On inference for fractional differential equations

Alexandra Chronopoulou, Samy Tindel

ON INFERENCE FOR FRACTIONAL DIFFERENTIAL EQUATIONS

ALEXANDRA CHRONOPOULOU AND SAMY TINDEL

Abstract. Based on Malliavin calculus tools and approximation results, we show how to compute a maximum likelihood type estimator for a rather general differential equation driven by a fractional Brownian motion with Hurst parameter $H > 1/2$. Rates of convergence for the approximation task are provided, and numerical experiments show that our procedure leads to good results in terms of estimation.

1. Introduction

In this introduction, we first try to motivate our problem and outline our results. We also argue that only a part of the question can be dealt with in a single paper. We briefly sketch a possible program for the remaining tasks in a second part of the introduction.

1.1. Motivations and outline of the results.

The inference problem for diffusion processes is now a fairly well understood problem. In particular, during the last two decades, several advances have allowed to tackle the problem of inference based on discretely observed diffusions [10, 36, 40], which is of special practical interest.

More specifically, consider a family of stochastic differential equations of the form

$$Y_t = a + \int_0^t \mu(Y_s;\theta)\, ds + \sum_{l=1}^{d} \int_0^t \sigma^l(Y_s;\theta)\, dB_s^l, \qquad t \in [0,T], \tag{1}$$

where $a \in \mathbb{R}^m$, $\mu(\cdot;\theta) : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma(\cdot;\theta) : \mathbb{R}^m \to \mathbb{R}^{m,d}$ are smooth enough functions, B is a d-dimensional Brownian motion and θ is a parameter varying in a subset $\Theta \subset \mathbb{R}^q$. If one wishes to identify θ from a set of discrete observations of Y, most of the methods which can be found in the literature are based on (or are closely linked to) the maximum likelihood principle. Indeed, if B is a Brownian motion and Y is observed at some equally distant instants $t_i = i\tau$ for $i = 0, \dots, n$, then the log-likelihood of a sample $(Y_{t_1}, \dots, Y_{t_n})$ can be expressed as

$$\ell_n(\theta) = \sum_{i=1}^{n} \ln\left(p\left(\tau, Y_{t_{i-1}}, Y_{t_i};\, \theta\right)\right), \tag{2}$$

Date: April 20, 2011.

2010 *Mathematics Subject Classification.* Primary 60H35; Secondary 60H07, 60H10, 65C30, 62M09.

Key words and phrases. Fractional Brownian motion, Stochastic differential equations, Malliavin calculus, Inference for stochastic processes.

S. Tindel is partially supported by the ANR grant ECRU. Both authors are part of the BIGS (Biology, Genetics and Statistics) team at INRIA Nancy.

where p stands for the transition semi-group of the diffusion Y. If Y enjoys some ergodic properties, with invariant measure ν_{θ_0} under $\mathbf{P}_{\theta_0}$, then we get

$$\text{a.s.}-\lim_{n\to\infty}\frac{1}{n}\ell_n(\theta)=\mathbf{E}_{\theta_0}\left[p\left(\tau,Z_1,Z_2;\,\theta\right)\right]\triangleq J_{\theta_0}(\theta), \tag{3}$$

where $Z_1\sim\nu_{\theta_0}$ and $\mathcal{L}(Z_2|\,Z_1)=p(\tau,Z_1,\cdot\,;\,\theta)$. Furthermore, it can be shown in a general context that $\theta\mapsto J_{\theta_0}(\theta)$ admits a maximum at $\theta=\theta_0$. This opens the way to a MLE analysis which is similar to the one performed in the case of i.i.d observations, at least theoretically.

However, in many interesting cases, the transition semi-group p is not amenable to explicit computations, and thus expression (2) has to be approximated in some sense. The most common approach, advocated for instance in [36], is based on a linearization of each $p(\tau,Y_{t_{i-1}},Y_{t_i};\,\theta)$, which transforms it into a Gaussian density

$$\mathcal{N}\left(Y_{t_{i-1}}+\mu(Y_{t_{i-1}};\theta)\,\tau,\,\sigma\sigma^*(Y_{t_{i-1}};\theta)\,\tau\right).$$

This linearization procedure is equivalent to the approximation of equation (1) by an Euler (first order) numerical scheme. Refinements of this procedure, based on Milstein type discretizations, are proposed in [10].

Some special situations can be treated differently (and often more efficiently): for instance, in case of a constant diffusion coefficient, the continuous time likelihood can be computed explicitly by means of Girsanov's theorem. When the dimension of the driving Brownian motion B is $d=1$, one can also apply Itô's formula in order to be back to an equation with constant diffusion coefficient, or use Doss-Sousman representation of solutions to (1). Let us also mention that statistical inference for SDEs driven by Lévy processes is currently intensively investigated, with financial motivations in mind.

The current article is concerned with the estimation problem for equations of the form (1), when the driving process B is a fractional Brownian motion. Let us recall that a fractional Brownian motion B with Hurst parameter $H\in(0,1)$, defined on a complete probability space $(\Omega,\mathcal{F},\mathbf{P})$, is a d-dimensional centered Gaussian process. Its law is thus characterized by its covariance function, which is given by

$$\mathbf{E}\left[B_t^iB_s^j\right]=\frac{1}{2}\left(t^{2H}+s^{2H}-|t-s|^{2H}\right)\mathbf{1}_{(i=j)},\qquad s,t\in\mathbb{R}_+.$$

The variance of the increments of B is then given by

$$\mathbf{E}\left[\left(B_t^i-B_s^i\right)^2\right]=|t-s|^{2H},\qquad s,t\in\mathbb{R}_+,\quad i=1,\dots,d,$$

and this implies that almost surely the fBm paths are γ-Hölder continuous for any $\gamma<H$. Furthermore, for $H=1/2$, fBm coincides with the usual Brownian motion, converting the family $\{B^H;\,H\in(0,1)\}$ into the most natural generalization of this classical process.

In the last decade, some important advances have allowed to solve [33, 43] and understand [19, 34] differential systems driven by fBm for $H\in(1/2,1)$. The rough paths machinery also allows to handle fBm with $H\in(1/4,1/2)$, as nicely explained in [11, 14, 27, 29]. However, the irregular situation $H\in(1/4,1/2)$ is not amenable to useful moments estimates for the solution Y to (1) together with its Jacobian (that is the derivative with respect to the initial condition). This is why we concentrate, in the sequel, on the simpler case $H>1/2$ for our estimation problem. In any case, many real world noisy systems are

currently modeled by equations like (1) driven by fBm, and this is particularly present in the Biophysics literature, as assessed by [25, 35], or for Finance oriented applications as in [5, 13, 20, 21, 39, 42]. This leads to a demand for rigorous estimation procedures for SDEs driven by fractional Brownian motion, which is the object of our paper.

Concerns about the inference problem for fractional diffusion processes started a decade ago with the analysis of fractional Ornstein-Uhlenbeck processes [23]. Then a more recent representative set of references on the topic includes [37, 41]. More specifically, [41] handles the case of a one-dimensional equation of the form

$$Y_t = a + \theta \int_0^t \mu(Y_s)\, ds + B_t, \qquad t \in [0, T], \tag{4}$$

where μ is regular enough, and where B is a fBm with $H \in (0, 1)$. The simple dependence on the parameter θ and the fact that an additive noise is considered enables the use of Girsanov's transform in order to get an exact expression for the MLE. Convergence of the estimator is then obtained through an extensive use of Malliavin calculus.

As far as [37] is concerned, it is focused on the case of a polynomial equation, for which the exact moments of the solution can be computed. The estimator relies then on a generalization of the moment method, which tries to fit empirical moments of the solution with their theoretical value. The range of application of this method is however confined to specific situations, for the following reasons:

- It assumes that N independent runs of equation (1) can be obtained, which is usually not the case.
- It hinges on multiple integrals computations, which are time consuming and are avoided in most numerical schemes.

As can be seen from this brief review, parameter estimation for rough equations is still in its infancy. We shall also argue that it is a hard problem.

Indeed, if one wishes to transpose the MLE methods used for diffusion processes to the fBm context, an equivalent of the log-likelihood functions (2) should first be produced. But the covariance structure of B is quite complex and the attempts to put the law of Y defined by (1) into a semigroup setting are cumbersome, as illustrated by [1, 17, 31]. We have thus decided to consider a highly simplified version of the log-likelihood. Namely, still assuming that Y is observed at a discrete set of instants $0 < t_1 < \cdots < t_n < \infty$, set

$$\ell_n(\theta) = \sum_{i=1}^{n} \ln\left(f(t_i, Y_{t_i}; \theta)\right), \tag{5}$$

where we suppose that under $\mathbf{P}_\theta$ the random variable Y_{t_i} admits a density $z \mapsto f(t_i, z; \theta)$. Notice that in case of an elliptic diffusion coefficient σ the density $f(t_i, \cdot; \theta)$ is strictly positive, and thus expression (5) makes sense by a straightforward application of [11, Proposition 19.6]. However, the successful replication of the strategy implemented for Brownian diffusions (that we have tried to summarize above) relies on some highly non trivial questions: existence of an invariant measure for equation (1), rate of convergence to this invariant measure, convergence of expressions like (5), characterization of the limit in terms of θ as in (3), to mention just a few. We shall come back to these considerations

in the next section, but let us insist at this point on the fact that all those questions would fit into a research program over several years.

Our aim in this paper is in a sense simpler: we assume that quantities like (5) are meaningful for estimation purposes. Then we shall implement a method which enables to compute $\ell_n(\theta)$ and optimize it in θ, producing thus a pseudo MLE estimator. We focused first on this specific aspect of the problem for the following reasons:

(1) From a statistical point of view, it is obviously essential to obtain a computationally efficient estimation procedure. This will allow us for instance to evaluate numerically the accuracy of our method.
(2) The procedure itself is nontrivial, and requires the use of advanced stochastic analysis tools: probabilistic representation of the density, Malliavin type integration by parts, Stratonovich-Skorohod correction terms, discretization of systems of pathwise stochastic differential equations for instance.

We have thus decided to tackle the implementation issues first. If it turns out to be satisfying, we shall then try to proceed to a full justification of our method.

Let us also mention that it might not be clear to the reader that $\ell_n(\theta)$ can be meaningful in terms of statistical estimation, since it only involves evaluations at single points Y_{t_i}. However our numerical experiments indicate that this quantity behaves nicely for our purposes. Moreover, it will become clear from the forthcoming computations that our methodology can be extended to handle quantities like

$$\tilde{\ell}_n(\theta) := \sum_{i=1}^{n} \ln\left(f(t_i, t_{i+1}, Y_{t_i}, Y_{t_{i+1}}; \theta)\right),$$

where $f(s,t,x,z;\theta)$ stands for the density of the couple (Y_s, Y_t). This kind of pseudo log-likelihood is obviously closer in spirit to the diffusion case. Densities of tuples could also be considered at the price of technical complications.

Let us now try to give a flavor of the kind of result we shall obtain in this article, in a very loose form:

Theorem 1.1. *Consider Equation (1) driven by a d-dimensional fractional Brownian motion B with Hurst parameter $H > 1/2$. Assume μ and σ are smooth enough coefficients, and that $\sigma\sigma^*$ is strictly elliptic. For a sequence of times $t_0 < \cdots < t_n < \infty$, let y_{t_i}, $i = 1, \dots, n$ be the corresponding observations. Then:*

(i) *The gradient of the log-likelihood function admits the following probabilistic representation: $\nabla_l \ell_n(\theta) = \sum_{i=1}^n \frac{V_i(\theta)}{W_i(\theta)}$, with*

$$W_i(\theta) = \mathbf{E}\left[\mathbf{1}_{(Y_{t_i}(\theta) > y_{t_i})}\, H_{(1,\dots,m)}\Big(Y_{t_i}(\theta)\Big)\right] \tag{6}$$

where $H_{(j_1,\dots,j_n)}(Y_{t_i}(\theta))$ is an expression involving Malliavin derivatives an Skorohod integrals of $Y(\theta)$. A similar expression is also available for $V_i(\theta)$.

(ii) *A computational procedure is constructed in order to obtain $H_{(1,\dots,m)}(Y_{t_i}(\theta))$ in a suitable way.*

(iii) *When Y_t is replaced by its Euler scheme approximation with step T/M and expected values in (6) are approximated thanks to N Monte Carlo steps, we show that*

- *N can be chosen in function of M in an optimal way (see Proposition 4.7).*
- *The corresponding approximation of $\nabla_l \ell_n(\theta)$ converges to the real one with rate $n^{-(2\gamma-1)}$ for any $1/2 < \gamma < H$.*

All those results are stated in a more rigorous way in the remainder of the article.

Here is how our article is structured: we give some preliminaries and notations on Young and Malliavin calculus for fractional Brownian motion at Section 2. The probabilistic representation for the log-likelihood is given at Section 3. Discretization procedures are designed at Section 4, and finally numerical examples are given at Section 5.

1.2. **Remaining open problems.** We emphasized above the fact that only a part of the problem at stake was going to be solved in the current article. We now briefly sketch the remaining tasks to be treated.

The most important obstacle in order to fully justify our methodology is to get a suitable convergence theorem for $\ell_n(\theta)/n$, where $\ell_n(\theta)$ is defined by (5). In a natural way, this should be based on some strong ergodicity properties for Y_t. After a glance at the literature on ergodicity for fractional systems, one can distinguish two cases:

(i) When $\sigma(\cdot;\theta)$ is constant, the convergence of $\mathcal{L}(Y_t)$ as $t \to \infty$ is established in [15], with a (presumably non optimal) rate of convergence $t^{-1/8}$.

(ii) For a general smooth and elliptic coefficient σ, only the uniqueness of the invariant measure is shown in [17], with an interesting extension to the hypoelliptic case in [18]. Nothing is known about the convergence of $\mathcal{L}(Y_t)$, not to mention rates.

This brief review already indicates that the convergence to invariant measures is still quite mysterious for fractional differential equations, at least for a non constant coefficient σ. Moreover, recall that if $\nu(\theta)$ stands for the invariant measure corresponding to the system with coefficients $\mu(\cdot;\theta), \sigma(\cdot;\theta)$, we also wish to retrieve some information on the dependence $\theta \mapsto \nu(\theta)$ (See [16] for some partial results in this direction).

Let us mention another concrete problem: even in the case of a constant σ, the convergence of $\mathcal{L}(Y_t)$ to an invariant measure $\nu(\theta)$ is proven in [15] in the total variation sense. In terms of the density $p(t,x;\theta)$ of Y_t, it means that $p(t,\cdot;\theta)$ converges to the density of ν in L^1 topology. However, in order to get a limit for $\ell_n(\theta)/n$, one expects to use at least a convergence in some Sobolev space $W^{\alpha,p}$ for α, p large enough.

One possibility in order to get this sharper convergence is to bound first the density $p(t,\cdot;\theta)$ in another Sobolev space $W^{\alpha',p'}$ and then to use interpolation theory. It seems thus sufficient to obtain Gaussian bounds on $p(t,\cdot;\theta)$, uniformly in t. In case of Brownian diffusions, these Gaussian bounds are obtained by analytic tools, thanks to the Markov property. This method being obviously not available for systems driven by fBm, a possible inspiration is contained in the upper Gaussian bounds for the stochastic wave equation which can be found in [6]. The latter technical results stem from an intensive use of Malliavin calculus, which should also be invoked in our case, and notice the recent efforts [2, 3] in this direction.

