Parametric estimation of several parameters in discretely-observed Stochastic Differential Equations with additive fractional noise

El Mehdi Haress, Alexandre Richard

Parametric estimation of several parameters in discretely-observed Stochastic Differential Equations with additive fractional noise

El Mehdi Haress *   Alexandre Richard†

April 4, 2023

Abstract

We investigate the problem of joint statistical estimation of several parameters for a stochastic differential equations driven by an additive fractional Brownian motion. Based on discrete-time observations of the model, we construct an estimator of the Hurst parameter, the diffusion parameter and the drift in a parametrised family of coercive drift coefficients. Our procedure is based on the assumption that the stationary distribution of the SDE and of its increments permit to identify the parameters of the model. We prove consistency results and derive a rate of convergence for the estimator under this assumption. Finally, we show that the identifiability assumption is satisfied in the case of a family of fractional Ornstein-Uhlenbeck processes and illustrate our results with some numerical experiments.

1 Introduction

Consider the following $\mathbb{R}^d$-valued stochastic differential equation

$$Y_t = Y_0 + \int_0^t b_{\xi_0}(Y_s)ds + \sigma_0 B_t, \tag{1.1}$$

where B is an $\mathbb{R}^d$-fractional Brownian motion (fBm) with Hurst parameter $H_0 \in (0, 1)$. The goal in this work is to estimate simultaneously the parameter ξ_0, the diffusion coefficient σ_0 and the Hurst parameter H_0 from discrete observations of the process Y. We will assume that the drift parameter ξ_0 lies in a set Ξ of $\mathbb{R}^m$ and $\{b_\xi(\cdot), \xi \in \Xi\}$ is a parametrised family of drift coefficients with $b_\xi(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, and σ_0 is an invertible $\mathbb{R}^{d\times d}$ matrix. The unknown parameters are denoted by $\theta_0 = (\xi_0, \sigma_0, H_0) \in \mathbb{R}^{q+1}$, where $q = m + d^2$.

In the framework of SDEs driven by fBm, many recent works have focused on the parametric estimation of the drift, mostly assuming that the process Y is observed continuously and that the parameters H and σ are known (see e.g [1, 15, 16, 25, 29]). These works propose estimators of ξ_0 which are strongly consistent, providing a rate of convergence towards ξ_0 and even sometimes a central limit theorem is obtained [15, 16]. However there are restrictions on the drift function, namely that it is of the form $b_\xi(y) = -\xi y$, i.e. a family of Ornstein-Uhlenbeck (OU) processes, or of the form $b_\xi(y) = \xi b(y)$ as in [29]. In practical situations though, we only have access to discrete-time observations of the process Y. Taking into account this constraint, two recent papers [17, 24] constructed estimators of ξ_0 which were proven to be strongly consistent. Their rate of convergence is studied and a central limit theorem is also proven in [17]: while [17] considers the fractional OU case, [24] treats general drift functions which satisfy a coercivity assumption.

The diffusion coefficient σ_0 is usually estimated using the quadratic variations, which is possible only when the process is either observed continuously or the step-size goes to zero (i.e high frequency

*Université Paris-Saclay, CentraleSupélec, MICS and CNRS FR-3487. el-mehdi.haress@centralesupelec.fr
†Université Paris-Saclay, CentraleSupélec, MICS and CNRS FR-3487. alexandre.richard@centralesupelec.fr
E.H. acknowledges the support of the Labex de Mathématique Hadamard. This work is supported by the SIMALIN project ANR-19-CE40-0016 from the French National Research Agency.

data), see [31] and [2]. The Hurst parameter H_0 is also estimated using quadratic variations, see e.g. [21], or by a direct access to discrete observations of a fractional Brownian motion path with a step-size that goes to zero as in [11].

When it comes to estimating all the parameters (ξ_0, σ_0, H_0), we refer to [4] where the observations are assumed to be made continuously, and [13] which is, to the best of our knowledge, the only work which estimates all the parameters of a fractional Ornstein-Uhlenbeck process in a discrete-time setting.

In this paper, we follow the approach of [24]. That is, we work with the assumption *the stationary distribution of Y identifies the parameters.* Hence we will use the stationary distribution of (1.1) to estimate the parameters. However, as illustrated by the authors of [13], in the simple case of a one-dimensional fractional OU process, this claim is false for more than one parameter to estimate. In fact, the stationary distribution of Y is Gaussian and therefore distinguished by its mean (which does not depend on the parameters) and its variance. In this case, the variance itself cannot identify the three parameters. In [13], this issue is circumvented by considering the increments of Y; the increments of the stationary solution are also Gaussian but have different variances. Thus, adding two increments, the authors have access to three functions and show that these functions are sufficient to estimate the parameters. We propose here a similar approach that generalizes the one presented in [13]. We add q linear transformations of the original process and assume that they are enough to identify the parameters. Therefore, our assumption (which is detailed later) will be that *the stationary distribution of Y and its increments identify the parameters (ξ, σ, H).*

Assume for simplicity that the observations are of the form $(Y^{\theta_0}_{kh})_{k=0,..,n+q}$ and consider q linear transformations $\{\ell^i(Y^{\theta_0}_{kh}, \ldots, Y^{\theta_0}_{kh+ih})\}_{k=0,..,n}$ where $i \in [\![1, q]\!]$. Hence, we now have access to $q+1$ paths, which we use to define one path of a higher-dimensional process X^{θ_0} that we call the augmented process associated to the SDE (1.1). With access to a path of X^{θ_0}, we construct the estimator of θ_0 by $\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmin}}\, d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \mu_\theta\right)$, where μ_θ is the stationary distribution of X^θ. We prove that $\hat{\theta}_n$ is a strongly consistent estimator of θ_0 and study its rate of convergence.

In [13], the authors do not prove the identifiability assumption but provide numerical experiments to support their claim. We prove here that in the setting of [13], i.e. of a fractional OU process, the aforementioned identifiability assumption holds. Also, as in [24], we consider tow variations of this assumption, a weak one which we will just call the identifiability assumption and a strong one. Moreover, to construct good estimators of the drift parameter ξ, the authors of [24] proved beforehand results on the regularity of Y with respect to ξ. This is a natural procedure, since the estimation method relies on minimizing a certain functional of Y, by showing that it has the right regularity properties so that its true minimum is in fact the true parameter ξ. Here, in view of estimating all the parameters, we will will study the regularity of Y with respect to σ and H as well.

Since we are interested in ergodic estimators, we need the regularity of Y in all the parameters to be uniform in time. In particular we need the regularity in H to hold uniformly in $t \geq 0$. To achieve this, the drift will be assumed to be contractive. Let us mention that the sensitivity in the Hurst parameter has been studied in various situations and is an important topic in modeling: in [20], the fractional Brownian motion is proven to be infinitely differentiable with respect to its Hurst parameter for a fixed time; in [18, 19], the law of (multiple) integrals with respect to the fBm are proven to be continuous in H; in [26], the Hölder continuity in H is studied for generalised fractional Brownian fields (over compact index sets); and in [10], the laws of quasilinear stochastic wave and heat equations with additive fractional noise are proven to be continuous in H. Finally in [28], the difference between functionals of a fractional stochastic differential equation (SDE) and its Markovian counterpart ($H = \frac{1}{2}$) are proven to be of order $|H - \frac{1}{2}|$, both for the law of the solution on a compact time interval and for the law of the first hitting time (see also [27] for a numerical approach and applications, in particular in neuroscience). In this work, new results on the Hurst regularity of fractional models were needed, and they have been gathered in our separate paper [14].

Organisation of the paper. The paper is organized as follows. In Section 2, we first detail the notations and some assumptions. Then, we describe how to construct the estimator and present the main results. In Section 3, we prove the consistency result and the rate of convergence. In Section 4, we show that our estimator can be practically implemented by estimating the stationary distribution through an Euler scheme. We prove consistency and obtain a rate of convergence with this additional layer of estimation. In Section 5.1, we prove the identifiability assumption for the fractional Ornstein-Uhlenbeck process, and in Section 5.2, we construct a more general family of SDEs that verifies the strong identifiability assumption. The construction is based on small perturbations of the fractional Ornstein-Uhlenbeck model. We also implement our method and run numerical simulations in Section 5.3. In the Appendix, we recall in Section A some results from our companion paper [14]. In Section B, we prove continuity and tightness results on Y and the solution of the Euler scheme associated to (1.1). Section C is dedicated to the proof of Proposition 4.1. Finally, in Section D, we obtain a crucial result for the proof of the identifiability assumption for the fractional Ornstein-Uhlenbeck process.

2 A general procedure

We first give some general notation. Then we state the assumptions on the coefficients of (1.1) and define the estimator. At the end of this section, we give an almost sure convergence for this estimator result as well as a convergence rate.

2.1 Notation and assumptions

Notations. Let $\mathcal{M}_1(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$. We will consider the p-Wasserstein distance, which is defined for every μ, ν in $\mathcal{M}_1(\mathbb{R}^d)$ as follows:

$$\mathcal{W}_p(\mu,\nu) = \inf\{(\mathbb{E}|X-Y|^p)^{\frac{1}{p}}\,; \mathcal{L}(X)=\mu,\ \mathcal{L}(Y)=\nu\}.$$

For any given p, we denote by $\mathcal{D}_p$ the set of distances dominated by the p-Wasserstein distance. Mainly, we shall work with the distance $d_{CF,p}$ defined as

$$d_{CF,p}(\mathcal{L}(X),\mathcal{L}(Y)) = \left(\int (\mathbb{E}[e^{i\langle \chi, X\rangle}] - \mathbb{E}[e^{i\langle \chi, Y\rangle}])^2 g_p(\chi) d\chi\right)^{1/2}, \tag{2.1}$$

where g_p is an integrable kernel for $p > (\frac{d}{2} \vee 1)$ of the form

$$g_p(\chi) = c_p(1+|\chi|^2)^{-p}, \tag{2.2}$$

and $c_p = (\int_{\mathbb{R}^d}(1+|\chi|^2)^{-p}d\chi)^{-1}$ is a normalizing constant. Finally, we denote by C a constant that can change from line to line and that does not depend on time and the parameters ξ, H, σ. When we want to make the dependence of C on some other parameter a explicit, we will write C_a.

