Bernoulli **20**(2), 2014, 878–918
DOI: 10.3150/13-BEJ509

Nonparametric inference for fractional diffusion

BRUNO SAUSSEREAU

Université de Franche-Comté, Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, 16 route de Gray, 25030 Besançon, France. E-mail: bruno.saussereau@univ-fcomte.fr

A non-parametric diffusion model with an additive fractional Brownian motion noise is considered in this work. The drift is a non-parametric function that will be estimated by two methods. On one hand, we propose a locally linear estimator based on the local approximation of the drift by a linear function. On the other hand, a Nadaraya–Watson kernel type estimator is studied. In both cases, some non-asymptotic results are proposed by means of deviation probability bound. The consistency property of the estimators are obtained under a one sided dissipative Lipschitz condition on the drift that insures the ergodic property for the stochastic differential equation. Our estimators are first constructed under continuous observations. The drift function is then estimated with discrete time observations that is of the most importance for practical applications.

Keywords: fractional Brownian motion; non-parametric fractional diffusion model; statistical inference; stochastic differential equation

1. Introduction

The inference problem for diffusion process is now a well understood problem. The inference based on discretely observed diffusion is very important from a practical point of view and it has also benefited from numerous studies. With the development of technology, differential equations driven by noise with memory is increasingly popular in the statistical community as a modelling device. The subject of this work concerns the non-parametric estimation problem of the drift coefficient of a fractional diffusion described by the scalar equation

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + B_t^H, \qquad t \geq 0, \tag{1}$$

where $x_0 \in \mathbf{R}$ is the initial value of the process $X = (X_t)_{t\geq 0}$, and $B^H = (B_t^H)_{t\geq 0}$ is a fractional Brownian motion (fBm in short) with Hurst parameter $H \in (0,1)$. This means that B^H is a Gaussian process, centered, starting from 0 and such that $\mathbf{E}(B_t^H - B_s^H)^2 = |t-s|^{2H}$. Therefore, the process B^H has $\mathfrak{h}$-Hölder continuous paths for all $\mathfrak{h} \in (0,H)$. If $H = 1/2$, then B^H is clearly a Brownian motion and we refer to [25], Chapter 5, for a survey about the fBm.

Stochastic differential equations driven by fBm have recently carried out a lot of development. The special case of a constant diffusion coefficient is more specifically treated in [26] where it is proved that equation (1) has a strong unique solution if we assume the linear growth condition

$|b(x)| \le c_b(1+|x|)$ for b when $H < 1/2$, and Hölder continuity of order $\mathfrak{b} \in (1-1/2H, 1)$ when $H > 1/2$. In this paper, we assume that these conditions are true.

If we suppose that the observation process is

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + \sigma B_t^H, \qquad t \ge 0$$

with some unknown diffusion coefficient σ, then one may estimate the unknown Hurst parameter H and the diffusion coefficient σ via the quadratic variation (see [1,5,14]). This is the reason why we restrict ourselves to the case $\sigma = 1$ and H known.

Almost all the existing articles relate to the parametric case when one consider the model

$$X_t = x_0 + \int_0^t \theta b(X_s)\,\mathrm{d}s + B_t^H, \qquad t \ge 0, \tag{2}$$

with θ is the unknown parameter. Let us briefly review the works that have been already done. When the drift is linear, X is the fractional Ornstein–Uhlenbeck process and the estimation of θ has attracted a lot of attention. This problem has been first tackled by [15] using a maximum likelihood procedure. Some least square estimates are proposed in [2,12]. See also [28] for other methods. When b is not necessary linear, the pioneering work is [32] (see also the extended electronic version [33]). In this paper, the maximum likelihood estimator of θ is studied both with continuous and discrete observations. One may also refer to [3], Chapter 6 and [22], Chapter 6. A moment matching estimation is done in [27] and let us finally mention that a general discrete data maximum likelihood is proposed in [4] and a least square method is studied in [23] for the parametric estimation problem for model described by the equation (2).

To our knowledge, there is only one paper dealing with non-parametric estimation. In [21], the authors consider the model

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + \varepsilon B_t^H, \qquad t \ge 0, \tag{3}$$

and proposed a kernel type estimator of the trend coefficient $b_s := b(x_s)$ where $(x_s)_{s\ge 0}$ is the solution of equation (3) when $\varepsilon = 0$. The asymptotic behaviour is discussed when $\varepsilon \to 0$ on a finite time horizon when $H > 1/2$.

Our problem is of different nature than the previous ones. We will investigate two procedures to estimate the unknown function b at a fixed point $x \in \mathbf{R}$, that is, $b(x)$. We start with estimators based on continuous observation of X. Then, using a discretization, we propose estimators based on discrete time data which are the most important for practical applications. It is difficult to work directly with the fBm B^H because it is not a semimartingale. So we use the fundamental martingale (so called in [15], see also [24]) that will have nicer properties. Then some simple and classical ideas lead us to the construction of two estimators of $b(x)$. First, we consider that the drift is a constant function in a neighbourhood of the point x and thus the problem becomes parametric. The form of the least square estimator in this parametric case is used to propose a kernel type estimator of Nadaraya–Watson type (see (16) in Definition 1). If the drift is assumed to be linear in a small vicinity of x, then by similar arguments we define a locally linear estimator of the drift function in the point x (see (40) in Definition 2). These local linear smoothers are known to avoid some undesirable edge effects.

In order to study our these two estimators, we apply the same strategy. We prove some deviation probability bounds using non-asymptotic approach. Our probability bounds are stated conditionally to a random set. This method has been employed in [31] for a diffusion process with drift and variance given as non-parametric functions of the state variable. So a first step will consist to mainly focus ourselves on a non-asymptotic approach for which there is no difference between the ergodic and non-ergodic cases. Similar probability bounds are valid for the discretization of our estimators.

Then we investigate the consistency. With kernel type estimators, it is clear that it is necessary to impose some conditions which provide that the observed process $(X_t)_{t\geq 0}$ returns to any vicinity of the point x infinitely many times. The ergodicity can guarantee this property in the classical Brownian case (see [17]). The null recurrence of X can also be invoked when $H = 1/2$ as in [20]. We refer to [19] for the case of Harris recurrent diffusion. In the fractional case such ergodic properties, will hold under the assumption that the drift has polynomial growth and satisfies a one-sided dissipative Lipschitz condition (see [8], e.g.). Starting from our conditional deviation probability bound, we shall prove that the probability of the random event with respect to which the results are stated converges to 1 under the cited above hypotheses on the drift b. Thus, the weak consistency (this mean that the convergence holds in probability) is proved for both estimators, for continuous and discrete observations.

The most important results of our paper are certainly the ones concerning the problem of the estimation of the unknown value of the drift b in a fixed point x under discrete observations of the process X. For simplicity, we describe the Nadaraya–Watson estimator. For equally spaced observation times $\{t_k\}_{0\leq k\leq n}$, we denote ε_n the mesh size defined by $\varepsilon_n = t_{k+1} - t_k = cn^{-\mathfrak{a}}$ with a positive constant c and $\mathfrak{a} \in (0,1)$. We may also observe that $n = t_n^{\gamma}$ (up to a multiplicative constant) with $\gamma = 1/(1-\mathfrak{a}) > 1$. The Nadaraya–Watson estimator of $b(x)$ with the bandwidth h is defined at time t_n by

$$\hat{b}^{\mathrm{NW}}_{t_n,h}(x) = \frac{\sum_{k=0}^{n-1}(t_n - t_k)^{1-2H} N((X_{t_k} - x)/h)(X_{t_{k+1}} - X_{t_k})}{\sum_{k=0}^{n-1}(t_n - t_k)^{1-2H} N((X_{t_k} - x)/h)(t_{k+1} - t_k)},$$

where the kernel N is a positive regular function with support in $[-1, +1]$. It is obtained via a discretization of the continuous version of the estimator (see Definition 1). Some deviation probability bounds are proved for the continuous and the discrete version of the Nadaraya–Watson estimator in Theorem 2. If we assume that the drift has polynomial growth of order $\mathfrak{m}$ and satisfies a one-sided dissipative Lipschitz, then we obtain ergodic properties for the solution of (1) (see Proposition 1). This means that there exists a random variable $\bar{X}$ such that the solution of equation (1) converges for $t \to \infty$ to the stationary and ergodic process $(\bar{X}_t)_{t\geq 0} = (\bar{X}(\theta_t(\omega)))_{t\geq 0}$ where θ_t is the appropriate shift operator on the canonical probability space associated to the fBm. Then we shall prove the consistency of the estimator:

$$\hat{b}^{\mathrm{NW}}_{t_n,h}(x) \xrightarrow[n\to\infty,h\to 0]{\text{in probability}} b(x)$$

under the additional assumption that the number of approximation points satisfies $n = t_n^{\gamma}$ with $\gamma > 1 + \mathfrak{m}H^2$ (see Remark 2 for a discussion about the dependence between n, $\mathfrak{m}$ and H) and another assumption on the non-degeneracy of the stationary solution (see Hypothesis 4).

Similar results are obtained for the locally linear estimator that we do not present in this Introduction because it would require further and heavy notation. Nevertheless, the approach is identical. We first construct a continuous time version of the estimator in Definition 2 and some deviation probability bounds are obtained in Theorem 7. Then a discrete version is proposed (see Definition 3) and the consistency is obtained in Theorem 9 which is the other main result of this work.

The remainder of this paper is structured as follows. In Section 2, we give some notation and we state our main assumptions. Then we recall the link between B^H and the fundamental martingale for which classical stochastic calculus is available. This allows us to introduce a new observable process having a semi-martingale decomposition (see (8)). Then we enunciate the ergodic properties of the solution. The ergodic property under discrete observations as it is stated in Proposition 1 is new under our assumptions on the discretization procedure. Its proof is postponed in Appendix B. Section 3 is devoted to the Nadaraya–Watson estimator of the drift whereas the study of the locally-linear estimator is done in Section 4. In the two sections cited above, we state deviation probability bounds and consistency of the estimators under continuous and discrete observations. Some proofs related to the locally-linear estimator are gathered in Section 5. Finally, we shall make use of a Fernique's type lemma that is stated and proved in Appendix A.

2. Preliminaries

We consider a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ on which a one dimensional fractional Brownian motion B^H is defined. We denote $\mathcal{F}_t = \sigma(B_s^H, s \leq t)$ the σ-field generated by B^H completed with respect to $\mathbf{P}$.

In a first subsection, we give some notation and we state our assumptions. Then we indicate how to associate to the observed process an auxiliary semi-martingale which is appropriate for the statistical analysis. For this, we will resume the notation of [15,24]. Thereafter, the ergodic properties of the stochastic differential equation (1) will be discussed under ad-hoc assumption on the drift.

2.1. Notation and assumptions

In all the sequel, we use the following notation. If f and g are two functions form $\mathbf{R}$ to $\mathbf{R}$, we write $f(t) \succeq g(t)$ when there exists a constant K such that $f(t)/g(t) \geq K$. When the ratio of f and g is constant, we write $f(t) \asymp g(t)$.

The drift b may satisfy one or several items of the following hypothesis.

Hypothesis 1.

(a) (*Local regularity*). *For any x, b is locally Hölder of order* $\mathfrak{b}$ *in the point x: there exists L_x such that*

$$\left|b(y)-b(y')\right| \le L_x\left|y-y'\right|$$

for any y, y' in a neighbourhood of x.

(b) (*Global regularity*). *The drift term b is continuously differentiable with a polynomial growth condition on its derivative and on b itself: there exists $c_b > 0$ and $\mathfrak{m} \in \mathbf{N}$ such that*

$$\left|b(x)\right| + \left|b'(x)\right| \le c_b\left(1+|x|^{\mathfrak{m}}\right), \qquad x \in \mathbf{R}.$$

(c) (*One-sided dissipative Lipschitz condition*). *There exists a constant $L > 0$ such that for*

$$\left(b(y)-b(y')\right) \times \left(y-y'\right) \le -L\left|y-y'\right|^2, \qquad y, y' \in \mathbf{R}.$$

We remark that the one-sided dissipative Lipschitz condition implies that b is Lipschitz with the same constant L. It has been proved in [23] that there exists a unique solution to equation (1) under Hypotheses 1(b) and (c).

The kernel function that we shall need satisfies the following usual properties.

Hypothesis 2. *The kernel function N is continuously differentiable, non-negative with support in $[-1,1]$. Without loss of generality, we may assume that it is bounded by* 1.

Hypothesis 2 is supposed to be fulfilled in all the rest of this paper.

We also need the following notation concerning the discretization of the time interval $[0, T]$:

Hypothesis 3. *For a given $n \in \mathbf{N}$, a time discretization $\{t_k\}_{0\le k\le n}$ is considered with equally spaced observation times $\varepsilon_n := t_{k+1} - t_k \asymp n^{-\mathfrak{a}}$ with $\mathfrak{a} \in (0,1)$.*

We observe that the number of approximation points n is related to the time horizon of the discrete observations t_n by $n \asymp t_n^{\gamma}$ with $\gamma = 1/(1-\mathfrak{a}) > 1$.

We remark that $\varepsilon_n \asymp n^{-(\gamma-1)/\gamma}$. The forthcoming discussions will be held by means of γ instead of $\mathfrak{a}$ because they lead to more readable expressions.

2.2. Preliminaries on fractional Brownian motion

It is difficult to work directly with the fBm B^H because it is not a semimartingale. Hence, we introduce some related processes that will have nicer properties. For this purpose, let w_H be the function defined by

$$w_H(t,s) = c_H s^{1/2-H}(t-s)^{1/2-H}\mathbf{1}_{(0,t)}(s), \tag{4}$$

where $c_H = (2H\Gamma(3/2-H)\Gamma(H+1/2))^{-1}$. Thanks to [15,24], the process $M^H = (M_t^H)_{t\geq 0}$ defined by

$$M_t^H = \int_0^t w_H(t,s)\,\mathrm{d}B_s^H \tag{5}$$

is a centered Gaussian process with independent increments. Its variance function is given by

$$\mathbf{E}\big((M_t^H)^2\big) = \frac{\Gamma(3/2-H)}{2H\Gamma(3-2H)\Gamma(H+1/2)} t^{2-2H} := \lambda_H t^{2-2H}.$$

Thus, $(M_t^H)_{t\geq 0}$ is a martingale (called the fundamental martingale in [15]). The natural filtration of the martingale M^H coincides with the natural filtration of the fBm B^H. Finally, the process $B = (B_t)_{t\geq 0}$ defined by

$$B_t = \frac{1}{\sqrt{\lambda_H(2-2H)}} \int_0^t s^{H-1/2}\,\mathrm{d}M_s^H$$

is a standard Brownian motion that generates the same filtration as B^H and M^H. The inverse relationship will also be helpful:

$$M_t^H = \big(\lambda_H(2-2H)\big)^{1/2} \int_0^t s^{1/2-H}\,\mathrm{d}B_s. \tag{6}$$

We introduce the observable process $Y = (Y_t)_{t\geq 0}$ defined by

$$\begin{aligned} Y_t &= x_0 + \int_0^t w_H(t,s)\,\mathrm{d}X_s \\ &= x_0 + \int_0^t w_H(t,s) b(X_s)\,\mathrm{d}s + \int_0^t w_H(t,s)\,\mathrm{d}B_s^H. \end{aligned} \tag{7}$$

By (5) and (6), we have the following alternative expressions:

$$\begin{aligned} Y_t &= x_0 + \int_0^t w_H(t,s) b(X_s)\,\mathrm{d}s + M_t^H \\ &= x_0 + \int_0^t w_H(t,s) b(X_s)\,\mathrm{d}s + \big(\lambda_H(2-2H)\big)^{1/2} \int_0^t s^{1/2-H}\,\mathrm{d}B_s. \end{aligned} \tag{8}$$

In order to use the martingale M^H, we remark that

$$w_H(t,s)\,\mathrm{d}s = \frac{c_H}{(2-2H)\lambda_H}(t-s)^{1/2-H} s^{H-1/2}\,\mathrm{d}\langle M^H\rangle_s,$$

thus if we let

$$\tilde{w}(t,s) = \frac{c_H}{(2-2H)\lambda_H}(t-s)^{1/2-H} s^{H-1/2} \mathbf{1}_{(0,t)}(s)$$

we may write

$$Y_t = x_0 + \int_0^t \tilde{w}_H(t,s) b(X_s)\, \mathrm{d}\langle M^H \rangle_s + M_t^H. \tag{9}$$

The above representations will be the starting point of the construction of our estimators.

2.3. Ergodic properties of the stochastic differential equation

In this subsection, we give details on the ergodic properties of the fractional SDE (1). We use the results of Section 4 in [8], and we borrow the presentation of [23]. However, we repeat it for conciseness and we give some precisions.

Without loss of generality, we work on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ associated to a fBm $B^H := (B_t^H)_{t \in \mathbf{R}}$ defined on $\mathbf{R}$ entirely. This means that B^H is a zero mean Gaussian process having the variance function equals to $\mathbf{E}(|B_t^H - B_s^H|^2) = |t-s|^{2H}$ for any $s, t \in \mathbf{R}$. The Wiener space Ω is the topological space $C_0(\mathbf{R}; \mathbf{R})$ equipped with the compact open topology and $\mathcal{F}$ is the associated Borel σ-algebra. The measure $\mathbf{P}$ is the distribution of the fBm B^H which now corresponds to the evaluation process $B_t^H(\omega) = \omega(t)$ for $t \in \mathbf{R}$. The law of the two-sided fBm is invariant to the shift operators with increment $t \in \mathbf{R}$. In other word, the operator θ_t defined from Ω to Ω by $\theta_t \omega(\cdot) = \omega(\cdot + t) - \omega(\cdot)$ is such that the shifted process $(B_s(\theta_t \cdot))_{s \in \mathbf{R}}$ is again a fBm.