Finally, let us mention that it seems possible to produce some reasonable convergent parametric estimators for equations driven by fBm in a rather general context. Among the methods which can be adapted from the diffusion case with the current stochastic analysis techniques, let us mention the least square estimator of [22], as well as the local

asymptotic normality property shown in [12]. However, it seems obvious that the road to a complete picture of parameter estimation for stochastic equations driven by fBm is still hard and long. We hope to complete it in some subsequent communications.

2. Preliminaries and notations

As mentioned in the introduction, we are concerned with equations driven by a d-dimensional fractional Brownian motion B. We recall here some basic facts about the way to solve those equations, and some Malliavin calculus tools which will be needed later on. Let us introduce first some general notation for Hölder type spaces:

Notation 2.1. *We will denote by $\mathcal{C}^\alpha(V)$ the set of V-valued α-Hölder functions for any $\alpha \in (0,1)$, and by $\mathcal{C}_b^n(U;V)$ the set of n times differentiable functions, bounded together with all their derivatives, from U to V. In the previous notation, U and V stand for two finite dimensional vector spaces. The state space V can be omitted for notational sake when its value is non ambiguous. When we want to stress the fact that we are working on a finite interval $[0,T]$, we write $\mathcal{C}_T^\alpha(V)$ for the space of α-Hölder functions f from $[0,T]$ to V. The corresponding Hölder norms shall be denoted by $\|f\|_{\alpha,T}$.*

2.1. **Differential equations driven by fBm.** Recall that the equation we are interested in is of the form (1). Before stating the assumptions on our coefficients we need an additional notation:

Notation 2.2. *For $n,p \geq 1$, a function $f \in \mathcal{C}^p(\mathbb{R}^n;\mathbb{R})$ and any tuple $(i_1,\dots i_p) \in \{1,\dots,d\}^p$, we set $\partial_{i_1\dots i_p} f$ for $\frac{\partial^p f}{\partial x_{i_1}\dots\partial x_{i_p}}$. Similarly, consider a function $g_\theta \in \mathcal{C}^p(\Theta;\mathbb{R})$, for $n,p \geq 1$ and a vector of parameters $\theta \in \Theta \subset \mathbb{R}^q$. For any tuple $(i_1,\dots i_p) \in \{1,\dots,q\}^p$, we set $\nabla_{i_1\dots i_p} g_\theta^i$ for $\frac{\partial^p g_\theta^i}{\partial\theta_{i_1}\dots\partial\theta_{i_p}}$, where $i = 1,\dots,n$.*

Using this notation, we work under the following set of assumptions:

Hypothesis 2.3. *For any $\theta \in \Theta$, we assume that $\mu(\cdot;\theta) : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma(\cdot;\theta) : \mathbb{R}^m \to \mathbb{R}^{m,d}$ are $\mathcal{C}_b^2$ coefficients. Furthermore, we have*

$$\sup_{\theta\in\Theta} \sum_{l=0}^{2} \sum_{1\leq i_1,\dots,i_l\leq q} \|\nabla^l_{i_1\cdots i_l}\mu(\cdot;\theta)\|_\infty + \|\nabla^l_{i_1\cdots i_l}\sigma(\cdot;\theta)\|_\infty < \infty.$$

When equation (1) is driven by a fBm with Hurst parameter $H > 1/2$ it can be solved, thanks to a fixed point argument, with the stochastic integral interpreted in the (pathwise) Young sense (see e.g. [14]). Let us recall that Young's integral can be defined in the following way:

Proposition 2.4. *Let $f \in \mathcal{C}^\gamma$, $g \in \mathcal{C}^\kappa$ with $\gamma + \kappa > 1$, and $0 \leq s \leq t \leq 1$. Then the integral $\int_s^t g_\xi \, df_\xi$ is well-defined as limit of Riemann sums along partitions of $[s,t]$. Moreover, the following estimation is fulfilled:*

$$\left| \int_s^t g_\xi \, df_\xi \right| \leq C\|f\|_\gamma \|g\|_\kappa |t-s|^\gamma, \tag{7}$$

where the constant C only depends on γ and κ. A sharper estimate is also available:

$$\left| \int_s^t g_\xi \, df_\xi \right| \leq |g_s| \, \|f\|_\gamma |t-s|^\gamma + c_{\gamma,\kappa} \|f\|_\gamma \|g\|_\kappa |t-s|^{\gamma+\kappa}. \tag{8}$$

With this definition in mind and under assumptions 2.3, we can solve our differential system of interest, and the following moments bounds are proven in [11, 19]:

Proposition 2.5. *Consider a fBm B with Hurst parameter $H > 1/2$. Then:*

(1) Under Hypothesis 2.3, equation (1) driven by B admits a unique β-Hölder continuous solution Y, for any $\beta < H$.

(2) Furthermore,

$$\|Y\|_{T,\beta} \leq |a| + c_{f,T} \|B\|_{\beta,T}^{1/\beta}.$$

(3) If we denote by Y^a the solution to (1) with initial condition a, then

$$\|Y^b - Y^a\|_{T,\beta} \leq |b-a| \, \exp\left(c_{f,T} \|B\|_{\beta,T}^{1/\beta} \right).$$

(4) If we only assume that f has linear growth, with $\nabla f, \nabla^2 f$ bounded, the following estimate holds true:

$$sup_{t\in[0,T]} |Y_t| \leq (1 + |a|) \, \exp\left(c_{f,T} \|B\|_{\beta,T}^{1/\beta} \right).$$

Remark 2.6. The framework of fractional integrals is used in [19] in order to define integrals with respect to B. It is however easily seen to be equivalent to the Young setting we have chosen to work with.

Some differential calculus rules for processes controlled by fBm will also be useful in the sequel:

Proposition 2.7. *Let B be a d-dimensional fBm with Hurst parameter $H > 1/2$. Consider $a, \hat{a} \in \mathbb{R}$, $b, \hat{b} \in \mathcal{C}_T^\alpha(\mathbb{R}^d)$ with $\alpha + H > 1$, and $c, \hat{c} \in \mathcal{C}_T(\mathbb{R})$ (all these assumptions are understood in the almost sure sense). Define two processes $z, \hat{z}$ on $[0,T]$ by*

$$z_t = a + \sum_{j=1}^d \int_0^t b_u^j \, dB_u^j + \int_0^t c_u \, du, \quad \text{and} \quad \hat{z}_t = \hat{a} + \sum_{j=1}^d \int_0^t \hat{b}_u^j \, dB_u^j + \int_0^t \hat{c}_u \, du.$$

Then for $t \in [0,T]$, one can decompose the product $z_t \hat{z}_t$ into

$$z_t \, \hat{z}_t = a \, \hat{a} + \sum_{j=1}^n \int_0^t \left[\hat{z}_u \, b_u^j + z_u \, \hat{b}_u^j \right] dB_u^j + \int_0^t \left[z_u \, \hat{c}_u + \hat{z}_u c_u \right] du,$$

where all the integrals with respect to B are understood in the Young sense.

The proof of this elementary and classical result is omitted here. See [28, Proposition 2.8] for the proof of a similar rule.

2.2. **Malliavin calculus techniques.** Our representation of the density for the solution to (1) obviously relies on Malliavin calculus tools that we proceed now to recall. As already mentioned in the introduction, on a finite interval $[0,T]$ and for some fixed $H \in (1/2,1)$, we consider $(\Omega, \mathcal{F}, P)$ the canonical probability space associated with a fractional Brownian motion with Hurst parameter H. That is, $\Omega = \mathcal{C}_0([0,T];\mathbb{R}^d)$ is the Banach space of continuous functions vanishing at 0 equipped with the supremum norm, $\mathcal{F}$ is the Borel sigma-algebra and P is the unique probability measure on Ω such that the canonical process $B = \{B_t,\ t \in [0,T]\}$ is a d-dimensional fractional Brownian motion with Hurst parameter H. Remind that this means that B has d independent coordinates, each one being a centered Gaussian process with covariance $R_H(t,s) = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H})$.

2.2.1. *Functional spaces.* Let $\mathcal{E}$ be the space of d-dimensional elementary functions on $[0,T]$:

$$\mathcal{E} = \Big\{ f = (f_1, \dots, f_d);\ f_j = \sum_{i=0}^{n_j-1} a_i^j \mathbf{1}_{[t_i^j, t_{i+1}^j)}\,, \quad 0 = t_0 < t_1^j < \cdots < t_{n_j-1}^j < t_{n_j}^j = T, \\ \text{for } j = 1, \dots, d \Big\}. \quad (9)$$

We call $\mathcal{H}$ the completion of $\mathcal{E}$ with respect to the semi-inner product

$$\langle f,\, g\rangle_{\mathcal{H}} = \sum_{i=1}^{d} \langle f_i,\, g_i\rangle_{\mathcal{H}_0}\,, \quad \text{where} \quad \langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]}\rangle_{\mathcal{H}_0} := R(s,t), \quad s,t \in [0,T].$$

Then, one constructs an isometry $K_H^* : \mathcal{H} \to L^2([0,1];\mathbb{R}^d)$ such that

$$K_H^* \left(\mathbf{1}_{[0,t_1]}, \dots, \mathbf{1}_{[0,t_d]}\right) = \left(\mathbf{1}_{[0,t_1]} K_H(t_1,\cdot), \dots, \mathbf{1}_{[0,t_d]} K_H(t_d,\cdot)\right),$$

where the kernel K_H is given by

$$K_H(t,s) = c_H s^{\frac{1}{2}-H} \int_s^t (u-s)^{H-\frac{3}{2}} u^{H-\frac{1}{2}}\, du$$

and verifies that $R_H(t,s) = \int_0^{s\wedge t} K_H(t,r) K_H(s,r)\, dr$, for some constant c_H. Moreover, let us observe that K_H^* can be represented in the following form: for $\varphi = (\varphi_1, \dots, \varphi_d) \in \mathcal{H}$, we have $K_H^*\varphi$

$$K_H^*\varphi = \left(K_H^*\varphi^1, \dots, K_H^*\varphi^d\right), \quad \text{where} \quad [K_H^*\varphi^i]_t = \int_t^1 \varphi_r^i \partial_r K_H(r,t)\, dr.$$

2.2.2. *Malliavin derivatives.* Let us start by defining the Wiener integral with respect to B: for any element f in $\mathcal{E}$ whose expression is given as in (9), we define the Wiener integral of f with respect to B as

$$B(f) := \sum_{j=1}^{d} \sum_{i=0}^{n_j-1} a_i^j (B^j_{t_{i+1}^j} - B^j_{t_i^j})\,.$$

We also denote this integral as $\int_0^T f_t dB_t$, since it coincides with a pathwise integral with respect to B.

For $\theta : \mathbb{R} \to \mathbb{R}$, and $j \in \{1, \dots, d\}$, denote by $\theta^{[j]}$ the function with values in $\mathbb{R}^d$ having all the coordinates equal to zero except the j-th coordinate that equals to θ. It is readily seen that

$$\mathbf{E}\left[B\left(\mathbf{1}_{[0,s)}^{[j]}\right) B\left(\mathbf{1}_{[0,t)}^{[k]}\right)\right] = \delta_{j,k} R_{s,t}.$$

This definition can be extended by linearity and closure to elements of $\mathcal{H}$, and we obtain the relation

$$\mathbf{E}\left[B(f)\, B(g)\right] = \langle f, g \rangle_{\mathcal{H}},$$

valid for any couple of elements $f, g \in \mathcal{H}$. In particular, $B(\cdot)$ defines an isometric map from $\mathcal{H}$ into a subspace of $L^2(\Omega)$.

We can now proceed to the definition of Malliavin derivatives. With this notation 2.2 in hand, let us consider $\mathcal{S}$ be the family of smooth functionals F of the form

$$F = f(B(h_1), \dots, B(h_n)), \tag{10}$$

where $h_1, \dots, h_n \in \mathcal{H}$, $n \geq 1$, and f is a smooth function with polynomial growth, together with all its derivatives. Then, the Malliavin derivative of such a functional F is the $\mathcal{H}$-valued random variable defined by

$$DF = \sum_{i=1}^{n} \partial_i f(B(h_1), \dots, B(h_n))\, h_i.$$

For all $p > 1$, it is known that the operator D is closable from $L^p(\Omega)$ into $L^p(\Omega; \mathcal{H})$ (see e.g. [32, Section 1]). We will still denote by D the closure of this operator, whose domain is usually denoted by $\mathbb{D}^{1,p}$ and is defined as the completion of $\mathcal{S}$ with respect to the norm

$$\|F\|_{1,p} := \left(E(|F|^p) + E(\|DF\|_{\mathcal{H}}^p)\right)^{\frac{1}{p}}.$$

It should also be noticed that partial Malliavin derivatives with respect to each component B^j of B will be invoked: they are defined, for a functional F of the form (10) and $j = 1, \dots, d$, as

$$D^j F = \sum_{i=1}^{n} \partial_i f(B(h_1), \dots, B(h_n)) h_i^{[j]},$$

and then extended by closure arguments again. We refer to [32, Section 1] for the definition of higher derivatives and Sobolev spaces $\mathbb{D}^{k,p}$ for $k > 1$. Another essential object related to those derivatives is the so-called Malliavin matrix of a $\mathbb{R}^m$-valued random variable $F \in \mathbb{D}^{1,2}$, defined by

$$\gamma_F = \left(\left\langle DF^i, DF^j \right\rangle\right)_{1 \leq i,j \leq m}. \tag{11}$$

2.2.3. *Skorohod integrals.* We will denote by δ the adjoint of the operator D (also referred to as the *divergence operator*). This operator is closed and its domain, denoted by $\mathrm{Dom}(\delta)$, is the set of $\mathcal{H}$-valued square integrable random variables $u \in L^2(\Omega; \mathcal{H})$ such that

$$\left|\mathbf{E}\left[\langle DF, u \rangle_{\mathcal{H}}\right]\right| \leq C \|F\|_2,$$

for all $F \in \mathbb{D}^{1,2}$, where C is some constant depending on u. Moreover, for $u \in \mathrm{Dom}(\delta)$, $\delta(u)$ is the element of $L^2(\Omega)$ characterized by the duality relationship:

$$\mathbf{E}\left[F\delta(u)\right] = \mathbf{E}\left[\langle DF, u \rangle_{\mathcal{H}}\right], \quad \text{for any} \quad F \in \mathbb{D}^{1,2}. \tag{12}$$

The quantity $\delta(u)$ is usually called *Skorohod integral* of the process u.

Skorohod integrals are obviously analytic objects, not suitable for easy numerical implementations. However, they can be related to the Young type integrals introduced at Proposition 2.4. For this, we need to define another functional space as follows:

Notation 2.8. *We call $|\mathcal{H}|$ the space of measurable functions $\varphi : [0,T] \to \mathbb{R}^d$ such that*

$$\|\varphi\|^2_{|\mathcal{H}|} := c_H \int_0^1 \int_0^1 |\varphi_r||\varphi_u||r-u|^{2H-2} dr du < +\infty,$$

where $c_H = H(2H-1)$, and we denote by $\langle \cdot, \cdot \rangle_{|\mathcal{H}|}$ the associated inner product. We also write $\mathbb{D}^{k,p}(|\mathcal{H}|)$ for the space of $\mathbb{D}^{k,p}$ functionals with values in $|\mathcal{H}|$.