Assumptions. First, we assume that the number of unknown parameters $q+1$ is such that $q \geq 1$ (we have at least two unknowns), which is decomposed into m parameters for the drift b_{ξ_0}, $\xi_0 \in \Xi \subset \mathbb{R}^m$, d^2 parameters for $\sigma \in \mathbb{R}^{d\times d}$ and the last one which is the Hurst parameter. The next assumption states the compactness of the spaces where the parameters lie.

$\mathbf{A_0}$. Ξ is compactly embedded in $\mathbb{R}^m$ for a given $m \geq 1$. H_0 belongs to $\mathcal{H}$, a compact subset of $(0,1)$. The diffusion matrix σ_0 belongs to Σ a compact set of $d \times d$-invertible matrices.

Therefore, we have that $\Theta = \Xi \times \Sigma \times \mathcal{H}$ is a compact subset of $\mathbb{R}^{q+1}$. We will also assume a coercivity assumption on the drift b.

A_1. $b \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Xi; \mathbb{R}^d)$ and there exist constants $\beta, K, c > 0$ and $r \in \mathbb{N}$ such that

(i) For every $x, y \in \mathbb{R}^d$ and $\xi \in \Xi$, we have

$$\langle b_\xi(x) - b_\xi(y), x - y\rangle \le -\beta |x-y|^2 \text{ and } |b_\xi(x) - b_\xi(y)| \le K|x-y|. \tag{2.3}$$

(ii) For every $x \in \mathbb{R}^d$ and $\xi_1, \xi_2 \in \Xi$, the following growth bound is satisfied:

$$|b_{\xi_1}(x) - b_{\xi_2}(x)| \le c(1+|x|^r). \tag{2.4}$$

For $\theta = (\xi, \sigma, H) \in \Theta$, we denote by Y^θ the unique solution of the following equation

$$Y_t^\theta = Y_0 + \int_0^t b_\xi(Y_s^\theta) ds + \sigma B_t, \tag{2.5}$$

where $Y_0 \in \mathbb{R}^d$ and B is an fBm of Hurst parameter H. Under A_1, [12] (see also [24, Remark 2.4] and the references therein) gives the existence and uniqueness of the invariant measure to (2.5). We denote by $\bar{Y}^\theta$ the unique stationary solution. We also denote by ν_θ the stationary distribution of $\bar{Y}^\theta$. For $i \in [\![1, q]\!]$, let ℓ^i denote a linear transformation from $\left(\mathbb{R}^d\right)^{i+1}$ to $\mathbb{R}^d$, which we will assume fixed for the rest of the paper.

Let us define the following processes for all $i \in [\![1, q]\!]$:

$$\begin{aligned}
Z^{i,\theta}_\cdot &= \ell^i(Y^\theta_\cdot, ..., Y^\theta_{\cdot+ih}) \\
\bar{Z}^{i,\theta}_\cdot &= \ell^i(\bar{Y}^\theta_\cdot, ..., \bar{Y}^\theta_{\cdot+ih}) \\
X^\theta_\cdot &= (Y^\theta_\cdot, Z^{1,\theta}_\cdot, ..., Z^{q,\theta}_\cdot) \qquad (2.6)\\
\bar{X}^\theta_\cdot &= (\bar{Y}^\theta_\cdot, \bar{Z}^{1,\theta}_\cdot, ..., \bar{Z}^{q,\theta}_\cdot). \qquad (2.7)
\end{aligned}$$

Observe that for all $\theta \in \Theta$ and $i \in [\![1, q]\!]$, the processes $\bar{Z}^{i,\theta}$ and $\bar{X}^\theta$ are stationary. Denote their distribution by $\eta_{i,\theta}$ and μ_θ respectively. For simplicity, we will not write the parameter θ on the processes when θ is the true parameter θ_0. Also, denote by X^θ the augmented process associated to the SDE (2.5). Simple triangle inequalities yield the following inequalities for all $\theta, \theta' \in \Theta$ and $p > 0$:

$$\begin{aligned}
|X^\theta_\cdot|^p &\le C_{p,q}\left(\sum_{i=0}^q |Y^\theta_{\cdot+ih}|^p\right) \\
|X^\theta_\cdot - X^{\theta'}_\cdot|^p &\le C_{p,q}\left(\sum_{i=0}^q |Y^\theta_{\cdot+ih} - Y^{\theta'}_{\cdot+ih}|^p\right) \qquad (2.8)\\
|X^\theta_\cdot - \bar{X}^\theta_\cdot|^p &\le C_{p,q}\left(\sum_{i=0}^q |Y^\theta_{\cdot+ih} - \bar{Y}^\theta_{\cdot+ih}|^p\right),
\end{aligned}$$

where $C_{p,q}$ is a constant that do not depend on θ or θ'. This means that upper bounds on X will be obtained by upper bounding Y, and the regularity of the process X will be studied through the regularity of the process Y.

As was highlighted previously in the introduction, the estimators are defined by assuming that μ_θ characterizes θ. This weak indentifiability hypothesis reads as follows:

I_w. For any θ in Θ,

$$\mu_\theta = \mu_{\theta_0} \Longleftrightarrow \theta = \theta_0, \tag{2.9}$$

where we recall that μ_θ is the stationary distribution of $\bar{X}^\theta$.

Remark 2.1. *A similar assumption is considered in [24]. However, they work with the stationary distribution of $\bar{Y}$. The assumption we make is weaker. In fact, assume that $\nu_\theta = \nu_{\theta_0}$ iff $\theta = \theta_0$. Now, let θ, θ_0 in Θ such that $d_{CF,p}(\mu_\theta, \mu_{\theta_0}) = 0$. Using the definition of $d_{CF,p}$, we have*

$$\textit{for almost all } \chi \in \mathbb{R}^{(q+1)d},\ \mathbb{E}\left[e^{i\langle \chi, X^\theta\rangle}\right] = \mathbb{E}\left[e^{i\langle \chi, X^{\theta_0}\rangle}\right],$$

which implies that

$$\text{for almost all } \chi \in \mathbb{R}^d,\ \mathbb{E}\left[e^{i\langle \chi, \bar{Y}^\theta \rangle}\right] = \mathbb{E}\left[e^{i\langle \chi, \bar{Y}^{\theta_0} \rangle}\right],$$

Hence, we have $d_{CF,p}(\nu_\theta, \nu_{\theta_0}) = 0$*, which means that* $\theta = \theta_0$.

2.2 Construction of the estimator

Assume that the solution Y is discretely observed at some times $\{kh; k = 1, ..., n+q\}$ for a fixed time step $h > 0$. Under Assumption $\mathbf{A_1}$, we show the following lemma (the proof is postponed to Section 3.2):

Lemma 2.2. *For any* $d \in \mathcal{D}_2$ *and any* $\theta \in \Theta$*, we have*

$$d\left(\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds, \mu_\theta\right) \underset{t\to+\infty}{\longrightarrow} 0 \quad a.s,$$

and

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{kh}^\theta}, \mu_\theta\right) \underset{n\to+\infty}{\longrightarrow} 0 \quad a.s.$$

Remark 2.3. *The integral* $\int_0^t \delta_{X_s} ds$ *is to be understood as the probability measure which associates to each Borel set* A *the value* $\int_0^t \delta_{X_s}(A) ds$.

Hence, under the identifiability assumption $\mathbf{I_w}$, we define the estimator

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmin}}\ d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \mu_\theta\right), \tag{2.10}$$

where d is a distance in $\mathcal{M}_1(\mathbb{R}^d)$. However, this means that we need to compute μ_θ which in most cases is not explicitly known. We discuss a way to overcome this problem in Section 4.

2.3 Main results

The first result states the strong consistency of the estimator (2.10) under the assumptions $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{I_w}$. The proof of the theorem below is detailed in Section 3.3.

Theorem 2.4. *Assume that* $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{I_w}$ *hold. Consider a distance* d *on* $\mathcal{M}_1(\mathbb{R}^d)$ *which belongs to* $\mathcal{D}_2$*. Then* $(\hat{\theta}_n)_{n\in\mathbb{N}}$ *defined in* (2.10) *is a strongly consistent estimator of* θ_0 *in the following sense:*

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \ a.s.$$

We will also establish a rate of convergence of this estimator when $d = d_{CF,p}$ for some $p \in \mathbb{N}^*$, under the strong identifiability assumption:

$\mathbf{I_s}$. There exists a constant $c_1 > 0$ and $\alpha \geq 2$, such that for every θ in Θ,

$$d_{CF,p}(\mu_\theta, \mu_{\theta_0})^\alpha \geq c_1|\theta - \theta_0|^2.$$

Under this assumption, we obtain a rate of convergence, which will be proved in Section 3.4.

Theorem 2.5. *Assume that* $\mathbf{A_0}$, $\mathbf{A_1}$, $\mathbf{I_s}$ *hold. There exists a constant* $C > 0$ *such that for all* $n \in \mathbb{N}$,

$$\mathbb{E}|\hat{\theta}_n - \theta_0|^2 \leq Cn^{-\alpha(1-(\max(\mathcal{H})\vee\frac{1}{2}))}.$$

3 Proof of consistency of the estimator and rate of convergence

To prove the almost sure convergence, we will use [24, Proposition 4.3] that we recall in Proposition 3.1 below for the reader's convenience. It concerns the limiting property of a collection of real-valued processes $\{L_r(\theta)\}_r$ indexed by a generic r which lies in a topological space and converges to a generic r_0. In this Section, we always have $r \equiv n \in \mathbb{N}$, and so $\lim_{r\to r_0}$ is to be understood as $\lim_{n\to\infty}$. In Section 4, we will take $r \equiv (\gamma, n, N)$ with $\gamma \to 0$ and $n, N \to \infty$, and therefore $\lim_{r\to r_0}$ will be understood as $\lim_{n\to\infty, N\to\infty, \gamma\to 0}$.

Proposition 3.1. *Let Θ be a compact set and $\{\theta \in \Theta \mapsto L_r(\theta)\}_r$ a family of non-negative stochastic processes. Assume that*

(i) With probability 1, $\lim_{r\to r_0} L_r(\theta) = L(\theta)$ uniformly in θ.