Moreover for all integrable real valued random variable F it holds

$$\lim_{t \to \infty} \frac{1}{T} \int_0^T F\big(\theta_t(\omega)\big)\, \mathrm{d}t = \mathbf{E}(F) \qquad \mathbf{P}\text{-almost surely.} \tag{10}$$

The ergodic properties of (1) will hold under the assumption that the drift b satisfies the polynomial growth condition 1(b) and the one-sided dissipative Lipschitz condition 1(c). Under these hypotheses, there exists a random variable $\bar{X}$ with finite moments of any order, and such that

$$\lim_{t \to \infty} \big| X_t(\omega) - \bar{X}\big(\theta_t(\omega)\big)\big| = 0 \tag{11}$$

for $\mathbf{P}$-almost-all $\omega \in \Omega$. Thus the solution of equation (1) converges when t goes to infinity to a stationary and ergodic process $(\bar{X}_t)_{t \geq 0}$ defined by $\bar{X}_t(\omega) = \bar{X}(\theta_t(\omega))$. By [10,11] the law of $(\bar{X}_t)_{t \geq 0}$ coincides with the attracting invariant measure for the solution of (1). The next proposition will be crucial when we will study consistency of our estimators.

Proposition 1. *Assume that Hypotheses* 1(b) *and* (c) *are true. Consider a continuously differentiable function* φ *such that*

$$\big|\varphi(y)\big| + \big|\varphi'(y)\big| \leq c_\varphi\big(1 + |y|^{\mathfrak{p}}\big), \qquad y \in \mathbf{R} \tag{12}$$

for some $c_\varphi > 0$ *and* $\mathfrak{p} \in \mathbf{N}$.

(i) *We have*

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(X_s)\, \mathrm{d}s = \mathbf{E}\big(\varphi(\bar{X})\big) \qquad \mathbf{P}\text{-}a.s. \tag{13}$$

(ii) *If* $\gamma > 1 + (\mathfrak{m}^2 + \mathfrak{p})H$ *and* $\gamma > \mathfrak{p} + 1$ *then*

$$\lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} \left\{ \sum_{k=0}^{n-1} \varphi(X_{t_k}) \mathbf{1}_{[t_k, t_{k+1})}(s) \right\} \mathrm{d}s = \mathbf{E}\big(\varphi(\bar{X})\big) \qquad \mathbf{P}\text{-}a.s., \tag{14}$$

where the observation times are defined in Hypothesis 3.

Let us make the following remarks and comments about the above proposition.

Remark 1. The proof of this result is partially contained in Proposition 2.3 and Lemma 3.1 in [23]. But in our result, we have a condition on the number of approximation points that depends on H and on the degrees of polynomial growth $\mathfrak{m}$ and $\mathfrak{p}$. We think that it is not possible to get rid of the fact that $n \asymp t_n^{\gamma}$ with $\gamma > 1 + \max((\mathfrak{m}^2 + \mathfrak{p})H, \mathfrak{p})$.

A proof of this result is proposed in Appendix B.

Remark 2. The above result is valid for a wide class of drift function since $\mathfrak{m} \in \{0, 1, 2, \dots\}$. We cover the case of bounded function as well than the case of linear and polynomial growing functions. Such a remark is valid for the function φ.

Assume that φ is bounded together wit hits derivative ($\mathfrak{p} = 0$). The condition on γ becomes $\gamma > 1 + \mathfrak{m}^2 H$ and thus the number of approximation points is related to the polynomial growth order $\mathfrak{m}$ and to the Hurst parameter H. When $\mathfrak{m}$ is fixed, we need more points in the time discretization when H grows. This is intuitively correct since the trajectories becomes more regular when H increases. Thus, the process is less oscillating and it is necessary to observe more often the diffusion in order to insure that X visits very often any neighbourhood of any fixed point.

When H is fixed, the number of approximation points is growing with the polynomial growth coefficients $\mathfrak{m}$ and $\mathfrak{p}$. It is surely related to the speed of convergence of the process X to the stationary ergodic process $\bar{X}$. It is intuitive to think that when we let the drift coefficient behaves like of polynomial function, the convergence must be slower when the degree is big. To our knowledge, such investigation has not yet been carried out.

3. The Nadaraya–Watson type estimator

Our first method for estimating the value of the drift b in a fixed point $x \in \mathbf{R}$ is inspired of the Nadaraya–Watson kernel regression. We construct this estimator in the following subsection. Thereafter, some deviation probability bounds are given and finally, the consistency will be stated under the ergodicity assumption.

3.1. Construction and decomposition of the Nadaraya–Watson estimator

First of all, we assume that the whole trajectory $(X_t)_{0\le t\le T}$ is observed between the times 0 and T. We will discuss a discretized version of our estimator in a moment.

The construction of a Nadaraya–Watson estimator is based on a simple idea. First, we think that the drift b is a constant function, that is $b(x) = \theta$ for any x. Hence an estimator of θ is an estimator of $b(x)$. We denote

$$X_t^\theta = x_0 + \int_0^t \theta \, \mathrm{d}s + B_t^H.$$

Similarly to (9), we introduce the observable process $Y^\theta = (Y_t^\theta)_{t\geq 0}$

$$Y_t^\theta = x_0 + \int_0^t \tilde{w}_H(t,s)\theta \, \mathrm{d}\langle M^H\rangle_s + M_t^H.$$

The unknown parameter θ can be estimated by the least squares method (see, e.g., [18]). The least squares estimator of θ obtained at time t is given by

$$\hat{\theta}(t) = \frac{\int_0^t \tilde{w}_H(t,s)\, \mathrm{d}Y_s^\theta}{\int_0^t (\tilde{w}_H(t,s))^2 \, \mathrm{d}\langle M^H\rangle_s}.$$

We denote

$$\alpha_H = \frac{c_H}{\sqrt{\lambda_H(2-2H)}}, \tag{15}$$

and we remark that

$$\int_0^t \bigl(\tilde{w}_H(t,s)\bigr)^2 \, \mathrm{d}\langle M^H\rangle_s = \int_0^t \alpha_H^2 (t-s)^{1-2H} \, \mathrm{d}s.$$

Thus, we obtain the alternative representation of $\hat{\theta}(t)$:

$$\hat{\theta}(t) = \frac{\int_0^t \tilde{w}_H(t,s)\, \mathrm{d}Y_s^\theta}{\int_0^t \alpha_H^2 (t-s)^{1-2H} \, \mathrm{d}s}.$$

In the context of our fractional diffusion (1), the drift b is not constant. Hence, we approximate it by a constant function θ in a neighbourhood $[x-h, x+h]$ of the point x. For this purpose, we consider a kernel function N satisfying Hypothesis 2. The above discussion leads to the following definition of an estimator of $b(x)$ by means of the observable process Y.

Definition 1. *The Nadaraya–Watson estimator of the drift b in a point x with the bandwidth h is defined at time t by*

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) = \frac{\int_0^t (\alpha_H^2/c_H)(t-s)^{1/2-H} s^{H-1/2} N((X_s - x)/h)\, \mathrm{d}Y_s}{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\, \mathrm{d}s} \tag{16}$$

with the convention that $a/0 := 0$. Equivalently, the more classical expression holds

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) = \frac{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\, \mathrm{d}X_s}{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\, \mathrm{d}s}. \tag{17}$$

Using the representation (8) and the fact that $N \leq 1$, we notice that the stochastic integral in (16) is well defined. The integral with respect to the process X in (17) is just an alternative writing of the one with respect to Y in (16). Moreover, starting from (16) and using (8), (4) and (15) we may express our estimator as

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) = \frac{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h) b(X_s)\,\mathrm{d}s}{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\,\mathrm{d}s} + \frac{\int_0^t \alpha_H (t-s)^{1/2-H} N((X_s - x)/h)\,\mathrm{d}B_s}{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\,\mathrm{d}s}.$$

Then we obtain the following decomposition of the error:

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) = b(x) + \xi_{x,h}(X_t) + r_{x,h}^{\mathrm{loc}}(X_t), \tag{18}$$

where

$$\xi_{x,h}(X_t) = \frac{\int_0^t \alpha_H (t-s)^{1/2-H} N((X_s - x)/h)\,\mathrm{d}B_s}{\int_0^t \alpha_H^2 (t-s)^{1-2H} N((X_s - x)/h)\,\mathrm{d}s},$$

$$r_{x,h}^{\mathrm{loc}}(X_t) = \frac{\int_0^t (t-s)^{1-2H} N((X_s - x)/h)[b(X_s) - b(x)]\,\mathrm{d}s}{\int_0^t (t-s)^{1-2H} N((X_s - x)/h)\,\mathrm{d}s}.$$

There are two kinds of errors in (18). The first one is a stochastic one (the term $\xi_{x,h}(X_t)$). The second one ($r_{x,h}^{\mathrm{loc}}(X_t)$) represents the accuracy of the local approximation of b by a constant function in a neighbourhood of the point x.

From a practical point of view, the real interest is the case when the observed data are discrete. So we provide now an effective estimation procedure. We assume that the process $(X_t)_{0 \leq t \leq T}$ is observed at times $(t_k)_{0 \leq k \leq n}$ (see Hypothesis 3). We discretize the expression of $\hat{b}_{t,h}^{\mathrm{NW}}(x)$ given in (17) by Riemann sums as

$$\hat{b}_{t_n,h}^{\mathrm{NW}}(x) = \frac{\sum_{k=0}^{n-1} (t_n - t_k)^{1-2H} N((X_{t_k} - x)/h)(X_{t_{k+1}} - X_{t_k})}{\sum_{k=0}^{n-1} (t_n - t_k)^{1-2H} N((X_{t_k} - x)/h)(t_{k+1} - t_k)}.$$

In order to have a decomposition of the error, we consider the simple process $(Q_s^n)_{s \geq 0}$ defined by

$$Q_s^n = \sum_{k=0}^{n-1} (t_n - t_k)^{1-2H} N\left(\frac{X_{t_k} - x}{h}\right) \mathbf{1}_{[t_k, t_{k+1})}(s).$$

With the help of equations (7) and (8), we may rewrite $\hat{b}^{\mathrm{NW}}_{t_n,h}(x)$ as

$$\begin{aligned}\hat{b}^{\mathrm{NW}}_{t_n,h}(x) &= \frac{1}{\int_0^{t_n} Q^n_s\,\mathrm{d}s}\int_0^{t_n} Q^n_s\,\mathrm{d}X_s\\ &= \frac{1}{\int_0^{t_n} Q^n_s\,\mathrm{d}s}\int_0^{t_n} \frac{Q^n_s}{w_H(t,s)}\,\mathrm{d}Y_s\\ &= \frac{1}{\int_0^{t_n} Q^n_s\,\mathrm{d}s}\left(\int_0^{t_n} Q^n_s b(X_s)\,\mathrm{d}s + \big(\lambda_H(2-2H)\big)^{1/2}\int_0^{t_n} \frac{Q^n_s}{w_H(t,s)} s^{1/2-H}\,\mathrm{d}B_s\right).\end{aligned}$$

Thus, we obtain a similar decomposition than (18)

$$\hat{b}^{\mathrm{NW}}_{t_n,h}(x) - b(x) = \xi_{x,h}\big(X^n_{t_n}\big) + r^{\mathrm{loc}}_{x,h}(X_{t_n}) + r^{\mathrm{traj}}_{x,h}(X_{t_n}) \tag{19}$$

with

$$\begin{aligned}\xi_{x,h}(X_{t_n}) &= \frac{\int_0^{t_n} \alpha_H(t-s)^{H-1/2} Q^n_s\,\mathrm{d}B_s}{\int_0^{t_n} \alpha^2_H Q^n_s\,\mathrm{d}s},\\ r^{\mathrm{loc}}_{x,h}(X_{t_n}) &= \frac{\int_0^{t_n}\sum_{k=0}^{n-1}(t_n-t_k)^{1-2H} N((X_{t_k}-x)/h)(b(X_{t_k})-b(x))\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s}{\int_0^{t_n} Q^n_s\,\mathrm{d}s},\\ r^{\mathrm{traj}}_{x,h}(X_{t_n}) &= \frac{\int_0^{t_n}\sum_{k=0}^{n-1}(t_n-t_k)^{1-2H} N((X_{t_k}-x)/h)(b(X_s)-b(X_{t_k}))\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s}{\int_0^{t_n} Q^n_s\,\mathrm{d}s}.\end{aligned}$$

We remark that $r^{\mathrm{loc}}_{x,h}(X_{t_k})$ represents again the accuracy of the local approximation of b by a constant function in a neighbourhood of the point x, but only in the discrete times $(t_k)_{0\le k\le n}$. It is worth to notice that a new term is involved: $r^{\mathrm{traj}}_{x,h}(X_{t_n})$. It represents the error made when one proceed to the discretization of the continuous process $(X_s)_{s\ge 0}$.

In the next subsection, we study deviation probability bounds for $\hat{b}^{\mathrm{NW}}_t(x)$ and $\hat{b}^{\mathrm{NW}}_{t,n}(x)$.

3.2. Deviation probability

In order to study the error from a probabilistic point of view, we need to introduce for some $\rho > 0$ and $\beta > 0$ the random sets

$$\mathcal{A}^{\mathrm{NW}}_{t,h} = \left\{\int_0^t \alpha_H(t-s)^{1-2H} N\left(\frac{X_s-x}{h}\right)\mathrm{d}s \ge \rho t^{1-H+\beta}\right\} \quad \text{and}$$

$$\mathcal{A}^{\mathrm{NW}}_{t_n,h} = \left\{\int_0^{t_n} \alpha_H \sum_{k=0}^{n-1}(t_n-t_k)^{1-2H} N\left(\frac{X_{t_k}-x}{h}\right)\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s \ge \rho t_n^{1-H+\beta}\right\}.$$

Some properties of the Nadaraya–Watson estimator are stated in the following theorem conditionally on the above events.

Theorem 2. *Under Hypothesis* 1(a), *when the trajectory is continuously observed*, *we have for any* $\zeta > 0$:

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t,h}(x) - b(x)\big| \geq L_x h^{\mathfrak{b}} + \zeta, \mathcal{A}^{\mathrm{NW}}_{t,h}\big) \leq 2\exp\big(-\rho^2(1-H)\zeta^2 t^{2\beta}\big). \tag{20}$$

We assume that b *satisfies Hypotheses* 1(b) *and* (c). *There exists* $\tau_0 \geq 1$, $c_{x,h,L} > 0$[1] *and a constant* $c_{H,\gamma}$ *such that for any* $t_n \geq \tau_0$, *the following conditional deviation probability bound holds*:

$$\begin{aligned}&\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t_n,h}(x) - b(x)\big| \geq Lh + c_{x,h,L}\varepsilon_n + \zeta, \mathcal{A}^{\mathrm{NW}}_{t_n,h}\big)\\ &\quad\leq 2\exp\bigg(-\frac{\rho^2(1-H)\zeta^2}{16\alpha_H^2} t_n^{2\beta}\bigg) + c_{H,\gamma} t_n^{H(\gamma-1)/(\gamma+1)} \exp\bigg(-\frac{\zeta}{4L} t_n^{(2H(\gamma-1))/(\gamma+1)}\bigg).\end{aligned} \tag{21}$$

It is interesting that the above results about the quality of our estimation are non-asymptotic and do not require any ergodic or mixing properties of the observed process. Clearly the event $\mathcal{A}^{\mathrm{NW}}_{t_n,h}$ is completely determined by the observed values of the trajectory of X. It is therefore always possible to check whether the path belongs or not to this set. If it is not the case, we are not able to guarantee a reasonable quality for the estimation of $b(x)$.

In the following remark, we discuss the rate of our approximation.

Remark 3.

1. If we choose a time dependent bandwidth h_t such that $h_t^{\mathfrak{b}} \asymp L_x^{-1} t^{-\beta/2}$, then the rate of estimation is of order $t^{-\beta/2}$:

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t,h_t}(x) - b(x)\big| \succeq t^{-\beta/2}, \mathcal{A}^{\mathrm{NW}}_{t,h_t}\big) \leq 2\exp\big(-\rho^2(1-H)t^{\beta}\big).$$

2. In the discrete case, for a fixed $\beta > 0$, we consider:
   - h_n a time dependant bandwidth with $h_n \asymp L^{-1} t_n^{-\beta/2}$;
   - $\gamma = (4H+\beta)/(4H-\beta)$;
   - $\varepsilon_n = t_n/n \asymp t_n^{-(\gamma-1)}$ with $\gamma - 1 = 2\beta/(4H-\beta) > \beta/2$.