The following proposition is then a slight extension of [32, Proposition 5.2.3]:

Proposition 2.9. *Let $\{u_t^{ij},\ t \in [0,1]\}$, for $i = 1, \dots, m$ and $j = 1, \dots, d$, be a stochastic process in $\mathbb{D}^{1,2}(|\mathcal{H}|)$ such that*

$$\sum_{j=1}^d \int_0^1 \int_0^1 |D_s^j u_t^{ij}|\, |t-s|^{2H-2} ds dt < +\infty \quad a.s. \tag{13}$$

We also assume that almost surely, u has β-Hölder paths with $\beta + H > 1$. Then the Young integral $\sum_{j=1}^d \int_0^T u_t^{ij}\, dB_t^j$ exists and for all $i = 1, \dots, m$ can be written as

$$\sum_{j=1}^d \int_0^T u_t^{ij}\, dB_t^j = \delta(u^i) + \sum_{j=1}^d \int_0^T \int_0^T D_s^j u_t^{ij} |t-s|^{2H-2} ds dt,$$

where $\delta(u)$ stands for the Skorohod integral of u.

3. Probabilistic expression for the log-likelihood

Recall that we are focusing on equation (1) driven by a d-dimensional fBm B, and that we have chosen to use expression (5) as a substitute to the log-likelihood function. We have thus reduced the initial maximization problem to the solution of $\nabla_l \ell_n(\theta) = 0$. This will be performed numerically by means of a root approximation algorithm.

Observe first that in order to define (5), the density of $Y_t(\theta)$ must exist for any $t > 0$. Let us thus recall the classical setting (given in [19]) under which Y_t admits a smooth density:

Hypothesis 3.1. *Let μ and σ be coefficients satisfying Hypothesis 2.3. For $\xi \in \mathbb{R}^m$ and $\theta \in \Theta$, set $\alpha(\xi) = \sigma(\xi,\theta)\sigma^*(\xi,\theta)$. Then we assume that*

(i) *For any $k \geq 0$ and $j_1, \dots, j_k \in \{1, \dots, m\}$ we have*

$$\sup_{\theta \in \Theta} \sum_{l=0}^2 \sum_{1 \leq p_1, \dots, p_l \leq q} \|\nabla^l_{p_1 \cdots p_l} \partial^k_{j_1, \dots, j_k} \mu(\cdot;\theta)\|_\infty + \|\nabla^l_{p_1 \cdots p_l} \partial^k_{j_1, \dots, j_k} \sigma(\cdot;\theta)\|_\infty \leq c_k,$$

for a strictly positive constant c_k.

(ii) *There exists a strictly positive constant ε such that $\langle \alpha(\xi;\theta)\eta, \eta \rangle_{\mathbb{R}^m} \geq \varepsilon |\eta|^2_{\mathbb{R}^m}$ for any couple of vectors $\eta, \xi \in \mathbb{R}^m$, uniformly in $\theta \in \Theta$.*

Then the density result for Y_t can be read as follows:

Theorem 3.2. *Consider the stochastic differential equation (1) with initial condition $a \in \mathbb{R}^m$. Assume Hypothesis 3.1 is satisfied. Then, for any $t > 0$ and $\theta \in \Theta$, the law of $Y_t(\theta)$ admits a $\mathcal{C}^\infty$ density, denoted by $f(t, \cdot; \theta)$, with respect to Lebesgue's measure.*

In the sequel, we shall suppose that the density $f(t, \cdot; \theta)$ exists without further mention, the aim of this section being to produce a probabilistic representation of $f(t, \cdot; \theta)$ for computational purposes. To this aim, we shall first give the equations governing the Malliavin derivatives of the processes $Y(\theta)$ and $\nabla Y(\theta)$, and then use a stochastic analysis formula in order to represent our log-likelihood. We separate these tasks in two different subsections.

3.1. **Some Malliavin derivatives.** This section is devoted to a series of preliminary lemmas which will enable to formulate our probabilistic representation of $f(t, \cdot; \theta)$. Let us first introduce a notation which will prevail until the end of the paper:

Notation 3.3. *For a set of indices or coordinates $(k_1, \dots, k_r)$ of length $r \geq 1$ and $1 \leq j \leq r$, we denote by $(k_1, \dots, \check{k}_j, \dots, k_r)$ the set of indices or coordinates of length $r - 1$ where k_j has been omitted.*

We now give a general expression for the higher order derivatives of Y_t, borrowed from [34].

Lemma 3.4. *Assume Hypothesis 2.3 and 3.1 hold true. For $n \geq 1$ and $(i_1, \dots, i_n) \in \{1, \dots, d\}^n$, denote by $D^{i_1,\dots,i_n} Y_t^i(\theta)$ the n^{th} Malliavin derivative of $Y_t^i(\theta)$ with respect to the coordinates $B^{i_1}, \dots, B^{i_n}$ of B. Then $D^{i_1,\dots,i_n} Y_t^i(\theta)$, considered as an element of $\mathcal{H}^{\otimes n}$, satisfies the following linear equation: for $t \geq r_1 \vee \cdots \vee r_n$,*

$$D^{i_1,\dots,i_n}_{r_1,\dots,r_n} Y_t^i(\theta) = \sum_{p=1}^{n} \alpha^i_{i_p, i_1 \dots, \check{i}_p, \dots, i_n}(r_p; r_1, \dots, \check{r}_p, \dots, r_n; \theta)$$

$$+ \int_{r_1 \vee \cdots \vee r_n}^t \beta^i_{i_1,\dots,i_n}(s; r_1, \dots, r_n; \theta)\, ds + \sum_{l=1}^{d} \int_{r_1 \vee \cdots \vee r_n}^t \alpha^i_{l, i_1, \dots, i_n}(s; r_1, \dots, r_n; \theta)\, dB_s^l, \quad (14)$$

where

$$\alpha^i_{j, i_1, \dots, i_n}(s; r_1, \dots, r_n; \theta) = \sum \sum_{k_1, \dots, k_\nu = 1}^{m} \partial^\nu_{k_1 \dots k_\nu} \sigma^{ij}(Y_s(\theta); \theta)\, D^{i(I_1)}_{r(I_1)} Y_s^{k_1}(\theta) \dots D^{i(I_\nu)}_{r(I_\nu)} Y_s^{k_\nu}(\theta)$$

$$\beta^i_{i_1, \dots, i_n}(s; r_1, \dots, r_n; \theta) = \sum \sum_{k_1, \dots, k_\nu = 1}^{m} \partial^\nu_{k_1 \dots k_\nu} \mu^i(Y_s(\theta); \theta)\, D^{i(I_1)}_{r(I_1)} Y_s^{k_1}(\theta) \dots D^{i(I_\nu)}_{r(I_\nu)} Y_s^{k_\nu}(\theta).$$

In the expressions above, the first sums are extended to the set of all partitions $I_1, \dots, I_\nu$ of $\{1, \dots, n\}$ and for any subset $K = \{i_1, \dots, i_\eta\}$ of $\{1, \dots, n\}$ we set $D^{i(K)}_{r(K)}$ for the derivative operator $D^{i_1,\dots,i_\eta}_{r_1,\dots,r_\eta}$. Notice that $D^{i_1,\dots,i_n}_{r_1,\dots,r_n} Y_t^i(\theta) = 0$ whenever $t < r_1 \vee \cdots \vee r_n$.

The formulas above might seem intricate. The following example illustrate their use in a simple enough situation:

Example 3.5. The second order derivative $D^{1,3}_{r_1,r_2}Y^2_t(\theta)$ can be computed as:

$$\begin{aligned} D^{1,3}_{r_1,r_2}Y^2_t(\theta) &= \alpha^2_{1,3}(r_1,r_2;\theta)+\alpha^2_{3,1}(r_2,r_1;\theta) \\ &\quad +\int_{r_1\vee r_2}^t \beta^2_{1,3}(s,r_1,r_2;\theta)\,ds+\sum_{l=1}^d\int_{r_1\vee r_2}^t \alpha^2_{l,1,3}(s,r_1,r_2;\theta)dB^l_s, \end{aligned}$$

where

$$\begin{aligned} \alpha^2_{1,3}(r_1,r_2;\theta) &= \sum_{k=1}^m \partial_k\sigma^{21}(Y_{r_2}(\theta);\theta)\; D^3_{r_2}Y^k_{r_1}(\theta), \\ \alpha^2_{3,1}(r_2,r_1;\theta) &= \sum_{k=1}^m \partial_k\sigma^{23}(Y_{r_1}(\theta);\theta)\; D^1_{r_1}Y^k_{r_2}(\theta) \end{aligned}$$

and

$$\begin{aligned} \beta^2_{1,3}(s,r_1,r_2;\theta) &= \partial^2_{kk}\mu^2(Y_s(\theta);\theta)D^{1,3}_{r_1,r_2}Y^k_s(\theta)+\partial^2_{k_1k_2}\mu^2(Y_s(\theta);\theta)D^1_{r_1}Y^{k_1}_s(\theta)D^3_{r_2}Y^{k_2}_s(\theta), \\ \alpha^2_{l,1,3}(s,r_1,r_2;\theta) &= \partial^2_{kk}\sigma^{2l}(Y_s(\theta);\theta)D^{1,3}_{r_1,r_2}Y^k_s(\theta)+\partial^2_{k_1k_2}\sigma^{2l}(Y_s(\theta);\theta)D^1_{r_1}Y^{k_1}_s(\theta)D^3_{r_2}Y^{k_2}_s(\theta), \end{aligned}$$

where we have used the convention of summation over repeated indices.

Our formula for the log-likelihood will also involve some derivatives of the process $Y(\theta)$ with respect to the parameter θ. The existence of this derivative is assessed below:

Proposition 3.6. *Under the same hypothesis as for Lemma 3.4, the random variable $Y^i_t(\theta)$ is a smooth function of θ for any $t\geq 0$. We denote by $\nabla_l Y^i_t(\theta)$ the derivative of $Y^i_t(\theta)$ with respect to the l^{th} element of the vector of parameters θ. This process satisfies the following SDE:*

$$\begin{aligned} \nabla_l Y^i_t(\theta) &= \int_0^t [\partial_i\mu^i(Y_u(\theta);\theta)\nabla_l Y^i_u(\theta)+\nabla_l\mu^i(Y_u(\theta);\theta)]du \\ &\quad +\sum_{j=1}^d\int_0^t [\partial\sigma^{ij}(Y_u(\theta);\theta)\nabla_l Y^i_u(\theta)+\nabla_l\sigma^{ij}(Y_u(\theta);\theta)]dB^j_u. \end{aligned}$$

Proof. The proof goes exactly along the same lines as for [34, Proposition 4], and the details are left to the reader.

□

We shall also need some equations governing the Malliavin derivatives of $\nabla_l Y(\theta)$. This is the aim of the following lemma:

Lemma 3.7. *For any $l\in\{1,\dots,q\}$ and $n\geq 1$, the process $\nabla_l D^{i_1,\dots,i_n}Y(\theta)$ is n-times differentiable in the Malliavin calculus sense. Moreover, taking up the notations of Lemma 3.4, the process $\nabla_l D^{i_1,\dots,i_n}Y^i_t(\theta)$ satisfies the following linear equation: for $t\geq r_1\vee$*

$\cdots \vee r_n$,

$$\nabla_l D^{i_1,\ldots,i_n}_{r_1,\ldots,r_n} Y^i_t(\theta) = \sum_{p=1}^{n} \hat{\alpha}^{i,l}_{i_p,i_1\ldots,\check{i}_p,\ldots,n}(r_{i_p}, r_1, \ldots, \check{r}_p, \ldots, r_n; \theta)$$
$$+ \int_{r_1\vee\cdots\vee r_n}^{t} \hat{\beta}^{i,l}_{i_1,\ldots,i_n}(s; r_1, \ldots, r_n; \theta)\, ds + \sum_{l=1}^{d} \int_{r_1\vee\cdots\vee r_n}^{t} \hat{\alpha}^{i,l}_{l,i_1,\ldots,i_n}(s; r_1, \ldots, r_n; \theta)\, dB^l_s,$$

where $\hat{\alpha}^{i,l}_{j,i_1,\ldots,i_n} = \nabla_l \alpha^i_{j,i_1,\ldots,i_n}$ *and* $\hat{\beta}^{i,l}_{j,i_1,\ldots,i_n} = \nabla_l \beta^i_{i_1,\ldots,i_n}$. *More specifically,* $\hat{\beta}^{i,p}_{j,i_1,\ldots,i_n}$ *is defined recursively by*

$$\hat{\beta}^{i,p}_{i_1,\ldots,i_n}(s; r_1, \ldots, r_n; \theta)$$
$$= \sum_{I_1\cup\ldots\cup I_\nu} \sum_{k_1,\ldots,k_\nu=1}^{m} \Big\{ \nabla_p[\partial^\nu_{k_1\ldots k_\nu}\mu^i(Y_s(\theta);\theta)]\; D^{i(I_1)}_{r(I_1)} Y^{k_1}_s(\theta) \cdots D^{i(I_\nu)}_{r(I_\nu)} Y^{k_\nu}_s(\theta)$$
$$+ \partial^\nu_{k_1\ldots k_\nu}\mu^i(Y_s(\theta);\theta) \sum_{p=1}^{\nu} \nabla_p D^{i(I_p)}_{r(I_p)} Y^{k_p}_s(\theta)\; D^{i(I_1)}_{r(I_1)} Y^{k_1}_s(\theta) \cdots D^{\check{i}(I_p)}_{\check{r}(I_p)} Y^{\check{k}_p}_s(\theta) \cdots D^{i(I_\nu)}_{r(I_\nu)} Y^{k_\nu}_s(\theta) \Big\},$$

where we have set

$$\nabla_p[\partial^\nu_{k_1\ldots k_\nu}\mu^i(Y_s(\theta);\theta)] = \nabla_p \partial^\nu_{k_1\ldots k_\nu}\mu^i(Y_s(\theta);\theta) + \partial\partial^\nu_{k_1\ldots k_\nu}\mu^i(Y_s(\theta);\theta)\nabla_p Y_s(\theta).$$

Notice that the same kind of equation (skipped here for sake of conciseness) holds true for the coefficients $\hat{\alpha}^{i,l}_{j,i_1,\ldots,i_n}$.