(ii) $\theta \mapsto L(\theta)$ is non-random and continuous in θ.

(iii) For any r, the set $\operatorname{argmin}\{L_r(\theta), \theta \in \Theta\}$ is non empty.

Let $\theta_r \in \operatorname{argmin}_{\theta\in\Theta} L_r(\theta)$. If A is a limit point of θ_r, then $A \in \operatorname{argmin}_{\theta\in\Theta} L(\theta)$.

In this Section, we always have $L_r(\theta) = d(\frac{1}{n}\sum_0^{n-1} \delta_{X_{kh}}, \mu_\theta)$, with $r \equiv n$ and $r_0 \equiv \infty$.

3.1 Continuity of $\theta \mapsto d(\mu_\theta, \mu_{\theta_0})$

First, we prove two lemmas that state the L^p-continuity with respect to θ of the solution to (2.5), and the exponential convergence of the law of X^θ (defined in (2.6)) towards its stationary distribution μ_θ. Then we deduce the continuity of the mapping $\theta \mapsto d(\mu_\theta, \mu_{\theta_0})$ in Proposition 3.4.

Lemma 3.2. *Assume $\mathbf{A_0}$ and $\mathbf{A_1}$ are satisfied. Let $T > 0$ and $p > 0$. There exists a constant $C_{T,p} > 0$ such that for any $\theta_1, \theta_2 \in \Theta$,*

$$\|Y_T^{\theta_1} - Y_T^{\theta_2}\|_{L^p} \le C_{T,p}|\theta_1 - \theta_2|.$$

Proof. Without any loss of generality, we assume $p \ge 2$. Up to introducing pivot terms, we can consider three different cases:

$$\begin{aligned}
&1)\ \theta_1 = (\xi, H_1, \sigma) \text{ and } \theta_2 = (\xi, H_2, \sigma)\\
&2)\ \theta_1 = (\xi, H, \sigma_1) \text{ and } \theta_2 = (\xi, H, \sigma_2)\\
&3)\ \theta_1 = (\xi_1, H, \sigma) \text{ and } \theta_2 = (\xi_2, H, \sigma).
\end{aligned}$$

In the first case, where only H changes, we get from the definition of $Y_t^{\theta_1}$ and $Y_t^{\theta_2}$ that for any $t \in [0, T]$,

$$Y_t^{\theta_1} - Y_t^{\theta_2} = \int_0^t [b_\xi(Y_t^{\theta_1}) - b_\xi(Y_t^{\theta_2})]ds + \sigma(B_t^{H_1} - B_t^{H_2}).$$

Since b is K-Lipschitz, we get

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^2 \le 2\left(\int_0^t K|Y_t^{\theta_1} - Y_t^{\theta_2}|ds\right)^2 + 2|\sigma|^2|B_t^{H_1} - B_t^{H_2}|^2.$$

By Jensen's inequality, we have

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^2 \le 2K^2 t\int_0^t |Y_t^{\theta_1} - Y_t^{\theta_2}|^2 ds + 2|\sigma|^2|B_t^{H_1} - B_t^{H_2}|^2$$

By Grönwall's lemma, we deduce that

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^2 \leq 2K^2 T \int_0^t |\sigma|^2 |B_s^{H_1} - B_s^{H_2}|^2 e^{2K^2 T(t-s)} ds + 2|\sigma|^2 |B_t^{H_1} - B_t^{H_2}|^2.$$

By Jensen's inequality, there exists a constant C_p such that

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^p \leq C_p \left(2^{p/2} K^p T^{p-1} \int_0^t |\sigma|^p |B_s^{H_1} - B_s^{H_2}|^p e^{K^2 T p(t-s)} ds + |\sigma|^p |B_t^{H_1} - B_t^{H_2}|^p \right).$$

Since $B_t^{H_1} - B_t^{H_2}$ is a Gaussian random variable, $\mathbb{E}|B_t^{H_1} - B_t^{H_2}|^p$ is proportional to $\left(\mathbb{E}|B_t^{H_1} - B_t^{H_2}|^2\right)^{p/2}$. Using [14, Proposition 2.1], the fractional Brownian motion verifies

$$\mathbb{E}|B_t^{H_1} - B_t^{H_2}|^p \leq C \left(t^{pH_1} \vee t^{pH_2}\right) (\log^2(t) + 1)^{p/2} |H_1 - H_2|^p.$$

Therefore,

$$\mathbb{E}|Y_t^{\theta_1} - Y_t^{\theta_2}|^p \leq C_p |\sigma|^p \left(2^{p/2} K^p T^p e^{K^2 T^2 p} + 1\right) \left(T^{pH_1} \vee T^{pH_2}\right) (\log^2(T) + 1)^{p/2} |H_1 - H_2|^p.$$

Since $\sigma \in \Sigma$, we conclude that

$$\begin{aligned} \|Y_t^{\theta_1} - Y_t^{\theta_2}\|_{L^p} &\leq C_{p,\sigma,K} (T e^{K^2 T^2} + 1)(T^{H_1} \vee T^{H_2})(\log^2(T) + 1)^{1/2} |H_1 - H_2| \\ &\leq C_{p,\sigma,K} (T e^{K^2 T^2} + 1)(1 + T^{\max(\mathcal{H})})(\log^2(T) + 1)^{1/2} |H_1 - H_2|. \end{aligned}$$

In the second case, since b is K-Lipschitz, using Jensen's inequality, we have

$$\begin{aligned} |Y_t^{\theta_1} - Y_t^{\theta_2}|^2 &= \left(\int_0^t [b_\xi(Y_t^{\theta_1}) - b_\xi(Y_t^{\theta_2})] ds + (\sigma_1 - \sigma_2) B_t \right)^2 \\ &\leq 2K^2 T \int_0^t |Y_s^{\theta_1} - Y_t^{\theta_s}|^2 ds + 2|\sigma_1 - \sigma_2|^2 |B_t|^2. \end{aligned}$$

By Grönwall's lemma, we get

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^2 \leq |\sigma_1 - \sigma_2|^2 \left(|B_t|^2 + 2K^2 T \int_0^t |B_s|^2 e^{2K^2 T(t-s)} ds \right).$$

Therefore, by Jensen's inequality, there exists a constant C_p such that

$$|Y_t^{\theta_1} - Y_t^{\theta_2}|^p \leq C_p |\sigma_1 - \sigma_2|^p \left(|B_t|^p + 2^{p/2} K^p T^{p-1} \int_0^t |B_s|^p e^{K^2 T p(t-s)} ds \right).$$

It follows that

$$\begin{aligned} \|Y_t^{\theta_1} - Y_t^{\theta_2}\|_{L^p} &\leq C_p |\sigma_1 - \sigma_2| (T^H + T^{1+H} e^{K^2 T^2}) \\ &\leq C_p |\sigma_1 - \sigma_2| (1 + T^{\max(\mathcal{H})})(T e^{K^2 T^2} + 1) \end{aligned}$$

Finally, in the third case, we have by [24, Proposition 3.5] that

$$\|Y_t^{\theta_1} - Y_t^{\theta_2}\|_{L^p} \leq C_{T,p} |\xi_1 - \xi_2|,$$

where it appears from the proof of [24, Proposition 3.5] that $C_{T,p}$ does not depend on H or σ. □

Lemma 3.3. *Assume* $\mathbf{A_0}$ *and* $\mathbf{A_1}$ *hold. Let* d *be a distance in* $\mathcal{D}_p$. *Then there exists a constant* $C > 0$ *such that for all* $\theta \in \Theta$ *and for all* $t \geq 0$, *we have*

$$d(\mathcal{L}(X_t^\theta), \mu_\theta) \leq C e^{-\frac{1}{C} t}. \tag{3.1}$$

Proof. Since $d \in \mathcal{D}_p$, we have:

$$\begin{aligned} d(\mathcal{L}(X_t^\theta), \mu_\theta) &\leq \mathbb{E}(|X_t^\theta - \bar{X}_t^\theta|^p)^{\frac{1}{p}} \\ &\leq C\left(\sum_{i=0}^{q} \mathbb{E}(|Y_{t+ih}^\theta - \bar{Y}_{t+ih}^\theta|^p)^{\frac{1}{p}}\right). \end{aligned}$$

Using $\mathbf{A_1}$, we have

$$\begin{aligned} \frac{d}{dt}|Y_t^\theta - \bar{Y}_t^\theta|^2 &= 2\langle Y_t^\theta - \bar{Y}_t^\theta, b_\xi(Y_t^\theta) - b_\xi(\bar{Y}_t^\theta)\rangle \\ &\leq -2\beta|Y_t^\theta - \bar{Y}_t^\theta|^2. \end{aligned}$$

It follows that

$$|Y_t^\theta - \bar{Y}_t^\theta|^2 \leq |Y_0^\theta - \bar{Y}_0^\theta|^2 e^{-\beta t},$$

which leads to

$$\mathbb{E}(|Y_t^\theta - \bar{Y}_t^\theta|^p) \leq |Y_0^\theta - \bar{Y}_0^\theta|^p e^{-p\beta t}.$$

Hence,

$$\begin{aligned} \|Y_t^\theta - \bar{Y}_t^\theta\|_{L^p} &\leq \|Y_0^\theta - \bar{Y}_0^\theta\|_{L^p} e^{-\beta t} \\ &\leq (\|Y_0^\theta\|_{L^p} + \|\bar{Y}_0^\theta\|_{L^p})\, e^{-\beta t}. \end{aligned} \tag{3.2}$$

Moreover, by Proposition B.1(i), we have

$$\left(\mathbb{E}|\bar{Y}_0^\theta|\right)^{\frac{1}{p}} = \lim_{t\to\infty}\left(\mathbb{E}|Y_t^\theta|\right)^{\frac{1}{p}} \leq \sup_{t\geq 1}\sup_{\theta\in\Theta}\left(\mathbb{E}|Y_t^\theta|\right)^{\frac{1}{p}} < \infty.$$

This concludes the proof. □

We can now state the main continuity result of this section.