   Then the approximation rate is again of order $t_n^{-\beta/2}$ since (21) implies

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t_n,h}(x) - b(x)\big| \succeq t^{-\beta/2}, \mathcal{A}^{\mathrm{NW}}_{t_n,h_n}\big) \preceq \exp\big(-C_{\rho,L,H} t_n^{\beta}\big) + t_n^{\mu_1} \exp\bigg(-\frac{t_n^{\mu_2}}{4L}\bigg),$$

   with $\mu_1, \mu_2 > 0$.

The stochastic integral that appears in the expression of $\xi_{x,h}(X_t)$ is a fractional martingale (so called in [13]). In order to study the asymptotic behaviour of the Nadaraya–Watson estimators, we need asymptotic properties of this fractional martingale. This will be done thanks to a straightforward exponential inequality for this kind of stochastic integral. One refers to [30] for related results on exponential inequalities for fractional martingales.

[1] See (32), (29) in the proof for an explicit expression of τ_0 and $c_{x,h,L}$.

Lemma 3. *We consider $K = (K_s)_{s\geq 0}$, an adapted process such that for a positive function v*

$$\sup_{0\leq u\leq t} \int_0^u (t-s)^{1-2H} |K_s|^2 \, \mathrm{d}s \leq v(t).$$

Then for any $\zeta \geq 0$ it holds that

$$\mathbf{P}\left(\left| \int_0^t (t-s)^{1/2-H} K_s \, \mathrm{d}B_s \right| \geq \zeta \right) \leq 2\exp\left(-\frac{\zeta^2}{2v(t)} \right). \tag{22}$$

Proof. For a fixed time t, we consider the true martingale $(Z_u^t)_{0\leq u\leq t}$ defined by

$$Z_u^t = \int_0^u (t-s)^{1/2-H} K_s \, \mathrm{d}B_s.$$

Here t is consider as a fixed parameter for the martingale Z^t. It holds that

$$\langle Z^t \rangle_u = \int_0^u (t-s)^{1-2H} |K_s|^2 \, \mathrm{d}s \leq v(t).$$

The classical exponential inequality (see [29], Exercice 3.16, Chapter 4) implies the result. □

Now we can prove Theorem 2.

Proof of Theorem 2. The proof is divided in several steps.

Step 1: *Proof of* (20). We use the decomposition (18). Obviously we have the following estimation

$$\left| r_{x,h}^{\mathrm{loc}} \right| \leq L_x h^{\mathfrak{b}}. \tag{23}$$

Thanks to the exponential inequality (22) of Lemma 3 we have for any $\zeta > 0$

$$\begin{aligned} &\mathbf{P}\left(\left|\xi_{x,h}(X_t)\right| \geq \zeta, \mathcal{A}_{t,h}^{\mathrm{NW}} \right) \\ &\quad \leq \mathbf{P}\left(\int_0^t (t-s)^{1/2-H} N\left(\frac{X_s - x}{h} \right) \mathrm{d}B_s \geq \rho\zeta t^{1-H+\beta} \right) \\ &\quad \leq 2\exp\left(-\rho^2(1-H)\zeta^2 t^{2\beta} \right). \end{aligned} \tag{24}$$

By (18), (23) and (24) we obtain

$$\begin{aligned} \mathbf{P}\left(\left|\hat{b}_{t,h}^{\mathrm{NW}}(x) - b(x)\right| \geq L_x h^{\mathfrak{b}} + \zeta, \mathcal{A}_{t,h}^{\mathrm{NW}} \right) &\leq \mathbf{P}\left(\left|\xi_{x,h}(X_t)\right| + \left|r_{x,h}^{\mathrm{loc}}(X_t)\right| \geq L_x h^{\mathfrak{b}} + \zeta, \mathcal{A}_{t,h}^{\mathrm{NW}} \right) \\ &\leq \mathbf{P}\left(\left|\xi_{x,h}(X_t)\right| \geq \zeta, \mathcal{A}_{t,h}^{\mathrm{NW}} \right) \\ &\leq 2\exp\left(-\rho^2(1-H)\zeta^2 t^{2\beta} \right), \end{aligned}$$

and (20) is proved.

Step 2: *Proof of* (21). We analyse separately the three terms in the decomposition (19). We begin with $\xi_{x,h}(X_{t_n})$ and we write

$$\begin{aligned}
&\mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big| \geq \zeta, \mathcal{A}^{\mathrm{NW}}_{t_n,h}\big) \\
&\quad = \mathbf{P}\left(\left|\int_0^{t_n} \alpha_H (t-s)^{H-1/2} Q^n_s \,\mathrm{d}B_s\right| \geq \zeta \int_0^{t_n} \alpha_H^2 Q^n_s \,\mathrm{d}s, \mathcal{A}^{\mathrm{NW}}_{t_n,h}\right) \\
&\quad \leq \mathbf{P}\left(\left|\int_0^{t_n} \alpha_H (t-s)^{H-1/2} Q^n_s \,\mathrm{d}B_s\right| \geq \zeta \rho t_n^{1-H+\beta}\right).
\end{aligned}$$

We fix t_n and we consider the martingale $Z^n := (Z^n_r)_{0\leq r\leq t_n}$ defined by

$$Z^n_r = \int_0^r \alpha_H (t_n - s)^{H-1/2} Q^n_s \,\mathrm{d}B_s.$$

Since $0 \leq N \leq 1$, the quadratic variation of the martingale Z^n satisfies:

$$\begin{aligned}
\langle Z^n \rangle_r &= \alpha_H^2 \int_0^r (t_n - s)^{2H-1} \left\{ \sum_{k=0}^{n-1} (t_n - t_k)^{2-4H} N^2\left(\frac{X_{t_k} - x}{h}\right) \mathbf{1}_{[t_k, t_{k+1})}(s) \right\} \mathrm{d}s \\
&\leq \alpha_H^2 \sum_{k=0}^{n-1} \int_0^{t_n} (t_n - s)^{2H-1} (t_n - t_k)^{2-4H} \mathbf{1}_{[t_k, t_{k+1})}(s) \,\mathrm{d}s.
\end{aligned}$$

When $H > 1/2$, it holds $(t_n - s)^{2H-1} \leq (t_n - t_k)^{2H-1}$ for $t_k \leq s < t_{k+1}$. Hence

$$\begin{aligned}
\langle Z^n \rangle_{t_n} &\leq \alpha_H^2 \sum_{k=0}^{n-1} \int_0^{t_n} (t_n - t_k)^{1-2H} \mathbf{1}_{[t_k, t_{k+1})}(s) \,\mathrm{d}s \\
&\leq \alpha_H^2 \int_0^{t_n} (t_n - s)^{1-2H} \,\mathrm{d}s \\
&\leq \frac{\alpha_H^2}{2-2H} t_n^{2-2H}.
\end{aligned}$$

For the second case when $H < 1/2$, the inequality $t_n - s > t_n - t_{k+1} = t_n - t_k - \Delta$ (valid for $t_k \leq s < t_{k+1}$) implies that

$$(t_n - t_k)^{2-4H} \leq (t_n - s)^{2-4H} \left(1 + \frac{\Delta}{t_n - s}\right)^{2+4H} \leq 4(t_n - s)^{2-4H}.$$

Therefore, we obtain

$$\langle Z^n \rangle_{t_n} \leq 4\alpha_H^2 \int_0^{t_n} (t_n - s)^{1-2H} \,\mathrm{d}s = \frac{2\alpha_H^2}{1-H} t_n^{2-2H}.$$

By Lemma 3, we conclude that

$$\mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big| \geq \zeta, \mathcal{A}^n_{t_n,h}\big) \leq 2\exp\biggl(-\frac{\rho^2(1-H)\zeta^2}{4\alpha_H^2} t_n^{2\beta}\biggr). \tag{25}$$

Now we study the error term $r^{\mathrm{loc}}_{x,h}(X_{t_n})$ from (19). The drift b is Lipschitz by Hypothesis 1(c). So we have

$$\big|r^{\mathrm{loc}}_{x,h}(X_{t_n})\big| \leq Lh. \tag{26}$$

The last term $r^{\mathrm{traj}}_{x,h}(X_{t_n})$ is more difficult to handle. At first, we write

$$\begin{aligned}&\big|r^{\mathrm{traj}}_{x,h}(X_{t_n})\big| \\ &\quad\leq \frac{\int_0^{t_n}\sum_{k=0}^{n-1}(t_n-t_k)^{1-2H}N((X_{t_k}-x)/h)L|X_s-X_{t_k}|\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s}{\int_0^{t_n}Q^n_s\,\mathrm{d}s}.\end{aligned}$$

By equation (1),

$$X_s - X_{t_k} = \int_{t_k}^s b(X_r)\,\mathrm{d}r + B^H_s - B^H_{t_k},$$

and when $|X_{t_k}-x|\leq h$ we may write for $t_k \leq s < t_{k+1}$:

$$\begin{aligned}|X_s - X_{t_k}| &\leq \int_{t_k}^s \big\{\big|b(X_r)-b(X_{t_k})\big| + \big|b(X_{t_k})-b(x)\big| + \big|b(x)\big|\big\} + \big|B^H_s - B^H_{t_k}\big| \\ &\leq \varepsilon_n\big\{Lh + c_b\big(1+|x|^{\mathfrak{m}}\big)\big\} + L\int_{t_k}^s |X_r - X_{t_k}|\,\mathrm{d}r + \varepsilon_n^{\mathfrak{h}}\big\|B^H\big\|_{0,t_n,\mathfrak{h}},\end{aligned}$$

where we have denoted for $0<\mathfrak{h}<H$:

$$\big\|B^H\big\|_{0,t_n,\mathfrak{h}} = \sup_{0\leq r,s\leq t_n}\frac{|B^H_s - B^H_r|}{|s-r|^{\mathfrak{h}}}.$$

When ε_n is small, the Gronwall inequality implies that for any $t_k \leq s < t_{k+1}$:

$$|X_s - X_{t_k}| \leq \varepsilon_n\big\{Lh + c_b\big(1+|x|^{\mathfrak{m}}\big)\big\} + \varepsilon_n^{\mathfrak{h}}\big\|B^H\big\|_{0,t_n,\mathfrak{h}}. \tag{27}$$

Therefore,

$$\big|r^{\mathrm{traj}}_{x,h}(X_{t_n})\big| \leq c_{x,h,L}\varepsilon_n + L\varepsilon_n^{\mathfrak{h}}\big\|B^H\big\|_{0,t_n,\mathfrak{h}}, \tag{28}$$

with

$$c_{x,h,L} = L\big\{Lh + c_b\big(1+|x|^{\mathfrak{m}}\big)\big\}. \tag{29}$$

Now we are able to end the proof of (21). Starting from the decomposition (19), using the estimations (25), (26) and (28), we obtain

$$\begin{aligned}
&\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t_n,h}(x)-b(x)\big|\geq Lh+C_{x,h,L}\varepsilon_n+\zeta,\mathcal{A}^{\mathrm{NW}}_{t_n,h}\big)\\
&\quad\leq \mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big|+\big|r^{\mathrm{loc}}_{x,h}(X_{t_n})\big|+\big|r^{\mathrm{traj}}_{x,h}(X_{t_n})\big|\geq Lh+C_{x,h,L}\varepsilon_n+\zeta,\mathcal{A}^{\mathrm{NW}}_{t_n,h}\big)\\
&\quad\leq \mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big|+Lh+C_{x,h,L}\varepsilon_n+L\varepsilon_n^{\mathfrak{h}}\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\geq Lh+C_{x,h,L}\varepsilon_n+\zeta,\mathcal{A}^{\mathrm{NW}}_{t_n,h}\big)\\
&\quad\leq \mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big|\geq\zeta/2,\mathcal{A}^{\mathrm{NW}}_{t_n,h}\big)+\mathbf{P}\big(L\varepsilon_n^{\mathfrak{h}}\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\geq\zeta/2\big)\\
&\quad\leq 2\exp\biggl(-\frac{\rho^2(1-H)\zeta^2}{16\alpha_H^2}t_n^{2\beta}\biggr)+\mathbf{P}\biggl(\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\geq\frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}}}\biggr).
\end{aligned}\tag{30}$$

We treat the last term in the right-hand side of the above inequality. We need a Fernique's type lemma for the exponential moment of the Hölder norm of the trajectories of the fBm B^H. Such a result is stated in Lemma 10 in the Appendix A. Chebyshev's exponential inequality yields

$$\begin{aligned}
\mathbf{P}\biggl(\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\geq\frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}}}\biggr)&\leq\exp\biggl(-\frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}}}\biggr)\mathbf{E}\big(\exp\big(\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\big)\big)\\
&\leq c_{H,\mathfrak{h}}\big(1+t_n^{H-\mathfrak{h}}\big)\exp\biggl(\frac{128H^2}{\mathfrak{h}^2}t_n^{2(H-\mathfrak{h})}-\frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}}}\biggr),
\end{aligned}\tag{31}$$

where we have used (59) from Lemma 10. We recall that $\varepsilon_n\asymp n^{-(\gamma-1)/\gamma}$ where n is the number of approximation points satisfying $n=t_n^{\gamma}$ with $\gamma>0$. We may write (31) as

$$\begin{aligned}
&\mathbf{P}\biggl(\big\|B^H\big\|_{0,t_n,\mathfrak{h}}\geq\frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}}}\biggr)\\
&\quad\leq c_{H,\mathfrak{h}}\big(1+t_n^{H-\mathfrak{h}}\big)\exp\biggl(-\frac{\zeta}{2L}t_n^{\mathfrak{h}(\gamma-1)}\biggl(1-\frac{256H^2L}{\zeta\mathfrak{h}^2}t_n^{2(H-\mathfrak{h})}t_n^{\mathfrak{h}(1-\gamma)}\biggr)\biggr).
\end{aligned}$$

If we choose $\mathfrak{h}$ such that $H>\mathfrak{h}>2H/(\gamma+1)$, then $2H-\mathfrak{h}-\gamma\mathfrak{h}<0$. For simplicity, we fix

$$\mathfrak{h}_0=\frac{H}{2}+\frac{H}{\gamma+1}=\frac{H(\gamma+3)}{2(\gamma+1)}.$$

When t_n is large, more precisely:

$$t_n\geq\tau_0:=\biggl(\frac{4608L}{\zeta}\biggr)^{2/(H(\gamma-1))}\vee 1,\tag{32}$$

we have

$$\biggl(1-\frac{256H^2L}{\zeta\mathfrak{h}_0^2}t_n^{2(H-\mathfrak{h}_0)}t_n^{\mathfrak{h}_0(1-\gamma)}\biggr)\geq\frac{1}{2}$$

and (31) yields

$$\mathbf{P}\left(\|B^H\|_{0,t_n,\mathfrak{h}_0} \geq \frac{\zeta}{2L\varepsilon_n^{\mathfrak{h}_0}} \right) \leq 2c_{H,\mathfrak{h}_0} t_n^{H-\mathfrak{h}_0} \exp\left(-\frac{\zeta}{4L} t_n^{\mathfrak{h}_0(\gamma-1)} \right). \tag{33}$$

With $\gamma > 1$ we have

$$\mathfrak{h}_0(\gamma-1) = \frac{H(\gamma-1)(\gamma+3)}{2(\gamma+1)} \geq \frac{2H(\gamma-1)}{\gamma+1}.$$

We report (33) in (30) and we deduce (21).

The proof of Theorem 2 is now complete. □

In the end, we get rid of the conditional result in the next subsection.

3.3. Consistency of the Nadaraya–Watson estimators

We start the investigation of the consistency of our estimators by the following proposition that is anecdotal but interesting in itself.

The strong consistency of $\hat{b}_{t,h}^{\mathrm{NW}}(x)$ is naturally related to the almost-sure convergence to 0 of $\xi_{x,h}(X_t)$ (see the decomposition (18)). If $H = 1/2$, by the strong law of large numbers for martingales, this convergence will holds as soon as $\int_0^\infty N^2(\frac{X_{t_k}-x}{h})\,\mathrm{d}s = +\infty$ almost-surely. When $H < 1/2$, such a condition will also ensure us the convergence of the fractional stochastic term $\xi_{x,h}(X_t)$.

Proposition 4. *Under the Hypothesis* 1(a), *when* $H < 1/2$ *and*

$$\int_0^\infty N^2\left(\frac{X_s - x}{h}\right) \mathrm{d}s = +\infty \qquad \mathbf{P}\text{-}a.s. \tag{34}$$

then the Nadaraya–Watson estimator is strongly consistent:

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) \xrightarrow[t\to\infty,h\to 0]{\mathbf{P}\text{-}a.s.} b(x). \tag{35}$$

Of course when the bandwidth is time dependent with $\lim_{t\to\infty} h_t = 0$, we have $\lim_{t\to\infty} \hat{b}_{t,h_t}^{\mathrm{NW}}(x) = b(x)$ almost-surely. The proof Proposition 4 is based on the following fractional version of the integral Toeplitz lemma.

Lemma 5. *Let* $\alpha > 0$. *Let* $(x_t)_{t\geq 0}$ *be a continuous real function such that* $\lim_{t\to\infty} x_t = x$ *and let* $(\gamma_t)_{t\geq 0}$ *be a measurable, positive and bounded function. Then it holds that*

$$\frac{\int_0^t (t-s)^{\alpha-1} (\int_0^s \gamma_r\,\mathrm{d}r) x_s\,\mathrm{d}s}{\int_0^t (t-s)^{\alpha-1} (\int_0^s \gamma_r\,\mathrm{d}r)\,\mathrm{d}s} \underset{t\to\infty}{\longrightarrow} x,$$

provided that $\lim_{t\to\infty} \int_0^t \gamma_s\,\mathrm{d}s = +\infty$.