The next object we need for our calculations is the inverse of the Malliavin matrix $\gamma_{Y_t(\theta)}$ of $Y_t(\theta)$. Recall that according to (11), the Malliavin matrix of $Y_t(\theta)$ is defined by

$$\gamma_t(\theta) := \gamma_{Y_t(\theta)} = \left(\left\langle D_\cdot Y^i_t(\theta), D_\cdot Y^j_t(\theta)\right\rangle\right)_{1\le i,j\le m}, \tag{15}$$

where we have set $\gamma_t(\theta) := \gamma_{Y_t(\theta)}$ for notational sake in the computations below. We shall now compute $\gamma_t^{-1}(\theta)$ as the solution to a SDE:

Proposition 3.8. *The matrix valued process* $\gamma_t^{-1}(\theta)$ *is the unique solution to the following linear equation in* η*:*

$$\begin{aligned}\eta_t(\theta) &= \tilde{\alpha}_0^{-1}(Y_t(\theta);\theta) - \sum_{l=1}^{d}\int_0^t [\eta_u(\theta)\tilde{\alpha}_l(Y_u(\theta);\theta) + \tilde{\alpha}_l^T(Y_u(\theta);\theta)\eta_u] dB^l_u \\ &\quad - \int_0^t [\eta_u(\theta)\tilde{\beta}(Y_u(\theta);\theta) + \tilde{\beta}^T(Y_u(\theta);\theta)\eta_u(\theta)] du,\end{aligned} \tag{16}$$

with

$$\tilde{\alpha}_0(Y_t(\theta);\theta) = \sum_{j=1}^{m}\int_0^t\int_0^t \sigma^{ij}(Y_r(\theta);\theta)\sigma^{i'j}(Y_{r'}(\theta);\theta)\, |r-r'|^{2H-2} dr\, dr', i, i' = 1, \ldots, m$$

and where the other coefficients $\tilde{\alpha}$ *and* $\tilde{\beta}$ *are defined by*

$$\tilde{\alpha}_l(Y_u(\theta);\theta) = \Big(\partial_k \sigma^{i'l}(Y_u(\theta);\theta)\Big)_{1\le i',k\le m} \quad \text{and} \quad \tilde{\beta}(Y_u(\theta);\theta) = \Big(\partial_k \mu^{i'}(Y_u(\theta);\theta)\Big)_{1\le i',k\le m}. \tag{17}$$

Proof. The proof of this fact is an adaptation of [19, Theorem 7] to the case of a SDE with drift. We include it here for sake of completeness, and we drop the dependence of Y on θ for notational sake in the computations below.

Let us start by invoking Proposition 2.7 and equation (14) in order to compute the product of two first-order Malliavin derivatives:

$$\begin{aligned} D^j_r Y^i_t \; D^j_{r'} Y^{i'}_t &= \sigma^{ij}(Y_r)\sigma^{i'j}(Y_{r'}) + \\ &+ \sum_{k=1}^m \Bigg\{ \int_0^t \sum_{l=1}^d \Big[\partial_k \sigma^{il}(Y_u)\; D^j_{r'} Y^{i'}_u \; D^j_r Y^k_u + \partial_k \sigma^{i'l}(Y_u)\; D^j_r Y^i_u \; D^j_{r'} Y^k_u \Big] \, dB^l_u \\ &+ \int_0^t \Big[\partial_k \mu^i(Y_u)\; D^j_{r'} Y^{i'}_u \; D^j_r Y^k_u \; + \partial_k \mu^{i'}(Y_u)\; D^j_r Y^i_u \; D^j_{r'} Y^k_u du \Big] \Bigg\}. \end{aligned} \tag{18}$$

Moreover, recall that γ_t is defined by (15). Thus, the covariance matrix becomes

$$\gamma^{ii'}_t = \sum_{j=1}^d \Big\langle D^j Y^i_t,\, D^j Y^{i'}_t \Big\rangle_{\mathcal{H}} = c_H \sum_{j=1}^d \int_0^t \int_0^t D^j_r Y^i_t(\theta)\; D^j_{r'} Y^{i'}_t(\theta)\; |r - r'|^{2H-2}\, dr\, dr'.$$

Plugging (18) into this relation, we end up with the following equation for $\gamma^{ii'}$:

$$\begin{aligned} \gamma^{ii'}_t = \tilde{\alpha}^{ii'}_0 + \sum_{l=1}^d \int_0^t \sum_{k=1}^m \Big[\partial_k \sigma^{i\,l}(Y_u)\; \gamma^{i'k}_u + \partial_k \sigma^{i'l}(Y_u)\; \gamma^{ik}_u \Big] dB^l_u \\ + \int_0^t \sum_{k=1}^m \Big[\partial_k \mu^i(Y_u)\; \gamma^{i'k}_u + \partial_k \mu^{i'}(Y_u)\; \gamma^{i\,k}_u \Big] du. \end{aligned}$$

Using our notation (17) and matrix product rules, we obtain that γ_t is solution to:

$$\gamma_t = \sum_{l=1}^d \int_0^t (\tilde{\alpha}_l(Y_u)\gamma_u + \gamma_u \tilde{\alpha}^T_l(Y_u)) dB^l_u + \int_0^t (\tilde{\beta}(Y_u)\gamma_u + \gamma_u \tilde{\beta}^T(Y_u)) du.$$

Consider now η solution to (16). Applying again Proposition 2.7, it is readily checked that $\gamma_t \eta_t = \mathrm{Id}$ for any $t \in [0, T]$, which ends the proof.

□

Remark 3.9. Gathering equation (16) and Proposition 2.5, it is easily seen that for any $t > 0$ and $\theta \in \Theta$, $Y_t(\theta)$ is a non degenerate random variable in the sense given at [32, Definition 2.1.2]: we have $\det(\gamma_t^{-1}) \in L^p(\Omega)$ for any $p > 1$.

Now that we have derived an equation for $\eta = \gamma^{-1}$, an equation for the Malliavin derivative of η is also available:

Proposition 3.10. *For any $l \in \{1, \dots, q\}$ and $n \geq 1$, the process $\eta_t = \gamma_t^{-1}$ is n-time differentiable in the Malliavin calculus sense. Moreover, the process $D^{i_1,\dots,i_n}\eta_t$ satisfies the*

following equation: for $t \geq r_1 \vee \cdots \vee r_n$,

$$D^{i_1,\ldots,i_n}_{r_1,\ldots,r_n}\eta^{ij}_t(\theta) = -\sum_{k_1=1}^{n}\sum_{k_2=1}^{k_1}(D^{i_1,\ldots,i_{k_2}}_{r_1,\ldots,r_{k_2}}\tilde{\alpha}_0^{-1}\ D^{i_1,i_2,\ldots,i_{k_1-k_2}}_{r_1,r_2,\ldots,r_{k_1-k_2}}\tilde{\alpha}_0\ D^{i_1,\ldots,i_{n-k_1}}_{r_1,\ldots,r_{n-k_1}}\tilde{\alpha}_0^{-1})^{ij}$$
$$-\sum_{\ell=1}^{d}\int_{r_1\vee\ldots\vee r_n}^{t} C^{ij}_{\ell,i_1,\ldots,i_n}(s;r_1,\ldots,r_n;\theta)dB^{\ell}_s - \int_{r_1\vee\ldots\vee r_n}^{t} A^{ij}_{i_1,\ldots,i_n}(s;r_1,\ldots,r_n;\theta)ds,$$

where

$$A^{ij}_{i_1,\ldots,i_n}(s;r_1,\ldots,r_n;\theta) = \sum_{k_1,\ldots,k_\nu}^{m}\sum_{k=1}^{m}$$
$$\Big\{[\partial^{\nu}_{k_1,\ldots,k_\nu}(\tilde{\beta}(Y_u(\theta);\theta))^{kj}\ D^{i(I_1)}_{r(I_1)}\eta^{ik}_s(\theta)\ldots D^{i(I_\nu)}_{r(I_\nu)}\eta^{ik}_s(\theta)\ \ D^{i(I_1)}_{r(I_1)}Y^{i}_s(\theta)\ldots D^{i(I_\nu)}_{r(I_\nu)}Y^{i}_s(\theta)]$$
$$+[\partial^{\nu}_{k_1,\ldots,k_\nu}(\tilde{\beta}(Y_u(\theta);\theta))^{ik}\ D^{i(I_1)}_{r(I_1)}\eta^{kj}_s(\theta)\ldots D^{i(I_\nu)}_{r(I_\nu)}\eta^{kj}_s(\theta)\ \ D^{i(I_1)}_{r(I_1)}Y^{j}_s(\theta)\ldots D^{i(I_\nu)}_{r(I_\nu)}Y^{j}_s(\theta)]\Big\}$$

and the same kind of equation holds for $C^{ij}_{l,i_1,\ldots,i_n}(s;r_1,\ldots,r_n;\theta)$*, with the coefficients* β *replaced by* α_l*.*

Proof. The proof of this proposition is based on Lemma 3.4 and the fact that $\frac{dA_\lambda^{-1}}{d\lambda} = -A_\lambda^{-1}\frac{dA_\lambda}{d\lambda}A_\lambda^{-1}$.

□

Finally, one can also differentiate η with respect to our standing parameter θ, which yields:

Lemma 3.11. *The derivative of the inverse of the Malliavin matrix* η_t *with respect to* θ *satisfies the following SDE*

$$\nabla_l\eta_t(\theta) = \nabla_l\tilde{\alpha}_0^{-1} - \sum_{\ell=1}^{d}\int_0^t\{\nabla_l\eta_u(\theta)\tilde{\alpha}_\ell(Y_u(\theta);\theta) + \eta_u(\theta)\nabla_l[\tilde{\alpha}_\ell(Y_u(\theta);\theta)]$$
$$+\nabla_l[\tilde{\alpha}^T_\ell(Y_u(\theta);\theta)]\eta_u(\theta) + \tilde{\alpha}^T_\ell(Y_u(\theta);\theta)\nabla_l\eta_u(\theta)\}dB^{\ell}_u - \int_0^t\{\nabla_l\eta_u(\theta)\tilde{\beta}(Y_u(\theta);\theta)$$
$$+\eta_u(\theta)\nabla_l[\tilde{\beta}(Y_u(\theta);\theta)] + \nabla_l[\tilde{\beta}^T(Y_u(\theta);\theta)]\eta_u(\theta) + \tilde{\beta}^T(Y_u(\theta);\theta)\nabla_l\eta_u(\theta)\}du,$$

where $\nabla_l[\tilde{\beta}_\ell(Y_u)] = \partial\tilde{\beta}_\ell(Y_u)\nabla_l Y_u + \nabla_l\tilde{\beta}_\ell(Y_u)$ *and* $\nabla_l[\tilde{\alpha}_\ell(Y_u)] = \partial\tilde{\alpha}_\ell(Y_u)\nabla_l Y_u + \nabla_l\tilde{\alpha}_\ell(Y_u)$.

3.2. **Probabilistic representation of the likelihood.** We have chosen to represent the log-likelihood of our sample thanks to the following formula borrowed from the stochastic analysis literature:

Proposition 3.12. *Let* F *be a* $\mathbb{R}^m$*-valued non degenerate random variable (see Remark 3.9 for references on this concept), and let* f *be the density of* F*. For* $n \geq 1$ *and* $(j_1,\ldots,j_n) \in \{1,\ldots,m\}^n$*, let* $H_{(j_1,\ldots,j_n)}(F)$ *be defined recursively by* $H_{(j_1)}(F) = \sum_{j=1}^m \delta((\gamma_F^{-1})^{j_1 j}DF^j)$ *and*

$$H_{(j_1,\ldots,j_n)}(F) = \sum_{j=1}^{m}\delta\left(\left(\gamma_F^{-1}\right)^{j_n j}DF^j H_{(j_1,\ldots,j_{n-1})}(F)\right), \tag{19}$$

where the Skorohod operator δ is defined at Section 2.2.3. Then one can write

$$f(x) = \mathbf{E}\left[\mathbf{1}_{(F>x)} H_{(1,\dots,m)}(F)\right] = \mathbf{E}\left[(F-x)_+ H_{(1,\dots,m,1,\dots,m)}(F)\right], \tag{20}$$

where $\mathbf{1}_{(F>x)} := \prod_{i=1}^m \mathbf{1}_{(F^i>x_i)}$ *and* $(F-x)_+ := \prod_{i=1}^m (F^i - x_i)_+$.

Proof. The first formula is a direct application of [32, Proposition 2.1.5]. The second one is obtained along the same lines, integrating by parts m additional times with respect to the first one.

□

The formula above can obviously be applied to $Y_t(\theta)$ for any strictly positive t, since we have noticed at Remark 3.9 that $Y_t(\theta)$ is a non-degenerate random variable. However, the expression of $H_{(j_1,\dots,j_n)}(Y_t(\theta))$ given by (19) is written in terms of Skorohod integrals, which are not amenable to numerical computations. We will thus recast this expression in terms of Young integrals plus some correction terms:

Proposition 3.13. *Under Hypothesis 2.3 and 3.1, let us define $Q_{st}^{pji} := (\gamma_s^{-1})^{pj} D_s^i Y_t^j(\theta)$ for $0 \le s < t \le T$, $p, j \in \{1,\dots,m\}$ and $i \in \{1,\dots,d\}$. Consider $p \in \{1,\dots,m\}$ and a real valued random variable G which is smooth in the Malliavin calculus sense. Set*

$$U_p(G) = \sum_{i=1}^m \sum_{j=1}^d G \int_0^t Q_{st}^{pji}\, dB_s^i - c_H \sum_{i=1}^m \sum_{j=1}^d \int_0^t \int_0^t D_s^i \left[G Q_{rt}^{pji}\right] |r-s|^{2H-2} dr ds, \tag{21}$$

where the integral with respect to B is understood in the Young sense. Then the quantities $H_{(j_1,\dots,j_n)}(Y_t(\theta))$ defined at Proposition 3.12 can be expressed as

$$H_{(j_1,\dots,j_n)}(Y_t(\theta)) = \sum_{j=1}^m U_{j_n} \circ \cdots \circ U_{j_1}\left(Y_t^j(\theta)\right). \tag{22}$$

Proof. It is an immediate consequence of Proposition 2.9, since we have noticed in our Remark 3.9 that $Y_t(\theta)$ is a non-degenerate random variable.

□

The previous proposition is still not sufficient to warranty an effective computation of the log-likelihood. Indeed, the right hand side of (21) contains terms of the form $D_s[GQ_{rt}^{pji}]$, which should be given in a more explicit form. This is the content of our next proposition.