Proposition 3.4. *Assume* $\mathbf{A_0}$ *and* $\mathbf{A_1}$ *hold and let* d *be a distance in* $\mathcal{D}_p$*. Then the mapping* $\theta \mapsto d(\mu_\theta, \mu_{\theta_0})$ *is continuous on* Θ*.*

Proof. Let now $\theta_1, \theta_2 \in \Theta$. Then for $t \geq 0$,

$$\begin{aligned} d(\mu_{\theta_1}, \mu_{\theta_2}) \leq C\mathcal{W}_p(\mu_{\theta_1}, \mu_{\theta_2}) &\leq C\mathcal{W}_p(\mu^{\theta_1}, \mathcal{L}(X_t^{\theta_1})) + C\mathcal{W}_p(\mu^{\theta_2}, \mathcal{L}(X_t^{\theta_2})) + C\|X_t^{\theta_1} - X_t^{\theta_2}\|_{L^p} \\ &\leq 2C\sup_{\theta\in\Theta}\mathcal{W}_p(\mathcal{L}(X_t^\theta), \mu_\theta) + C\|X_t^{\theta_1} - X_t^{\theta_2}\|_{L^p}. \end{aligned}$$

Let $\epsilon > 0$. By Lemma 3.3 there exists t_0 such that

$$2C\sup_{\theta\in\Theta}\mathcal{W}(\mathcal{L}(X_{t_0}^\theta), \mu_\theta) \leq \frac{\epsilon}{2}.$$

Now in view of (2.8) and Lemma 3.2, there exists a constant $C_{t_0,p}$ such that $\|X_{t_0}^{\theta_1} - X_{t_0}^{\theta_2}\|_{L^p} \leq C_{t_0,p}|\theta_1 - \theta_2|$. Let $\delta > 0$ be such that $C_{t_0,p}\delta \leq \epsilon/2$. Then for $|\theta_1 - \theta_2| \leq \delta$, we have

$$d(\mu_{\theta_1}, \mu_{\theta_2}) \leq \epsilon,$$

and this proves the continuity of $\theta \mapsto d(\mu_\theta, \mu_{\theta_0})$. □

3.2 Convergence of the contrast: proof of Lemma 2.2

Let $\theta = (\xi, H, \sigma) \in \Theta$. We will first prove that almost surely, the random measure $\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds$ converges in law to μ_θ. This implies that $\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds$ converges to μ_θ in the Prokhorov distance. To extend this result to distances d in $\mathcal{D}_2$ (i.e dominated by the 2-Wasserstein distance), we use the fact that the 2-Wasserstein distance is dominated by the Prokorov distance d_P as follows (see [9, Theorem 2]):

$$d\left(\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds, \mu_\theta\right) \leq C_p \sup_{t\geq 0}\left(\max\left(\frac{1}{t}\int_0^t |X_s^\theta|^2 ds \vee \mathbb{E}|\bar{X}_t^\theta|^2\right) + 1\right) d_P\left(\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds, \mu_\theta\right).$$

By definition of the process X^θ, we have that

$$\left(\frac{1}{t}\int_0^t |X_s^\theta|^2 ds \vee \mathbb{E}|\bar{X}_t^\theta|^2\right) \leq C_q \sum_{i=0}^{q}\left(\frac{1}{t}\int_0^t |Y_{s+ih}^\theta|^2 ds \vee \mathbb{E}|\bar{Y}_{t+ih}^\theta|^2\right). \tag{3.3}$$

Therefore, we conclude thanks to Proposition B.1 that in the present case, the convergence in law (i.e. in Prokhorov distance) implies the convergence for the 2-Wasserstein distance. Let us now prove the convergence in law. The proof of the convergence in law follows the same steps as [24, Proposition 3.3] and relies on a tightness argument. While we can show that the family $\{\frac{1}{t}\int_0^t \delta_{X_s^\theta} ds\}_{t\geq 0}$ is tight, it is not easy to identify the limit points. That is why we consider a family of probability measures on the set of continuous functions for which the identification of the limit is easier, namely $\{\pi_t = \frac{1}{t}\int_0^t \delta_{X_{s+.}^\theta} ds\}_{t\geq 0}$. A classical criterion (see e.g. [3, Corollary p.83]) ensures that $\{\pi_t^\theta; t \geq 0\}$ is a.s. tight if for every positive T, η and ε, there exists $\delta > 0$ such that for all $t_0 \in [0, T]$,

$$\limsup_{t\to+\infty} \frac{1}{t}\int_0^t \frac{1}{\delta}\mathbb{1}_{\{\sup_{u\in[t_0,t_0+\delta]}|X_{s+u}^\theta - X_{s+t_0}^\theta| \geq \varepsilon\}} ds \leq \eta \quad \text{a.s.}$$

Moreover, the above inequality holds true as long as there exist some positive r and ρ such that

$$\begin{aligned}&\forall T > 0,\ \exists \delta > 0,\ r > 0,\ \rho > 0 \text{ s.t } \forall t_0 \in [0, T],\\ &\qquad \limsup_{t\to\infty} \frac{1}{t}\int_0^t \sup_{u\in[t_0,t_0+\delta]} |X_{s+u}^\theta - X_{s+t_0}^\theta|^r ds \leq C_{r,T}\delta^{1+\rho} \quad \text{a.s.}\end{aligned} \tag{3.4}$$

For $T, r, \delta > 0$, by definition of X^θ and (2.8), we have

$$\begin{aligned}\limsup_{t\to\infty} \frac{1}{t}\int_0^t \sup_{u\in[t_0,t_0+\delta]} |X_{s+u}^\theta - X_{s+t_0}^\theta|^r ds &\leq \sum_{i=0}^{q} \limsup_{t\to\infty} \frac{1}{t}\int_0^t \sup_{u\in[t_0,t_0+\delta]} |Y_{s+u+ih}^\theta - Y_{s+t_0+ih}^\theta|^r ds\\ &\leq \sum_{i=0}^{q} \limsup_{t\to\infty} \frac{t+ih}{t}\frac{1}{t+ih}\int_0^{t+ih} \sup_{u\in[t_0,t_0+\delta]} |Y_{s+u}^\theta - Y_{s+t_0}^\theta|^r ds\\ &\leq C_q \limsup_{t\to\infty} \frac{1}{t+ih}\int_0^{t+ih} \sup_{u\in[t_0,t_0+\delta]} |Y_{s+u}^\theta - Y_{s+t_0}^\theta|^r ds\\ &\leq C_q \limsup_{t\to\infty} \frac{1}{t}\int_0^{t} \sup_{u\in[t_0,t_0+\delta]} |Y_{s+u}^\theta - Y_{s+t_0}^\theta|^r ds.\end{aligned}$$

By [24, Eq A.19], we can further bound the right-hand side above by $C\delta^{r-1} + C_r\delta^{Hr}$. Choosing $\delta < 1$ and $r > \max(2, \frac{1}{\min(\mathcal{H})})$, we get (3.4).

Hence, let $(t_n)_{n\geq 1}$ be an increasing sequence going to $+\infty$ such that $\{\frac{1}{t_n}\int_0^{t_n} \delta_{X_{s+.}^\theta} ds\}_{n\geq 1}$ converges (pathwise) to a limiting distribution γ. We first show that γ is the law of a stationary process. For any bounded functional $F : \mathcal{C}\left([0,\infty), \mathbb{R}^d\right) \to \mathbb{R}$, we have

$$\gamma\left(F \circ \theta_T\right) - \gamma(F) = \lim_{n\to+\infty} \frac{1}{t_n}\int_0^{t_n} F\left(X_{s+T+.}^\theta\right) - F\left(X_{s+.}^\theta\right) ds.$$

By a simple change of variables, we have

$$\lim_{n\to+\infty}\frac{1}{t_n}\int_0^{t_n}F\left(X^\theta_{s+T+.}\right)-F\left(X^\theta_{s+.}\right)\,ds=\lim_{n\to+\infty}\frac{1}{t_n}\left(\int_{t_n}^{t_n+T}F\left(X^\theta_{s+.}\right)ds-\int_0^T F\left(X^\theta_{s+.}\right)ds\right)=0.$$

We thus get that γ is stationary. Let us now prove that γ is the law of $\bar{X}^\theta$. A process $x_t=(y_t,z^1_t,..,z^q_t)$ has the law of X^θ if

$$y_.-y_0-\int_0^{\cdot} b_\xi(y_u)du \text{ has the law of } \sigma B \text{ where B has Hurst parameter } H;$$

$$z^i_.-\ell^i\left(\int_0^{\cdot} b_\xi(y_u)du,\dots,\int_0^{\cdot+ih} b_\xi(y_u)du\right) \text{ has the law of } \sigma\ell^i(B_.,\dots,B_{.+ih}) \text{ for all } i\in[\![1,q]\!].$$

Let us define

$$G(x_.)=\begin{pmatrix} y_.-y_0-\int_0^{\cdot} b_\xi(y_u)du \\ z^1_.-\ell^1\left(\int_0^{\cdot} b_\xi(y_u)du,\int_0^{\cdot+h} b_\xi(y_u)du\right) \\ \cdot \\ \cdot \\ \cdot \\ z^q_.-\ell^q\left(\int_0^{\cdot} b_\xi(y_u)du,\dots,\int_0^{\cdot+qh} b_\xi(y_u)du\right)\end{pmatrix}$$

and

$$\mathbf{B}_.=(\sigma B_.,\dots,\sigma\ell^q(B_.,\dots,B_{qh+.}))\,.$$

Hence we have to prove that

$$\gamma\circ G^{-1} \text{ is the law of } \mathbf{B}_..$$

Using that G is continuous for the u.s.c topology, we have

$$\gamma\circ G^{-1}=\lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}\delta_{G(X^\theta_{s+.})}ds.$$

Let $T>0$ and $F:\mathcal{C}([0,T],\mathbb{R}^{d(q+1)})\mapsto\mathbb{R}$ be a bounded measurable function. We want to show that

$$\lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}F\left(G(X^\theta_{s+.})\right)ds=\mathbb{E}F\left(\mathbf{B}_.\right).$$