Proof. Let $\varepsilon > 0$ and A be such that $|x_s - x| < \varepsilon$ for $s > A$. We denote $C_A = \sup_{s\le A} |x_s - x|$. By Fubini's theorem

$$\int_0^t (t-s)^\alpha \gamma_s \,\mathrm{d}s = \alpha \int_0^t (t-s)^{\alpha-1}\left(\int_0^s \gamma_r \,\mathrm{d}r\right)\mathrm{d}s,$$

and we write for $t > A$

$$\begin{aligned}\left|\frac{\int_0^t (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r) x_s \,\mathrm{d}s}{\int_0^t (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r)\,\mathrm{d}s} - x\right| &\le \frac{\int_0^t (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r)|x_s - x|\,\mathrm{d}s}{\int_0^t (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r)\,\mathrm{d}s} \\ &\le \varepsilon + C_A \frac{\int_0^A (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r)\,\mathrm{d}s}{\int_0^t (t-s)^{\alpha-1}(\int_0^s \gamma_r \,\mathrm{d}r)\,\mathrm{d}s}.\end{aligned} \tag{36}$$

Another application of Fubini's theorem implies that

$$\begin{aligned}\frac{\int_0^A (\int_r^A (t-s)^{\alpha-1}\,\mathrm{d}s)\gamma_r \,\mathrm{d}r}{\int_0^t (\int_r^t (t-s)^{\alpha-1}\,\mathrm{d}s)\gamma_r \,\mathrm{d}r} &= \frac{\int_0^A \gamma_r[(t-r)^\alpha - (t-A)^\alpha]\,\mathrm{d}r}{\int_0^t (t-r)^\alpha \gamma_r \,\mathrm{d}r} \\ &\le \frac{\int_0^A (t-r)^\alpha \gamma_r \,\mathrm{d}r}{\int_0^t (t-r)^\alpha \gamma_r \,\mathrm{d}r} \\ &\le \frac{A t^\alpha (\sup_{s\ge 0} |\gamma_s|)}{\int_0^{t/2} (t-r)^\alpha \gamma_r \,\mathrm{d}r} \\ &\le \frac{A t^\alpha (\sup_{s\ge 0} |\gamma_s|)}{t^\alpha \int_0^{t/2} \gamma_r \,\mathrm{d}r}\end{aligned}$$

and the last term tends to 0 as $t \to \infty$. We report this convergence in (36) and we obtain the result. □

Now we prove (35).

Proof of equation (35). By (18) and (23), we have

$$\left|\hat{b}_{t,h}^{\mathrm{NW}}(x) - b(x)\right| \le L_x h^{\mathfrak{b}} + \left|\xi_{x,h}(X_t)\right|.$$

Let $\alpha = 1/2 - H > 0$. By the stochastic Fubini theorem,

$$\begin{aligned}&\int_0^t (t-s)^\alpha N\left(\frac{X_s - x}{h}\right)\mathrm{d}B_s \\ &\quad = \alpha \int_0^t (t-s)^{\alpha-1}\left(\int_0^s N\left(\frac{X_r - x}{h}\right)\mathrm{d}B_r\right)\mathrm{d}s\end{aligned}$$

and we write $\xi_{x,h}(X_t) = T^1_{x,h}(X_t) \times T^2_{x,h}(X_t)$ with

$$T^1_{x,h}(X_t) = \left(\alpha \int_0^t (t-s)^{\alpha-1} \left(\int_0^s N^2((X_r - x)/h)\,\mathrm{d}r\right) \left(\left(\int_0^s N((X_r - x)/h)\,\mathrm{d}B_r\right) \Big/ \left(\int_0^s N^2((X_r - x)/h)\,\mathrm{d}r\right)\right) \mathrm{d}s\right) \Big/ \left(\int_0^t \alpha_H (t-s)^{\alpha-1} \left(\int_0^s N^2((X_r - x)/h)\,\mathrm{d}r\right) \mathrm{d}s\right)$$

and

$$T^2_{x,h}(X_t) = \frac{\int_0^t (t-s)^{\alpha} N^2((X_s - x)/h)\,\mathrm{d}s}{\int_0^t (t-s)^{2\alpha} N((X_s - x)/h)\,\mathrm{d}s}.$$

Since $\int_0^\infty N^2(\frac{X_s - x}{h})\,\mathrm{d}s = +\infty$ almost-surely,

$$\frac{\int_0^s N((X_r - x)/h)\,\mathrm{d}B_r}{\int_0^s N^2((X_r - x)/h)\,\mathrm{d}r} \xrightarrow[t\to\infty]{\text{a.s.}} 0$$

and the generalized Toeplitz Lemma 5 yields that $\lim_{t\to\infty} T^1_{x,h}(X_t) = 0$ almost-surely. Now we prove that the second term $T^2_{x,h}(X_t)$ is bounded when t is large. Since the kernel function N satisfies $0 \le N^2 \le N$, we have

$$\begin{aligned}
\left|T^2_{x,h}(X_t)\right| &\le \frac{\int_0^t (t-s)^{\alpha} N((X_s - x)/h)\,\mathrm{d}s}{\int_0^t (t-s)^{2\alpha} N((X_s - x)/h)\,\mathrm{d}s} \\
&\le \frac{(\int_0^t (t-s)^{2\alpha} N((X_s - x)/h)\,\mathrm{d}s)^{1/2} (\int_0^t N((X_s - x)/h)\,\mathrm{d}s)^{1/2}}{\int_0^t (t-s)^{2\alpha} N((X_s - x)/h)\,\mathrm{d}s},
\end{aligned}$$

where we have used the Cauchy–Schwarz inequality. For t big enough in such a way that $\int_0^{t-1} N(\frac{X_s - x}{h})\,\mathrm{d}s \ge 2$, we may write

$$\begin{aligned}
\left|T^2_{x,h}(X_t)\right|^2 &\le \frac{\int_0^t N((X_s - x)/h)\,\mathrm{d}s}{\int_0^{t-1} (t-s)^{2\alpha} N((X_s - x)/h)\,\mathrm{d}s} \\
&\le \frac{\int_0^{t-1} N((X_s - x)/h)\,\mathrm{d}s + \int_{t-1}^t N((X_s - x)/h)\,\mathrm{d}s}{\int_0^{t-1} N((X_s - x)/h)\,\mathrm{d}s} \\
&\le 1 + \frac{1}{2}.
\end{aligned}$$

Therefore, $\lim_{t\to\infty} \xi_{x,h}(X_t) = 0$ almost-surely and the proof is completed. □

Remark 4. Let $M^{(\alpha)} = (M_t^{(\alpha)})_{t\geq 0}$ with $\alpha = 1/2 - H > 0$ be the fractional martingale (as so called in [13]) defined by $M_t^{(\alpha)} = \int_0^t (t-s)^\alpha N(\frac{X_s - x}{h})\,\mathrm{d}B_s$. We have seen in the previous proof that (34) insures us that the fractional martingale $M^{(\alpha)}$ satisfies the strong law of large numbers:

$$\frac{M_t^{(\alpha)}}{\prec M^{(\alpha)} \succ_t} \xrightarrow[t\to\infty]{\mathbf{P}\text{-a.s.}} 0.$$

with a "fractional bracket" defined by $\prec M^{(\alpha)} \succ_t = \int_0^t (t-s)^{2\alpha} N^2(\frac{X_s - x}{h})\,\mathrm{d}s$. This is to our knowledge the first result of asymptotic behaviour for fractional martingales. Unfortunately the technics we employed to prove this convergence are not adapted to prove a similar result for a fractional martingale with $\alpha < 0$. See also [30] for further discussions on this topic.

Now we will work under the one-sided dissipative Lipschitz condition that ensures the ergodic properties of the observation process X. Before stating the main result of this paper, we make the following non-degeneracy assumption of the stationary solution.

Hypothesis 4. *The law of $\bar{X}$ is non-degenerate in a neighbourhood of x: for any small bandwidth h it holds*

$$\mathbf{E}\left[N\left(\frac{\bar{X} - x}{h}\right)\right] > 0.$$

Remark 5. It seems important to understand when the law of the stationary solution is non-degenerate. This will be certainly the subject of future works. It is obviously true if the distribution of $\bar{X}$ has full support. Nevertheless, the above hypothesis is satisfied in the case of the ergodic fractional Ornstein–Uhlenbeck process.

Theorem 6. *We assume that Hypotheses* 1(b) *and* (c) *hold true.*

When the whole trajectory is observed, the Nadaraya–Watson estimator is consistent:

$$\hat{b}_{t,h}^{\mathrm{NW}}(x) \xrightarrow[t\to\infty, h\to 0]{} b(x), \qquad \begin{cases} \textit{almost-surely when } H < 1/2; \\ \textit{in probability when } H > 1/2. \end{cases} \tag{37}$$

Its discretized version is also consistent when we assume that the number of approximation points satisfies $n \asymp t_n^\gamma$ with $\gamma > 1 + \mathfrak{m}^2 H$:

$$\hat{b}_{t_n,h}^{\mathrm{NW}}(x) \xrightarrow[n\to\infty, h\to 0]{\text{in probability}} b(x). \tag{38}$$

We observe that the number of approximation points depends on the regularity of b in the above result. This has already been discussed in Remark 2.

Proof of Theorem 6. In the following arguments, we will make use of Proposition 1 with $\varphi = N$ or $\varphi = N^2$. Obviously, (12) is satisfied with $\mathfrak{p} = 0$.

Step 1: *Proof of* (37). When $H < 1/2$, by (13) we obtain

$$\frac{1}{t}\int_0^t N^2\Big(\frac{X_s - x}{h}\Big)\,\mathrm{d}s \xrightarrow[t\to\infty]{\mathbf{P}\text{-a.s.}} \mathbf{E}\Big(N^2\Big(\frac{\bar{X} - x}{h}\Big)\Big).$$

By Cauchy–Schwarz inequality and Hypothesis 4

$$\mathbf{E}\Big[N^2\Big(\frac{\bar{X} - x}{h}\Big)\Big] \geq \Big(\mathbf{E}\Big[N\Big(\frac{\bar{X} - x}{h}\Big)\Big]\Big)^{1/2} > 0,$$

and (34) is satisfied. Thereby (37) is a consequence of (35) from Proposition 4.

Now $H > 1/2$ and let $\varepsilon > 0$. We use the probability deviation bound (20) with $\zeta = t^{-\beta/2}$. For t large enough and small h, we have

$$\begin{aligned}\mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t,h}(x) - b(x)\big| \geq \varepsilon\big) &\leq \mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t,h}(x) - b(x)\big| \geq Lh^{\mathfrak{b}} + t^{-\beta/2}\big)\\ &\leq \mathbf{P}\big(\big|\hat{b}^{\mathrm{NW}}_{t,h}(x) - b(x)\big| \geq Lh^{\mathfrak{b}} + t^{-\beta/2}, \mathcal{A}^{\mathrm{NW}}_{t,h}\big) + \mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t,h}\big)\\ &\leq 2\exp\big(-\rho^2(1-H)t^{\beta}\big) + \mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t,h}\big).\end{aligned}$$

The consistency (37) will be proved as soon as

$$\mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t,h}\big) = \mathbf{P}\Big(\int_0^t \alpha_H (t-s)^{1-2H} N\Big(\frac{X_s - x}{h}\Big)\,\mathrm{d}s \leq \rho t^{1-H+\beta}\Big) \xrightarrow[t\to\infty, h\to 0]{} 0. \tag{39}$$

Since $(t-s)^{1-2H} \geq t^{1-2H}$,

$$\begin{aligned}\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t,h} &\subseteq \Big\{\int_0^t N\Big(\frac{X_s - x}{h}\Big)\,\mathrm{d}s \leq \frac{\rho}{\alpha_H} t^{H+\beta}\Big\}\\ &\subseteq \Big\{\frac{1}{t}\int_0^t N\Big(\frac{X_s - x}{h}\Big)\,\mathrm{d}s \leq \frac{\rho}{\alpha_H} t^{H+\beta-1}\Big\} := \mathcal{A}_{t,h}.\end{aligned}$$

By the ergodic result (13),

$$\frac{1}{t}\int_0^t N\Big(\frac{X_s - x}{h}\Big)\,\mathrm{d}s \xrightarrow[t\to\infty]{\mathbf{P}\text{-a.s.}} \mathbf{E}\Big[N\Big(\frac{\bar{X} - x}{h}\Big)\Big] > 0$$

thus $\mathbf{1}_{\mathcal{A}_{t,h}}$ tends to 0 almost-surely when we choose β such that $H + \beta - 1 < 0$. This implies (39) and (37) is proved.

Step 2: *Consistency under discrete observations*. When $H > 1/2$, the proof is identical to the above one. We use the probability deviation bound (21) instead of (20) and the discrete ergodic property (14) is invoked in place of (13).

When $H < 1/2$, we use the deviation bound (21) and it remains to prove that

$$\mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t_n,h}\big) = \mathbf{P}\Big(\frac{\alpha_H}{n}\sum_{k=0}^{n-1}(t_n - t_k)^{1-2H} N\Big(\frac{X_{t_k} - x}{h}\Big) \leq \rho t_n^{1-H+\beta}\Big) \xrightarrow[n\to\infty, h\to 0]{} 0.$$

Let m be such that $t_{m-1} < \frac{t_n}{2} \le t_m$. Since $s \mapsto (t_n - s)^{1-2H}$ is a decreasing function, $(t_n - t_k)^{1-2H} \ge (t_n - \frac{t_n}{2})^{1-2H} = (\frac{t_n}{2})^{1-2H}$ for $k \le m-1$ and we deduce

$$\begin{aligned}
&\int_0^{t_n} \sum_{k=0}^{n-1} (t_n - t_k)^{1-2H} N\left(\frac{X_{t_k} - x}{h}\right) \mathbf{1}_{[t_k, t_{k+1})}(s)\, \mathrm{d}s \\
&\quad \ge \left(\frac{t_n}{2}\right)^{1-2H} \int_0^{t_n} \sum_{k=0}^{m-1} N\left(\frac{X_{t_k} - x}{h}\right) \mathbf{1}_{[t_k, t_{k+1})}(s)\, \mathrm{d}s \\
&\quad \ge \left(\frac{t_n}{2}\right)^{1-2H} \int_0^{t_n/2} \sum_{k=0}^{m-1} N\left(\frac{X_{t_k} - x}{h}\right) \mathbf{1}_{[t_k, t_{k+1})}(s)\, \mathrm{d}s \\
&\quad \ge \left(\frac{t_n}{2}\right)^{1-2H} \int_0^{t_n/2} \sum_{k=0}^{m-1} N\left(\frac{X_{\tilde{t}_k} - x}{h}\right) \mathbf{1}_{[\tilde{t}_k, \tilde{t}_{k+1})}(s)\, \mathrm{d}s,
\end{aligned}$$

where $\tilde{t}_k = t_k$ for $k \le m-1$ and $\tilde{t}_m = t_n/2$. We notice that the mesh size of this new time discretization is less that ε_n. The observation times are no more equally spaced but it is easy to convince ourselves that it does not affect the results of Proposition 1. Thereby

$$\begin{aligned}
\Omega \setminus \mathcal{A}^{\mathrm{NW}}_{t_n,h} &\subseteq \left\{ \left(\frac{t_n}{2}\right)^{1-2H} \int_0^{t_n/2} \sum_{k=0}^{m-1} N\left(\frac{X_{\tilde{t}_k} - x}{h}\right) \mathbf{1}_{[\tilde{t}_k, \tilde{t}_{k+1})}(s)\, \mathrm{d}s \le \frac{\rho}{\alpha_H} t_n^{1-H+\beta} \right\} \\
&\subseteq \left\{ \frac{1}{\tilde{t}_m} \int_0^{\tilde{t}_m} \sum_{k=0}^{m-1} N\left(\frac{X_{\tilde{t}_k} - x}{h}\right) \mathbf{1}_{[\tilde{t}_k, \tilde{t}_{k+1})}(s)\, \mathrm{d}s \le \frac{4\rho}{\alpha_H} (\tilde{t}_m)^{H+\beta-1} \right\}
\end{aligned}$$

and accordingly of the inequality (14) from Proposition 1, we may conclude as in the first step. □

4. The locally linear estimate

We follow the same structure than Section 3. Nevertheless, since the notation become more heavy, we distinguish the case of continuous observations from the discrete one.