Proposition 3.14. *Set $H_{(j_1,\dots,j_n)}(Y_t(\theta)) := K_{j_1\dots j_n}$. Then the term $D_s[K_{j_1\dots j_n} Q_{rt}^{pji}]$ in (21) can be computed inductively as follows:*

(i) *We have $D_s[K_{j_1\dots j_n} Q_{rt}^{pji}] = D_s K_{j_1\dots j_n}\, Q_{rt}^{pji} + K_{j_1\dots j_n}\, D_s Q_{rt}^{pji}$, and $D_s Q_{rt}^{pji}$ is computed by invoking Proposition 3.10 for the derivative of γ_t^{-1} and Lemma 3.4 for the derivative of $Y_t(\theta)$. We are thus left with the computation of $D_s K_{j_1\dots j_n}$.*

(ii) *Assume now that we can compute $n-r$ Malliavin derivatives of $K_{j_1\dots j_r}$. Notice that this condition is met for $r = 0$, since $Y_t(\theta)$ itself can be differentiated n times in an explicit way according to Lemma 3.4 again. Then for any $j_1,\dots,j_{r+1}$ and $k \le n-r-1$,*

the quantity $K_{j_1\dots j_{r+1}}$ can be differentiated k times, with a Malliavin derivative given by

$$D^{i_1,\dots,i_k}_{\rho_1\dots\rho_k} K_{j_1\dots j_{r+1}} = \sum_{\ell=1}^{k} D^{i_1,\dots,\check{i}_\ell,\dots,i_k}_{\rho_1\dots\check{\rho}_\ell\dots\rho_k}(K_{j_1\dots j_r}\, Q^{pji}_{\rho_\ell t}) + \sum_{j=1}^{d}\int_0^t D^{i_1,\dots,i_k}_{\rho_1\dots\rho_k}(K_{j_1\dots j_r}\, Q^{pji}_{st})dB^j_s$$
$$-c_H \int_0^t\int_0^t D^{k+1}_{r_1\rho_1\dots\rho_k}(K_{j_1\dots j_r}\, Q^{pji}_{r_2 t})|r_1-r_2|^{2H-2}dr_1dr_2. \tag{23}$$

Proof. We focus on the induction step (ii), the other one being straightforward: for a smooth random variable W, one easily gets by induction that

$$D^{i_1,\dots,i_p}_{r_1\dots r_p}\delta(W) = \sum_{\ell=1}^{p} D^{i_1,\dots,\check{i}_\ell\dots,i_p}_{r_1\dots\check{r}_\ell\dots r_p} W_{r_\ell} + \delta(D^{i_1,\dots,i_p}_{r_1\dots r_p}W). \tag{24}$$

Suppose we know the $n-r$ Malliavin derivatives for $U_{j_r}\circ\cdots\circ U_{j_1}(F) := K_{j_1\dots j_r}$. Recall moreover that

$$K_{j_1\dots j_{r+1}} = U_{j_{r+1}}(K_{j_r\dots j_1}) = \delta(K_{j_r\dots j_1}\, Q_{\cdot t})$$

Applying directly relation (24) we thus get, for $k \le m-1$:

$$D^{i_1,\dots,i_k}_{\rho_1\dots\rho_k}\delta(K_{j_r\dots j_1}\, Q_{\cdot t}) = \sum_{\ell=1}^{k} D^{i_1,\dots,\check{i}_\ell,\dots,i_{k-1}}_{\rho_1\dots\check{\rho}_\ell\dots\rho_k}(K_{j_r\dots j_1}\, Q_{\rho_\ell t}) + \delta(D^{i_1,\dots,i_k}_{\rho_1\dots\rho_k}(K_{j_r\dots j_1}\, Q_{\cdot t})).$$

Our formula (23) is now obtained by applying Proposition 2.9 to the Skorohod integral $\delta(D^k_{\rho_1\dots\rho_k}(K_{j_r\dots j_1}\, Q_{\cdot t}))$ above.

□

Example 3.15. As an illustration of the proposition above, we compute $U_2\circ U_1(F)$ for $F = Y^i_t$, $i\in\{1,\dots,m\}$ and our d-dimensional fBm B.

Write first $U_1(Y^i_t) = \delta(Y^i_t\ (\gamma^{-1})^{1j_1}\ D^{j_1}Y^i_t)$, and since this quantity has to be expressed in a suitable way for numerical approximations, we have

$$U_1(Y^i_t) = \sum_{j_1=1}^{d} Y^i_t\int_0^t Q^{1ij_1}_{ut}dB^{j_1}_u - c_H\sum_{j_1=1}^{d}\int_0^t\int_0^t D^{j_1}_{u_1}[Y^i_t Q^{1ij_1}_{u_2 t}]|u_1-u_2|^{2H-2}du_1du_2,$$

where Q is defined at Proposition 3.13 and where the first integral in the right hand side is understood in the Young sense. In order to compute the second one, we have to compute Malliavin derivatives. This is done through Lemma 3.4 for Y and Proposition 3.10 for Q.

We now have to differentiate $U_1(Y^i_t)$: the derivation rules for Skorohod integrals immediately yield

$$D^{j_2}_{u_2}[U_1(Y^i_t)] = \sum_{j_2=1}^{d} Y^i_t Q^{2ij_2}_{u_2 t} + \sum_{j_2=1}^{d}\delta(D^{j_2}_{u_2}Y^i_t Q^{2ij_2}_{\cdot t}).$$

Once again, the Skorohod integral above is not suitable for numerical approximations. Write thus

$$D^{j_2}_{u_2}[U_1(Y^i_t)] = \sum_{j_2=1}^d Y^i_t Q^{2ij_2}_{u_2t} + \sum_{j_2=1}^d \int_0^t D^{j_2}_{u_2}[Y^i_t Q^{2ij_2}_{rt}]dB^{j_2}_r \\ - c_H \sum_{j_2=1}^d \int_0^t \int_0^t D^{j_2}_{u_2} D^{j_1}_{u_1}[Y^i_t Q^{2ij_2}_{u_2t}]|u_2-u_1|^{2H-2}du_1du_2,$$

and compute the Malliavin derivatives of the products YQ thanks to Lemma 3.4 for Y and Proposition 3.10 for Q. Once this is done, just write

$$\begin{aligned} & U_2(U_1(Y^i_t)) = \delta(U_1(Y^i_t)Q^i_{\cdot t}) \\ = & \sum_{j_2=1}^d U_1(Y^i_t)\int_0^t Q^{2ij_2}_{ut}dB^{j_2}_u - c_H \sum_{j_2=1}^d \int_0^t \int_0^t D^{j_2}_{u_2}[U_1(Y^i_t)Q^{2ij_2}_{u_1t}]|u_2-u_1|^{2H-2}du_1du_2. \end{aligned}$$

In order to give our formula for the derivative of the log-likelihood, we still need to compute the derivative with respect to θ of $H_{(j_1,\dots,j_n)}(Y_t(\theta))$. For this we state the following lemma

Lemma 3.16. *The derivative with respect to θ of $U_p(Y_t(\theta))$ can be written as*

$$\begin{aligned} \nabla_l U_p(Y^i_t(\theta)) &= \sum_{j=1}^d [\nabla_l Y^i_t(\theta)\int_0^t Q^{pij}_{st}(\theta)\,dB^j_s + Y^i_t(\theta)\int_0^t \nabla_l[Q^{pij}_{st}(\theta)]\,dB^j_s] \\ & -c_H \sum_{j=1}^d \int_0^t\int_0^t \nabla_l[D^j_s Y^i_t(\theta) Q^{pij}_{rt}(\theta)]|r-s|^{2H-2}drds, \end{aligned}$$

where $\nabla_l Y^i_t(\theta)$ is computed according to Proposition 3.6 and $\nabla_l[D^j_s Y^i_t]$ is given by Lemma 3.7. As far as $\nabla_l[Q^{pj}_{st}(\theta)]$ is concerned, it is obtained through the following equation:

$$\nabla_l[Q^{pj}_{st}(\theta)] = \nabla_l \eta^{pj}_s(\theta)\; D_s Y^j_t(\theta) \;+\; \eta^{pj}_s(\theta)\nabla_l[D_s Y^j_t(\theta)],$$

where the expression for $\nabla_l \eta^{pj}_s(\theta)$ is a consequence of Lemma 3.11.

We are now ready to state our probabilistic expression for the log-likelihood function (5).

Theorem 3.17. *Assume Hypothesis 2.3 and 3.1 hold true. Let y_{t_i}, $i = 1,\dots,n$ be the observation arriving at time t_i. Let also Y_{t_i} be the solution to the SDE (1) at time t_i. Then, the gradient of the log-likelihood function admits the following probabilistic representation: $\nabla_l \ell_n(\theta) = \sum_{i=1}^n \frac{V_i(\theta)}{W_i(\theta)}$, with*

$$W_i(\theta) = \mathbf{E}\Big[\mathbf{1}_{(Y_{t_i}(\theta)>y_{t_i})}\; H_{(1,\dots,m)}\Big(Y_{t_i}(\theta)\Big)\Big] \tag{25}$$

and

$$V_i(\theta) = \mathbf{E}\Big[\nabla_l Y_{t_i}(\theta)\ \mathbf{1}_{(Y_{t_i}(\theta)>y_{t_i})}\ H_{(1,\dots,m,1,\dots,m)}\Big(Y_{t_i}(\theta)\Big) + \Big(Y_{t_i}(\theta) - y_{t_i}\Big)_+ \nabla_l H_{(1,\dots,m,1,\dots,m)}\Big(Y_{t_i}(\theta)\Big)\Big], \quad (26)$$

where (i) $H_{(j_1,\dots,j_n)}(Y_{t_i}(\theta))$ *is given recursively by (22) and computed at Proposition 3.14* (ii) $\nabla_l Y_{t_i}(\theta)$ *is given by Proposition 3.6* (iii) $\nabla_l H_{(1,\dots,m,1,\dots,m)}$ *is obtained by applying Lemma 3.16.*

Proof. Recall that under Hypothesis 2.3 and 3.1, $Y_t(\theta)$ admits a $\mathcal{C}^\infty$ density $f(t,\cdot;\theta)$ for any $t>0$ and $\theta\in\Theta$. Moreover, we have defined $\ell_n(\theta)$ as $\ell_n(\theta)=\sum_{i=1}^n \ln(f(t_i,y_{t_i};\theta))$. Thus

$$\nabla_l \ell_n(\theta) = \sum_{i=1}^n \frac{\nabla_l f(t_i,y_{t_i};\theta)}{f(t_i,y_{t_i};\theta)} := \sum_{i=1}^n \frac{V_i(\theta)}{W_i(\theta)}.$$

Now $W_i(\theta)$ can be expressed like (25) by a direct application of (20), first relation. As far as $V_i(\theta)$ is concerned, write

$$f(t_i,y_{t_i};\theta) = \mathbf{E}\left[(Y_{t_i}(\theta)-y_{t_i})_+ H_{(1,\dots,m,1,\dots,m)}(Y_{t_i}(\theta))\right],$$

according to the second relation in (20). By using standard arguments, one is allowed to differentiate this expression within the expectation, which directly yields (26).

□

4. Discretization of the log-likelihood

The expression of the log-likelihood that we derived in Proposition 3.17 is a fraction of two expectations that do not have explicit formulas even in the one-dimensional case. In addition, our goal is to find the root of this non-explicit expression, the ML estimator, which is an even harder task. To solve this problem in practice we first use a stochastic approximation algorithm in order to find the root of $\nabla_l \ell_n(\theta)$. In each iteration of the algorithm we compute the value of the expression using Monte-Carlo (MC) simulations. For each Monte-Carlo simulation, since we do not have available an exact way of simulating the kernels of the expectation, we use an Euler approximation scheme. More specifically, we simulate using Euler approximation terms such as Y_t, DY_t, which are solutions to fractional stochastic differential equations.

Therefore, in our approach we have three types of error in the computation of the MLE: the error of the stochastic approximation algorithm, the Monte-Carlo error and the discretization bias introduced by the Euler approximation for the stochastic differential equations. Our aim here is to combine the Monte Carlo and Euler approximations in an optimal way in order to get a global error bound for the computation of $\nabla_l \ell_n(\theta)$.

4.1. Pathwise convergence of the Euler scheme.

The Euler scheme is the main source of error in our computations. There is always a trade-off between the number of Euler steps and the number of simulations, but what is usually computationally costly is the number of Euler steps. This is even worse when we deal with fractional SDEs, since the rate of convergence depends on H and the closer the value of H to $1/2$, the more steps are required for the simulation.

In this section, we compute the magnitude of the discretization error we introduce. We measure the bias of the Euler scheme via the root mean square error. That is, we want to estimate the quantity $\sup_{\tau\in[0,T]}(\mathbf{E}|Y_\tau(\theta)-\bar{Y}^M_\tau(\theta)|^2)^{1/2}$, where $Y_t(\theta)$ is the solution to the SDE (1) and $\bar{Y}^M_\tau(\theta)$ is the Euler approximation of $Y_\tau(\theta)$ given on the grid $\{\tau_k;\, k\leq M\}$ by

$$\bar{Y}^M_{\tau_{k+1}}(\theta)=\bar{Y}^M_{\tau_k}(\theta)+\mu(\bar{Y}^M_{\tau_k}(\theta);\theta)(\tau_{k+1}-\tau_k)+\sum_{j=1}^d\sigma^j(\bar{Y}^M_{\tau_k}(\theta);\theta)\delta B^{M,j}_{\tau_k\tau_{k+1}}, \tag{27}$$

in which we denote $\delta B^{M,j}_{\tau_k\tau_{k+1}}=B^{M,j}_{\tau_{k+1}}-B^{M,j}_{\tau_k}$ and $\tau_k=\frac{kT}{M}$ for $k=0,\dots,M-1$. Notice that those estimates can be found in [8, 11, 30]. We include their proof here because it is simple enough, and also because they can be easily generalized to the case of a linear equation. This latter case is of special interest for us, since it corresponds to Malliavin derivatives, and is not included in the aforementioned references.

Notation 4.1. *For simplicity, in this section we write $Y:=Y(\theta)$.*

Proposition 4.2. *Let $T>0$ and recall that $\bar{Y^M}$ is defined by equation (27). Then, there exists a random variable C with finite L^p moments such that for all $\gamma<H$ and $H>1/2$ we have*

$$\|Y_t-\bar{Y}\|_{\gamma,T}\leq C_T\, M^{1-2\gamma} \tag{28}$$

Consequently, we obtain that the MSE is of order $\mathcal{O}(M^{1-2\gamma})$.

Proof. In order to prove (28) we apply techniques of the classical numerical analysis for the flow of an ordinary differential equation driven by a smooth path. Namely, the exact flow of (1) is given by $\Phi(y;s,t):=Y_t$, where Y_t is the unique solution of (1) when $t\in[s,T]$ and the initial condition is $Y_s=y$. Introduce also the numerical flow

$$\Psi(y;\tau_k,\tau_{k+1}):=y+\mu(y)(\tau_{k+1}-\tau_k)+\sum_{j=1}^d\sigma^j(y)\delta B^{M,j}_{\tau_k\tau_{k+1}}, \tag{29}$$

where $\tau_k=\frac{kT}{M}$, $k=0,\dots,M-1$. Thus, we can write that

$$\begin{aligned}\bar{Y}^M_{\tau_{k+1}} &= \Psi\Big(\bar{Y}^M_{\tau_k};\ \tau_k,\tau_{k+1}\Big),\ k=0,\dots,M-1\\ Y^M_0 &= \alpha.\end{aligned}$$

For $q>k$ we also have that

$$\Psi(y;\tau_k,\tau_q):=\Psi(\cdot;\tau_{q-1},\tau_q)\circ\Psi(\cdot;\tau_{q-2},\tau_{q-1})\circ\ldots\circ\Psi(y;\tau_k,\tau_{k+1}).$$

The one-step error computes as

$$\begin{aligned}r_k &= \Phi(y;\tau_k,\tau_{k+1})-\Psi(y;\tau_k,\tau_{k+1})\\ &= \int_{\tau_k}^{\tau_{k+1}}\Big[\mu(Y_s)-\mu(y)\Big]ds+\int_{\tau_k}^{\tau_{k+1}}\Big[\sigma(Y_s)-\sigma(y)\Big]dB_s\end{aligned} \tag{30}$$

Furthermore, since $Y \in \mathcal{C}^\gamma$ and $B \in \mathcal{C}^\gamma$ for $\gamma > 1/2$, using (7) we have

$$\begin{aligned}\Big|\int_{\tau_k}^{\tau_{k+1}} \Big[\sigma(Y_s) - \sigma(y)\Big] dB_s\Big| &\leq c_\gamma \, \|\partial\sigma\|_\infty \|Y\|_\gamma \, \|B\|_\gamma \, \left|\frac{T}{M}\right|^{2\gamma} \\ &\leq c_{\gamma,\sigma} \, \|\partial\sigma\|_\infty \, \|B\|_\gamma^{1/\gamma} \, \|B\|_\gamma \, \left|\frac{T}{M}\right|^{2\gamma},\end{aligned}$$

where we used the fact that $\|Y\|_\gamma \leq c_\sigma \|B\|_\gamma^{1/\gamma}$ (see Proposition 2.5). Similarly, for the drift part we have

$$\begin{aligned}\Big|\int_{\tau_k}^{\tau_{k+1}} \Big[\mu(Y_s) - \mu(y)\Big] ds\Big| &\leq c_\gamma \, \|\partial\mu\|_\infty \, \|Y\|_\gamma \, \left|\frac{T}{M}\right|^{\gamma+1} \\ &\leq c_{\gamma,\mu} \, \|\partial\mu\|_\infty \, \|B\|_\gamma^{1/\gamma} \, \left|\frac{T}{M}\right|^{\gamma+1}.\end{aligned}$$