It is sufficient to check the convergence for the finite dimensional distributions. For any $N\geq 1$, $\{u_1,\dots,u_N\}\in\mathbb{R}^N$ and a measurable and bounded $f:\mathbb{R}^{d(q+1)N}\mapsto\mathbb{R}$, we want to show that

$$\lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}f\left(G(X^\theta_{s+u_1}),\dots,G(X^\theta_{s+u_N})\right)ds=\mathbb{E}f\left(\mathbf{B}_{u_1},\dots,\mathbf{B}_{u_N}\right).$$

By construction, we have

$$G(X_{s+.})=\begin{pmatrix}\sigma(B_{s+.}-B_s) \\ \sigma\ell^1(B_{s+.}-B_s,B_{s+h\cdot}-B_s) \\ \cdot \\ \cdot \\ \cdot \\ \sigma\ell^q(B_{s+.}-B_s,\dots,B_{s+qh+.}-B_s)\end{pmatrix}.$$

Therefore, we can write

$$f(G(X^{\theta}_{s+u_1}),\dots,G(X^{\theta}_{s+u_N})) = \tilde{f}\left(\{B_{s+u_1+ih}-B_s\}_{i=0,\dots,q},\dots,\{B_{s+u_N+ih}-B_s\}_{i=0,\dots,q}\right)$$
$$f\left(\mathbf{B}_{u_1},\dots,\mathbf{B}_{u_N}\right) = \tilde{f}\left(\{B_{u_1+ih}\}_{i=0,\dots,q},\dots,\{B_{u_N+ih}\}_{i=0,\dots,q}\right),$$

where $\tilde{f} = f\circ\lambda$ for some linear transformation λ, so $\tilde{f}$ is still a bounded measurable function. By the ergodicity of the increments of the fractional Brownian motion [6, Eq 5], we have

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n} f\left(G(X^{\theta}_{s+u_1}),\dots,G(X^{\theta}_{s+u_N})\right)ds \\
&= \lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}\tilde{f}\left(\{B_{s+u_1+ih}-B_s\}_{i=0,\dots,q},\dots,\{B_{s+u_N+ih}-B_s\}_{i=0,\dots,q}\right)ds \\
&= \mathbb{E}\tilde{f}\left(\{B_{u_1+ih}\}_{i=0,\dots,q},\dots,\{B_{u_N+ih}\}_{i=0,\dots,q}\right) = \mathbb{E}f(\mathbf{B}_{u_1},\dots,\mathbf{B}_{u_N}).
\end{aligned}$$

Hence, $\gamma\circ G^{-1}$ has the law of $\mathbf{B}$ and we conclude that γ is the law of $\bar{X}^{\theta}$.

The same analysis presented in this Section still holds if we replace $\frac{1}{t}\int_0^t |X^{\theta}_s|^p ds$ by $\frac{1}{n}\sum_{k=0}^{n-1}|X^{\theta}_{kh}|^p$. This is mostly due to the fact that in Proposition B.1, we also proved that the moments $\frac{1}{n}\sum_{k=0}^{n-1}|X^{\theta}_{kh}|^p$ are finite uniformly in n, and therefore the right-hand side in (3.3) is finite even when the integral is replaced by a discrete sum.

3.3 Proof of Theorem 2.4

Let d be a distance that belongs to $\mathcal{D}_p$. We want to apply Proposition 3.1 to $r\equiv n$ and

$$L_n(\theta) = d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_\theta\right).$$

In view of Lemma 2.2, we know that for each θ, $L_n(\theta)$ converges a.s. to $L(\theta)=d(\mu_{\theta_0},\mu_\theta)$. Besides, the continuity of $L(\theta)$ comes from Proposition 3.4. If we prove the uniform convergence, then we can finally apply Proposition 3.1 to get that the limit points of $\hat{\theta}_n$ are included in the set $\operatorname{argmin}\{L(\theta),\theta\in\Theta\}$, which under assumption $\mathbf{I_w}$ is reduced to $\{\theta_0\}$.

Now to prove the uniform convergence, it is sufficient to show that the family

$$\left\{\theta\mapsto d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_\theta\right), n\geq 1,\ \theta\in\Theta\right\}$$

is equicontinuous. Actually, for any θ_1 and θ_2 in Θ, we have

$$\left|d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_{\theta_1}\right) - d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_{\theta_2}\right)\right| \leq d(\mu_{\theta_1},\mu_{\theta_2}).$$

In view of Proposition 3.4, the term on the right-hand side goes to 0 as $|\theta_1-\theta_2|\to 0$. This proves the equicontinuity and thus the uniform convergence.

3.4 Proof of Theorem 2.5

For the rest of this section, d always refer to the distance $d_{CF,p}$. In view of the strong identifiability assumption $\mathbf{I_s}$, it suffices to bound $\mathbb{E}d(\mu_{\hat{\theta}_n},\mu_{\theta_0})^\alpha$ to obtain a rate of convergence on $\mathbb{E}|\hat{\theta}_n-\theta_0|^2$. Our strategy is in line with the Section 5 of [24], with adaptations due to the estimation of σ and H. It is based on the following decomposition: since $\hat{\theta}_n$ minimizes the function $\theta\mapsto d(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}},\mu_{\theta_0})$, we have

$$
\begin{aligned}
d(\mu_{\hat{\theta}_n},\mu_{\theta_0}) &\le d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}},\mu_{\theta_0}\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}},\mu_{\hat{\theta}_n}\right)\\
&\le 2d\left(\frac{1}{n}\sum_{0}^{n-1}\delta_{X_{kh}},\mu_{\theta_0}\right)\\
&=: 2D_n^{(1)}.
\end{aligned}
$$

L^α bound on $D_n^{(1)}$. Following the proof of [24, Section 5.1], we obtain a bound on $D_n^{(1)}$.

Lemma 3.5. *Assume that* $\mathbf{I_s}$ *holds with p and α satisfying $p > \frac{\alpha+d}{2}$. There exist positive constants $C_{\alpha,q}$ and C_α such that for any $n \in \mathbb{N}$,*

$$
\mathbb{E}[|D_n^{(1)}|^\alpha] \le \frac{C_{\alpha,q}}{n^\alpha} + C_\alpha(n+q)^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1))}),
$$

where we recall that q is the number of linear transformations added to construct the augmented process X^{θ_0}.

Proof. Decompose $D_n^{(1)}$ as $D_n^{(1)} \le D_n^{(11)} + D_n^{(12)}$ where

$$
\begin{aligned}
D_n^{(11)} &:= d\left(\mu_{\theta_0},\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}[\delta_{X_{kh}^{\theta_0}}]\right),\\
D_n^{(12)} &:= d\left(\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}[\delta_{X_{kh}^{\theta_0}}],\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}\right).
\end{aligned}
$$

The expectation of the random measure $\mathbb{E}[\delta_{X_t}]$ is understood as a deterministic measure given by $\mathbb{E}[\delta_{X_t}](f) = \mathbb{E}[f(X_t)]$ for any bounded measurable f.

Let us first bound $D_n^{(12)}$. Recall the concentration result [30, Theorem 2.3]: There exists a constant $C > 0$ such that for $d_n(x,y) = \frac{1}{n}\sum_{k=0}^{n-1}|x_i - y_i|$, any Lipschitz functions $F : ((\mathbb{R}^d)^n, d_n) \to (\mathbb{R}, |\,.\,|)$ and any $r > 0$,

$$
\mathbb{P}(F_Y - \mathbb{E}(F_Y) \ge r) \le \exp\left(-\frac{r^2 n^{2-\max(2H,1)}}{C\,\|F\|_{Lip}^2}\right),
$$

where $F_Y = F(Y_h, Y_{2h}, ..., Y_{nh})$. Hence in view of $\mathbb{E}(X^\alpha) = \int_0^\infty \alpha x^{\alpha-1}\mathbb{P}(X \ge x)dx$, we get that

$$
\mathbb{E}[|F_Y - \mathbb{E}(F_Y)|^\alpha] \le C_\alpha\,\|F\|_{Lip}^\alpha\, n^{-\frac{\alpha}{2}(2-\max(2H,1))},
$$

for some positive constant C_α that depends on α. Using the definition of $d_{CF,p}$, Jensen's inequality and the notation $f_\chi(x) = e^{i\langle\chi,x\rangle}$, we get

$$
\mathbb{E}[|D_n^{(12)})|^\alpha] \le \int \mathbb{E}\left[\left|\frac{1}{n}\sum_{k=0}^{n-1}f_\chi(X_{kh}^{\theta_0}) - \mathbb{E}(f_\chi(X_{kh}^{\theta_0}))\right|^\alpha\right] g_p(\chi)d\chi.
$$

Since $\|f_\chi\|_{Lip} \le |\chi|$, we deduce by taking $F_Y = F(Y_h^{\theta_0}, ..., Y_{(n+q)h}^{\theta_0}) = \frac{1}{n}\sum_{k=0}^{n-1}f_\chi(X_{kh}^{\theta_0})$ the following bound on $D_n^{(12)}$:

$$
\mathbb{E}[|D_n^{(12)}|^\alpha] \le C_\alpha(n+q)^{-\frac{\alpha}{2}(2-\max(2H,1))}\int |\chi|^\alpha g_p(\chi)d\chi. \tag{3.5}
$$

The integral on the right-hand side is finite if we choose $p > \frac{\alpha+d}{2}$.

We now bound $D_n^{(11)}$. Since Y^θ converge exponentially fast to $\bar{Y}^\theta$ as $t \to \infty$ (see (3.2)), and so does X^θ towards $\bar{X}^\theta$ in view of (2.8), and since f_χ is Lipschitz we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}f_\chi\left(X_{kh}^{\theta_0}\right)-\mu_{\theta_0}(f_\chi)\right| &= \left|\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\left[f_\chi\left(X_{kh}^{\theta_0}\right)-f_\chi\left(\bar{X}_{kh}^{\theta_0}\right)\right]\right|\\ &\leq \frac{1}{n}\|f_\chi\|_{Lip}\sum_{k=0}^{n-1}\mathbb{E}\left[\left|X_{kh}^{\theta_0}-\bar{X}_{kh}^{\theta_0}\right|\right]\leq \frac{C}{n}\|f_\chi\|_{Lip}. \end{aligned} \tag{3.6}$$

Thus by the definition (2.1) of $d_{CF,p}$, we get

$$d\left(\mu_{\theta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}[\delta_{X_{kh}^{\theta_0}}]\right) \leq \frac{C}{n},$$

which leads to

$$\mathbb{E}[|D_n^{(1,1)}|^\alpha] \leq \frac{C}{n^\alpha}. \tag{3.7}$$

We conclude the proof by combining the bounds (3.5) and (3.7). □

4 A practical implementation of the estimators

In the formula (2.10), the stationary distribution is in general not known, except in some simple cases like for Ornstein-Uhlenbeck processes. This means that the estimator cannot be computed in practice. In this section, we solve this problem by considering numerical approximations of μ_θ. On the other hand, this increases the complexity that is required to compute the estimator. However, we still obtain similar results (consistency and rate of convergence).