4.1. Heuristic approach

The idea is similar to the one used Section 3 and also follows the one developed in [31]. At first we discuss the case of a linear drift coefficient b of the form $b_{\theta_0,\theta_1}(z) = \theta_0 + \theta_1(z - x)/h$. Hence, b depends on two parameters θ_0 and θ_1 ($h > 0$ is fixed). Since $b_{\theta_0,\theta_1}(x) = \theta_0$, an estimator of θ_0 is an estimator of the value of the drift at the point x. We denote

$$X_t^\theta = x_0 + \int_0^t b_{\theta_0,\theta_1}\left(X_s^\theta\right) \mathrm{d}s + B_t^H.$$

Similarly to (9), we introduce the observable process $Y^\theta = (Y_t^\theta)_{t\geq 0}$ defined by

$$\begin{aligned} Y_t^\theta &= x_0 + \int_0^t \tilde{w}_H(t,s) b_{\theta_0,\theta_1}\left(X_s^\theta\right) \mathrm{d}\langle M^H\rangle_s + M_t^H \\ &= x_0 + \int_0^t \rho_s^{\mathrm{T}} \times \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \mathrm{d}\langle M^H\rangle_s + M_t^H, \end{aligned}$$

with $\rho = (\rho_s)_{s\geq 0}$ is the process with values in $\mathbf{R}^2$ defined by

$$\rho_s = \tilde{w}_H(t,s) \begin{pmatrix} 1 \\ \left(X_s^\theta - x\right)/h \end{pmatrix},$$

and for a matrix A, A^{T} denotes its transpose. Intuitively, the values θ_0 and θ_1 can be estimated by the least squares method (see, e.g., [18]). If the 2×2-matrix

$$\Pi_t = \int_0^t \rho_s \rho_s^{\mathrm{T}} \, \mathrm{d}\langle M^H\rangle_s$$

is not singular, the least squares estimator of $(\theta_0,\theta_1)^{\mathrm{T}}$ obtained at time t is given by

$$\hat{\theta}(t) = \begin{pmatrix} \hat{\theta}_0(t) \\ \hat{\theta}_1(t) \end{pmatrix} = \Pi_t^{-1} \int_0^t \rho_s \, \mathrm{d}Y_s^\theta.$$

With the constant α_H defined in (15), we may write

$$\begin{aligned} \Pi_t &= \int_0^t \left(\tilde{w}_H(t,s)\right)^2 \begin{pmatrix} 1 & \left(X_s^\theta - x\right)/h \\ \left(X_s^\theta - x\right)/h & \left(X_s^\theta - x\right)^2/h^2 \end{pmatrix} \mathrm{d}\langle M^H\rangle_s \\ &= \int_0^t \alpha_H^2 (t-s)^{1-2H} \begin{pmatrix} 1 & \left(X_s^\theta - x\right)/h \\ \left(X_s^\theta - x\right)/h & \left(X_s^\theta - x\right)^2/h^2 \end{pmatrix} \mathrm{d}s \end{aligned}$$

and we obtain the following expression of $\hat{\theta}_0(t)$:

$$\hat{\theta}_0(t) = \frac{m_2(t)}{\delta(t)} \int_0^t \tilde{w}_H(t,s) \, \mathrm{d}Y_s^\theta - \frac{m_1(t)}{\delta(t)} \int_0^t \tilde{w}_H(t,s) \left(\frac{X_s^\theta - x}{h}\right) \mathrm{d}Y_s^\theta,$$

where for $i \in \{0,1,2\}$:

$$m_i(t) = \int_0^t \alpha_H^2 (t-s)^{1-2H} \left(\frac{X_s^\theta - x}{h}\right)^i \mathrm{d}s$$

and

$$\delta(t) = m_0(t) m_2(t) - m_1^2(t).$$

In the context of the fractional diffusion (1), the drift b is not linear. Hence, we approximate it by a linear function $\theta_0 + \theta_1 (z-x)/h$ in a neighbourhood $[x-h, x+h]$ of the point x. For this purpose, we use a kernel function N that satisfies Hypothesis 2. The above discussion is the starting point of the construction of a locally linear estimator of b under continuous observations.

4.2. Observations based on the whole trajectory

4.2.1. *Construction and decomposition of the error*

We give the following definition of a locally linear estimator of $b(x)$ by means of the observable processes Y and X.

Definition 2. *The locally linear estimators at time t of $b(x)$ with the kernel N and a bandwidth h is defined by*

$$\hat{b}^{\mathrm{ll}}_{t,h}(x) = \frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} \int_0^t \tilde{w}_H(t,s) N\left(\frac{X_s - x}{h}\right) \mathrm{d}Y_s - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)} \int_0^t \tilde{w}_H(t,s) \left(\frac{X_s - x}{h}\right) N\left(\frac{X_s - x}{h}\right) \mathrm{d}Y_s \tag{40}$$

$$= \int_0^t \left[\frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)} \left(\frac{X_s - x}{h}\right)\right] \alpha_H^2 (t-s)^{1-2H} N\left(\frac{X_s - x}{h}\right) \mathrm{d}X_s, \tag{41}$$

where for $j = 0, 1, 2$:

$$\begin{cases} \mathrm{v}_j(t) = \int_0^t \alpha_H^2 (t-s)^{1-2H} \left(\frac{X_s - x}{h}\right)^j N\left(\frac{X_s - x}{h}\right) \mathrm{d}s, \\ \mathrm{d}(t) = \mathrm{v}_0(t)\mathrm{v}_2(t) - \left(\mathrm{v}_1(t)\right)^2. \end{cases} \tag{42}$$

The alternative expression (41) is obtained thanks to the definition of the process Y given in (7) and the relation

$$w_H(t,s)\tilde{w}_H(t,s) = \alpha_H^2 (t-s)^{1-2H} \qquad \text{for } t > s.$$

Moreover, the representation (8), the facts that $N \le 1$ and that for all $z \in \mathbf{R}$, $|zN(z)| \le 1$, we notice that the stochastic integrals in (40) are well defined. Moreover, we remark that $\mathrm{d}(t) > 0$ by the Cauchy–Schwarz inequality.

In order to understand what kind of quantities will appear in the deviation probability bound, we write a decomposition of the error $\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)$. Using (8) and (40), we rewrite $\hat{b}^{\mathrm{ll}}_{t,h}(x)$ as

$$\begin{aligned} \hat{b}^{\mathrm{ll}}_{t,h}(x) = {} & \frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} \int_0^t \alpha_H (t-s)^{1/2-H} N\left(\frac{X_s - x}{h}\right) \mathrm{d}B_s \\ & - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)} \int_0^t \alpha_H (t-s)^{1/2-H} \left(\frac{X_s - x}{h}\right) N\left(\frac{X_s - x}{h}\right) \mathrm{d}B_s \\ & + \int_0^t \left[\frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)} \left(\frac{X_s - x}{h}\right)\right] \alpha_H^2 (t-s)^{1-2H} N\left(\frac{X_s - x}{h}\right) b(X_s)\, \mathrm{d}s. \end{aligned}$$

Now we consider the local error functions $\delta_{x,h}$ defined by

$$\delta_{x,h}(z) = b(z) - \big(b(x) + b'(x) \times (z - x)\big).$$

By the definitions of the functions v_j it holds that

$$\int_0^t \left[\frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)}\left(\frac{X_s - x}{h}\right)\right]\alpha_H^2 (t-s)^{1-2H} N\left(\frac{X_s - x}{h}\right) b(X_s)\,\mathrm{d}s$$
$$= b(x) + \int_0^t \left[\frac{\mathrm{v}_2(t)}{\mathrm{d}(t)} - \frac{\mathrm{v}_1(t)}{\mathrm{d}(t)}\left(\frac{X_s - x}{h}\right)\right]\alpha_H^2 (t-s)^{1-2H} N\left(\frac{X_s - x}{h}\right) \delta_{x,h}(X_s)\,\mathrm{d}s.$$

Now for $j = 0, 1$ we denote

$$\nu_j(t,s) = \alpha_H (t-s)^{1/2-H}\left(\frac{X_s - x}{h}\right)^j N\left(\frac{X_s - x}{h}\right),$$
$$\tilde{\nu}_j(t,s) = \alpha_H^2 (t-s)^{1-2H}\left(\frac{X_s - x}{h}\right)^j N\left(\frac{X_s - x}{h}\right)$$

and $(\mathrm{V}_t)_{t\geq 0}$ is the process with values in $\mathbf{R}^{2\times 2}$ defined by

$$\mathrm{V}_t = \begin{pmatrix} \mathrm{v}_0(t) & \mathrm{v}_1(t) \\ \mathrm{v}_1(t) & \mathrm{v}_2(t) \end{pmatrix}. \tag{43}$$

Thus, we have again the expression

$$\hat{b}^{\mathrm{ll}}_{t,h}(x) = b(x) + \xi^1_{x,h}(X_t) + r^{\mathrm{loc}}_{x,h}(X_t), \tag{44}$$

where $\xi^1_{x,h}(X_t)$ and $r^{\mathrm{loc}}_{x,h}(X_t)$ are the first components of the following two dimensional vectors

$$\Xi_{x,h}(X_t) = (\mathrm{V}_t)^{-1}\int_0^t \begin{pmatrix} \nu_0(t,s) \\ \nu_1(t,s) \end{pmatrix} \mathrm{d}B_s,$$
$$R^{\mathrm{loc}}_{x,h}(X_t) = (\mathrm{V}_t)^{-1}\int_0^t \begin{pmatrix} \tilde{\nu}_0(t,s) \\ \tilde{\nu}_1(t,s) \end{pmatrix} \delta_{x,h}(X_s)\,\mathrm{d}s.$$

The interpretation of $\xi^1_{x,h}(X_t)$ and $r^{\mathrm{loc}}_{x,h}(X_t)$ and is the same one than in the Nadaraya–Watson procedure. It is important to notice that when $H = 1/2$ we obtain the same decomposition of $\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)$ as the one in [31], equation (5.3).

4.2.2. *Deviation probability and consistency*

In view of the term $r^{\mathrm{loc}}_{x,h}(X_t)$ in (44), the accuracy of the locally linear estimate will be expressed thanks to the quality of the approximation of b by a linear function. Under Hypothesis 1(a) it is natural to introduce in the neighbourhood $[x-h, x+h]$ of the point x the quantity

$$\Delta_{x,h} = \sup_{|z-x|\leq h}\big|b(z) - \big(b(x) + b'(x) \times (z-x)\big)\big|.$$

In order to study the error from a probabilistic point of view (see $\xi^1_{x,h}(X_t)$ in (44)), we make the following comments. If the kernel function N satisfies $N^2 = N$ and if $H = 1/2$, the process $V = (\mathrm{V}_t)_{t\geq 0}$ defined in (43) is the quadratic variation process of the two-dimensional martingale $M = (M_t)_{t\geq 0}$ defined by

$$M_t = \int_0^t \begin{pmatrix} 1 \\ \frac{X_s - x}{h} \end{pmatrix} N\left(\frac{X_s - x}{h}\right) \mathrm{d}B_s.$$

If we investigate the strong consistency of our estimator we shall use a strong law of large numbers for multivariate martingales (see [6,16,18,34]). Therefore, the strong consistency is a consequence of asymptotic properties as t goes to infinity of the eigenvalues of the matrix V_t. In the fractional framework, this kind of asymptotic has not yet been studied. Nevertheless, the eigenvalues of V_t play a crucial role in the following.

Thus, we introduce for some $\rho > 0$ and $\beta > 0$ the random set

$$\mathcal{A}^{\mathrm{ll}}_{t,h} = \left\{\lambda_m(t) \geq \rho t^{1-H+\beta}\right\},$$

where $\lambda_m(t)$ is the smallest eigenvalue of the matrix V_t. The properties of $\hat{b}^{\mathrm{ll}}_{t,h}(x)$ are first studied restricted to the event $\mathcal{A}^{\mathrm{ll}}_{t,h}$. The consistency is proved under Hypotheses 1(b) and (c) insuring the ergodicity of (1) and under the following non-degeneracy condition on the law of the stationary solution.

Hypothesis 5. *The law of $\bar{X}$ is (strongly) non-degenerate in a neighbourhood of x: for any small bandwidth h it holds*:

$$\mathbf{E}\left[\left(\frac{\bar{X} - x}{h}\right)^2 N\left(\frac{\bar{X} - x}{h}\right)\right] > 0.$$

Notice that we use the terminology "strongly non-degenerate" because Hypothesis 5 implies Hypothesis 4.

Theorem 7. *Let x be fixed.*

(i) *If b satisfies Hypothesis* 1(a), *then for any $\zeta > 0$, we have*

$$\mathbf{P}\left(\left|\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)\right| \geq c_{\rho,H}\Delta_{x,h}t^{1-H-\beta} + \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) \leq 4\exp\left(-\frac{(1-H)\rho^2}{2\alpha_H^2}\zeta^2 t^{2\beta}\right), \tag{45}$$

with $c_{\rho,H} = \sqrt{2}c_H^2/(\rho\lambda_H)$.

(ii) *When we assume Hypotheses* 1(b) *and* (c) *and Hypothesis* 5, *the fractional diffusion is then ergodic and we have the consistency of the locally linear estimator*:

$$\hat{b}^{\mathrm{ll}}_{t,h}(x) \xrightarrow[t\to\infty, h\to 0]{\text{in probability}} b(x).$$

The proof of this result is postponed in Section 5.

If we consider that $\zeta = t^{-\beta/2}$, (45) implies that

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)\big| \geq c_{\rho,H}\Delta_{x,h}t^{1-H-\beta} + t^{-\beta/2}, \mathcal{A}^{\mathrm{ll}}_{t,h}\big) \leq 4\exp\biggl(-\frac{(1-H)\rho^2}{2c_H^2}t^{\beta}\biggr).$$

The quality of the approximation of b by a linear function is measured by $\Delta_{x,h}$ in a vicinity of x. Under Hypothesis 1(a), we have $\Delta_{x,h} \leq 2L_x h$. Assume also that b is twice differentiable in a neighbourhood of x with second derivative bounded by L_x, then $\Delta_{x,h} \leq L_x h^2/2$. Now we are able to choose a time dependent bandwidth h_t. Clearly if $h_t^2 \asymp L_x^{-1} t^{H-1+\beta/2}$ (remind that the symbol $\asymp$ means that the ratio of the functions are bounded), we obtain that the rate of estimation is of order $t^{-\beta/2}$:

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)\big| \geq \bar{c}_{\rho,H}t^{-\beta/2}, \mathcal{A}^{\mathrm{ll}}_{t,h}\big) \leq 4\exp\biggl(-\frac{(1-H)\rho^2}{2c_H^2}t^{\beta}\biggr),$$

where $\bar{c}_{\rho,H} = c_{\rho,H}/2 + 1$ and of course β has been chosen such that $\beta < 2(H-1)$.

Since b is unknown, we have no reason to have information about L_x. As usual in such a non-parametric context, we have two choices. On the one hand, we can restrict our problem to a class of drift function b satisfying the above hypotheses with constants L_x such that $L_x \in (L_{\min}, L_{\max})$. On the other hand, an adaptive (data-driven) choice of the bandwidth may be considered (see [31] and the references therein). Unfortunately, the analysis of the error given in Theorem 2 seems to be not adapted to this powerful method of bandwidth's choice.

To continue the comparison with the work of Spokoiny (see [31]), we may relate our random set and the one that appears in [31]. Indeed, very simple calculations allow us to write the exact expression of the smallest eigenvalue of the matrix V_t as

$$\lambda_m(t) = \tfrac{1}{2}\big(\mathrm{v}_0(t) + \mathrm{v}_2(t) - \big(\big(\mathrm{v}_0(t) + \mathrm{v}_2(t)\big)^2 - 4\mathrm{d}(t)\big)^{1/2}\big). \tag{46}$$

The above expression employs analogous quantities that the one appearing in the random set $\mathcal{A}_h$ (see [31], page 819). Despite these analogies, it is not easy to compare the two events. Moreover, the discussion about the accuracy of the approximation and the "stochastic error" is different from the one made in [31]. This is due to the fact that the stochastic error is hidden in the random set $\mathcal{A}^{\mathrm{ll}}_{t,h}$ whereas it appears explicitly as a "conditional variance" in the work of Spokoiny.

Now we give an effective way to estimate the drift when we consider discrete observations.

4.3. Discrete observations

We consider the discretization of the quantities that are defined in (42). For $j = 0, 1, 2$,

$$\begin{cases} \mathrm{v}_j^n(t_n) = \displaystyle\int_0^{t_n} \sum_{k=1}^{n-1} \alpha_H^2 (t_n - t_k)^{1-2H} \biggl(\frac{X_{t_k} - x}{h}\biggr)^j N\biggl(\frac{X_{t_k} - x}{h}\biggr) \mathbf{1}_{[t_k, t_{k+1})}(s)\,\mathrm{d}s, \\ \mathrm{d}^n(t_n) = \mathrm{v}_0^n(t_n)\mathrm{v}_2^n(t_n) - \big(\mathrm{v}_1^n(t_n)\big)^2 \end{cases}$$

and we denote $V_{t_n}^n$ the matrix

$$\mathrm{V}_{t_n}^n = \begin{pmatrix} \mathrm{v}_0^n(t_n) & \mathrm{v}_1^n(t_n) \\ \mathrm{v}_1^n(t_n) & \mathrm{v}_2^n(t_n) \end{pmatrix}.$$

Considering (41), we propose the following estimator of $b(x)$ based on discrete observations.

