 1/2} \left[(n - (k-1)\Delta_n^{-1})^{H-1/2} - \left(\lceil k\Delta_n^{-1} \rceil - (k-1)\Delta_n^{-1} \right)^{H-1/2} \right] \\
&\leq 1 + \Delta_n^{H-3/2} \times \begin{cases} \frac{1}{H-1/2} \left(n - (k-1)\Delta_n^{-1} \right)^{H-1/2} & \text{if} \quad H > 1/2 \\ \frac{1}{1/2-H} \left(\lceil k\Delta_n^{-1} \rceil - (k-1)\Delta_n^{-1} \right)^{H-1/2} & \text{if} \quad H < 1/2 \end{cases} \\
&\leq 1 + \Delta_n^{H-3/2} \times \begin{cases} \frac{1}{H-1/2} \left(n - (k-1)\Delta_n^{-1} \right)^{H-1/2} & \text{if} \quad H > 1/2 \\ \frac{1}{1/2-H} \left(1 + \Delta_n^{-1} \right)^{H-1/2} & \text{if} \quad H < 1/2. \end{cases}
\end{aligned} \tag{27}$$

From (26) and (27), we thus deduce that there exists $c_1 = c_1(H) > 0$ such that

$$\psi_{n,k} \leq \Delta_n^{-1} + 1 + c_1 \begin{cases} \Delta_n^{-1} \left(t_n - (k-1) \right)^{H-1/2} & \text{if} \quad H > 1/2 \\ \Delta_n^{-1} & \text{if} \quad H < 1/2. \end{cases}$$

We can now move on the estimation of $\sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2$. From the inequality above, it follows that there exists $c_2 = c_2(H) > 0$ such that

$$\sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2 \leq c_2(\Delta_n^{-2} + 1)\lceil t_n \rceil + c_2 \left\{ \begin{array}{lll} \Delta_n^{-2} \sum_{k=1}^{\lceil t_n \rceil} \left(t_n - (k-1)\right)^{2H-1} & \text{if} & H > 1/2 \\ \Delta_n^{-2} \lceil t_n \rceil & \text{if} & H < 1/2. \end{array} \right. \quad (28)$$

It remains to estimate $\sum_{k=1}^{\lceil t_n \rceil} \left(t_n - (k-1)\right)^{2H-1}$ when $H > 1/2$. It is readily checked that

$$\sum_{k=1}^{\lceil t_n \rceil} \left(t_n - (k-1)\right)^{2H-1} \leq \sum_{k=1}^{\lceil t_n \rceil} \left(\lceil t_n \rceil - (k-1)\right)^{2H-1} = \sum_{k=1}^{\lceil t_n \rceil} k^{2H-1} \leq \lceil t_n \rceil^{2H}. \quad (29)$$

Finally, from (28) and (29), we get the existence of $c_3 = c_3(H) > 0$ and $c_4 = c_4(H) > 0$ such that

$$\begin{aligned} \sum_{k=1}^{\lceil t_n \rceil} \psi_{n,k}^2 &\leq c_3 \left\{ \begin{array}{lll} \Delta_n^{-2}\lceil t_n \rceil + \Delta_n^{-2}\lceil t_n \rceil^{2H} & \text{if} & H > 1/2 \\ \Delta_n^{-2}\lceil t_n \rceil & \text{if} & H < 1/2 \end{array} \right. \\ &\leq c_4 \left\{ \begin{array}{lll} \Delta_n^{-2}\lceil t_n \rceil^{2H} & \text{if} & H > 1/2 \\ \Delta_n^{-2}\lceil t_n \rceil & \text{if} & H < 1/2. \end{array} \right. \end{aligned}$$

This concludes the proof since $t_n = n\Delta_n$. □

6.3 Part 2: $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$

We now turn to the bound of $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$. First, let us remark that

$$\mathbf{E}[\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]] = 0$$

so that we can use again Lemma 1 in order to deduce exponential bounds. The related result is stated in Proposition 7.

We introduce notations related to the conditioning with respect to $\mathcal{F}_0$. Let $(W_t)_{t\in\mathbb{R}}$ denote the two-sided Brownian Motion induced by Mandelbrot-Van Ness representation (see (3)) and set $W^- = (W_t)_{t\leq 0}$. For $\varepsilon \in (0, 1/2)$ and $\varepsilon' > 0$, set

$$\Omega_- := \Big\{ w : (-\infty, 0] \to \mathbb{R}^d;\, w(0) = 0\,,\, \lim_{t\to-\infty} \frac{w(t)}{|t|^{\frac{1}{2}+\varepsilon'}} = 0 \text{ and } w \text{ is } (\frac{1}{2} - \varepsilon)\text{-Hölder continuous on compact intervals} \Big\}.$$

Owing to some classical properties on the Wiener process, this subspace is of Wiener measure 1 for any fixed $\varepsilon \in (0, 1/2)$ and $\varepsilon' > 0$. In other words,

$\mathbf{P}_{W^-}(\Omega_-) = 1$. Then, for any $w \in \Omega_-$,

$$\mathcal{L}((B_t^H)_{t\geq 0}|W^- = w) = \mathcal{L}((Z_t + D_t(w))_{t\geq 0})$$

where $Z_0 = D_0(w) = 0$ and for all $t > 0$,

$$Z_t = c_H \int_0^t (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} dW_s$$

and

$$D_t(w) = \alpha_H \int_{-\infty}^0 (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} dw_s.$$

$(Z_t)_{t\geq 0}$ and $(D_t(w))_{t\geq 0}$ are continuous processes on $[0, +\infty)$ (see Lemma 4 below) and for any $w \in \Omega_-$, the (additive) SDE

$$dY_t = b(Y_t)dt + \sigma(dZ_t + dD_t^w) \tag{30}$$

has a unique solution denoted by $(X_t^{x,w})_{t\geq 0}$.