4.1 Estimating the stationary distribution

To approximate μ_θ, we consider the Euler scheme of the stochastic process Y^θ, solution to (2.5). For a time-step $\gamma > 0$, the Euler scheme $Y^{\theta,\gamma}$ is then defined by $Y_0^{\theta,\gamma} = y_0 \in \mathbb{R}^d$ and

$$\begin{aligned} Y_{(k+1)\gamma}^{\theta,\gamma} &= Y_{k\gamma}^{\theta,\gamma} + \gamma b_\theta(Y_{k\gamma}^{\theta,\gamma}) + \sigma(\widehat{B}_{(k+1)\gamma} - \widehat{B}_{k\gamma})\\ Y_t^{\theta,\gamma} &= Y_{t_\gamma}^{\theta,\gamma} = Y_{k\gamma}^{\theta,\gamma} \text{ for } t \in [k\gamma,(k+1)\gamma), \end{aligned} \tag{4.1}$$

where $t_\gamma = \gamma\lfloor t/\gamma \rfloor$ and $\widehat{B}$ is a simulated fractional Brownian motion, which is *a priori* different from the process B in (2.5), since B is unobserved. In practice, this means that we will not be able to compare pathwise the observed process and the simulated one. When necessary, to mark the dependence of $Y_\cdot^{\theta,\gamma}$ on $\widehat{B}$, we write $Y_\cdot^{\theta,\gamma}(\widehat{B})$. We will say that $\left(\bar{Y}_t^{\theta,\gamma}\right)_{t\geq 0}$ is a discrete stationary solution to (4.1) if it is a solution of (4.1) satisfying

$$\left(\bar{Y}_{t_1+k\gamma}^{\theta,\gamma},\ldots,\bar{Y}_{t_n+k\gamma}^{\theta,\gamma}\right) \overset{\mathcal{L}}{=} \left(\bar{Y}_{t_1}^{\theta,\gamma},\ldots,\bar{Y}_{t_n}^{\theta,\gamma}\right) \quad \forall\, 0 < t_1 < \cdots < t_n,\, \forall n,k \in \mathbb{N}.$$

By [24, Proposition 3.4], there exists $\gamma_0 > 0$ such that for any $\gamma \in (0,\gamma_0]$ and $\theta \in \Theta$, (4.1) admits a unique stationary solution $\bar{Y}^{\theta,\gamma}$. As in Section 2.1, we define the augmented Euler scheme $X^{\theta,\gamma}$ by

$$X_\cdot^{\theta,\gamma} = \left(Y_\cdot^{\theta,\gamma}, \ell^1(Y_\cdot^{\theta,\gamma}, Y_{\cdot+h}^{\theta,\gamma}), ..., \ell^q(Y_\cdot^{\theta,\gamma}, ..., Y_{\cdot+qh}^{\theta,\gamma})\right).$$

Similarly, we write $X_\cdot^{\theta,\gamma}(\widehat{B})$ to insist on the dependence on $\widehat{B}$ when necessary. We also define the stationary augmented Euler scheme $\bar{X}^{\theta,\gamma}$ and denote its distribution by μ_θ^γ. We construct the estimator based on the following result (see Appendix C for the proof).

Proposition 4.1. *Let $(X_{k\gamma}^{\theta,\gamma})_{k\geq 0}$ be the augmented Euler scheme with time-step γ. Assume that* $\mathbf{A_0}$ *and* $\mathbf{A_1}$ *hold. Then for any distance $d \in \mathcal{D}_2$, there exists $\gamma_0 > 0$ such that for all $\theta \in \Theta$ and $\gamma \in (0, \gamma_0]$, we have*

$$\lim_{N\to\infty} d\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta,\gamma}}, \mu_\theta^\gamma\right) = 0.$$

In Proposition 4.5(i), we show that for any $\theta \in \Theta$, $d(\mu_\theta, \mu_\theta^\gamma)$ goes to 0 as $\gamma \to 0$. This suggests to define the estimator

$$\hat{\theta}_{N,n,\gamma} = \underset{\theta\in\Theta}{\operatorname{argmin}}\ d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta,\gamma}}\right). \tag{4.2}$$

4.2 Consistency and convergence results

Strong Consistency. The following result states the strong consistency of our estimators under the assumptions $\mathbf{A_1}$, $\mathbf{I_w}$ and $\mathbf{A_0}$. We discuss its proof in this section and provide details in the Appendix B.

Theorem 4.2. *Consider a distance d on $\mathcal{M}_1(\mathbb{R}^d)$ which belongs to $\mathcal{D}_2$. Assume that the exponent r in the sub-linear growth of b_ξ in (2.4) satisfies $r \leq 1$. Then the family $\{\hat{\theta}_{N,n,\gamma},\ (n, N, \gamma) \in \mathbb{N}^2 \times \mathbb{R}_+\}$ is a strong consistent estimator of θ_0 in the following sense:*

$$\lim_{\substack{n\to\infty\\ N\to\infty\\ \gamma\to 0}} \hat{\theta}_{N,n,\gamma} = \theta_0 \ \text{a.s.}$$

To prove the strong consistency, we use again Proposition 3.1 with

$$L_r(\theta) = d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta,\gamma}}\right),$$

this time with $r = (N, n, \gamma)$. We will prove in Section 4.3 that the contrast $L_r(\theta)$ converges uniformly as $(n, N, \gamma) \to (\infty, \infty, 0)$ to $L(\theta) = d(\mu_\theta, \mu_{\theta_0})$, by first proving pointwise convergence and then using an equicontinuity argument. Since $L(\theta)$ is the same as in Section 3, we have by Proposition 3.4 that $L(\theta)$ is continuous. Then we apply Proposition 3.1 to conclude.

Rate of Convergence. A rate of convergence is obtained for the estimators under the strong identifiability assumption $\mathbf{I_s}$.

Theorem 4.3. *Assume* $\mathbf{A_1}$, $\mathbf{A_0}$ *and* $\mathbf{I_s}$ *for $d = d_{CF,p}$ with $p > \frac{\alpha+d}{2}$. Assume that the exponent r in the sub-linear growth of b_ξ in (2.4) satisfies $r \leq 1$. Let $\varepsilon \in (0, \min(\mathcal{H}))$ and let $\hat{\theta}_{N,n,\gamma}$ be the estimator defined by (4.2). There exists positive constants C, γ_0 such that for all $\gamma \in (0, \gamma_0]$ and $n, N \in \mathbb{N}$,*

$$\mathbb{E}\left|\hat{\theta}_{N,n,\gamma} - \theta_0\right|^2 \leq C\left(n^{-1+\max(\mathcal{H})\vee\frac{1}{2}} + N^{-1+\max(\mathcal{H})\vee\frac{1}{2}} + \gamma^{\min(\mathcal{H})-\varepsilon} + (N\gamma)^{-\frac{\beta\alpha^2}{2(\beta\alpha+2d)}(2-2\max(\mathcal{H}\vee 1))}\right).$$

Proof. To prove the convergence above, we proceed similarly as in Section 3.4. We decompose the term $\mathbb{E}d(\mu_{\hat{\theta}_{N,n,\gamma}}, \mu_{\theta_0})^\alpha$ slightly differently. First we use a triangle inequality to get the following bound,

$$\begin{aligned} d\left(\mu_{\theta_0}, \mu_{\hat{\theta}_{N,n,\gamma}}\right) \leq{}& d\left(\mu_{\theta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\hat{\theta}_{N,n,\gamma},\gamma}}\right) \\ &+ d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\hat{\theta}_{N,n,\gamma},\gamma}}, \mu_{\hat{\theta}_{N,n,\gamma}}\right). \end{aligned}$$

Now, notice that since $\hat{\theta}_{N,n,\gamma}$ minimizes the function $\theta \mapsto d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta,\gamma}_{k\gamma}}\right)$, we can further bound $d(\mu_{\theta_0}, \mu_{\hat{\theta}_{N,n,\gamma}})$ as

$$d\left(\mu_{\theta_0}, \mu_{\hat{\theta}_{N,n,\gamma}}\right) \leq d\left(\mu_{\theta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right) + \sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta,\gamma}_{k\gamma}}, \mu_\theta\right). \tag{4.3}$$

To allow pathwise comparison, let us define the following processes. For any $\theta \in \Theta$, define $Y^{\theta,\gamma}_{\cdot}(B)$, an Euler scheme of Y^θ, computed with the same fBm B. Namely, $Y^{\theta,\gamma}_{\cdot}(B)$ is defined by (4.1) where $\widehat{B}$ is replaced by B. Similarly as in Section 2.1, we define $X^{\theta,\gamma}(B)$ by,

$$X^{\theta,\gamma}(B) = \left(Y^{\theta,\gamma}_{\cdot}(B)_{\cdot}, \ell^1(Y^{\theta,\gamma}_{\cdot}(B), Y^{\theta,\gamma}_{\cdot+h}(B)), ..., \ell^q(Y^{\theta,\gamma}_{\cdot}(B)_{\cdot}, ..., Y^{\theta,\gamma}_{\cdot+qh}(B))\right).$$

We also define $Y(\widehat{B})$ which is the solution to (2.5) with the fBm $\widehat{B}$, and similarly we define $X(\widehat{B})$. Now, we can do pathwise comparison between X^θ and $X^{\theta,\gamma}(B)$, and between $X^{\theta,\gamma}$ and $X^\theta(\widehat{B})$.