Therefore, the one-step error (30) satisfies

$$|r_k| \leq c_{\mu,\sigma} \, \|B\|_\gamma^{1+1/\gamma} \, \left|\frac{T}{M}\right|^{2\gamma}. \tag{31}$$

Now, we can write the classical decomposition of the error in terms of the exact and numerical flow. Since $\bar{Y}^M_{\tau_k} = \Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_k)$ and $Y_{\tau_k} = \Phi(\bar{Y}^M_{\tau_0}; \tau_0, \tau_k)$ we have

$$\bar{Y}^M_{\tau_q} - Y_{\tau_q} = \Phi(\bar{Y}_{\tau_0}; \tau_0, \tau_k) - \Phi(\bar{Y}_{\tau_q}; \tau_q, \tau_q) = \sum_{k=0}^{q-1} \Big(\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_q) - \Phi(\bar{Y}_{\tau_{k+1}}; \tau_{k+1}, \tau_q) \Big). \tag{32}$$

Since $\Phi\Big(\bar{Y}^M_{\tau_k}; \tau_k, \tau_q\Big) = \Phi\Big(\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_{k+1}); \tau_{k+1}, \tau_q\Big)$ we obtain

$$\begin{aligned}\Big|\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_q) - \Phi(\bar{Y}^M_{\tau_{k+1}}; \tau_{k+1}, \tau_q)\Big| &= \Big|\Phi\Big(\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_q); \tau_{k+1}, \tau_q\Big) - \Phi(\bar{Y}^M_{\tau_{k+1}}; \tau_{k+1}, \tau_q)\Big| \\ &\leq C_T(\|B\|_\gamma) \, |\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_{k+1}) - \bar{Y}^M_{\tau_{k+1}}|,\end{aligned}$$

where we have used the fact that

$$|\Phi(\alpha; t, s) - \Phi(\beta; t, s) \leq C_T(\|B\|_\gamma)|\alpha - \beta|,$$

where C_T is a subexponential function (see Proposition 2.5 again). Moreover, owing to relation (31),

$$|\Phi(\bar{Y}^M_{\tau_k}; \tau_k, \tau_q) - \bar{Y}^M_{\tau_{k+1}}| = |r_k| \leq c_{\mu,\sigma} \, \|B\|_\gamma^{1+1/\gamma} \, \left|\frac{T}{M}\right|^{2\gamma}. \tag{33}$$

Therefore, replacing (33) in (32) for any $q \leq n$ we obtain

$$|\bar{Y}^M_{\tau_q} - Y_{\tau_q}| \leq c_{\mu,\sigma} \, \|B\|_\gamma^{1+1/\gamma} \, \sum_{k=0}^{q-1} \left|\frac{T}{M}\right|^{2\gamma}$$

Let us push forward this analysis to Hölder type norms on the grid $0 \leq \tau_1 < \ldots < \tau_n = T$. We have for $q \geq p$

$$\begin{aligned}
& \delta\Big(Y - \bar{Y}^M\Big)_{\tau_p\tau_q} \\
= \; & \Big(\Phi(Y_{\tau_p};\tau_p,\tau_q) - Y_{\tau_p}\Big) - \Big(\Psi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q) - \bar{Y}^n_{\tau_p}\Big) \\
= \; & \Big(\Phi(Y_{\tau_p};\tau_p,\tau_q) - Y_{\tau_p}\Big) - \Big(\Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q) - \bar{Y}^M_{\tau_p}\Big) - \Big(\Psi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q) - \Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q)\Big) \\
= \; & \Big(\Big(\Phi(Y_{\tau_p};\tau_p,\tau_q) - \Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q)\Big) - \Big(Y_{\tau_p} - \bar{Y}^M_{\tau_p}\Big)\Big) - \Big(\Psi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q) - \Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q)\Big).
\end{aligned}$$

Similar to the calculations leading to (33) we obtain

$$\Big|\Psi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q) - \Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q)\Big| \leq c_{\mu,\sigma}\, \|B\|_\gamma^{1+1/\gamma} \sum_{k=p}^{q-1} \left|\frac{T}{M}\right|^{2\gamma}.$$

Moreover, owing to Proposition 2.5 part (2), observe that

$$\frac{\Big|\Big(\Phi(Y_{\tau_p};\tau_p,\tau_q) - \Phi(\bar{Y}^M_{\tau_p};\tau_p,\tau_q)\Big) - \Big(Y_{\tau_p} - \bar{Y}^M_{\tau_p}\Big)\Big|}{|\tau_q - \tau_p|^\gamma} \leq c(\|B\|_\gamma)\, |Y_{\tau_p} - \bar{Y}^M_{\tau_p}|.$$

Consequently, we have that for $0 \leq p < q \leq M$

$$\Big|\delta\Big(Y - \bar{Y}^M\Big)_{\tau_p\tau_q}\Big| \leq c'(\|B\|_\gamma^{1+1/\gamma})\Big\{\sum_{k=p}^{q-1}\left|\frac{T}{M}\right|^{2\gamma} + |\tau_q - \tau_p|^\gamma \sum_{k=0}^{q}\left|\frac{T}{M}\right|^{2\gamma}\Big\}$$

which easily yields that

$$\sup_{p,q=0,1,\ldots,M-1, p\neq q} \frac{\Big|\delta\Big(Y - \bar{Y}^M\Big)_{\tau_p\tau_q}\Big|}{|\tau_p - \tau_q|^\gamma} \leq c(\|B\|_\gamma)\, M^{1-2\,\gamma}.$$

By "lifting" this error estimate to $[0,T]$ and since $|t-s| \leq T/M$,

$$\|Y_t - \bar{Y}\|_{\gamma,\infty,T} \leq C\, M^{1-2\,\gamma}, \tag{34}$$

which concludes the first part of the proof.

Regarding the order of the Mean Square Error, it suffices to note that the constant C has finite L^p moments. □

As mentioned before, an elaboration of Proposition 4.2 is needed in the sequel. Indeed, in the expression of the log-likelihood in Proposition 3.17 we need to discretize more complicated quantities of the underlying process, such as (14) or (16). To this aim, let us notice first that all those equations can be written under the following generic form:

$$Z_t = \alpha + \int_0^t \xi^2_u Z_u du + \int_0^t \xi^{1,j}_u Z_t dB^j_u, \tag{35}$$

where ξ^1, ξ^2 are stochastic processes with bounded moments of any order. The corresponding Euler discretization is

$$\bar{Z}^M_{\tau_k} = \bar{Z}^M_{\tau_k} + \xi^2_{\tau_k}\bar{Z}^M_{\tau_k}(\tau_{k+1}-\tau_k) + \sum_{j=1}^{d}\xi^{1,j}_{\tau_k}\bar{Z}_{\tau_k}\,\delta B^{j,M}_{\tau_k\tau_{k+1}}, \tag{36}$$

and we give first an approximation result in this general context:

Proposition 4.3. *Let $T>0$, and consider the $\mathbb{R}^q$-valued solution Z to equation (35), where $\alpha\in\mathbb{R}^q$, $\xi^2,\xi^{1,j}\in\mathbb{R}^{q,q}$ and we suppose that $\|\xi^2\|_\gamma$ and $\|\xi^{1,j}\|_\gamma$ belong to $L^p(\Omega)$ for any value of $p\geq 1$. Let $\bar{Z}^M$ be defined by equation (36). Then, there exists a random variable $C^{'}$ with L^p finite moments, such that for all $\gamma<H$ and $H>1/2$ we have*

$$\|Z-\bar{Z}\|_{\gamma,T}\leq C^{'}_T\,M^{1-2\gamma} \tag{37}$$

Consequently, we obtain that the Mean Square Error is of order $\mathcal{O}(M^{1-2\gamma})$.

Proof. We follow a similar approach as in the previous proposition. Thus, the exact flow is equal to $\Phi(\zeta;s,t):=Z_t$, where Z_t is the unique solution of equation (35) when $t\in[s,T]$ and the initial condition is $Z_s=\zeta$. Consider also the numerical flow

$$\Psi(\zeta;\tau_k,\tau_{k+1}):=\zeta+\xi^2_u\zeta(\tau_{k+1}-\tau_k)+\sum_{j=1}^{d}\xi^{1,j}_u\zeta\delta B^{j,M}_{\tau_k\tau_{k+1}},$$

where $\tau_k=kT/M$, $n=0,\ldots,M-1$. Thus, we have

$$\begin{aligned}\bar{Z}^M_{\tau_{k+1}} &= \Psi(\bar{Z}^M_{\tau_k};\tau_{k+1},\tau_k),\ k=0,\ldots,M-1\\ \bar{Z}^M_0 &= \alpha.\end{aligned}$$

In this case, the one-step error can be written as

$$\begin{aligned}r_k &= \Phi(\zeta;\tau_k,\tau_{k+1})-\Psi(\zeta;\tau_k,\tau_{k+1})\\ &= \int_{\tau_k}^{\tau_{k+1}}\xi^2_u(Z_s-\zeta)du+\int_{\tau_k}^{\tau_{k+1}}\xi^1_u(Z_s-\zeta)dB_u\end{aligned}$$

We now treat each term separately. Therefore, using the fact that $\|Z\|_\gamma\leq\exp(c\|B\|^{1/\gamma}_\gamma)$, which is recalled at Proposition 2.5 point (4) in a slightly different context, we have that

$$\begin{aligned}\Big|\int_{\tau_k}^{\tau_{k+1}}\xi^1_s(Z_s-\zeta)dB_s\Big| &\leq c_\gamma\,\|Z\xi^1\|_\gamma\,\|B\|_\gamma\,\left|\frac{T}{M}\right|^{2\gamma}\\ &\leq c_\gamma\,\|\exp(\|B\|^{1/\gamma}_\gamma)\,\|B\|_\gamma\,\left|\frac{T}{M}\right|^{2\gamma}.\end{aligned}$$

Similarly, we also have

$$\begin{aligned}\Big|\int_{\tau_k}^{\tau_{k+1}}\xi^2_s(Z_s-\zeta)ds\Big| &\leq c_\gamma\,\|Z\xi^2\|_\gamma\,\|B\|_\gamma\,\left|\frac{T}{M}\right|^{2\gamma}\\ &\leq c_\gamma\,\|\exp(\|B\|^{1/\gamma}_\gamma)\,\|B\|_\gamma\,\left|\frac{T}{M}\right|^{2\gamma}.\end{aligned}$$

Therefore, the one-step error satisfies the following inequality

$$|r_k| \leq c_\gamma \exp(\|B\|_\gamma^{1/\gamma})\ \|B\|_\gamma\ \left|\frac{T}{M}\right|^{2\gamma}.$$

Along the same lines as for Proposition 4.2, the decomposition of the error in terms of the exact and numerical flow becomes

$$\bar{Z}^M_{\tau_q} - Z_{\tau_q} = \Phi(\bar{Z}^M_{\tau_q};\tau_q,\tau_q) - \Phi(\bar{Z}^M_{\tau_0};\tau_0,\tau_k) = \sum_{k=0}^{q-1}\Big(\Phi(\bar{Z}^M_{\tau_{k+1}};\tau_{k+1},\tau_q) - \Phi(\bar{Z}^M_{\tau_k};\tau_k,\tau_q)\Big),$$

and the same inequalities allowing to go from (32) to (33) yield

$$|\bar{Z}_{\tau_q} - Z_{\tau_q}| \ \leq\ c_\gamma\ \left|\frac{T}{M}\right|^{2\gamma}.$$

The claim of the proposition follows now as in Proposition 4.2.

□

We now use the previous proposition in order to approximate the kernels of the expectations in $\nabla_l \ell_n(\theta)$. Let us first introduce the following notation:

Notation 4.4. *Let $W_i(\theta)$, $V_i(\theta)$ as in (25) and (26) respectively and define $w_i(\theta)$ and $v_i(\theta)$ as*

$$w_i(\theta) \ = \ \mathbf{1}_{(Y_{t_i}(\theta)>y_{t_i})}\ H_{(1,\dots,m)}\Big(Y_{t_i}(\theta)\Big) \tag{38}$$

$$v_i(\theta) \ = \ \nabla_l Y_{t_i}(\theta)\ \mathbf{1}_{(Y_{t_i}(\theta)>y_{t_i})}\ H_{(1,.,m,1,.,m)} + \Big(Y_{t_i}(\theta) - y_{t_i}\Big)_+ \nabla_l H_{(1,.,m,1,.,m)}. \tag{39}$$

Let also $\bar{w}_i^M$ and $\bar{v}_i^M$ to be the Euler discretized versions of (38) and (39) using 4.3, and set $\bar{W}_i^M(\theta) = \mathbf{E}[\bar{w}_i^M]$ and $\bar{V}_i^M(\theta) = \mathbf{E}[\bar{v}_i^M]$.

Our convergence result for $\nabla_l \ell_n(\theta)$ can be read as follows:

Theorem 4.5. *Recall from Theorem 3.17 that $\nabla_l \ell_n(\theta)$ can be decomposed as $\nabla_l \ell_n(\theta) = \sum_{i=1}^n \frac{V_i(\theta)}{W_i(\theta)}$. Then the following approximation result holds true:*

$$\big|V_i(\theta) - \bar{V}_i^M(\theta)\big| + \big|W_i(\theta) - \bar{W}_i^M(\theta)\big| \leq \frac{c}{M^{2\gamma-1}},$$

for a strictly positive constant c.

Proof. We focus on the bound for $|V_i(\theta) - \bar{V}_i^M(\theta)|$, the other one being very similar. Now, applying Proposition 4.3 to the particular case of the equations governing Malliavin derivatives, we easily get

$$\|v_t - \bar{v}\|_{\gamma,T} \leq C_2 M^{1-2\gamma},$$

for an integrable random variable C_2. The proof is now easily finished by invoking the inequality

$$\big|V_i(\theta) - \bar{V}_i^M(\theta)\big| \leq \mathbf{E}\left[\|v_t - \bar{v}\|_{\gamma,T}\right].$$

□

Remark 4.6. We have given two separate approximations for $V_i(\theta)$ and $W_i(\theta)$. In order to fully estimate $(V_i(\theta)/W_i(\theta)) - (\bar{V}_i^M(\theta)/\bar{W}_i^M(\theta))$, one should also prove that $W_i(\theta)$ is bounded away from 0. This requires a lower bound for densities of differential equations driven by fractional Broawnian motion, which are out of the scope of the current article.

4.2. **Efficiency of the Monte Carlo simulation.** In this section we aim to study the computational tradeoff between the length of a time period in the Euler discretization (i.e. $1/M$) and the number of Monte Carlo simulations of the sample path (i.e. N). In order to do so we consider $\bar{w}_i^M$ and $\bar{v}_i^M$ as above.