Since F is Lipschitz continuous with respect to d_n, we can also remark that

$$|\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]| \leq \|F\|_{\mathrm{Lip}} \sum_{k=1}^n \int_{\mathbb{R}^d \times \Omega_-} |X_{t_k}^{X_0, W^-} - X_{t_k}^{y,w}| \nu(dy, dw), \tag{31}$$

where $\nu = \mathcal{L}(X_0, W^-)$ is an initial condition for the dynamical system which is such that the process is stationary. This involves that we will use bounds on $|X_t^{x,W^-} - X_t^{y,w}|$ to deduce the result for $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$. To this end, we first state a technical result about $D(w)$:

Lemma 4. *Let $w \in \Omega^-$ with $\varepsilon \in (0, H)$ and $\varepsilon' \in (0, 1-H)$. Then, $(D_t(w))_{t\geq 0}$ is continuous on $[0, \infty)$ and differentiable on $(0, +\infty)$. Furthermore, for any $\delta_1 \in (0, 1-H-\varepsilon']$ and $\delta_2 \in\leq H-1-\varepsilon$, there exist some positive constants C_{δ_1} and C_{δ_2} such that for any $t > 0$,*

$$\begin{aligned} &|D_t'(w)| \leq \Phi_{\varepsilon,\varepsilon'}(w) \left(C_{\delta_1} t^{-\delta_1} 1_{t>1} + C_{\delta_2} t^{-\delta_2} 1_{t\in(0,1]}\right) \\ &\text{with } \Phi_{\varepsilon,\varepsilon'}(w) = \sup_{r\in(0,1]} \frac{|w(-r)|}{r^{\frac{1}{2}-\varepsilon}} + \sup_{r>1} \frac{|w(-r)|}{r^{\frac{1}{2}+\varepsilon'}}. \end{aligned} \tag{32}$$

Proof. Let $w \in \Omega^-$. By an integration by parts, one checks that the process $(D_t(w))$ is well-defined for any $t > 0$ and admits the following alternative representation:

$$D_t(w) = \alpha_H \left(H - \frac{1}{2}\right) \int_{-\infty}^0 \left((t-r)^{H-\frac{3}{2}} - (-r)^{H-\frac{3}{2}}\right) w(r)\, dr \quad \text{if } t \in (0,1]$$

It easily follows that $(D_t(w))_{t>0}$ is smooth on $(0, +\infty]$ and that for all $t > 0$,

$$D_t'(w) = \alpha_H \left(H - \frac{1}{2}\right)\left(H - \frac{3}{2}\right) \int_{-\infty}^0 (t-r)^{H-\frac{5}{2}} w(r)\, dr.$$

On the one hand, for any $\bar{\varepsilon} \in [\varepsilon', 1-H)$,

$$|\int_{-\infty}^{-1} (t-r)^{H-\frac{5}{2}} w(r)\, dr| \leq \sup_{r>1} \frac{|w(-r)|}{r^{\frac{1}{2}+\bar{\varepsilon}}} \int_{-\infty}^{-1} (t-r)^{H-\frac{5}{2}} r^{\frac{1}{2}+\bar{\varepsilon}} dr$$
$$\leq \Phi_{\varepsilon,\varepsilon'}(w) \int_{-\infty}^{-1} (t-r)^{H-2+\bar{\varepsilon}} dr \leq \Phi_{\varepsilon,\varepsilon'}(w)(1+t)^{H-2+\bar{\varepsilon}}.$$

On the other hand, for any $\tilde{\varepsilon} \geq \varepsilon$,

$$\int_{-1}^{0} (t-r)^{H-\frac{5}{2}} w(r)\, dr \leq \sup_{r\in(0,1]} \frac{|w(-r)|}{r^{\frac{1}{2}-\tilde{\varepsilon}}} \int_{-1}^{0} (t-r)^{H-2-\tilde{\varepsilon}} dr \leq \Phi_{\varepsilon,\varepsilon'}(w) t^{H-1-\tilde{\varepsilon}}.$$

Inequality (32) easily follows from what precedes. In particular, since $\varepsilon \in (0,H)$, $t \mapsto D'_t(w)$ is integrable near 0 and hence, $t \mapsto D_t(w)$ is continuous on $\mathbb{R}_+$. □

In view of (31), we now provide a control of the evolution of two paths of the fractional SDE (30) starting from initial conditions $(x, \tilde{w})$ and (y, w).

Lemma 5. *Suppose that assumptions* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$ *are in force. Let* w *and* $\tilde{w}$ *belong to* Ω^- *with* $\varepsilon \in (0,H)$ *and* $\varepsilon' \in (0, 1-H)$*. Then, there exist some positive constants* $C_1 = C_1(H, L, \alpha, \|\sigma\|)$ *and* $C_2 = C_2(H, L, \alpha, \|\sigma\|)$ *such that for every* $t \geq 0$*,*

$$|X_t^{x,\widetilde{w}} - X_t^{y,w}|^2 \leq C_1 e^{-\alpha t}|x-y|^2 + C_2 \Phi_{\varepsilon,\varepsilon'}(\tilde{w}-w)(t \vee 1)^{2H-2+\varepsilon'}$$

where $\Phi_{\varepsilon,\varepsilon'}$ *is defined by* (32)*.*

Proof. For two paths w and $\widetilde{w}$ in Ω^-,

$$B_t^{\widetilde{w}} - B_t^w = D_t(\widetilde{w} - w).$$

Thus, for any $t_0 > 0$,

$$X_t^{x,\widetilde{w}} - X_t^{y,w} = X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w} + \int_{t_0}^{t} b(X_s^{x,\widetilde{w}}) - b(X_s^{x,w^-}) ds + \sigma D_t(\widetilde{w} - w).$$

As a consequence,

$$e^{\alpha t}|X_t^{x,\widetilde{w}} - X_t^{y,w}|^2 = e^{\alpha t_0}|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}|^2$$
$$+ \int_0^t 2e^{\alpha s}\langle X_s^{x,\widetilde{w}} - X_s^{x,w}), \sigma D'_s(\widetilde{w} - w)\rangle ds$$
$$+ \int_0^t e^{\alpha s}\left(\alpha|X_s^{x,\widetilde{w}} - X_s^{y,w}|^2 + 2\langle X_s^{x,\widetilde{w}} - X_s^{x,w}, b(X_s^{x,\widetilde{w}}) - b(X_s^{x,w})\rangle\right) ds.$$

By Assumption $(\mathbf{H_1})$,

$$\langle X_s^{x,\widetilde{w}} - X_s^{x,w}, b(X_s^{x,\widetilde{w}}) - b(X_s^{x,w})\rangle \leq -\alpha|X_s^{x,\widetilde{w}} - X_s^{y,w}|^2$$

whereas by the elementary inequality $|uv| \leq (\varepsilon/2)|u|^2 + 1/(2\varepsilon)|v|^2$ applied with $\varepsilon = \alpha$,

$$2\langle X_s^{x,\widetilde{w}} - X_s^{x,w}), \sigma D_s'(\widetilde{w} - w)\rangle \leq \alpha |X_s^{x,\widetilde{w}} - X_s^{y,w}|^2 + \frac{\|\sigma\|^2}{\alpha} |D_s'(\widetilde{w} - w)|^2.$$