Definition 3. *The discretized locally linear estimator at time t_n with a bandwidth h is*

$$\hat{b}_{t_n,h}^{\mathrm{ll}}(x) = \sum_{k=1}^{n-1} \left[\frac{\mathrm{v}_2^n(t_n)}{\mathrm{d}^n(t_n)} - \frac{\mathrm{v}_1^n(t_n)}{\mathrm{d}^n(t_n)} \left(\frac{X_{t_k} - x}{h} \right) \right] \alpha_H^2 (t_n - t_k)^{1-2H} N\left(\frac{X_{t_k} - x}{h} \right) (X_{t_{k+1}} - X_{t_k}).$$

As in (19), We will decompose $\hat{b}_{t_n,h}^{\mathrm{ll}}(x)$ into a sum of three terms:

$$\hat{b}_{t_n,h}^{\mathrm{ll}}(x) - b(x) = \xi_{x,h}(X_{t_n}) + r_{x,h}^{\mathrm{loc}}(X_{t_n}) + r_{x,h}^{\mathrm{traj}}(X_{t_n}). \tag{47}$$

For this purpose, we consider the simple process

$$T_s^n = \sum_{k=1}^{n-1} \left[\frac{\mathrm{v}_2^n(t_n)}{\mathrm{d}^n(t_n)} - \frac{\mathrm{v}_1^n(t_n)}{\mathrm{d}^n(t_n)} \left(\frac{X_{t_k} - x}{h} \right) \right] \alpha_H^2 (t_n - t_k)^{1-2H} N\left(\frac{X_{t_k} - x}{h} \right) \mathbf{1}_{[t_k, t_{k+1})}(s).$$

Since $\hat{b}_{t_n,h}^{\mathrm{ll}}(x) = \int_0^{t_n} T_s^n \, \mathrm{d}X_s$, we use (7) and (8) to write that

$$\begin{aligned} \hat{b}_{t_n,h}^{\mathrm{ll}}(x) &= \int_0^{t_n} \frac{T_s^n}{w_H(t_n, s)} \, \mathrm{d}Y_s \\ &= \int_0^{t_n} b(X_s) T_s^n \, \mathrm{d}s + \big(\lambda_H (2 - 2H)\big)^{1/2} \int_0^{t_n} \frac{T_s^n}{w_H(t_n, s)} s^{1/2-H} \, \mathrm{d}B_s \\ &= \int_0^{t_n} b(X_s) T_s^n \, \mathrm{d}s + \alpha_H^{-1} \int_0^{t_n} T_s^n (t_n - s)^{H-1/2} \, \mathrm{d}B_s. \end{aligned}$$

As in the case of continuous observations,

$$\xi_{x,h}(X_{t_n}) = \alpha_H^{-1} \int_0^{t_n} T_s^n (t_n - s)^{H-1/2} \, \mathrm{d}B_s$$

is the first component of the two dimensional random vector

$$\Xi_{x,h}^n = \big(\mathrm{V}_{t_n}^n\big)^{-1} \int_0^{t_n} \begin{pmatrix} \mu^0(t_n, s) \\ \mu^1(t_n, s) \end{pmatrix} \mathrm{d}B_s,$$

where for $j = 0, 1$:

$$\mu^j(t_n, s) = (t_n - s)^{H-1/2} \sum_{k=1}^{n-1} \alpha_H (t_n - t_k)^{1-2H} \left(\frac{X_{t_k} - x}{h} \right)^j N\left(\frac{X_{t_k} - x}{h} \right) \mathbf{1}_{[t_k, t_{k+1})}(s).$$

The last two terms in (47) come from the equality:

$$\int_0^{t_n} T_s^n b(X_s)\,\mathrm{d}s = \int_0^{t_n} T_s^n\big(b(X_s)-b(X_{t_k})\big)\,\mathrm{d}s + \int_0^{t_n} T_s^n\big(b(X_{t_k})-b(x)\big)\,\mathrm{d}s$$
$$:= r_{x,h}^{\mathrm{traj}}(X_{t_n}) + r_{x,h}^{\mathrm{loc}}(X_{t_n}).$$

Easy but tedious computations yield that $r_{x,h}^{\mathrm{traj}}(X_{t_n})$, respectively, $r_{x,h}^{\mathrm{loc}}(X_{t_n})$, is the first component of

$$R_{x,h}^{\mathrm{traj}}(X_{t_n}) = \big(\mathrm{V}_{t_n}^n\big)^{-1}\int_0^{t_n}\begin{pmatrix}\varpi_{\mathrm{traj}}^0(t_n,s)\\ \varpi_{\mathrm{traj}}^1(t_n,s)\end{pmatrix}\mathrm{d}s,$$

respectively,

$$R_{x,h}^{\mathrm{loc}}(X_{t_n}) = \big(\mathrm{V}_{t_n}^n\big)^{-1}\int_0^{t_n}\begin{pmatrix}\varpi_{\mathrm{loc}}^0(t_n,s)\\ \varpi_{\mathrm{loc}}^1(t_n,s)\end{pmatrix}\mathrm{d}s,$$

where we have denoted for $j=0,1$:

$$\varpi_{\mathrm{traj}}^j(t_n,s) = \sum_{k=1}^{n-1}\alpha_H(t_n-t_k)^{1-2H}\left(\frac{X_{t_k}-x}{h}\right)^j N\left(\frac{X_{t_k}-x}{h}\right)\big(b(X_s)-b(X_{t_k})\big)\mathbf{1}_{[t_k,t_{k+1})}(s),$$

$$\varpi_{\mathrm{loc}}^j(t_n,s) = \sum_{k=1}^{n-1}\alpha_H(t_n-t_k)^{1-2H}\left(\frac{X_{t_k}-x}{h}\right)^j N\left(\frac{X_{t_k}-x}{h}\right)\delta_{x,h}(X_{t_k})\mathbf{1}_{[t_k,t_{k+1})}(s).$$

Remark 6. As it has been already noticed in Nadaraya–Watson's type estimation procedure (see Section 3.1), three terms appear in the decomposition (47). Each of them have a precise meaning:

- the first term, $\xi_{x,h}(X_t)$, is a "stochastic error term";
- $r_{x,h}^{\mathrm{loc}}(X_{t_k}))$ represents again the accuracy of the local approximation of b by a constant function in a neighbourhood of the point x in the discrete times $(t_k)_{0\le k\le n}$;
- $r_{x,h}^{\mathrm{traj}}(X_{t_n})$ is the error due to the discretization of the continuous process $(X_s)_{s\ge 0}$.

The next result establishes the probability deviation bound for the discrete locally linear estimator of $b(x)$.

Theorem 8. *We assume that b is Lipschitz with Lipschitz's constant L. There exists $\mathfrak{u}_1,\mathfrak{u}_2>0$ such that for any t_n large enough, we have the conditional deviation probability bound*:

$$\begin{aligned}&\mathbf{P}\big(\big|\hat b_{t_n,h}^{\mathrm{ll}}(x)-b(x)\big| \ge c_{\rho,H}\Delta_{x,h}t_n^{1-H-\beta} + c_1\varepsilon_n t_n^{1-H-\beta} + \zeta, \mathcal{A}_{t_n,h}^{\mathrm{NW}}\big)\\ &\quad\le 4\exp\left(-\frac{\rho^2(1-H)\zeta^2}{16\alpha_H^2}t_n^{2\beta}\right) + c_H t_n^{\mathfrak{u}_1}\exp\left(-\frac{\zeta}{4c_2}t_n^{\mathfrak{u}_2}\right),\end{aligned} \tag{48}$$

where we have set:

- $\mathcal{A}^{\mathrm{ll}}_{t_n,h} = \{\lambda^n_m(t_n) \geq \rho t_n^{1-H+\beta}\}$ *with* β *a positive real number such that* $1-\beta < \gamma H$;
- $\lambda^n_m(t_n)$ *denotes the smallest eigenvalue of the matrix* $\mathrm{V}^n_{t_n}$;
- $c_1 > 0$ *depends on* ρ, L, x, h *and* H;
- $c_2 > 0$ *depends on* ρ, H *and* L.

The proof of this result is done in Section 5.

All the constants in (48) are known explicitly. The interested reader shall find them in the proof. The next theorem is one of the most important result of this work since it sets convergence of the discretized estimator toward the unknown value $b(x)$.

Theorem 9. *Assume that Hypotheses* 1(b), (c) *and Hypothesis* 5 *hold. If moreover the number of approximation points satisfies* $n \asymp t_n^{\gamma}$ *with* $\gamma > \max(1 + (\mathfrak{m}^2 + 2)H; 3)$, *then the locally linear estimator of* $b(x)$ *is consistent*:

$$\hat{b}^{\mathrm{ll}}_{t_n,h}(x) \xrightarrow[t_n \to \infty, h \to 0]{\text{in probability}} b(x).$$

Proof. Since we apply Proposition 1 with $z \mapsto z^2 N(z)$ that satisfies (12) with $\mathfrak{p} = 2$, the condition on γ is justified. Then we just have to use (48) with $1 - \gamma H < \beta < 1 - H$ and we argue as in the proof of (38) form Theorem 6. □

5. Proofs

5.1. Proof of Theorem 7

We split its proof into separate steps. The starting point is the decomposition (44). We treat each term of $\hat{b}^{\mathrm{ll}}_{t_n,h}(x) - b(x)$ separately.

We recall some basic facts from linear algebra. We denote for a vector $\mathrm{z} = (z_1, z_2)^{\mathrm{T}} \in \mathbf{R}^2$, $\|\mathrm{z}\|_\infty = |z_1| \vee |z_2|$ and $\|\mathrm{z}\|_2$ its Euclidian norm. For any $t \geq 0$, $0 < \lambda_m(t) \leq \lambda_M(t)$ are the eigenvalues of the symmetric matrix V_t. For $\mathrm{y} = (y_1, y_2)^{\mathrm{T}} \in \mathbf{R}^2$, we denote $\mathrm{z} = (z_1, z_2)^{\mathrm{T}} = (\mathrm{V}_t)^{-1}\mathrm{y}$ and it holds

$$\|\mathrm{z}\|_\infty \leq \|\mathrm{z}\|_2 = \left(\frac{|y_1|^2}{\lambda_m(t)^2} + \frac{|y_2|^2}{\lambda_M(t)^2}\right)^{1/2} \leq \sqrt{2}\left(\frac{|y_1|}{\lambda_m(t)} \vee \frac{|y_2|}{\lambda_M(t)}\right). \tag{49}$$

5.1.1. *Proof of Theorem* 7(i)

We study $r^{\mathrm{loc}}_{x,h}(X_t)$. Since for any real z, $0 \leq N(z) \leq 1$ and $z^2 N(z) \leq N(z)$, we have the inequality $\mathrm{v}_2(t) \leq \mathrm{v}_0(t)$. By the Cauchy–Schwarz inequality, we obtain

$$\left|\int_0^t \alpha_H^2 (t-s)^{1-2H} \left(\frac{X_s - x}{h}\right) N\left(\frac{X_s - x}{h}\right) \delta_{x,h}(X_s)\,\mathrm{d}s\right| \leq \Delta_{x,h}\big(\mathrm{v}_0(t)\mathrm{v}_2(t)\big)^{1/2} \leq \Delta_{x,h}\mathrm{v}_0(t),$$

and thus

$$\left|\int_0^t \tilde{\nu}_1(t,s)\delta_{x,h}(X_s)\,\mathrm{d}s\right| \le \Delta_{x,h}\mathrm{v}_0(t). \tag{50}$$

The relation (49) yields

$$\begin{aligned}\left|r^{\mathrm{loc}}_{x,h}(X_t)\right| &\le \left\|R_{x,h}(X_t)\right\|_\infty \\ &\le \sqrt{2}\left(\frac{|\int_0^t \tilde{\nu}_0(t,s)\delta_{x,h}(X_s)\,\mathrm{d}s|}{\lambda_m(t)} \vee \frac{|\int_0^t \tilde{\nu}_1(t,s)\delta_{x,h}(X_s)\,\mathrm{d}s|}{\lambda_M(t)}\right) \\ &\le \sqrt{2}\Delta_{x,h}\left(\frac{\mathrm{v}_0(t)}{\lambda_m(t)} \vee \frac{\mathrm{v}_0(t)}{\lambda_M(t)}\right)\end{aligned}$$

and consequently

$$\left|r^{\mathrm{loc}}_{x,h}(X_t)\right| \le \sqrt{2}\Delta_{x,h}\frac{\mathrm{v}_0(t)}{\lambda_m(t)}.$$

Since $\mathrm{v}_0(t) \le (c_H^2/\lambda_H)t^{2-2H}$, we deduce the following bound on the random set $\mathcal{A}^{\mathrm{ll}}_{t,h}$

$$\left|r^{\mathrm{loc}}_{x,h}(X_t)\right| \le c_{\rho,H}\Delta_{x,h}t^{1-H-\beta} \tag{51}$$

with $c_{\rho,H} = \sqrt{2}c_H^2/(\rho\lambda_H)$.

For the analysis of $\xi^1_{x,h}(X_t)$, we consider $\zeta > 0$ and by (49) we may write

$$\begin{aligned}&\mathbf{P}\left(\left|\xi^1_{x,h}(X_t)\right| \ge \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) \\ &\quad\le \mathbf{P}\left(\left\|\Xi^1_{x,h}(X_t)\right\|_\infty \ge \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) \\ &\quad\le \mathbf{P}\left(\frac{|\int_0^t \nu_0(t,s)\,\mathrm{d}B_s|}{\lambda_m(t)} \vee \frac{|\int_0^t \nu_1(t,s)\,\mathrm{d}B_s|}{\lambda_M(t)} \ge \zeta/\sqrt{2}, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) \\ &\quad\le \mathbf{P}\left(\frac{|\int_0^t \nu_0(t,s)\,\mathrm{d}B_s|}{\lambda_m(t)} \ge \zeta/\sqrt{2}, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) + \mathbf{P}\left(\frac{|\int_0^t \nu_1(t,s)\,\mathrm{d}B_s|}{\lambda_M(t)} \ge \zeta/\sqrt{2}, \mathcal{A}^{\mathrm{ll}}_{t,h}\right).\end{aligned}$$

Since $\lambda_M(t) \ge \lambda_m(t) \ge \rho t^{1-H+\beta}$ on the random set $\mathcal{A}^{\mathrm{ll}}_{t,h}$, it follows that:

$$\begin{aligned}\mathbf{P}\left(\left|\xi^1_{x,h}(X_t)\right| \ge \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) &\le \mathbf{P}\left(\left|\int_0^t \nu_0(t,s)\,\mathrm{d}B_s\right| \ge \frac{\rho\zeta}{\sqrt{2}}t^{1-H+\beta}\right) \\ &\quad + \mathbf{P}\left(\left|\int_0^t \nu_1(t,s)\,\mathrm{d}B_s\right| \ge \frac{\rho\zeta}{\sqrt{2}}t^{1-H+\beta}\right).\end{aligned} \tag{52}$$

For $j = 0, 1$, $|(\frac{X_s-x}{h})^j N(\frac{X_s-x}{h})| \le 1$. Then we may apply the exponential inequality (22) and we obtain

$$\mathbf{P}\left(\left|\xi^1_{x,h}(X_t)\right| \ge \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\right) \le 4\exp\left(-\frac{(1-H)\rho^2}{2\alpha_H^2}\zeta^2 t^{2\beta}\right). \tag{53}$$

Thanks to the decomposition (44) and the bounds (51) and (53), we deduce that

$$\begin{aligned}
&\mathbf{P}\big(\big|\hat{b}^{\mathrm{ll}}_{t,h}(x) - b(x)\big| \geq c_{\rho,H}\Delta_{x,h}t^{1-H-\beta} + \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\big) \\
&\quad \leq \mathbf{P}\big(\big|\xi^1_{x,h}(X_t)\big| + \big|r^{\mathrm{loc}}_{x,h}(X_t)\big| \geq c_{\rho,H}\Delta_{x,h}t^{1-H-\beta} + \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\big) \\
&\quad \leq \mathbf{P}\big(\big|\xi^1_{x,h}(X_t)\big| \geq \zeta, \mathcal{A}^{\mathrm{ll}}_{t,h}\big) \\
&\quad \leq 4\exp\biggl(-\frac{(1-H)\rho^2}{2\alpha_H^2}\zeta^2 t^{2\beta}\biggr)
\end{aligned}$$

and the proof of (45) is now completed.