Bounding the second term in (4.3). We split the second term in the right hand side of (4.3) as follows

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right) \leq d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \mu_{\theta_0}\right) + d\left(\mu_{\theta_0}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right). \tag{4.4}$$

Furthermore, we split the last term above as,

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right) \leq d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}, \mu_{\theta_0}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}, \mu_{\theta_0}\right).$$

Moreover,

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}\right) \leq 2d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \mu_{\theta_0}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}, \mu_{\theta_0}\right).$$

Injecting the above bound into (4.4), we get

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{L^{\theta_0}_{k\gamma}}\right) \leq d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}}, \mu_{\theta_0}\right) + 2d\left(\mu_{\theta_0}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}, \mu_{\theta_0}\right) + 2d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}\right). \tag{4.5}$$

Bounding the third term in (4.3). We split the third term in (4.3) as follows

$$\sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta,\gamma}_{k\gamma}},\mu_\theta\right) \le \sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta,\gamma}_{k\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}\right) + \sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})},\mu_\theta\right). \tag{4.6}$$

Final bound on $d\left(\mu_{\theta_0},\mu_{\hat\theta_{N,n,\gamma}}\right)$. Using (4.5) and (4.6) in (4.3), we get

$$\begin{aligned} d\left(\mu_{\theta_0},\mu_{\hat\theta_{N,n,\gamma}}\right) \le{}& 2d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_{\theta_0}\right) + 2d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}},\mu_{\theta_0}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})},\mu_{\theta_0}\right)\\ &+ \sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})},\mu_\theta\right)\\ &+ 2\sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}\right) + 2\sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}\right). \end{aligned} \tag{4.7}$$

The first three terms on the right-hand side can be bounded exactly as the term $D_n^{(11)}$ in the proof of Lemma 3.5, we get

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X^{\theta_0}_{kh}},\mu_{\theta_0}\right) \le \frac{C\alpha,q}{n^\alpha} + C_\alpha(n+q)^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}) \tag{4.8}$$

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}},\mu_{\theta_0}\right) \le \frac{C\alpha,q}{N^\alpha} + C_\alpha(N+q)^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}) \tag{4.9}$$

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})},\mu_{\theta_0}\right) \le \frac{C\alpha,q}{N^\alpha} + C_\alpha(N+q)^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}). \tag{4.10}$$

Remark 4.4. *For the term $d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})},\mu_{\theta_0}\right)$, notice that μ_{θ_0} is also the law of $\bar X^{\theta_0,\hat B}$, the stationary augmented process associated to (1.1) with the fBm $\hat B$ instead of B, so (3.6) in the proof of Lemma 3.5 still holds since we compare two solutions with the same noise, and therefore we know that they converge exponentially to each other as $t\to\infty$ by Proposition 3.3.*

Let us define

$$\begin{aligned} D^{(21)}_{N,\gamma}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}(B)}\right)\\ D^{(22)}_{N,\gamma}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0,\gamma}_{k\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_0}_{k\gamma}(\widehat{B})}\right)\\ D^{(3)}_{N,\gamma}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta}_{k\gamma}(\widehat{B})},\mu_\theta\right). \end{aligned} \tag{4.11}$$

In Section 4.4, we show how to bound the moments of $\sup_{\theta\in\Theta} D^{(21)}_{N,\gamma}(\theta)$, $\sup_{\theta\in\Theta} D^{(22)}_{N,\gamma}(\theta)$ and $\sup_{\theta\in\Theta} D^{(3)}_{N,\gamma}(\theta)$. Namely, we prove that for any $\varepsilon < \alpha\min(\mathcal{H})$ and any $\varpi\in(0,1)$, there exist constants $C_{\alpha,\varepsilon}$ and $C_{\alpha,\varepsilon,\varpi}$ such that for any N and $\gamma \leq \gamma_0$, the following bounds hold:

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(21)}(\theta)\right|^{\alpha} \leq C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon} \tag{4.12}$$

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(22)}(\theta)\right|^{\alpha} \leq C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon} \tag{4.13}$$

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(3)}(\theta)\right|^{\alpha} \leq C_{\alpha,\varepsilon,\varpi}\left(\gamma^{\alpha\min(\mathcal{H})-\varepsilon} + T^{-\bar{\eta}}\right), \tag{4.14}$$

with $\bar{\eta} = \frac{\varpi\alpha^2}{2(\alpha\varpi+2d)}(2-(2\max(\mathcal{H})\vee 1))$ and $T = N\gamma$. Injecting the bounds (4.8), (4.9), (4.10), (4.12), (4.13) and (4.14) into the decomposition (4.7) concludes the proof. □

4.3 Proof of Theorem 4.2

As explained after Theorem 4.2, we just have to prove the uniform convergence of the contrast $(N,n,\gamma)\mapsto d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}}\right)$ towards $d(\mu_{\theta_0},\mu_\theta)$. First we prove that almost surely, we have convergence as $(n,N,\gamma)\to(\infty,\infty,0)$ for each fixed θ. We have already proved in Section 3.2 that $d(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}},\mu_\theta)$ converges to $d(\mu_{\theta_0},\mu_\theta)$ as n goes to infinity. By Proposition 4.1, we have that $d(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}},\mu_\theta^\gamma)$ converges to 0 as $N\to\infty$. Finally, we prove in Proposition 4.5(i) that $d(\mu_\theta,\mu_\theta^\gamma)$ converges to 0 as $\gamma\to 0$. Therefore we conclude that

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}}\right) \underset{(n,N,\gamma)\to(\infty,\infty,0)}{\longrightarrow} d(\mu_{\theta_0},\mu_\theta).$$

We extend the convergence result to a uniform convergence in θ in the following Proposition.

Proposition 4.5. *Let $0 < p \leq 2$ and $d \in \mathcal{D}_p$. Under the assumptions* $\mathbf{A_1}$, $\mathbf{I_w}$, $\mathbf{A_0}$, *there exists $\gamma_0 > 0$ such that for all $\gamma\in(0,\gamma_0]$, the following assertions hold true.*

(i) $\lim_{\gamma\to 0}\sup_{\theta\in\Theta} d(\mu_\theta,\mu_\theta^\gamma) = 0.$

(ii) $\lim_{N\to\infty}\sup_{\theta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}},\mu_\theta^\gamma\right) = 0.$

(iii) $\lim_{\gamma\to 0}\lim_{n,N\to\infty}\sup_{\theta\in\Theta}\left|d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}}\right) - d(\mu_{\theta_0},\mu_\theta)\right| = 0.$

Proof. Notice that (iii) is a simple consequence of the previous statements (i) and (ii).

Proof of (i). By the triangle inequality we have

$$d(\mu_\theta,\mu_\theta^\gamma) \leq d(\mu_\theta,\mathcal{L}(X_{N\gamma}^\theta)) + d(\mu_\theta^\gamma,\mathcal{L}(X_{N\gamma}^{\theta,\gamma})) + d(\mathcal{L}(X_{N\gamma}^{\theta,\gamma}),\mathcal{L}(X_{N\gamma}^\theta)).$$

Since d is bounded by the 2-Wasserstein distance, for all $N \geq 1$ we have

$$\begin{aligned} d(\mu_\theta,\mu_\theta^\gamma) &\leq \mathcal{W}_2(\mu_\theta,\mathcal{L}(X_{N\gamma}^\theta)) + \mathcal{W}_2(\mu_\theta^\gamma,\mathcal{L}(X_{N\gamma}^{\theta,\gamma})) + \mathcal{W}_2(\mathcal{L}(X_{N\gamma}^{\theta,\gamma}),\mathcal{L}(X_{N\gamma}^\theta)) \\ &=: W^{(1)} + W^{(2)} + W^{(3)}. \end{aligned} \tag{4.15}$$

As for $W^{(1)}$, we have

$$W^{(1)} = \mathcal{W}_2(\mu_\theta,\mathcal{L}(X_{N\gamma}^\theta)) \leq \left(\mathbb{E}|X_{N\gamma}^\theta - \bar{X}_{N\gamma}^\theta|^2\right)^{\frac{1}{2}}.$$

By Proposition 3.3, the right-hand side term converges to 0 as $N \to \infty$ uniformly in θ. We now look at the second term:

$$W^{(2)} = \mathcal{W}_2(\mu_\theta^\gamma, \mathcal{L}(X_{N\gamma}^{\theta,\gamma})) \leq \left(\mathbb{E}|\bar{X}_{N\gamma}^{\theta,\gamma} - X_{N\gamma}^{\theta,\gamma}|^2\right)^{\frac{1}{2}}$$
$$\leq C_q \sum_{i=0}^{q} \left(\mathbb{E}|\bar{Y}_{N\gamma+ih}^{\theta,\gamma} - Y_{N\gamma+ih}^{\theta,\gamma}|^2\right)^{\frac{1}{2}}. \tag{4.16}$$

By [24, Equation (4.2)], we have for any $k \in \mathbb{N}$,

$$\left|\bar{Y}_{k\gamma}^{\theta,\gamma} - Y_{k\gamma}^{\theta,\gamma}\right|^2 \leq \left(1 - 2\gamma\beta + \gamma^2 K^2\right)^k \left|\bar{Y}_0^{\theta,\gamma} - Y_0^{\theta,\gamma}\right|^2. \tag{4.17}$$

Furthermore, for any $i \in [\![0, q]\!]$, there exists $j \in \mathbb{N}$ such that $Y_{N\gamma+ih}^{\theta,\gamma} = Y_{j\gamma}^{\theta,\gamma}$ and $\bar{Y}_{N\gamma+ih}^{\theta,\gamma} = \bar{Y}_{j\gamma}^{\theta,\gamma}$. Therefore, the bound (4.17) holds for all the terms in (4.16). We conclude that there exists $\gamma_0 > 0$, such that for $\gamma \leq \gamma_0$, the second term goes to 0 uniformly in θ when $N \to \infty$. Now for the last term in (4.15), by definition of the Wasserstein distance, we have

$$W^{(3)} = \mathcal{W}_2(\mathcal{L}(X_{N\gamma}^{\theta}), \mathcal{L}(X_{N\gamma}^{\theta,\gamma})) \leq \left(\mathbb{E}|X_{N\gamma}^{\theta} - X_{N\gamma}^{\theta,\gamma}(B)|^2\right)^{\frac{1}{2}}$$
$$\leq C_q \sum_{i=0}^{q} \left(\mathbb{E}|Y_{N\gamma+ih}^{\theta} - Y_{N\gamma+ih}^{\theta,\gamma}(B)|^2\right)^{\frac{1}{2}}.$$