Thus, for any $\varepsilon' > 0$, we have for any $t \geq t_0$,

$$e^{\alpha t}|X_t^{x,\widetilde{w}} - X_t^{y,w}|^2 \leq e^{\alpha t_0}|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}|^2 + \frac{\|\sigma\|^2}{\alpha}\int_{t_0}^t e^{\alpha s}|D_s'(\widetilde{w} - w)|^2 ds.$$

By Lemma 4, we deduce that a positive constant C exists such that:

$$\int_{t_0}^t e^{\alpha s}|D_s'(\widetilde{w} - w)|^2 ds \leq C\Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w)\left(1 + \int_1^{t\vee 1} e^{\alpha s} s^{2H-2+2\varepsilon'} ds\right).$$

By an integration by parts, it follows that

$$\begin{aligned}\frac{e^{-\alpha t}}{\alpha}\int_{t_0}^t e^{\alpha s}|D_s'(\widetilde{w} - w)|^2 ds &\leq \frac{C\Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w)}{\alpha}\left(e^{-\alpha t} + \int_1^{t\vee 1} e^{\alpha(s-t)} s^{2H-2+2\varepsilon'} ds\right) \\ &\leq C\Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w))(t \vee 1)^{2H-2+2\varepsilon'}.\end{aligned}$$

Thus,

$$|X_t^{x,\widetilde{w}} - X_t^{y,w}|^2 \leq e^{-\alpha(t-t_0)}|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}|^2 + C\|\sigma\|^2\Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w)(t \vee 1)^{2H-2+2\varepsilon'}.$$

Let us finally control $|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}|^2$. Since b is L-Lipschitz continuous, for every $t \geq 0$,

$$|X_t^{x,\widetilde{w}} - X_t^{y,w}| \leq |x - y| + L\int_0^t |X_s^{x,\widetilde{w}} - X_s^{y,w}| ds + \|\sigma\| \sup_{s\in(0,t]} |D_s(\widetilde{w} - w)|,$$

and the Gronwall Lemma yields:

$$|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}| \leq \left(|x - y| + \|\sigma\| \sup_{s\in(0,t_0]} |D_s(\widetilde{w} - w)|\right) e^{Lt_0}.$$

Now, assume that $t_0 \in (0, 1]$. By Lemma 4,

$$\sup_{s\in(0,t_0]} |D_s(\widetilde{w} - w)| \leq \int_0^{t_0} |D_s'(\widetilde{w} - w)| ds \leq C\Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w) t_0^{H-\varepsilon}.$$

As a consequence,

$$|X_{t_0}^{x,\widetilde{w}} - X_{t_0}^{y,w}|^2 \leq C\left(|x - y|^2 + \Phi_{\varepsilon,\varepsilon'}(\tilde{w} - w)\right).$$

The result follows. □

Before stating Proposition 7 (which provides the exponential bound for $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$), we finally obtain bounds on the moments of $\Phi_{\varepsilon,\varepsilon'}(W^-)$ and of the invariant distribution.

Lemma 6. *Let $\varepsilon, \varepsilon'$ be some positive numbers. Then, there exists $C_{\varepsilon'} > 0$ such that for every $\lambda > 0$,*

$$\mathbf{E}[|\Phi_{\varepsilon,\varepsilon'}(W^-)|^p] \leq C_{\varepsilon'}^p \mathbf{E}[|\mathcal{Z}|^p]$$

where $\mathcal{Z}$ has $\mathcal{N}(0,1)$-distribution.

Proof. It is enough to consider the one-dimensional case and by a symmetry argument, it is certainly equivalent to prove the result for a Brownian motion on $\mathbb{R}_+$. Furthermore, using that for a Brownian motion W on $\mathbb{R}_+$, $t \mapsto tW_{\frac{1}{t}}$ (with initial value equal to 0) is also a Brownian motion, we deduce that we have only to prove that for any $\varepsilon' > 0$,

$$\mathbf{E}\left[\sup_{t\geq 1}\left|\frac{W_t}{t^{\frac{1}{2}+\varepsilon'}}\right|^p\right] \leq C_{\varepsilon'}^p \mathbf{E}[|\mathcal{Z}|^p].$$

By the Itô formula,

$$\frac{W_t}{t^{\frac{1}{2}+\varepsilon'}} = W_1 + \int_1^t \frac{1}{s^{\frac{1}{2}+\varepsilon'}} dW_s - (\frac{1}{2}+\varepsilon')\int_1^t \frac{W_s}{s^{\frac{3}{2}+\varepsilon'}} ds. \tag{33}$$

For the first right-hand side term, there is nothing to prove. For the second one, we remark that it is a Gaussian process and it follows that a Brownian Motion $\tilde{W}$ exists such that

$$\sup_{t\geq 1}\left|\int_1^t \frac{1}{s^{\frac{1}{2}+\varepsilon'}} dW_s\right| \stackrel{(d)}{=} \sup_{t\in[0,\sigma_\infty)} |\tilde{W}_t| \quad \text{where} \quad \sigma_\infty = \int_1^{+\infty} \frac{1}{s^{1+2\varepsilon'}} ds < +\infty,$$

and "$\stackrel{(d)}{=}$" stands for the equality in distribution. But

$$\sup_{t\in[0,\sigma_\infty)} |\tilde{W}_t| \leq \sup_{t\in[0,\sigma_\infty)} \tilde{W}_t - \sup_{t\in[0,\sigma_\infty)} (-\tilde{W}_t)$$

and hence

$$\mathbf{E}[\sup_{t\in[0,\sigma_\infty)} |\tilde{W}_t|^p] \leq 2^p \mathbf{E}[|\sup_{t\in[0,\sigma_\infty)} \tilde{W}_t|^p] \leq (2\sqrt{\sigma_\infty})^p \mathbf{E}[|\mathcal{Z}|^p]$$

since $\sup_{t\in[0,\sigma_\infty)} \tilde{W}_t$ has the same distribution as $\sqrt{\sigma_\infty}|\mathcal{Z}|$ where $\mathcal{Z}$ has $\mathcal{N}(0,1)$-distribution. Let us now consider the last term of (33). We have

$$\sup_{t\geq 1}\left|\int_1^t \frac{W_s}{s^{\frac{3}{2}+\varepsilon'}} ds\right| \leq \int_1^{+\infty} \frac{|W_s|}{s^{\frac{3}{2}+\varepsilon'}} ds.$$