Recall that, given t units of computer time, the Monte-Carlo estimators for $W_i(\theta)$ and $V_i(\theta)$ can be written as

$$\frac{1}{c_1(t,\frac{1}{M})}\sum_{k=1}^{c_1(t,\frac{1}{M})} w_{i,k}^M, \quad \frac{1}{c_2(t,\frac{1}{M})}\sum_{k=1}^{c_2(t,\frac{1}{M})} v_{i,k}^M$$

where $\{w_{i,\ell}^M;\, \ell \geq 1\}$ (resp. $\{v_{i,\ell}^M;\, \ell \geq 1\}$) is a sequence of i.i.d. copies of w_i^M (resp. of v_i^M), and $c_1(t,\frac{1}{M}), c_2(t,\frac{1}{M})$ are the maximal number of runs one is allowed to consider with t units of computer time. Using the result by [10] we can state the following proposition:

Proposition 4.7. *Let N be the number of Monte Carlo simulations and M the number of steps of the Euler scheme, then the tradeoff between N and M for computing $W_i(\theta)$ (and similarly $V_i(\theta)$) is*

$$N \asymp M^{\frac{\tilde{\gamma}}{2\gamma-1}-3},$$

for all $1/2 < \gamma < H$ and $\tilde{\gamma} = Tm(d+1)$, where T is the time horizon, m the dimension of the observed process and d the dimension of the noise process.

Proof. We discuss the proof only for W_i, by following exactly the same steps we can obtain the same result for V_i.

We only need to check that our process w satisfies the conditions of Theorem 1 in [10].

(i) We can easily see that the discretized $\bar{w}_{t_i}^M$ converges uniformly to w_{t_i}.
(ii) In addition, we have bounded moments of w_{t_i}, thus $\mathbf{E}[\bar{W}_{t_i}^2] \to \mathbf{E}[w_{t_i}^2]$.
(iii) From Proposition 4.5 we have that the rate of convergence of the Euler scheme of $\bar{w}_{t_i}^M$ is $M^{1-2\gamma}$, for $1/2 < \gamma < H$.
(iv) The computer time required to generate $\bar{w}_{t_i}^M$ is given by $\tau(1/M)$, which satisfies:

$$\tau(1/M) = Tm(d+1)M = \tilde{\gamma}M$$

where T is the length of the time period, m is the dimension of the SDE, d is the dimension of the fBm and M is the number of Euler steps.

By applying Theorem 1 (by [10]) the optimal rule for choosing the number of Monte-Carlo simulations and the number of Euler steps is chosen such that the asymptotic error is minimized. Therefore, for t the total budget of computer time, as t increases, then the Euler step should converge to zero with order $\frac{1-2\gamma}{\tilde{\gamma}+2-4\gamma}$ or equivalently:

$$\frac{1}{M} \asymp t^{\frac{1-2\gamma}{\tilde{\gamma}+2-4\gamma}} \text{ thus } t \asymp M^{-\frac{\tilde{\gamma}+2-4\gamma}{1-2\gamma}}.$$

But the number of operations needed for an arbitrary Monte Carlo simulation t_0 is equal to $\tilde{\gamma}MN$. Thus, we finally obtain that $N \asymp M^{-\frac{\tilde{\gamma}+2-4\gamma}{1-2\gamma}-1}$. □

4.3. **Discretization of the score function.** Consider the following discretized version of the score function, i.e. $\nabla_l \ell_n(\theta)$:

$$\hat{\nabla}_l \ell_n(\theta) = \frac{\hat{V}_i}{\hat{W}_i} := \frac{\frac{1}{N}\sum_{k=1}^N \bar{v}_{i,k}^M}{\frac{1}{N}\sum_{k=1}^N \bar{w}_{i,k}^M}, \tag{40}$$

where $\bar{w}_{1,k}^M, \bar{w}_{2,k}^M, \dots$ and $\bar{v}_{1,k}^M, \bar{v}_{2,k}^M, \dots$ are iid copies of $\bar{w}_i^M$ and $\bar{v}_i^M$ respectively. Our aim in this section is to give a global bound for the mean square error obtained by approximating $\nabla_l \ell_n(\theta)$ by $\hat{\nabla}_l \ell_n(\theta)$.

Proposition 4.8. *The discretized score function $\hat{\nabla}_l \ell_n(\theta)$ converges to the continuous score function $\nabla_l \ell_n(\theta)$ with rate of convergence of order $M^{-(2\gamma-1)}$, where $1/2 < \gamma < H$ and M is the number of Euler steps used in the discretization.*

Proof. We discuss the idea of the proof for the W_i term first:

$$\begin{aligned}
\mathbf{E}\left(\hat{W}_i - W_i\right)^2 &= \mathbf{E}\left(\frac{1}{N}\sum_{k=1}^N \bar{w}_{i,k}^M - \mathbf{E}[w_i(\theta)]\right)^2 \\
&= \mathbf{E}\left(\frac{1}{N}\sum_{k=1}^N \bar{w}_{i,k}^M - \frac{1}{N}\sum_{k=1}^N w_{i,k} + \frac{1}{N}\sum_{k=1}^N w_{i,k} - \mathbf{E}[w_i(\theta)]\right)^2.
\end{aligned}$$

Thanks now to the independence property between Monte Carlo runs, we get

$$\begin{aligned}
\mathbf{E}\left(\hat{W}_i - W_i\right)^2 &\le \frac{2}{N}\sum_{k=1}^N \mathbf{E}(\bar{w}_{i,k}^M - w_{i,k})^2 \;+\; 2\,\mathbf{E}\left(\frac{1}{N}\sum_{k=1}^N w_{i,k} - \mathbf{E}[w_i(\theta)]\right)^2 \\
&= \frac{1}{N}\sum_{k=1}^N (\text{Euler MSE})^2 \;+\; (\text{Monte Carlo MSE})^2 \asymp (M^{1-2\gamma})^2 + \frac{1}{N},
\end{aligned}$$

and thus

$$\text{MSE}\left(\hat{W}_i - W_i\right) \asymp \sqrt{(M^{1-2\gamma})^2 + \frac{1}{N}}.$$

Now, if we use Proposition 4.7, i.e. $N \asymp M^{-\frac{\tilde{\gamma}+2-4\gamma}{1-2\gamma}-1}$, for all $1/2 < \gamma < H$, and $\tilde{\gamma} = Tm(d+1)$, where T is the time horizon, m the dimension of the observed process and d the dimension of the noise process, we have

$$MSE\left(\hat{W}_i - W_i\right) \asymp \sqrt{M^{2-4\gamma} + M^{\frac{\tilde{\gamma}}{1-2\gamma}+3}} \asymp M^{1-2\gamma},$$

since the first is the dominant term above.

Following the same procedure, we can show that $\text{MSE}(\hat{V}_i - V_i) \asymp M^{1-2\gamma}$ and thus the claim of the proposition follows easily.

□

Remark 4.9. In Proposition 4.8 the rate of convergence is independent of the dimension of the problem, i.e. it is independent of the parameter $\tilde{\gamma} = Tm(d+1)$.

5. Numerical Examples

In this section our aim is to investigate the performance of the suggested maximum likelihood method in practice. We study the one-dimensional fractional Ornstein-Uhlenbeck process, a linear two-dimensional system of fractional SDEs and then some real data given by a financial time series. Before presenting our results, we first discuss some technical issues raised by the algorithmic implementation of our method.

The goal is to find the root of the quantity $\nabla_l \ell_n(\theta)$ with respect to θ. We can divide this procedure in two parts. The first part consists in computing the root of the log-likelihood using a stochastic approximation algorithm. This is a stochastic optimization technique firstly introduced by Robbins and Monro (1951) that is used when only noisy observations of the function are available. In our case it is appropriate, since we want to solve

$$\nabla_l \ell_n(\theta) = 0,$$

where $\nabla_l \ell_n(\theta)$ is given by Theorem 3.17 and has to be approximated by $\hat{\nabla}_l \ell_n(\theta)$. Thus, the recursive procedure is of the following form

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{\nabla}_l \ell_n(\hat{\theta}_k). \tag{41}$$

where $\hat{\nabla}_l \ell_n$ is the estimate of $\nabla_l \ell_n$ at the k-th iteration based on the observations and a_k is a sequence of real numbers such that $\sum_{k=1}^{\infty} a_k = \infty$ and $\sum_{k=1}^{\infty} a_k^2 < \infty$. Under appropriate conditions (see for example [4]), the iteration in (41) converges to θ almost surely. The step sizes satisfy $a_k > 0$ and the way that we choose them can be found in [26].

The second part consists of the computation of $\hat{\nabla}_l \ell_n(\hat{\theta}_k)$ at each step of the stochastic approximation algorithm. Thus, for a given value of θ_k (the one computed at the k-th iteration) we want to compute $\hat{\nabla}_l \ell_n(\theta_k)$ when we are given n discrete observations of the process: y_{t_i}, $i = 1, \ldots, n$. Here, we describe the main idea of the algorithm we use for only one step. Thus, assume that we are at $[t_{i-1}, t_i]$, and at time t_i we obtain the i-th observation. We want to compute $W_i(\theta)$ and $V_i(\theta)$ according to expressions (25) and (26) respectively. To compute the expectations we use simple Monte-Carlo simulations.Therefore, we discretize the time interval into N steps

$$t_{i-1} = s_0 < s_1 < \cdots < s_N = t_i.$$

From each simulated path (apart from that of fBm) we only need to keep the terminal value which is the value of the process at time t_i. The algorithm is the following

(1) Simulate N values of fBm in the interval $[t_{i-1}, t_i]$ using for example the circulant matrix method (any exact -preferably- simulation technique can be used).
(2) Using the simulated values from step 1 and an Euler scheme for the SDE (1), simulate the value of the process at time t_i. For example, for $k = 0, \ldots, N$

$$\bar{Y}^M_{s_k} = \bar{Y}^M_{s_{k-1}} + \mu(\bar{Y}^M_{s_k})(s_k - s_{k-1}) + \sum_{j=1}^{d} \sigma^{(j)}(\bar{Y}^M_{s_{k-1}})(B^{(j)}_{s_k} - B^{(j)}_{s_{k-1}}).$$

(3) Using step 2 and the observation at time t_i, compute the indicator function $\mathbf{1}_{(Y_{t_i}(\theta) > y_{t_i})}$.

(4) Using step 1 and an Euler scheme simulate $D_{t_i}Y^i_\tau$, as given in Lemma 3.3 for $n = 1$ -first Malliavin derivative-.
(5) Using step 1 and an Euler scheme simulate $\eta^{kj}_{t_i}$, $k, j = 1, \dots, m$, as given in Proposition 3.7.
(6) Steps 4 and 5 are used to compute $Q^{pj}_{st_i}$, $p \in \{1, \dots, m\}$, $j \in \{1, \dots, d\}$ as defined in Propositions 3.12 and 3.14.
(7) Simulate the Malliavin derivative of the product $D_s[Y_t Q^{pj}_{rt}]$.
(8) Using the previous steps, numerical integration for the double integral and numerical integration for the stochastic integral we compute $U_p(Y_{t_i}(\theta))$ as defined in Proposition 3.12.
(9) Recursively compute $H_{(1,\dots,m)}(Y_{t_i}(\theta))$ as given in (19).
(10) Combine steps 3 and 9 to obtain the kernel $W_i(\theta)$.
(11) We repeat steps 1 through 10 N times and we average these values to obtain an estimate for the expectation $W_i(\theta)$.

Using a similar procedure we can obtain an estimate for the expectation $V_i(\theta)$. Finally, for each $i = 1, \dots, n$ we compute $V_i(\theta)/W_i(\theta)$ and sum over i to obtain the desired value of the log-likelihood at θ_k.

We have completed the study of our numerical approximation of the log-likelihood, and are now ready for the analysis of some numerical examples.

5.1. **Fractional Ornstein-Uhlenbeck process.** Though our method can be applied to highly nonlinear contexts, we focus here on some linear situations, which allow easier comparisons with existing methods or exact computations. Let us first study the one-dimensional fractional Ornstein-Uhlenbeck process, i.e.

$$dY_t = -\lambda Y_t dt \; + \; dB_t, \tag{42}$$

where the solution is given $Y_t(\lambda) = \int_0^t e^{-\lambda(t-s)} dB_s$ (notice the existence of an explicit solution here). In this case our methodology is quite simplified. The log-likelihood can be written as follows:

$$\partial_\lambda \ell(\lambda; y) = \sum_{i=1}^n \frac{\mathbf{E}\Big[\partial_\lambda Y_t(\lambda)\, \mathbf{1}_{(Y_t(\lambda)>y)}\; H_{(1,1)}(\lambda) + \Big(Y_t(\lambda) - y\Big)_+ \partial_\lambda H_{(1,1)}(\lambda)\Big]}{\mathbf{E}\Big[\mathbf{1}_{(Y_t(\lambda)>y)}\; H_{(1)}\Big(Y_t(\lambda), 1\Big)\Big]}.$$

The Malliavin derivative of $Y_t(\lambda)$ satisfies the following ODE

$$D_s Y_t(\lambda) = 1 - \lambda \int_s^t D_s Y_u(\lambda) du,$$

with solution $D_s Y_t(\lambda) = e^{-\lambda\, t}\; \mathbf{1}_{\{s \le t\}}$. The corresponding norm is

$$\|D_\cdot Y_t(\lambda)\|^2 = c_H \int_s^t \int_s^t e^{-\lambda(u+v)} |u - v|^{2H-2} du dv.$$

The higher order derivatives of $Y_t(\lambda)$ are equal to zero. Therefore,

$$H_{(1)}\Big(Y_t(\lambda)\Big) = \frac{1}{\|D_\cdot Y_t(\lambda)\|^2} \int_s^t e^{-\lambda u} dB_u$$

and thus

$$H_{(1,1)}(\lambda) = \frac{1}{\|D.Y_t(\lambda)\|^4}\int_s^t\int_s^t e^{-\lambda(u+v)}dB_u dB_v - c_H\,\|D.Y_t(\lambda)\|^{-2}.$$

The derivative with respect to the unknown parameter λ satisfies

$$\partial_\lambda Y_t(\lambda) = -\int_0^t Y_s(\lambda) - \lambda\partial_\lambda Y_s(\lambda)ds$$

with solution $\partial_\lambda Y_t(\lambda) = \int_0^t (t-s)e^{-\lambda(t-s)}dB_s$. The last term we need to compute is:

$$\begin{aligned}\partial_\lambda H_{(1,1)}(\lambda) = {} & \frac{1}{\|D.Y_t(\lambda)\|^8}\Bigg[\|D.Y_t(\lambda)\|^4\int_s^t\int_r^t -(u+v)e^{-\lambda(u+v)}dB_u dB_v \\ & -2c_H\|D.Y_t(\lambda)\|^2\int_s^t\int_r^t -(u+v)e^{-\lambda(u+v)}|u-v|^{2H-2}dudv\Bigg] \\ & -\frac{c_H^2\int_s^t\int_r^t -(u+v)e^{-\lambda(u+v)}|u-v|^{2H-2}dudv}{\|D.Y_t(\lambda)\|^4}.\end{aligned}$$

Now, we compute the MLE following the algorithm we described above. The results we obtained are summarized in the following table:

True λ	MLE $\hat{\lambda}$	Standard Error
0.5	0.497	0.00369
4	3.861	0.00127

Remark 5.1. The value of H used for the simulation of the process is 0.6. The number of observations is $n = 50$, the number of Euler steps is $M = 500$, the number of stochastic approximation steps is $K = 50$ and the number of MC simulations $N = 500$.