5.1.2. *Proof of Theorem* 7(ii)

We follow the same arguments that the ones used in the proof of Theorem 6. With $\beta < 1 - H$, we need to show that

$$\mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{ll}}_{t,h}\big) = \mathbf{P}\big(\lambda_m(t) \leq \rho t^{1-H+\beta}\big) \underset{t\to\infty}{\longrightarrow} 0. \tag{54}$$

Since $0 \leq N \leq 1$ and $z^2 N(z) \leq N(z)$ for any real z, $\lambda_m(t) \geq \mathrm{v}_2(t)$ by (46). Thus,

$$\begin{aligned}
\mathbf{P}\big(\Omega \setminus \mathcal{A}^{\mathrm{ll}}_{t,h}\big) &\leq \mathbf{P}\big(\mathrm{v}_2(t) \leq \rho t^{1-H+\beta}\big) \\
&\leq \mathbf{P}\biggl(\int_0^t \alpha_H^2 (t-s)^{1-2H}\biggl(\frac{X_s - x}{h}\biggr)^2 N\biggl(\frac{X_s - x}{h}\biggr)\,\mathrm{d}s \leq \rho t^{1-H+\beta}\biggr).
\end{aligned}$$

When $H > 1/2$, $(t-s)^{1-2H} \geq t^{1-2H}$ and consequently

$$\begin{aligned}
&\biggl\{\int_0^t \alpha_H^2 (t-s)^{1-2H}\biggl(\frac{X_s - x}{h}\biggr)^2 N\biggl(\frac{X_s - x}{h}\biggr)\,\mathrm{d}s \leq \rho t^{1-H+\beta}\biggr\} \\
&\quad \subseteq \biggl\{\frac{1}{t}\int_0^t \biggl(\frac{X_s - x}{h}\biggr)^2 N\biggl(\frac{X_s - x}{h}\biggr)\,\mathrm{d}s \leq \frac{\rho}{\alpha_H^2} t^{H+\beta-1}\biggr\}.
\end{aligned}$$

Since Hypothesis 5 holds, we may apply Proposition 1. So (54) is true and the result is proved for $H > 1/2$. When $H < 1/2$, we use $(t-s)^{1-2H} \geq (t - \frac{t}{2})^{1-2H} = (\frac{t}{2})^{1-2H}$ and we write

$$\begin{aligned}
&\biggl\{\int_0^t \alpha_H^2 (t-s)^{1-2H}\biggl(\frac{X_s - x}{h}\biggr)^2 N\biggl(\frac{X_s - x}{h}\biggr)\,\mathrm{d}s \leq \rho t^{1-H+\beta}\biggr\} \\
&\quad \subseteq \biggl\{\frac{1}{t/2}\int_0^{t/2} \biggl(\frac{X_s - x}{h}\biggr)^2 N\biggl(\frac{X_s - x}{h}\biggr)\,\mathrm{d}s \leq \frac{4\rho}{\alpha_H^2}\biggl(\frac{t}{2}\biggr)^{H+\beta-1}\biggr\}.
\end{aligned}$$

The concluding arguments are unchanged.

5.2. Proof of Theorem 8

We treat separately each term of the decomposition (47). Repeating the arguments that led us to (52) yields that on $\mathcal{A}^{\mathrm{ll}}_{t_n,h}$

$$\mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big| \geq \zeta, \mathcal{A}^{\mathrm{ll}}_{t_n,h}\big) \leq \mathbf{P}\left(\left|\int_0^{t_n} \mu^0(t_n,s)\,\mathrm{d}B_s\right| \geq \frac{\rho\zeta}{\sqrt{2}} t_n^{1-H+\beta}\right) + \mathbf{P}\left(\left|\int_0^{t_n} \mu^1(t_n,s)\,\mathrm{d}B_s\right| \geq \frac{\rho\zeta}{\sqrt{2}} t_n^{1-H+\beta}\right).$$

We follow the arguments that led us to (25). We fix t_n and for $j = 0, 1$, we consider the martingales $Z^{n,j} := (Z_r^{n,j})_{0\leq r\leq t_n}$ defined by

$$Z_r^{n,j} = \int_0^r \mu^j(t_n,s)\,\mathrm{d}B_s.$$

Since $|(\frac{X_s-x}{h})^j N(\frac{X_s-x}{h})| \leq 1$, the quadratic variations of the martingales $Z^{n,j}$ satisfy

$$\langle Z^{n,j}\rangle_r \leq \frac{2\alpha_H^2}{1-H} t_n^{2-2H}.$$

Then by Lemma 3 we obtain that

$$\mathbf{P}\big(\big|\xi_{x,h}(X_{t_n})\big| \geq \zeta, \mathcal{A}^{\mathrm{ll}}_{t_n,h}\big) \leq 4\exp\left(-\frac{\rho^2(1-H)\zeta^2}{8\alpha_H^2} t_n^{2\beta}\right). \tag{55}$$

Now we deal with $r^{\mathrm{traj}}_{x,h}(X_{t_n})$ and $r^{\mathrm{loc}}_{x,h}(X_{t_n})$. For $j = 0, 1$ we have a discrete version of (50)

$$\left|\int_0^t \varpi^j_{\mathrm{loc}}(t_n,s)\,\mathrm{d}s\right| \leq \Delta_{x,h}\mathrm{v}_0^n(t_n).$$

Arguing as in the proof of (51),

$$\big|r^{\mathrm{loc}}_{x,h}(X_{t_n})\big| \leq c_{\rho,H}\Delta_{x,h} t_n^{1-H-\beta} \tag{56}$$

on the random set $\mathcal{A}^{\mathrm{ll}}_{t_n,h}$ (we recall that $c_{\rho,H} = \sqrt{2}c_H^2/(\rho\lambda_H)$). We use (27) and we obtain similarly,

$$\left|\int_0^t \varpi^j_{\mathrm{traj}}(t_n,s)\,\mathrm{d}s\right| \leq \big(c_{x,h,L}\varepsilon_n + L\varepsilon_n^{\mathfrak{h}}\|B^H\|_{0,t_n,\mathfrak{h}}\big)\mathrm{v}_0^n(t_n),$$

with $c_{x,h,L} = L\{Lh + c_b(1+|x|^{\mathrm{m}})\}$. Finally on the set $\mathcal{A}^{\mathrm{ll}}_{t_n,h}$ it holds

$$\big|r^{\mathrm{traj}}_{x,h}(X_{t_n})\big| \leq \big(c_1\varepsilon_n + c_2\varepsilon_n^{\mathfrak{h}}\|B^H\|_{0,t_n,\mathfrak{h}}\big) t_n^{1-H-\beta} \tag{57}$$

with $c_1 = c_{\rho,H} c_{x,h,L}$ and $c_2 = c_{\rho,H} L$. Like in the proof of (21) (see Section 3.2), we combine the inequalities (55), (56) and (57) and we deduce:

$$\mathbf{P}\big(\big|\hat{b}^{\mathrm{ll}}_{t_n,h}(x) - b(x)\big| \geq c_{\rho,H}\Delta_{x,h} t_n^{1-H-\beta} + c_1 \varepsilon_n t_n^{1-H-\beta} + \zeta, \mathcal{A}^{\mathrm{NW}}_{t_n,h}\big) \\ \leq 4\exp\left(-\frac{\rho^2(1-H)\zeta^2}{16\alpha_H^2} t_n^{2\beta}\right) + \mathbf{P}\left(\|B^H\|_{0,t_n,\mathfrak{h}} \geq \frac{\zeta}{2c_2 \varepsilon_n^{\mathfrak{h}} t_n^{1-H-\beta}}\right). \tag{58}$$

By Lemma 10, we have

$$\mathbf{P}\left(\|B^H\|_{0,t_n,\mathfrak{h}} \geq \frac{\zeta}{2c_2\varepsilon_n^{\mathfrak{h}} t_n^{1-H-\beta}}\right) \\ \leq c_{H,\mathfrak{h}}\big(1 + t_n^{H-\mathfrak{h}}\big)\exp\left(-\frac{\zeta}{2c_2} t_n^{\mathfrak{h}(\gamma-1)+H-1+\beta}\left(1 - \frac{256H^2 c_2}{\zeta\mathfrak{h}^2} t_n^{2(H-\mathfrak{h})} t_n^{-\mathfrak{h}(\gamma-1)-H+1-\beta}\right)\right) \\ \leq c_{H,\mathfrak{h}}\big(1 + t_n^{H-\mathfrak{h}}\big)\exp\left(-\frac{\zeta}{2c_2} t_n^{\mathfrak{h}(\gamma-1)+H-1+\beta}\left(1 - \frac{256H^2 c_2}{\zeta\mathfrak{h}^2} t_n^{-\mathfrak{h}(\gamma+1)+H+1-\beta}\right)\right).$$

With $\beta < 1 - \gamma H$ one may choose $\mathfrak{h}$ such that

$$\max\left(\frac{1-H-\beta}{\gamma-1}, \frac{1+H-\beta}{\gamma+1}\right) < \mathfrak{h} < H.$$

When t_n is large enough, we obtain

$$\mathbf{P}\left(\|B^H\|_{0,t_n,\mathfrak{h}} \geq \frac{\zeta}{2c_2\varepsilon_n^{\mathfrak{h}} t_n^{1-H-\beta}}\right) \leq 2c_{H,\mathfrak{h}} t_n^{H-\mathfrak{h}} \exp\left(-\frac{\zeta}{4c_2} t_n^{\mathfrak{h}(\gamma-1)+H-1+\beta}\right).$$

We report the above estimation in (58) and the proof is complete when we set $\mathfrak{u}_1 = H - \mathfrak{h}$ and $\mathfrak{u}_1 = \mathfrak{h}(\gamma-1) + H - 1 + \beta$.

Appendix A: A Fernique's type lemma

The exponential moments of the Hölder norm of the trajectories of a fBm are classical results from the theory of Gaussian processes (see [7], e.g.). Nevertheless we are interested in the large time behaviour of this moment. So we prove in this Appendix the following Fernique's type lemma in which we give precision on the time dependence of the estimation.

Lemma 10. *Let $T > 0$, $0 < \mathfrak{h} < H < 1$. We denote*

$$\|B^H\|_{0,T,\mathfrak{h}} = \sup_{0\leq s,t\leq T} \frac{|B_t^H - B_s^H|}{|t-s|^{\mathfrak{h}}}.$$

Then for any $T > T_0 = (\mathfrak{h}/(8H))^{1/(H-\mathfrak{h})}$ *there exists a constant* $c_{H,\mathfrak{h}}$ *such that*

$$\mathbf{E}\big[\exp\big(\|B^H\|_{0,T,\mathfrak{h}}\big)\big] \le c_{H,\mathfrak{h}}\big(1+T^{H-\mathfrak{h}}\big)\exp\left(\frac{128H^2}{\mathfrak{h}^2}T^{2(H-\mathfrak{h})}\right). \tag{59}$$

The explicit form of $c_{H,\mathfrak{h}}$ is given in (64).

Proof of Lemma 10. We denote $[z]$ the integer part of a non-negative real z. First, we prove that

$$\big|B_t^H - B_s^H\big| \le \xi_{H,\mathfrak{h},T}|t-s|^{\mathfrak{h}}, \tag{60}$$

where $\xi_{H,\mathfrak{h},T}$ is a positive random variable such that for any $p > p_0 := [2/(H-\mathfrak{h})]$

$$\mathbf{E}\big(\xi_{H,\mathfrak{h},T}^p\big) \le \left(16\frac{H}{\mathfrak{h}}\right)^p T^{p(H-\mathfrak{h})}(p-1)!!. \tag{61}$$

The double factorial of a positive integer p is defined by

$$\begin{cases} (2k-1)!! = \displaystyle\prod_{i=1}^{k}(2i-1) = \frac{(2k)!}{2^k k!}; \\ (2k)!! = \displaystyle\prod_{i=1}^{k}(2i) = 2^k k!. \end{cases}$$

We remark that when $p \le p_0$, we obtain easily that

$$\mathbf{E}\big(\xi_{H,\mathfrak{h},T}^p\big) \le \left(16\frac{H}{\mathfrak{h}}\right)^p T^{p(H-\mathfrak{h})}p_0!!. \tag{62}$$

In order to prove (60) and (61), we proceed as follows. With $\psi(u) = u^{2/(H-\mathfrak{h})}$ and $p(u) = u^H$ in Lemma 1.1 of [9], the Garsia–Rodemich–Rumsey inequality reads

$$\big|B_t^H - B_s^H\big| \le 8\int_0^{|t-s|}\left(\frac{4\mathfrak{D}}{u^2}\right)^{(H-\mathfrak{h})/2} Hu^{H-1}\,\mathrm{d}u,$$

where the random variable $\mathfrak{D}$ is

$$\mathfrak{D} = \int_0^T\int_0^T \frac{|B_t^H - B_s^H|^{2/(H-\mathfrak{h})}}{|t-s|^{2H/(H-\mathfrak{h})}}\,\mathrm{d}t\,\mathrm{d}s.$$

We have

$$\begin{aligned}\big|B_t^H - B_s^H\big| &\le 8(4\mathfrak{D})^{(H-\mathfrak{h})/2}\int_0^{|t-s|} Hu^{\mathfrak{h}-1}\,\mathrm{d}u \\ &\le 8\frac{H}{\mathfrak{h}}(4\mathfrak{D})^{(H-\mathfrak{h})/2}|t-s|^{\mathfrak{h}}.\end{aligned}$$

We denote $\xi_{H,\mathfrak{h},T} = 8\frac{H}{\mathfrak{h}}(4\mathfrak{D})^{(H-\mathfrak{h})/2}$. By Jensen's inequality, for $p \geq 2/(H-\mathfrak{h})$ it holds

$$\begin{aligned}
\mathbf{E}\big(\xi_{H,\mathfrak{h},T}^p\big) &\leq \left(84^{(H-\mathfrak{h})/2}\frac{H}{\mathfrak{h}}\right)^p \mathbf{E}\left(\int_0^T\int_0^T \frac{|B_t^H - B_s^H|^{2/(H-\mathfrak{h})}}{|t-s|^{2H/(H-\mathfrak{h})}}\,\mathrm{d}t\,\mathrm{d}s\right)^{p(H-\mathfrak{h})/2} \\
&\leq \left(16\frac{H}{\mathfrak{h}}\right)^p T^{p(H-\mathfrak{h})}\int_0^T\int_0^T \frac{\mathbf{E}(|B_t^H - B_s^H|^{2p})}{|t-s|^{2pH}}\frac{\mathrm{d}t\,\mathrm{d}s}{T^2} \\
&\leq \left(16\frac{H}{\mathfrak{h}}\right)^p T^{p(H-\mathfrak{h})}\mathbf{E}\big(|Z|^p\big),
\end{aligned}$$

where Z is a Gaussian random variable with zero mean and unit variance. Since

$$\mathbf{E}\big(|Z|^p\big) = \begin{cases} \sqrt{2/\pi}\,(p-1)!!, & \text{when } p \text{ is odd;} \\ (p-1)!!, & \text{when } p \text{ is even,} \end{cases}$$

we deduce (60) and (61).

What remains to be shown can be tediously deduced from Theorem 1.3.2 in [7]. We can also make the following direct computations. Using (60), (61) and (62), we have

$$\begin{aligned}
\mathbf{E}\big(\exp\big(\|B^H\|_{0,T,\mathfrak{h}}\big)\big) &\leq \mathbf{E}\big(\exp(\xi_{H,\mathfrak{h},T})\big) \\
&\leq \sum_{p=0}^{\infty}\frac{\mathbf{E}(\xi_{H,\mathfrak{h},T}^p)}{p!} \\
&\leq p_0!!\sum_{p=0}^{p_0}\frac{c_{H,\mathfrak{h},T}^p}{p!} + \sum_{p=p_0+1}^{\infty} c_{H,\mathfrak{h},T}^p\frac{(p-1)!!}{p!} \\
&\leq p_0!!\exp(c_{H,\mathfrak{h},T}) + \sum_{p=0}^{\infty} c_{H,\mathfrak{h},T}^p\frac{(p-1)!!}{p!}
\end{aligned}$$

where we have denoted $c_{H,\mathfrak{h},T} = 16(H/\mathfrak{h})T^{H-\mathfrak{h}}$. We notice that

$$(p-1)!!/p! = \begin{cases} 1/\big(2^k k\big)!, & \text{when } p = 2k; \\ 1/(2k+1)!!, & \text{when } p = 2k+1. \end{cases}$$

Since $(2k+1)!! \geq \prod_{i=1}^k 2k = 2^k k!$, we obtain

$$\begin{aligned}
\mathbf{E}\big(\exp\big(\|B^H\|_{0,T,\mathfrak{h}}\big)\big) &\leq p_0!!\exp(c_{H,\mathfrak{h},T}) + \sum_{k=0}^{\infty}\frac{c_{H,\mathfrak{h},T}^{2k}}{2^k k!} + \sum_{k=0}^{\infty}\frac{c_{H,\mathfrak{h},T}^{2k+1}}{2^k k!} \\
&\leq p_0!!\exp(c_{H,\mathfrak{h},T}) + \sum_{k=0}^{\infty}\left(\frac{c_{H,\mathfrak{h},T}^2}{2}\right)^k\frac{1}{k!} + c_{H,\mathfrak{h},T}\sum_{k=0}^{\infty}\left(\frac{c_{H,\mathfrak{h},T}^2}{2}\right)^k\frac{1}{k!} \qquad (63) \\
&\leq p_0!!\exp(c_{H,\mathfrak{h},T}) + (1 + c_{H,\mathfrak{h},T})\exp\big(c_{H,\mathfrak{h},T}^2/2\big).
\end{aligned}$$

With

$$c_{H,\mathfrak{h}} = \left[\frac{2}{H-\mathfrak{h}}\right] + \frac{16H}{\mathfrak{h}}, \tag{64}$$

the lemma is proved because when $T^{H-\mathfrak{h}} \geq \mathfrak{h}/(8H)$, $c_{H,\mathfrak{h},T} \leq c^2_{H,\mathfrak{h},T}/2$. □

We remark that if T do not satisfy the condition $T \geq \mathfrak{h}/(8H)$, then one may replace (59) by (63).

Remark 7. Thanks to (60) and (61), we have obtained the following estimation for the moments of the Hölder norm of the trajectories of the fBm:

$$\mathbf{E}\left(\|B^H\|^p_{0,T,\mathfrak{h}}\right) \leq \left(16\frac{H}{\mathfrak{h}}\right)^p T^{p(H-\mathfrak{h})}(p-1)!! \tag{65}$$

for any $p > p_0 := [2/(H-\mathfrak{h})]$.