Hence, by the Jensen inequality applied with the probability measure $\mu_{\varepsilon'}(ds) = Cs^{-(1+\varepsilon')}ds$, we get:

$$\mathbf{E}\left[\sup_{t\geq 1}\left|\int_1^t \frac{W_s}{s^{\frac{3}{2}+\varepsilon'}}ds\right|^p\right] \leq C\int_1^{+\infty}\mathbf{E}\left|\frac{W_s}{\sqrt{s}}\right|^p \frac{1}{s^{1+\varepsilon'}}ds.$$

Thus, by the scaling property, it follows that

$$\mathbf{E}\left[\sup_{t\geq 1}\left|\int_1^t \frac{W_s}{s^{\frac{3}{2}+\varepsilon'}}ds\right|^p\right] \leq C\mathbf{E}[|\mathcal{Z}|^p]$$

where $\mathcal{Z}$ has $\mathcal{N}(0,1)$-distribution. □

Lemma 7. *Assume* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$ *and let* $\bar{\nu}$ *denote the marginal invariant distribution. Then, there exists* $C = C(H, L, \alpha, |b(0_{\mathbb{R}^d})|, \|\sigma\|) > 0$ *such that for any* $p > 1$,

$$\int_{\mathbb{R}^d}|y|^p\bar{\nu}(dy) \leq C^p\mathbf{E}[|\mathcal{Z}|^p]$$

where $\mathcal{Z}$ *denotes a random variable with* $\mathcal{N}(0,1)$*-distribution.Furthermore,* $(L, \alpha, b_0, s) \mapsto C(H, L, \alpha, b_0, s)$ *is bounded on every compact set of* $\mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+ \times \mathbb{R}_+$.

Remark 7. *The dependency of* C *with respect to* $H, L, \alpha, b(0_{\mathbb{R}^d})$ *and* σ *is explicit and is given in the following proof.*

Proof. The proof uses some arguments of (Hairer, 2005, Proposition 3.12) by controlling the distance between the solution to the SDE with the one of a Ornstein-Uhlenbeck process for which the announced property holds. For the sake of completeness, let us give some details. Let $(U_t)_{t\geq 0}$ denote a solution to $dU_t = -U_t dt + \sigma dB_t^H$ and $(X_t)_{t\geq 0}$ a solution to (1). Assume that U and X are built with the same fBm and start from the same starting point x. Then,

$$X_t - U_t = \int_0^t b(X_s) - U_s ds$$

so that

$$|X_t - U_t|^2 = 2\int_0^t \langle b(X_s) - U_s, X_s - U_s\rangle ds.$$

For any x and $u \in \mathbb{R}^d$,

$$\langle b(x) - u, x - u\rangle = \langle b(x) - b(u), x - u\rangle + \langle b(u) - u, x - u\rangle.$$

Now, by $(\mathbf{H_1})$ and $(\mathbf{H_2})$ (which implies that $|b(u)| \leq (|b(0_{\mathbb{R}^d})| \vee L)(1 + |u|)$), we get

$$\langle b(x) - u, x - u\rangle \leq -\frac{\alpha}{2}|x - u|^2 + \frac{(|b(0_{\mathbb{R}^d})| \vee L + 1)^2}{\alpha}(1 + |u|^2).$$

Thus, with similar arguments as in the proof of Lemma 5 (based on Gronwall-type arguments), we deduce that for any $t \geq 0$,

$$|X_t - U_t|^2 \leq \frac{2(|b(0_{\mathbb{R}^d})| \vee L + 1)^2}{\alpha} \int_0^t e^{\alpha(s-t)}(1 + |U_s|^2)ds.$$

Thus, denoting by $\|.\|_p$, the L^p-norm we deduce from Jensen inequality that

$$\|X_t^x\|_p^p \leq 2^{p-1} \left(\frac{4(|b(0_{\mathbb{R}^d})| \vee L + 1)^2}{\alpha} \times \left(\frac{1}{\alpha} \vee 1 \right) \right)^{p/2} \int_0^t e^{\alpha(s-t)}(1 + \|U_s^x\|_p^p)ds + 2^{p-1}\|U_t^x\|_p^p.$$

Denote by $\bar{\nu}$ and π_σ the (marginal) invariant distributions of X and U. Owing to uniform integrability arguments and to the convergence in distribution of $(X_t)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ towards $\bar{\nu}$ and π_σ, we get:

$$\int_{\mathbb{R}^d} |y|^p \bar{\nu}(dy) = \lim_{t\to+\infty} \|X_t^x\|_p^p \leq \bar{C}^p \lim_{t\to+\infty} \|U_t^x\|_p^p = \bar{C}^p \int_{\mathbb{R}^d} |y|^p \pi_\sigma(dy)$$

where

$$\bar{C} := 2 \left(\frac{4(|b(0_{\mathbb{R}^d})| \vee L + 1)^2}{\alpha} \times \left(\frac{1}{\alpha} \vee 1 \right) \right)^{1/2} + 2.$$

Finally, let us recall that by a standard integration by parts,

$$U_t = xe^{-t} + \sigma \int_0^t e^{s-t} dB_s^H$$

and it follows that $\pi_\sigma = \pi_{I_d} \circ \varphi_\sigma^{-1}$ where $\varphi_\sigma = \sigma x$. Thus,

$$\int_{\mathbb{R}^d} |y|^p \pi_\sigma(dy) \leq \|\sigma\|^p \int |y|^p \pi_{I_d}(dy).$$

But by (Hairer, 2005, Proposition 3.12), π_{Id} has Gaussian distribution $\mathcal{N}(0_{\mathbb{R}^d}, c_0 I_d)$ where $c_0 \leq \Gamma(2H+1)$, so that the result follows with $C = \|\sigma\|(d\Gamma(2H+1))^{\frac{1}{2}}\bar{C}$ (which has the local boundedness property announced in the lemma). □

We are now in position to provide an exponential bound for $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$:

Proposition 7. *Assume* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$. *Suppose that* $n\Delta_n \geq 1$. *Then, for any* $\varepsilon' \in (0, 1-H)$, *a constant* $C = C(H, L, \alpha, b(0_{\mathbb{R}^d}), \sigma, \varepsilon')$ *exists such that*

$$\mathbf{E}[\exp(\lambda(\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]))] \leq \exp\left(\lambda^2 u_0^{(n)}\right),$$

where

$$u_0^{(n)} = C\|F\|_{\mathrm{Lip}}^2 \frac{(n\Delta_n)^{H+\varepsilon'}}{\Delta_n}.$$

In particular, with $\varepsilon' = \frac{1}{2}$ *and* $\varepsilon' = \frac{1-H}{2}$ *when* $H < 1/2$ *and* $H > 1/2$ *respectively,*

$$u_0^{(n)} \leq C\|F\|_{\mathrm{Lip}}^2 \begin{cases} n\Delta_n^{-1} & \text{if } H < 1/2, \\ n^{H+\frac{1-H}{2}} \Delta_n^{\frac{H-1}{2}} & \text{if } H > 1/2. \end{cases}$$