5.2. **Two-dimensional fractional SDE.** In this section we study the following system of fractional OU processes:

$$\begin{aligned}dY_t^{(1)} &= -\alpha Y_t^{(2)}dt \;+\; \beta dB_t^{(1)} \\ dY_t^{(2)} &= -\beta Y_t^{(1)}dt \;+\; \beta dB_t^{(2)}.\end{aligned} \tag{43}$$

In this case, the computations are more involved even though the SDEs are linear functions of Y. Furthermore, the parameter we want to estimate is two-dimensional as well ($\theta = (\alpha,\beta)^T$), which complicated the optimization procedure. Therefore, instead of computing only one derivative, we need to compute both derivatives with respect to α and β and then compute the solution of the system of two equations

$$\nabla_\alpha \ell(\alpha,\beta;y) = 0, \qquad \nabla_\beta \ell(\alpha,\beta;y) = 0,$$

where

$$\begin{aligned}\nabla_l \ell(\alpha,\beta;y) = \sum_{i=1}^n & [\mathbf{E}[\mathbf{1}_{(Y_t(\alpha,\beta)>y)}\; H_{(1,2)}(Y_t(\alpha,\beta))]^{-1} \\ & \times\{\mathbf{E}[\nabla_l Y_t(\alpha,\beta)\;\mathbf{1}_{(Y_t(\alpha,\beta)>y)}\; H_{(1,2,1,2)}(\alpha,\beta) + (Y_t(\alpha,\beta)-y)_+\nabla_l H_{(1,2,1,2)}(\alpha,\beta)]\}\end{aligned}$$

and $l = \alpha$ or β. The Malliavin derivative of Y_t computes as follows:

$$D_s Y_t^{(1)} = \beta - \alpha \int_s^t D_s Y_u^{(2)} du \qquad D_s Y_t^{(2)} = \beta - \beta \int_s^t D_s Y_u^{(1)} du.$$

The covariance matrix γ_t is given by $(\langle D.Y_t^i, D.Y_t^j \rangle)_{1 \le i,j \le 2}$. The inverse of the covariance matrix satisfies the following SDE

$$\gamma_t^{-1} = -\int_0^t [\gamma_u^{-1} M + M^T \gamma_u^{-1}] du,$$

where

$$M = \begin{bmatrix} 0 & \alpha \\ \beta & 0 \end{bmatrix}$$

Now, it remains to compute the quantities $H_{(1,2)}$ and $H_{(1,2,1,2)}$. This can be done using the recursive formulas in Proposition 3.12, but we need to keep in mind that higher order derivatives of Y are equal to zero, thus they will be simplified. Indeed,

$$H_{(1)}(Y_t) = \sum_{j=1}^{2} Y_t \int_0^t (\gamma_s^{-1})^{1j} D_s Y_t^j dB_s^j - c_H \int_0^t \int_0^t D_s Y_t^j Q_{rt} |r-s|^{2H-2} dr ds.$$

Moreover, we can easily see that

$$H_{(1,2)}(Y_t) = H_{(1)}(Y_t) \int_0^t Q_{st} dB_s - c_H \int_0^t \int_0^t D_s H_{(1)}(Y_t) Q_{rt} |r-s|^{2H-2} dr ds$$

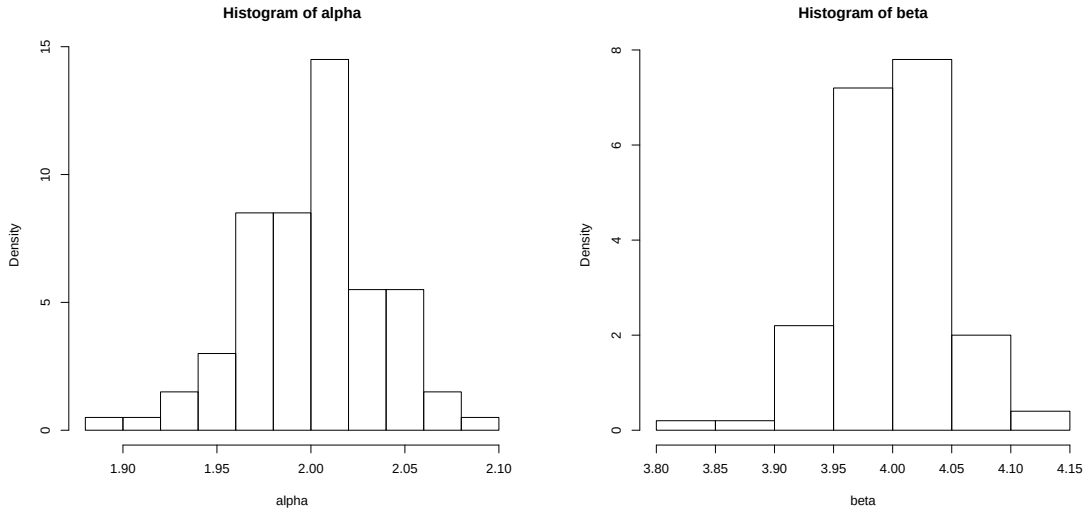
$$H_{(1,2,1,2)}(Y_t) = H_{(1,2,1)}(Y_t) \int_0^t Q_{st} dB_s - c_H \int_0^t \int_0^t D_s H_{(1,2,1)}(Y_t) Q_{rt} |r-s|^{2H-2} dr ds$$

Of course, recall that $Q_{st}^{pj} = (\gamma_s^{-1})^{pj} D_s Y_t^j$. In practice, these quantities are computed recursively. The last step is to compute the derivative of $H_{(1,2,1,2)}(Y_t)$ with respect to α and β, which in this case is not as complicated and compute the MLEs using the algorithm discussed in the previous section. The table below summarizes our results, and we have plotted the corresponding histograms in Figure 1.

Parameter	True Value	MLE	Standard Error
α	2	2.003	0.0518
β	4	3.987	0.0157

Remark 5.2. The value of H used for the simulation of the process is 0.6. The number of observations is $n = 50$, the number of Euler steps is $N = 500$, the number of stochastic approximation steps is $K = 50$ and the number of MC simulations $M = 500$.

5.3. **Application to financial data.** One of the most popular applications of fractional SDEs is in finance. Hu and Oksendal, [20], introduced the fractional Black-Scholes model in order to account for inconsistencies of the existing models in practice. More specifically, the stock price is described therein by a fractional geometric Brownian motion with Hurst parameter $1/2 < H < 1$. The choice of this model is based on empirical studies that displayed the presence of long-range dependence on stock prices, for example in Willinger, Taqqu and Teverovsky, [42].



Drift Parameter. Diffusion Parameter.

FIGURE 1. Empirical Distribution of the estimators for α and β.

However, the presence of fractional Brownian motion in the model allows for arbitrage in the general setting. It has been shown that arbitrage opportunities can be avoided in a number of ways, for example the reader can refer to Rogers [39], Dasgupta and Kallianpur [7] and Cheridito [5]. We choose to model the stock price as as follows:

$$dS_t = \mu S_t dt + \sigma dB_t, \tag{44}$$

where B is a fractional Brownian motion with Hurst index $1/2 < H < 1$. For this SDE (as well as for a more general class of fractional SDEs) Guasoni, [13], proved that there is no arbitrage when transaction costs are present.

Our goal is to estimate the unknown parameters μ and σ based on daily observations of the S&P 500 index (data from June 2010 until December 2010). Since the Hurst parameter is piece-wise constant, we devide the data in three groups (of 50 daily observations each) and we compute for each one the Hurst index using the Rescaled-Range (R/S) statistic. We obtain that for the first group of data $\hat{H}_1 = 0.59$, for the second $\hat{H}_2 = 0.63$ and for the third one $\hat{H}_3 = 0.61$. For the different groups, we apply our maximum likelihood approach in order to estimate μ and σ. The estimates are summarized in the following table:

Estim. Parameters	Group 1: $\hat{H}_1 = 0.59$	Group 2: $\hat{H}_2 = 0.63$	Group 3: $\hat{H}_3 = 0.61$
$\hat{\mu}$	0.015 (0.0123)	0.019 (0.0144)	0.011 (0.0214)
$\hat{\sigma}$	0.352 (0.058)	0.339 (0.046)	0.341 (0.024)

Remark 5.3. The volatility is computed in years. In addition, during this period of time the historical volatility is around 0.38, which is coherent with our own estimation.

References

[1] F. Baudoin, L. Coutin: Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stoch. Proc. Appl.* **117** (2007), no. 5, 550–574.

[2] F. Baudoin, C. Ouyang: Gradient Bounds for Solutions of Stochastic Differential Equations Driven by Fractional Brownian Motions. *Arxiv* Preprint (2011).

[3] F. Baudoin, C. Ouyang, S. Tindel: Gaussian bounds for the density of solutions of stochastic differential equations driven by fractional Brownian motions. In preparation.

[4] J. R. Blum: Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics.* **25**(1954), no. 4, 737–744.

[5] P. Cheridito: Arbitrage in fractional Brownian motion models. *Finance Stoch.*, **7** (2003), no. 4, 533–553.

[6] R. Dalang, E. Nualart: Potential theory for hyperbolic SPDEs. *Ann. Probab.* **32** (2004), no. 3A, 2099–2148.

[7] A. Dasgupta and G. Kallianpur: Arbitrage opportunities for a class of Gladyshev processes, *Appl. Math. Optim.*, **41** (2000), no. 3, 377-385.

[8] A. Deya, A. Neuenkirch, S. Tindel: A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Arxiv* preprint (2010). To appear in *Ann. Inst. Henri Poincar Probab. Stat.*.

[9] D. Duffie, P. Glynn: Efficient Monte Carlo simulation of security prices. *Ann. Appl. Probab.* **5** (4) (1995), 897–905.

[10] G. Durham, A. Gallant: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** (2002), no. 3, 297–338.

[11] P. Friz and N. Victoir: *Multidimensional stochastic processes as rough paths. Theory and applications.* Cambridge University Press, 2010.

[12] E. Gobet: Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli* **7** (2001), no. 6, 899–912.

[13] P. Guasoni: No arbitrage under transaction costs, with fractional Brownian motion and beyond. *Mathematical Finance*, **16** (2006), no. 3, 569–582.

[14] M. Gubinelli: Controlling rough paths. *J. Funct. Anal.* **216**, 86-140 (2004).

[15] M. Hairer: Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.* **33** (2005), no. 2, 703–758.

[16] M. Hairer, A. Majda: A simple framework to justify linear response theory. *Nonlinearity* **23** (2010), no. 4, 909–922.

[17] M. Hairer, A. Ohashi: Ergodicity theory of SDEs with extrinsic memory. *Ann. Probab.* **35** (2007), no. 5, 1950–1977.

[18] M. Hairer, S. Pilai: Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Arxiv* Preprint (2009).

[19] Y. Hu, D. Nualart: Differential equations driven by Hölder continuous functions of order greater than 1/2. *Abel Symp.* **2** (2007), 349-413.

[20] Y. Hu and B. Oksendal: Fractional white noise calculus and applications to finance, *Infnite dimentional analysis, quantum probbility and related topics*, **6** (2003), no. 1, 1–32.

[21] Y. Hu, B. Oksendal and A. Sulem: Optimal sonsumption and portfolio in a Black-Scholes market driven by fractional Brownian motion, *Infnite dimentional analysis, quantum probbility and related topics*, **6** (2003), no. 4, 519–536.

[22] R. Kasonga: Maximum likelihood theory for large interacting systems. *SIAM J. Appl. Math.* **50** (1990), no. 3, 865–875.

[23] M. Kleptsyna, A. Le Breton: Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Stat. Inference Stoch. Process.* **5** (2002), no. 3, 229–248.

[24] A. Kohatsu-Higa: Lower bounds for densities of uniformly elliptic random variables on Wiener space. *Probab. Theory Related Fields* **126** (2003), no. 3, 421–457.

[25] S. Kou, X. Sunney-Xie: Generalized Langevin equation with fractional Gaussian noise: subdiffusion within a single protein molecule. *Phys. Rev. Lett.* **93**, no. 18 (2004).

[26] H. J. Kushner, G.G. Yin: *Stochastic approximation algorithms and applications.* Springer-Verlag, 1997.
[27] A. Lejay (2003): *An Introduction to Rough Paths.* Séminaire de probabilités 37, vol. **1832** of Lecture Notes in Mathematics, 1-59.
[28] J. León, S. Tindel: Malliavin calculus for fractional delay equations. *Arxiv* Preprint (2009).
[29] T. Lyons, Z. Qian (2002): *System control and rough paths.* Oxford University Press.
[30] Y. Mishura, G. Shevchenko: The rate of convergence for Euler approximations of solutions of stochastic differential equations driven by fractional Brownian motion. *Stochastics* **80**(5) (2008), 489–511.
[31] A. Neuenkirch, I. Nourdin, A. Rößler, S. Tindel : Trees and asymptotic developments for fractional diffusion processes. *Ann. Inst. Henri Poincar Probab. Stat.* **45** (2009), no. 1, 157–174.
[32] D. Nualart: *Malliavin Calculus and Related Topics.* Springer-Verlag, 2006.
[33] D. Nualart, A. Răşcanu: Differential equations driven by fractional Brownian motion. *Collect. Math.* **53** no. 1 (2002), 55-81.
[34] D. Nualart, B. Saussereau: Malliavin calculus for stochastic differential equations driven by a fractional Brownian motion. *Stochastic Process. Appl.* **119** (2009), no. 2, 391–409.
[35] D. Odde, E. Tanaka, S. Hawkins, H. Buettner: Stochastic dynamics of the nerve growth cone and its microtubules during neurite outgrowth. *Biotechnology and Bioengineering* **50**, no. 4, 452-461 (1996).
[36] A. Pedersen: Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1** (1995), no. 3, 257–279.
[37] A. Papavasiliou, C. Ladroue: Parameter estimation for rough differential equations. *Arxiv* preprint (2009).
[38] Ll. Quer, D. Nualart: Gaussian density estimates for solutions to quasi-linear stochastic partial differential equations. *Stochastic Process. Appl.* **119** (2009), no. 11, 3914–3938.
[39] L. C. G. Rogers: Arbitrage with fractional Brownian motion, *Math. Finance*, **7** (1997), no. 1, 95–105.
[40] M. Sorensen: Parametric inference for discretely sampled stochastic differential equations. In Andersen, T.G. Davis, R.A., Kreiss, J.-P. and Mikosch, T. (eds.): Handbook of Financial Time Series, Springer (2009), 531 - 553.
[41] C. Tudor, F. Viens: Statistical aspects of the fractional stochastic calculus. *Ann. Statist.* **35** (2007), no. 3, 1183–1212.
[42] W. Willinger, M. S. Taqqu and V. Teverovsky: Stock market prices and long-range dependence, *Finance Stoch.*, **3** (1999), no. 1, 1–13.
[43] M. Zähle: Integration with respect to fractal functions and stochastic calculus I. *Probab. Theory Relat. Fields* **111** (1998), 333-374.

Alexandra Chronopoulou, Samy Tindel: Institut Élie Cartan Nancy, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France. *Email:* `Alexandra.Chronopoulou@iecn.u-nancy.fr, tindel@iecn.u-nancy.fr`