Appendix B: Proof of Proposition 1

Step 1: *Proof of* (13). We use the inequality

$$\begin{aligned}\left|\frac{1}{T}\int_0^T \varphi(X_t)\,\mathrm{d}t - \mathbf{E}\big(\varphi(\bar{X})\big)\right| &\leq \frac{1}{T}\int_0^T \left|\varphi(X_t) - \varphi\big(\bar{X}(\theta_t)\big)\right|\mathrm{d}t \\ &\quad + \left|\frac{1}{T}\int_0^T \varphi\big(\bar{X}(\theta_t)\big)\,\mathrm{d}t - \mathbf{E}\big(\varphi(\bar{X})\big)\right|.\end{aligned} \tag{66}$$

Since φ has polynomial growth and $\bar{X}$ has moment of any order, $\varphi(\bar{X})$ is an integrable random variable and (10) implies that the second term in the right-hand side of (66) tends to 0 almost-surely. Now we treat the first one. The inequalities

$$\begin{aligned}\left|\varphi(X_t) - \varphi\big(\bar{X}(\theta_t)\big)\right| &\leq c_\varphi\big(1 + |X_t|^{\mathfrak{p}} + \left|\bar{X}(\theta_t)\right|^{\mathfrak{p}}\big)\left|X_t - \bar{X}(\theta_t)\right| \\ &\leq c_{\varphi,\mathfrak{p}}\big(1 + \left|X_t - \bar{X}(\theta_t)\right|^{\mathfrak{p}} + 2\left|\bar{X}(\theta_t)\right|^{\mathfrak{p}}\big)\left|X_t - \bar{X}(\theta_t)\right|\end{aligned}$$

imply that

$$\begin{aligned}\frac{1}{T}\int_0^T \left|\varphi(X_t) - \varphi\big(\bar{X}(\theta_t)\big)\right|\mathrm{d}t &\leq \frac{c_{\varphi,\mathfrak{p}}}{T}\int_0^T \left|X_t - \bar{X}(\theta_t)\right|^{\mathfrak{p}+1}\mathrm{d}t \\ &\quad + \frac{c_{\varphi,\mathfrak{p}}}{T}\int_0^T \left|\bar{X}(\theta_t)\right|^{\mathfrak{p}}\left|X_t - \bar{X}(\theta_t)\right|\mathrm{d}t.\end{aligned} \tag{67}$$

By (11), $|X_t - \bar{X}(\theta_t)|^{\mathfrak{p}+1}$ tends to 0 almost-surely and an integral version of the Toeplitz lemma implies that the first term in the right-hand side of (67) tends to 0 almost-surely. For the second one, by the Cauchy–Schwarz inequality,

$$\left(\frac{1}{T}\int_0^T \left|\bar{X}(\theta_t)\right|^{\mathfrak{p}}\left|X_t-\bar{X}(\theta_t)\right| \mathrm{d}t\right)^2 \le \left(\frac{1}{T}\int_0^T \left|\bar{X}(\theta_t)\right|^{2\mathfrak{p}} \mathrm{d}t\right) \times \left(\frac{1}{T}\int_0^T \left|X_t-\bar{X}(\theta_t)\right|^2 \mathrm{d}t\right)$$

and thus it tends to 0 by the same arguments that we employed before. The proof of (13) is complete.

Step 2: *Proof of* (14). First, we write

$$\left|\frac{1}{t_n}\int_0^{t_n}\left\{\sum_{k=0}^{n-1}\varphi(X_{t_k})\mathbf{1}_{[t_k,t_{k+1})}(s)\right\}\mathrm{d}s-\mathbf{E}\big(\varphi(\bar{X})\big)\right| \le I_n^1+I_n^2$$

with

$$I_n^1=\frac{1}{t_n}\int_0^{t_n}\left\{\sum_{k=0}^{n-1}\left|\varphi(X_{t_k})-\varphi(X_s)\right|\mathbf{1}_{[t_k,t_{k+1})}(s)\right\}\mathrm{d}s \quad \text{and}$$

$$I_n^2=\left|\frac{1}{t_n}\int_0^{t_n}\varphi(X_s)\,\mathrm{d}s-\mathbf{E}\big(\varphi(\bar{X})\big)\right|.$$

By (13), $\lim_{n\to\infty} I_n^2=0$ almost-surely. We estimate I_n^1 as follows. First, of all we write:

$$\begin{aligned} I_n^1 &\le \frac{c_\varphi}{t_n}\int_0^{t_n}\left\{\sum_{k=0}^{n-1}\big(1+\left|X_{t_k}\right|^{\mathfrak{p}}+\left|X_s\right|^{\mathfrak{p}}\big)|X_{t_k}-X_s|\mathbf{1}_{[t_k,t_{k+1})}(s)\right\}\mathrm{d}s \\ &\le \frac{c_\varphi}{t_n}\Big(1+\sup_{0\le u\le t_n}|X_u|^{\mathfrak{p}}\Big)\int_0^{t_n}\left\{\sum_{k=0}^{n-1}|X_{t_k}-X_s|\mathbf{1}_{[t_k,t_{k+1})}(s)\right\}\mathrm{d}s. \end{aligned} \tag{68}$$

Since b satisfies the polynomial growth condition, it holds for $t_k\le s\le t_{k+1}$ that

$$\begin{aligned} |X_s-X_{t_k}| &\le \int_{t_k}^s\left|b(X_u)\right|\mathrm{d}u+\left|B_s^H-B_{t_k}^H\right| \\ &\le c_b\int_{t_k}^s\big(1+|X_u|^{\mathfrak{m}}\big)\,\mathrm{d}u+\left\|B^H\right\|_{0,t_n,\mathfrak{h}}|s-t_k|^{\mathfrak{h}} \\ &\le c_b\Big(1+\Big(\sup_{0\le u\le t_n}|X_u|\Big)^{\mathfrak{m}}\Big)\varepsilon_n+\left\|B^H\right\|_{0,t_n,\mathfrak{h}}\varepsilon_n^{\mathfrak{h}}. \end{aligned} \tag{69}$$

Under the one-sided dissipative Lipschitz condition, Proposition 1 in [8] establishes that

$$\begin{aligned} \sup_{0\le u\le t_n}|X_u| &\le c_b\Big(1+|x_0|+\Big(\sup_{0\le u\le t_n}\left|B_u^H\right|\Big)^{\mathfrak{m}}\Big) \\ &\le c_b\big(1+|x_0|+\left\|B^H\right\|_{0,t_n,\mathfrak{h}}^{\mathfrak{m}}t_n^{\mathfrak{m}\mathfrak{h}}\big), \end{aligned}$$

with $0 < \mathfrak{h} < H$ that will be fixed later. We report the above inequality in (69):

$$|X_s - X_{t_k}| \le c_{b,x_0}\varepsilon_n + c_b \|B^H\|_{0,t_n,\mathfrak{h}}^{\mathfrak{m}^2} t_n^{\mathfrak{m}^2\mathfrak{h}}\varepsilon_n + \|B^H\|_{0,t_n,\mathfrak{h}}\varepsilon_n^{\mathfrak{h}}, \qquad t_k \le s \le t_{k+1}. \tag{70}$$

Using (70) in (68), we deduce that there exist a constant C that depends on b, φ and x_0 such that

$$\begin{aligned} I_n^1 &\le C\big(1 + \|B^H\|_{0,t_n,\mathfrak{h}}^{\mathfrak{p}} t_n^{\mathfrak{p}\mathfrak{h}}\big) \times \big(\varepsilon_n + \|B^H\|_{0,t_n,\mathfrak{h}}^{\mathfrak{m}^2} t_n^{\mathfrak{m}^2\mathfrak{h}}\varepsilon_n + \|B^H\|_{0,t_n,\mathfrak{h}}\varepsilon_n^{\mathfrak{h}}\big) \\ &\le C\big(\varepsilon_n + \|B^H\|_{0,t_n,\mathfrak{h}}\varepsilon_n^{\mathfrak{h}} + \|B^H\|_{0,t_n,\mathfrak{h}}^{\mathfrak{p}+1} t_n^{\mathfrak{p}\mathfrak{h}}\varepsilon_n^{\mathfrak{h}} + \|B^H\|_{0,t_n,\mathfrak{h}}^{\mathfrak{m}^2+\mathfrak{p}} t_n^{(\mathfrak{m}^2+\mathfrak{p})\mathfrak{h}}\varepsilon_n\big). \end{aligned}$$

The almost-sure convergence of I_n^1 to 0 will follow from a Borel–Cantelli argument. Indeed for any $\eta > 0$ and for an integer q that will be chosen later, it holds that

$$\begin{aligned} \mathbf{P}\big(|I_n^1| \ge \eta\big) \le \frac{C}{\eta}\big(&\varepsilon_n^q + \varepsilon_n^{\mathfrak{h}q}\mathbf{E}\big(\|B^H\|_{0,t_n,\mathfrak{h}}^q\big) + t_n^{\mathfrak{h}q\mathfrak{p}}\varepsilon_n^{q\mathfrak{h}}\mathbf{E}\big(\|B^H\|_{0,t_n,\mathfrak{h}}^{q(\mathfrak{p}+1)}\big) \\ &+ t_n^{\mathfrak{h}q(\mathfrak{m}^2+\mathfrak{p})}\varepsilon_n^q\mathbf{E}\big(\|B^H\|_{0,t_n,\mathfrak{h}}^{q(\mathfrak{m}^2+\mathfrak{p})}\big)\big) \end{aligned}$$

and by (65) (see Remark 7 in Appendix A) we obtain

$$\begin{aligned} \mathbf{P}\big(|I_n^1| \ge \eta\big) &\le \frac{C}{\eta}\big(\varepsilon_n^q + \varepsilon_n^{\mathfrak{h}q} t_n^{q(H-\mathfrak{h})} + t_n^{\mathfrak{h}q\mathfrak{p}}\varepsilon_n^{q\mathfrak{h}} t_n^{q(\mathfrak{p}+1)(H-\mathfrak{h})} + t_n^{\mathfrak{h}q(\mathfrak{m}^2+\mathfrak{p})}\varepsilon_n^q t_n^{q(\mathfrak{m}^2+\mathfrak{p})(H-\mathfrak{h})}\big) \\ &\le \frac{C}{\eta}\big(\varepsilon_n^q + \varepsilon_n^{\mathfrak{h}q} t_n^{q(H-\mathfrak{h})} + \varepsilon_n^{q\mathfrak{h}} t_n^{q(\mathfrak{p}+1)H-q\mathfrak{h}} + \varepsilon_n^q t_n^{q(\mathfrak{m}^2+\mathfrak{p})H}\big). \end{aligned}$$

Since $t_n^\gamma = n$ and $\varepsilon_n = n^{-(\gamma-1)/\gamma}$, we have $\sum_{n\ge1}\mathbf{P}(|I_n^1| \ge \eta) \le \frac{C}{\eta}(S_1 + S_2 + S_3 + S_4)$ with

$$\begin{aligned} S_1 &= \sum_{n\ge1}\frac{1}{n^{q(\gamma-1)/\gamma}}, \\ S_2 &= \sum_{n\ge1}\frac{1}{n^{q(\mathfrak{h}\gamma-H)/\gamma}}, \\ S_3 &= \sum_{n\ge1}\frac{1}{n^{q(\mathfrak{h}\gamma-(\mathfrak{p}+1)H)/\gamma}} \quad \text{and} \\ S_4 &= \sum_{n\ge1}\frac{1}{n^{q(\gamma-1-(\mathfrak{m}^2+\mathfrak{p})H)/\gamma}}. \end{aligned}$$

It is supposed that $\gamma > 1 + (\mathfrak{m}^2 + \mathfrak{p})H$. We choose $\mathfrak{h}$ close to H in such a way that $\mathfrak{h}\gamma - H > 0$. Moreover, since $\gamma > \mathfrak{p} + 1$, one may choose $\mathfrak{h}$ such that it satisfies additionally $\gamma H > \gamma\mathfrak{h} > (\mathfrak{p} + 1)H$. Now it is clear that we may find an integer q in such a way that the three above sums converge. The Borel–Cantelli lemma yields that I_n^1 converges to 0 almost-surely.

References

[1] Begyn, A. (2005). Quadratic variations along irregular subdivisions for Gaussian processes. *Electron. J. Probab.* **10** 691–717 (electronic). MR2164027

[2] Belfadli, R., Es-Sebaiy, K. and Ouknine, Y. (2011). Parameter estimation for fractional Ornstein–Uhlenbeck processes: Non-ergodic case. Available at arXiv:1102.5491.

[3] Bishwal, J.P.N. (2008). *Parameter Estimation in Stochastic Differential Equations*. *Lecture Notes in Math.* **1923**. Berlin: Springer. MR2360279

[4] Chronopoulou, A. and Tindel, S. (2013). On inference for fractional differential equations. *Stat. Inference Stoch. Process.* **16** 29–61. MR3029332

[5] Coeurjolly, J.F. (2001). Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.* **4** 199–227. MR1856174

[6] Crimaldi, I. and Pratelli, L. (2005). Convergence results for multivariate martingales. *Stochastic Process. Appl.* **115** 571–577. MR2128630

[7] Fernique, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École D'Été de Probabilités de Saint-Flour, IV*-1974 1–96. *Lecture Notes in Math.* **480**. Berlin: Springer. MR0413238

[8] Garrido-Atienza, M.J., Kloeden, P.E. and Neuenkirch, A. (2009). Discretization of stationary solutions of stochastic systems driven by fractional Brownian motion. *Appl. Math. Optim.* **60** 151–172. MR2524684

[9] Garsia, A.M., Rodemich, E. and Rumsey, H. Jr. (1970). A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Univ. Math. J.* **20** 565–578.

[10] Hairer, M. (2005). Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.* **33** 703–758. MR2123208

[11] Hairer, M. and Pillai, N.S. (2011). Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 601–628. MR2814425

[12] Hu, Y. and Nualart, D. (2010). Parameter estimation for fractional Ornstein–Uhlenbeck processes. *Statist. Probab. Lett.* **80** 1030–1038. MR2638974

[13] Hu, Y., Nualart, D. and Song, J. (2009). Fractional martingales and characterization of the fractional Brownian motion. *Ann. Probab.* **37** 2404–2430. MR2573562

[14] Istas, J. and Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Henri Poincaré Probab. Stat.* **33** 407–436. MR1465796

[15] Kleptsyna, M.L. and Le Breton, A. (2002). Statistical analysis of the fractional Ornstein–Uhlenbeck type process. *Stat. Inference Stoch. Process.* **5** 229–248. MR1943832

[16] Küchler, U. and Sørensen, M. (1999). A note on limit theorems for multivariate martingales. *Bernoulli* **5** 483–493. MR1693604

[17] Kutoyants, Y.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. London: Springer. MR2144185

[18] Le Breton, A. and Musiela, M. (1986). Une loi des grands nombres pour les martingales locales continues vectorielles et son application en régression linéaire stochastique. *C. R. Acad. Sci. Paris Sér. I Math.* **303** 421–424. MR0862208

[19] Löcherbach, E. and Loukianova, D. (2008). On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stochastic Process. Appl.* **118** 1301–1321. MR2427041

[20] Loukianova, D. and Loukianov, O. (2008). Uniform deterministic equivalent of additive functionals and non-parametric drift estimation for one-dimensional recurrent diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 771–786. MR2446297

[21] Mishra, M.N. and Prakasa Rao, B.L.S. (2011). Nonparametric estimation of trend for stochastic differential equations driven by fractional Brownian motion. *Stat. Inference Stoch. Process.* **14** 101–109. MR2794957

[22] Mishura, Y.S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. *Lecture Notes in Math*. **1929**. Berlin: Springer. MR2378138
[23] Neuenkirch, A. and Tindel, S. (2011). A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise. Available at arXiv:1111.1816.
[24] Norros, I., Valkeila, E. and Virtamo, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli* **5** 571–587. MR1704556
[25] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. *Probability and Its Applications* (*New York*). Berlin: Springer. MR2200233
[26] Nualart, D. and Ouknine, Y. (2002). Regularization of differential equations by fractional noise. *Stochastic Process*. *Appl*. **102** 103–116. MR1934157
[27] Papavasiliou, A. and Ladroue, C. (2011). Parameter estimation for rough differential equations. *Ann*. *Statist*. **39** 2047–2073. MR2893861
[28] Prakasa Rao, B.L.S. (2010). *Statistical Inference for Fractional Diffusion Processes*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2778592
[29] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Berlin: Springer. MR1725357
[30] Saussereau, B. (2012). Deviation probability bounds for fractional martingales and related remarks. *Statist*. *Probab*. *Lett*. **82** 1610–1618. MR2930666
[31] Spokoiny, V.G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann*. *Statist*. **28** 815–836. MR1792788
[32] Tudor, C.A. and Viens, F.G. (2007). Statistical aspects of the fractional stochastic calculus. *Ann*. *Statist*. **35** 1183–1212. MR2341703
[33] Tudor, C.A. and Viens, F.G. (2007). Statistical aspects of the fractional stochastic calculus. Available at http://hal.archives-ouvertes.fr/hal-00130622/.
[34] van Zanten, H. (2000). A multivariate central limit theorem for continuous local martingales. *Statist*. *Probab*. *Lett*. **50** 229–235. MR1792301

Received May 2012 and revised December 2012