Remark 8. *Following carefully the constants involved in the proof below (induced by the previous lemmas), one checks that for every $H \in (0,1)$ and $\varepsilon' \in (0, 1-H)$, $(L, \alpha, b_0, s) \mapsto C(H, L, \alpha, b_0, s, \varepsilon')$ is bounded on every compact set of $\mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}_+ \times \mathbb{R}_+$. Thus, since the proof of Theorem 1 is obtained as a combination of Propositions 4 and 7, this property combined with Remark 6 implies that the constant $\mathfrak{C}$ of Theorem 1 has the local boundedness property announced in Remark 2.*

Remark 9. *The above bound easily involves that the contribution of $\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]$ is always less constraining (up to a multiplicative constant) than the one obtained in Proposition 4.*

Proof. By (31) and Lemma 5, for any $p > 1$,

$$|\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]|^p \leq (2C_1\|F\|_{\mathrm{Lip}})^p \left(\int_{\mathbb{R}^d} |X_0 - y| \bar{\nu}(dy) \sum_{k=1}^n e^{-\frac{\alpha}{2} t_k} \right)^p + (2C_2\|F\|_{\mathrm{Lip}})^p \left(\int \left| \Phi_{\varepsilon,\varepsilon'}(W^- - w) \right|^p \mathbf{P}_{W^-}(dw) \right) \left(\sum_{k=1}^n (t_k \vee 1)^{H-1+\varepsilon'} \right)^p$$

where $\bar{\nu}$ stands for the "marginal" invariant distribution, *i.e.* the projection of ν on the first coordinate and $\tilde{w}$ denotes the p Let us consider the two right-hand side terms separately. On the one hand, using Jensen inequality, we get

$$\left(\int_{\mathbb{R}^d} |X_0 - y| \bar{\nu}(dy) \sum_{k=1}^n e^{-\frac{\alpha}{2} t_k} \right)^p \leq \int_{\mathbb{R}^d} |X_0 - y|^p \bar{\nu}(dy) \left(\frac{2}{\alpha \Delta_n} \right)^p.$$

Thus, using that $|X_0 - y|^p \leq 2^p \left(|X_0|^p + |y|^p \right)$ and Lemma 7, we get

$$\mathbf{E} \left(\int_{\mathbb{R}^d} |X_0 - y| \bar{\nu}(dy) \sum_{k=1}^n e^{-\frac{\alpha}{2} t_k} \right)^p \leq \left(\frac{4}{\alpha \Delta_n} \right)^p C^p \mathbf{E}[|\mathcal{Z}|^p]. \tag{34}$$

On the other hand, since $H - 1 + \varepsilon' > -1$,

$$\sum_{k=1}^n (t_k \vee 1)^{H-1+\varepsilon'} \leq C \left(\frac{1}{\Delta_n} + n^{H+\varepsilon'} {\Delta_n}^{H-1+\varepsilon'} \right) \leq C \frac{(n\Delta_n)^{H+\varepsilon}}{\Delta_n}.$$

Furthermore,

$$\mathbf{E} \int_{\Omega_-} \left| \Phi_{\varepsilon,\varepsilon'}(W^- - w) \right|^p \mathbf{P}_{W^-}(dw) \leq 2^p \mathbf{E} |\Phi_{\varepsilon,\varepsilon'}(W^-)|^p.$$

Thus, by Lemma 6, it follows that

$$\mathbf{E} \int_{\Omega_-} \left| \Phi_{\varepsilon,\varepsilon'}(W^- - w) \right|^p \mathbf{P}_{W^-}(dw) \left(\sum_{k=1}^n (t_k \vee 1)^{H-1+\varepsilon'} \right)^p \leq \left(C \frac{(n\Delta_n)^{H+\varepsilon}}{\Delta_n} \right)^p \mathbf{E}[|\mathcal{Z}|^p]. \tag{35}$$

Combining (34), (35) and the fact that $\mathbf{E}|\mathcal{Z}|^p \leq C^p \Gamma\left(\frac{p+1}{2}\right) \leq C^p p \Gamma\left(\frac{p}{2}\right)$, we get: there exists a constant $C > 0$ such that for all $p \geq 2$,

$$\mathbf{E}|\mathbf{E}[F_X \mid \mathcal{F}_0] - \mathbf{E}[F_X]|^p \leq \left(C\|F\|_{\mathrm{Lip}} \frac{(n\Delta_n)^{H+\varepsilon}}{\Delta_n}\right)^p p\Gamma\left(\frac{p}{2}\right).$$

To conclude, we apply Lemma 1. □

7 Proofs of Statistical properties

7.1 Proof of Proposition 2

Step 1. Below we denote $x_0 = (t_1, \ldots, t_d)$ and we define for $h \in \mathcal{H}$ and η such that $\eta_i \in \{0, h_i\}$:

$$v_i(u) = (t_1 - \eta_1 u_1, \ldots, t_{i-1} - \eta_{i-1} u_{i-1}, \quad t_i \quad , t_{i+1} - h_{i+1} u_{i+1}, \ldots, t_d - h_d u_d).$$

We can write:

$$\begin{aligned} f(x_0 - h \cdot u) - f(x_0 - \eta \cdot u) &= \sum_{i=1}^{d} f(v_i(u) - h_i u_i e_i) - f(v_i(u) - \eta_i u_i e_i) \\ &= \sum_{i \in I} f(v_i(u) - h_i u_i e_i) - f(v_i(u)), \end{aligned}$$

where $I = \{i = 1, \ldots, d : \eta_i = 0\}$. Now fix $i \in I$.

If $\lfloor s_i \rfloor = 0$ then we obtain:

$$|f(v_i(u) - h_i u_i e_i) - f(v_i(u))| \leq L_i |h_i u_i|^{s_i}$$

which leads to (since $f \in \Sigma_d(\boldsymbol{s}, \mathbf{L})$)

$$\left| \int_{\mathbb{R}} K(u_i) \left(f(v_i(u) - h_i u_i e_i) - f(v_i(u))\right) du_i \right| \leq L_i h_i^{s_i} \int_{\mathbb{R}} \left|u_i^{s_i} K(u_i)\right| du_i.$$

Otherwise, using a Taylor expansion of the function $z \in \mathbb{R} \mapsto f(v_i(u) + z e_i)$ around 0, we obtain:

$$\begin{aligned} f(v_i(u) - h_i u_i e_i) - f(v_i(u)) = &\sum_{k=1}^{\lfloor s_i \rfloor} D_i^k f(v_i(u)) \frac{(-h_i u_i)^k}{k!} \\ &+ \frac{(-h_i u_i)^{\lfloor s_i \rfloor}}{\lfloor s_i \rfloor !} \left[\frac{\partial^{\lfloor s_i \rfloor} f}{\partial x_i^{\lfloor s_i \rfloor}}(v_i(u) - \tau h_i u_i) - \frac{\partial^{\lfloor s_i \rfloor} f}{\partial x_i^{\lfloor s_i \rfloor}}(v_i(u)) \right], \end{aligned}$$

where $\tau \in (0, 1)$. This implies that, using that K is a kernel of order larger than $\lfloor s_i \rfloor$ combined with the fact that $v_i(u)$ does not depend on u_i,

$$\left| \int_{\mathbb{R}} K(u_i) \left(f(v_i(u) - h_i u_i e_i) - f(v_i(u))\right) du_i \right| \leq \frac{L_i h_i^{s_i}}{\lfloor s_i \rfloor !} \int_{\mathbb{R}} \left|u_i^{s_i} K(u_i)\right| du_i.$$

Combining the above results we obtain:

$$\left|\int_{\mathbb{R}^d}\prod_{i=1}^d K(u_i)\left(f(x_0-h\cdot u)-f(x_0-\eta\cdot u)\right)du\right|\le\sum_{i\in I}\frac{L_ih_i^{s_i}}{\lfloor s_i\rfloor!}\int_{\mathbb{R}}\left|u_i^{s_i}K(u_i)\right|du_i.$$

Step 2. Let $h,\mathfrak{h}\in\mathcal{H}$. Taking $\eta=0$ in step 1 we obtain (13) Taking $\eta_i=0$ if $h_i=\mathfrak{h}_i\vee h_i$ and $\eta_i=\mathfrak{h}_i=h_i\vee\mathfrak{h}_i$ otherwise, we obtain:

$$\begin{aligned}\left|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0)-\mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right|&\le\left|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0)-\mathbf{E}\hat{f}_{\eta}(x_0)\right|+\left|\mathbf{E}\hat{f}_{\eta}(x_0)-\mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right|\\&\le 2\sum_{i\in I}\frac{L_ih_i^{s_i}}{\lfloor s_i\rfloor!}\int_{\mathbb{R}}\left|u_i^{s_i}K(u_i)\right|du_i\\&\le 2\sum_{i=1}^d\frac{L_ih_i^{s_i}}{\lfloor s_i\rfloor!}\int_{\mathbb{R}}\left|u_i^{s_i}K(u_i)\right|du_i.\end{aligned}$$

This implies (14).

7.2 Proof of Proposition 3

We have

$$\mathbf{E}\left|\hat{f}_h(x_0)-\mathbf{E}\hat{f}_h(x_0)\right|^p=\int_0^\infty\mathbf{P}\left(\left|\hat{f}_h(x_0)-\mathbf{E}\hat{f}_h(x_0)\right|^p>t\right)dt.$$

We now use Corollary 1 with the functional $g(u)=K_h(x_0-u)$. We obtain that

$$\begin{aligned}\mathbf{E}\left|\hat{f}_h(x_0)-\mathbf{E}\hat{f}_h(x_0)\right|^p&=p\int_0^\infty u^{p-1}\mathbf{P}\left(\left|\hat{f}_h(x_0)-\mathbf{E}\hat{f}_h(x_0)\right|>u\right)du\\&\le 2p\int_0^\infty u^{p-1}\exp\left(-\frac{u^2(n\Delta)^{\beta_H}}{4\mathfrak{C}\|g\|_{\mathrm{Lip}}^2}\right)du\\&=2p\left(\frac{4\mathfrak{C}\|g\|_{\mathrm{Lip}}^2}{(n\Delta_n)^{\beta_H}}\right)^{p/2}\int_0^{+\infty}u^{p-1}\exp\left(-u^2\right)du\\&=p\Gamma\left(\frac{p+1}{2}\right)\left(\frac{4\mathfrak{C}\|g\|_{\mathrm{Lip}}^2}{(n\Delta_n)^{\beta_H}}\right)^{p/2}\\&=p\Gamma\left(\frac{p+1}{2}\right)\mathfrak{C}^{p/2}\left(\frac{4\|g\|_{\mathrm{Lip}}^2}{(n\Delta_n)^{\beta_H}}\right)^{p/2}\end{aligned}$$

Since $\|g\|_{\mathrm{Lip}}\le\sqrt{d}\mathfrak{L}_KV_h^{-1}$, we obtain (15).

7.3 Proof of oracle inequality

We split the proof of Theorem 3 into several steps.

Step 1. Let $h \in \mathcal{H}$ be an arbitrary bandwidth. Using triangular inequality we have:

$$|\hat{f}(x_0) - f(x_0)| \leq |\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| + |\hat{f}_{h\vee\hat{h}}(x_0) - \hat{f}_h(x_0)| + |\hat{f}_h(x_0) - f(x_0)|.$$

Note that

$$\begin{aligned}
|\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| &\leq \left\{|\hat{f}_{\hat{h}}(x_0) - \hat{f}_{h\vee\hat{h}}(x_0)| - M_n(h,\hat{h})\right\}_+ + M_n(h,\hat{h}) \\
&\leq \max_{\mathfrak{h}\in\mathcal{H}} \left\{|\hat{f}_{\mathfrak{h}}(x_0) - \hat{f}_{h\vee\mathfrak{h}}(x_0)| - M_n(h,\mathfrak{h})\right\}_+ + M_n(h,\hat{h}) \\
&\leq B(h,x_0) + M_n(\hat{h}) + M_n(h).
\end{aligned}$$

Applying the same reasoning to the term $|\hat{f}_{h\vee\hat{h}}(x_0) - \hat{f}_h(x_0)|$ and using (12), this leads to

$$\begin{aligned}
|\hat{f}(x_0) - f(x_0)| &\leq B(h,x_0) + 2M_n(h) + B(\hat{h},x_0) + 2M_n(\hat{h}) + |\hat{f}_h(x_0) - f(x_0)| \\
&\leq 3B(h,x_0) + 4M_n(h) + |\hat{f}_h(x_0) - f(x_0)|
\end{aligned}$$

This implies that:

$$R_n(\hat{f},f) \leq 3\left(\mathbf{E}B^p(h,x_0)\right)^{1/p} + 4M_n(h) + R_n(\hat{f}_h,f)$$

Step 2. Now, we upper bound $B(h,x_0)$. Using basic inequalities we have:

$$\begin{aligned}
B(h,x_0) &\leq \max_{\mathfrak{h}\in\mathcal{H}} \left\{\left|\mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right|\right\}_+ + \max_{\mathfrak{h}\in\mathcal{H}} \left\{\left|\hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right| - M_n(\mathfrak{h})\right\}_+ \\
&\quad + \max_{\mathfrak{h}\in\mathcal{H}} \left\{\left|\hat{f}_{h\vee\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{h\vee\mathfrak{h}}(x_0)\right| - M_n(h\vee\mathfrak{h})\right\}_+ \\
&\leq E_h(x_0) + 2T,
\end{aligned}$$

where

$$T = \max_{\mathfrak{h}\in\mathcal{H}} \left\{\left|\hat{f}_{\mathfrak{h}}(x_0) - \mathbf{E}\hat{f}_{\mathfrak{h}}(x_0)\right| - M_n(\mathfrak{h})\right\}_+.$$

This leads to:

$$\left(\mathbf{E}B^p(h,x_0)\right)^{1/p} \leq E_h(x_0) + 2\left(\mathbf{E}T^p\right)^{1/p}.$$

Step 3. We have:

$$\mathbf{E}T^p \leq \sum_{\mathfrak{h}\in\mathcal{H}} \int_0^\infty \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_{t_i})\right| \geq M_n(\mathfrak{h}) + t^{1/p}\right) dt,$$

where

$$g_{\mathfrak{h}}(X_{t_i}) = K_{\mathfrak{h}}(x_0 - X_{t_i}) \qquad \text{and} \qquad \bar{g}_{\mathfrak{h}}(X_{t_i}) = g_{\mathfrak{h}}(X_{t_i}) - \mathbf{E}g_{\mathfrak{h}}(X_{t_i}).$$

We obtain, using Corollary 1:

$$\begin{aligned}
\mathbf{E}T^p &\leq p\sum_{\mathfrak{h}\in\mathcal{H}}\int_0^\infty u^{p-1}\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \bar{g}_{\mathfrak{h}}(X_{t_i})\right| \geq M_n(\mathfrak{h})+u\right)du \\
&\leq 2p\sum_{\mathfrak{h}\in\mathcal{H}}\int_0^\infty u^{p-1}\exp\left(-\frac{(u+M_n(\mathfrak{h}))^2(n\Delta_n)^{\beta_H}}{4\mathfrak{C}\|g_{\mathfrak{h}}\|^2_{\mathrm{Lip}}}\right)du \\
&\leq 2p\sum_{\mathfrak{h}\in\mathcal{H}}\int_0^\infty u^{p-1}\exp\left(-\frac{1}{\mathfrak{C}}\left(\frac{u+M_n(\mathfrak{h})}{\varphi_n(\mathfrak{h})}\right)^2\right)du \\
&\leq 2p\sum_{\mathfrak{h}\in\mathcal{H}}\int_0^\infty u^{p-1}\exp\left(-\frac{1}{\mathfrak{C}}\left(\frac{u}{\varphi_n(\mathfrak{h})}\right)^2\right)\exp\left(-p\frac{\mathfrak{K}}{\mathfrak{C}}|\log V_{\mathfrak{h}}|\right)du
\end{aligned}$$

Since $\mathfrak{K} > \mathfrak{C}$ we obtain:

$$\begin{aligned}
\mathbf{E}T^p &\leq p\Gamma\left(\frac{p+1}{2}\right)\sum_{\mathfrak{h}\in\mathcal{H}}\left(\mathfrak{C}^{1/2}\varphi_n(\mathfrak{h})V_{\mathfrak{h}}\right)^p V_{\mathfrak{h}}^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)} \\
&\leq p\Gamma\left(\frac{p+1}{2}\right)(4d\mathfrak{L}_K^2)^{p/2}\left(\frac{\mathfrak{C}}{(n\Delta_n)^{\beta H}}\right)^{p/2}\sum_{\mathfrak{h}\in\mathcal{H}}V_{\mathfrak{h}}^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)} \\
&\leq p\left(\frac{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)}}{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)}-1}\right)^d\Gamma\left(\frac{p+1}{2}\right)(4d\mathfrak{L}_K^2)^{p/2}\left(\frac{\mathfrak{C}}{(n\Delta_n)^{\beta H}}\right)^{p/2}
\end{aligned}$$

Finally we obtain the following upper bound:

$$(\mathbf{E}T^p)^{1/p} \leq 2\mathfrak{L}_K\sqrt{d}\left(p\Gamma\left(\frac{p+1}{2}\right)\left(\frac{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)}}{e^{p\left(\frac{\mathfrak{K}}{\mathfrak{C}}-1\right)}-1}\right)^d\right)^{1/p}\times\left(\frac{\mathfrak{C}}{(n\Delta_n)^{\beta H}}\right)^{1/2}.$$

This allows us to obtain Theorem 3.

7.4 Proof of Theorem 4

Set $\mathbf{s}\in(0,M+1]^d$, $\mathbf{L}\in(0,+\infty)^d$.To prove this result, we construct a specific bandwidth vector h^* that belongs to $\mathcal{H}$. This allows to apply Propositions 2 and 3 and to bound, in (16), the minimum over $h\in\mathcal{H}$ by the value for $h=h^*$. Let

$$l_i^* = \left\lfloor \frac{\gamma(\mathbf{s})}{s_i}\left(\log\left((n\Delta_n)^{\beta_H}\right)-\log\log\left((n\Delta_n)^{\beta_H}\right)\right)\right\rfloor.$$

Since $\gamma(\mathbf{s})/s_i \leq 1/2$ we have

$$0 \leq l_i^* \leq \frac{1}{2}\log\left((n\Delta_n)^{\beta_H}\right) \leq \frac{\beta_H}{2}\log\left(n\Delta_n\right). \tag{36}$$

Now, denote for $i = 1, \dots, d$:

$$h_i^* = e^{-l_i^*} \quad \text{and} \quad h_i(\mathbf{s}) = \left((n\Delta_n)^{-\beta_H} \log\left((n\Delta_n)^{\beta_H}\right)\right)^{\gamma(\mathbf{s})/s_i}.$$

Remark that, using these notations we have $h_i(\mathbf{s}) \leq h_i^* \leq e h_i(\mathbf{s})$. If we consider $h^* = (h_1^*, \dots, h_d^*)$ and $h(\mathrm{s}) = (h_1(\mathrm{s}), \dots, h_d(\mathrm{s}))$, then:

$$\begin{aligned}
V_{h^*} &\geq V_{h(\mathbf{s})} \\
&\geq \left((n\Delta_n)^{-\beta_H} \log\left((n\Delta_n)^{\beta_H}\right)\right)^{\gamma(\mathbf{s})(1/\bar{s}+1/s_{\min})} \\
&\geq \left((n\Delta_n)^{-\beta_H}\right)^{\gamma(\mathbf{s})(1/\bar{s}+1/s_{\min})} \\
&= \left((n\Delta_n)^{-\beta_H}\right)^{1/2-\gamma(\mathbf{s})} \\
&\geq (n\Delta_n)^{-\beta_H/2},
\end{aligned}$$

where $s_{\min} = \min_j s_j$ and using that $(n\Delta_n)^{\beta_H} \geq e$. This implies, in combination with (36), that $h^* \in \mathcal{H}$. In (16) we can bound the right hand side by taking $h = h^*$. Let us consider each term separately.

First, using Propositions 2 and 3, since $0 < s_i \leq M + 1$ for each i, we have

$$\begin{aligned}
R_n(\hat{f}_{h^*}, f) &\leq \left(e^{M+1}\Lambda_1 + \frac{\Lambda_2}{(\log((n\Delta_n)^{\beta_H}))^{1/2}}\right)\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})} \\
&\leq \left(e^{M+1}\Lambda_1 + \Lambda_2\right)\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}.
\end{aligned}$$

Secondly, using Proposition 2 we obtain:

$$E_{h^*}(x_0) \leq 2e^{M+1}\Lambda_1\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}.$$

Finally, using (10) and (11) we have:

$$\begin{aligned}
M_n(h^*) &= \left(\frac{4d\mathfrak{L}_K^2}{V_{h^*}^2(n\Delta_n)^{\beta_H}}\right)^{1/2}\sqrt{\mathfrak{K}p|\log V_{h^*}|} \\
&\leq 2\sqrt{\mathfrak{K}dp}\mathfrak{L}_K\left(\frac{1}{V_{h(\mathbf{s})}^2(n\Delta_n)^{\beta_H}}\right)^{1/2}\left(|\log(e^{d+1}V_{h(\mathbf{s})})|\right)^{1/2} \\
&\leq 2\sqrt{\mathfrak{K}dp}\mathfrak{L}_K\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}\left(\frac{|\log(e^{d+1}V_{h(\mathbf{s})})|}{\log((n\Delta_n)^{\beta_H})}\right)^{1/2} \\
&\leq 2\sqrt{\mathfrak{K}dp}\mathfrak{L}_K\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}\left(\frac{d+1+(1/2-\gamma(\boldsymbol{s}))\log((n\Delta_n)^{\beta_H})}{\log((n\Delta_n)^{\beta_H})}\right)^{1/2} \\
&\leq 2\sqrt{\mathfrak{K}d(d+3/2)p}\mathfrak{L}_K\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}
\end{aligned}$$

This allows us to conclude that $\min_{h\in\mathcal{H}}\left\{R_n(\hat{f}_h,f)+4M_n(h)+3E_h(x_0)\right\}$ is bounded up to a multiplicative constant by $\left(\frac{\log((n\Delta_n)^{\beta_H})}{(n\Delta_n)^{\beta_H}}\right)^{\gamma(\boldsymbol{s})}$. To conclude, only note that $(n\Delta_n)^{-\beta_H/2}\leq(n\Delta_n)^{-\gamma(\boldsymbol{s})\beta_H}$ since $\gamma(\boldsymbol{s})\leq 1/2$.

Acknowledgements. The authors have been supported by Fondecyt projects 1171335 and 1190801, and Mathamsud 19-MATH-06 and 20-MATH-05.

References

Baudoin, F. and Hairer, M. (2007). A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields*, 139(3-4):373–395.

Bertin, K. and Klutchnikoff, N. (2017). Pointwise adaptive estimation of the marginal density of a weakly dependent process. *J. Statist. Plann. Inference*, 187:115–129.

Bertin, K., Klutchnikoff, N., León, J., and Prieur, C. (2018). Adaptive density estimation on bounded domains under mixing conditions. *HAL*.

Besalú, M., Kohatsu-Higa, A., and Tindel, S. (2016). Gaussian-type lower bounds for the density of solutions of SDEs driven by fractional Brownian motions. *Ann. Probab.*, 44(1):399–443.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.

Bosq, D. et al. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes. *The Annals of Statistics*, 25(3):982–1000.

Castellana, J. and Leadbetter, M. (1986). On smoothed probability density estimation for stationary processes. *Stochastic processes and their applications*, 21(2):179–193.

Comte, F. and Marie, N. (2018). Nonparametric estimation in fractional sde. *arXiv preprint arXiv:1806.00115*.

Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM: Probability and Statistics*, 6:211–238.

Comte, F. and Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. *Stochastic processes and their applications*, 115(5):797–826.

Dalalyan, A. (2001). *Estimation non-paramétrique asymptotiquement efficace pour des processus de diffusion ergodiques*. PhD thesis, Le Mans.

Djellout, H., Guillin, A., and Wu, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732.

Goldenshluger, A. and Lepski, O. (2011a). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632.

Goldenshluger, A. and Lepski, O. (2011b). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632.

Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on r^ d. *Probability Theory and Related Fields*, 159(3-4):479–543.

Hairer, M. (2005). Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.*, 33(2):703–758.

Klutchnikoff, N. (2014). Pointwise adaptive estimation of a multivariate function. *Math. Methods Statist.*, 23(2):132–150.

Kutoyants, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, 1(2):131–155.

Lepskiĭ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470.

Mishra, M. and Prakasa Rao, B. L. (2011). Nonparametric estimation of trend for stochastic differential equations driven by fractional brownian motion. *Statistical inference for stochastic processes*, 14(2):101–109.

Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.

Rigollet, P. and Hütter, J.-C. (2017). High dimensional statistics. *Lecture notes (MIT)*.

Rosenblatt, M. et al. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Saussereau, B. (2012). Transportation inequalities for stochastic differential equations driven by a fractional brownian motion. *Bernoulli*, 18(1):1–23.

Schmisser, E. (2013). Nonparametric estimation of the derivatives of the stationary density for stationary processes. *ESAIM: Probability and Statistics*, 17:33–69.

Tribouley, K. and Viennet, G. (1998). l_p adaptive density estimation in a *beta* mixing framework. In *Annales de l'IHP Probabilités et statistiques*, volume 34, pages 179–208.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469.

Varvenne, M. (2019). Concentration inequalities for stochastic differential equations with additive fractional noise. *Electron. J. Probab.*, 24:1–22.