In [24, Proposition 3.7 (i)], it was proved that there exists positive constants C and ρ that depend only the Lipschitz constant K from $\mathbf{A_1}$ such that for any $m \in \mathbb{N}$,

$$|Y_{m\gamma}^{\theta} - Y_{m\gamma}^{\theta,\gamma}(B)|^2 \leq C \sum_{j=0}^{m-1} \phi_j(Y_{j\gamma}^{\theta,\gamma}(B)) e^{-\rho\gamma(m-j+1)},$$

where

$$\phi_j(z) = \gamma^3 |b_\xi(z)|^2 + \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt.$$

Note that this pathwise comparison is possible because the two processes are defined with the same noise B. Since b_ξ is uniformly sub-linear, it follows that

$$|Y_{m\gamma}^{\theta} - Y_{m\gamma}^{\theta,\gamma}(B)|^2 \leq C \sum_{j=0}^{m-1} \left(\gamma^3(1 + |Y_{j\gamma}^{\theta,\gamma}(B)|^{2r}) + \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt\right) e^{-\rho\gamma(m-j+1)}. \tag{4.18}$$

Now for $i \in [\![0, q]\!]$ and $k \in \mathbb{N}$, since the process $Y^{\theta,\gamma}$ is constant over intervals of size γ, we can always write

$$Y_{k\gamma+ih}^{\theta} - Y_{k\gamma+ih}^{\theta,\gamma}(B) = \left(Y_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}}^{\theta} - Y_{(k\gamma+ih)_\gamma}^{\theta}\right) + \left(Y_{(k\gamma+ih)_\gamma}^{\theta} - Y_{(k\gamma+ih)_\gamma}^{\theta,\gamma}(B)\right), \tag{4.19}$$

where

$$\varepsilon_{k,i} = k\gamma + ih - (k\gamma + ih)_\gamma < \gamma.$$

For the first term in (4.19), using the sub-linear growth of b, we write

$$|Y_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}}^{\theta} - Y_{(k\gamma+ih)_\gamma\gamma}^{\theta}| \leq C \left(\int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma+\varepsilon} (1 + |Y_s^\theta|^r) ds + |B_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}} - B_{(k\gamma+ih)_\gamma}|\right).$$

It follows from Jensen's inequality that

$$|Y_{j(k\gamma+ih)_\gamma+\varepsilon_{k,i}}^{\theta} - Y_{(k\gamma+ih)_\gamma\gamma}^{\theta}|^2 \leq C \left(\varepsilon_{k,i} \int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma+\varepsilon_{k,i}} (1 + |Y_s^\theta|^{2r}) ds + |B_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}} - B_{(k\gamma+ih)_\gamma}|^2\right). \tag{4.20}$$

The second term in (4.19) can be bounded using (4.18) with $m \equiv \frac{(k\gamma+ih)_\gamma}{\gamma}$. Combining this and (4.20) in (4.19), we get that for any $k \in \mathbb{N}$,

$$\sum_{i=0}^{q} |Y^{\theta}_{k\gamma+ih} - Y^{\theta,\gamma}_{k\gamma+ih}(B)|^2 \leq C \sum_{i=0}^{q} \gamma \int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma+\varepsilon_{k,i}} (1+|Y^{\theta}_s|^{2r}) ds + |B_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}} - B_{(k\gamma+ih)_\gamma}|^2$$
$$+ C \sum_{i=0}^{q} \sum_{j=0}^{\lfloor \frac{k\gamma+ih}{\gamma} \rfloor -1} \Big(\gamma^2(1+|Y^{\theta,\gamma}_{j\gamma}(B)|^{2r}) + \gamma^{-1} \int_0^{\gamma} |B_{j\gamma+t} - B_{j\gamma}|^2 dt\Big) \gamma e^{-\rho(\lfloor \frac{k\gamma+ih}{\gamma} \rfloor - j+1)}. \tag{4.21}$$

Taking the expectation, using

$$\limsup_{\substack{n\to\infty \\ \gamma\to 0}} \gamma \sum_{j=0}^{n} e^{-\rho(n-j+1)} < +\infty \quad \text{and} \quad r \leq 1,$$

we get

$$\sup_{\theta\in\Theta, \gamma\in(0,\gamma_0)} \limsup_{k\to\infty} \gamma^{-2\max(\mathcal{H})} \sum_{i=0}^{q} \mathbb{E} \left| Y^{\theta}_{N\gamma+ih} - Y^{\theta,\gamma}_{N\gamma+ih}(B) \right|^2$$
$$\leq C \left(1 + \sup_{\theta\in\Theta, \gamma\in(0,\gamma_0)} \limsup_{k\to\infty} \mathbb{E}|Y^{\theta,\gamma}_{k\gamma}(B)|^{2r} + \sup_{\theta\in\Theta} \limsup_{t\to\infty} \mathbb{E}|Y^{\theta}_t|^{2r} \right).$$

Using Proposition B.2(i) and Proposition B.1(i), it follows that there exists γ_0 such that for $\gamma \leq \gamma_0$, the right-hand side is finite. This conclude the proof of (i).

Proof of (ii)**.** We already know that the convergence is true for fixed θ. In order to extend the result to uniform convergence, we show that the family $\{\theta \mapsto d(\frac{1}{N}\sum_{k=0}^{N} \delta_{X^{\theta,\gamma}_{k\gamma}}, \mu^{\gamma}_{\theta}); N \geq 1; \theta \in \Theta\}$ is equicontinuous for a fixed $\gamma \in (0, \gamma_0]$. For some give θ_1 and θ_2 in Θ, there is

$$\left| d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_1,\gamma}_{k\gamma}}, \mu^{\gamma}_{\theta_1} \right) - d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_2,\gamma}_{k\gamma}}, \mu^{\gamma}_{\theta_2} \right) \right| \leq d(\mu^{\gamma}_{\theta_1}, \mu^{\gamma}_{\theta_2}) + d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_1,\gamma}_{k\gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_2,\gamma}_{k\gamma}} \right).$$

Decompose the second term to get

$$\begin{aligned} d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_1,\gamma}_{k\gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_2,\gamma}_{k\gamma}} \right)^2 &\leq C\mathcal{W}_2 \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_1,\gamma}_{k\gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X^{\theta_2,\gamma}_{k\gamma}} \right)^2 \\ &\leq C \frac{1}{N} \sum_{k=0}^{N-1} |X^{\theta_1,\gamma}_{k\gamma} - X^{\theta_2,\gamma}_{k\gamma}|^2 \\ &\leq C_q \frac{1}{N} \sum_{k=0}^{N-1} \sum_{i=0}^{q} |Y^{\theta_1,\gamma}_{k\gamma+ih} - Y^{\theta_2,\gamma}_{k\gamma+ih}|^2 \\ &\leq C_q \sum_{i=0}^{q} \frac{1}{N} \sum_{k=0}^{N-1} |Y^{\theta_1,\gamma}_{k\gamma+ih} - Y^{\theta_2,\gamma}_{k\gamma+ih}|^2. \end{aligned}$$

Let $\beta \in (0,1)$ and $p \geq 1$. By Proposition B.3, we get that there exists a random variable $\mathbf{C}$ with finite moments of order p such that for all $\theta_1, \theta_2 \in \Theta$,

$$\frac{1}{N} \sum_{k=0}^{N-1} |Y^{\theta_1,\gamma}_{k\gamma} - Y^{\theta_2,\gamma}_{k\gamma}|^2 \leq \mathbf{C} \left(1 \wedge |\theta_1 - \theta_2|^{\beta} \right).$$

These results still hold when we replace $Y^{\theta,\gamma}_{k\gamma}$ by $Y^{\theta,\gamma}_{k\gamma+ih}$, since we compare two piecewise constant processes. Therefore, we have that $d(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_1,\gamma}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^{\theta_2,\gamma}_{k\gamma}})$ goes to 0 as $|\theta_1-\theta_2|\to 0$ uniformly in N. The same goes for $d(\mu^\gamma_{\theta_1},\mu^\gamma_{\theta_2})$ by taking the limit $N\to\infty$. This concludes the proof of the equicontinuity and therefore the proof of (ii). □

4.4 Proof of the bounds (4.12), (4.13) and (4.14)

We prove here the bounds (4.12), (4.13) and (4.14) on $D^{(21)}_{N,\gamma}$, $D^{(22)}_{N,\gamma}$ and $D^{(3)}_{N,\gamma}$ that were defined in (4.11). In this Section, d always refer to the distance $d_{CF,p}$.

Proposition 4.6. *Recall that α is the exponent in the strong identifiability assumption* $\mathbf{I_s}$*. Assume that the exponent r in the sub-linear growth of b_ξ in (2.4) satisfies $r\le 1$. For any $\varepsilon\in(0,\alpha\min(\mathcal{H}))$ and any $\varpi\in(0,1)$, there exists constants $C_{\alpha,\varepsilon}>0$ and $C_{\alpha,\varepsilon,\varpi}>0$ such that for all $\gamma\in(0,\gamma_0]$ and $N\ge 1$,*

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D^{(21)}_{N,\gamma}(\theta)\right|^\alpha \le C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}$$
$$\mathbb{E}\sup_{\theta\in\Theta}\left|D^{(22)}_{N,\gamma}(\theta)\right|^\alpha \le C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}$$
$$\mathbb{E}\sup_{\theta\in\Theta}\left|D^{(3)}_{N,\gamma}(\theta)\right|^\alpha \le C_{\alpha,\varepsilon,\varpi}\left(\gamma^{\alpha\min(\mathcal{H})-\varepsilon}+(N\gamma)^{-\frac{\varpi\alpha^2}{2(\alpha\varpi+2d)}(2-(2\max(\mathcal{H})\vee 1))}\right).$$

Proof. First, observe that in both the terms $D^{(21)}_{N,\gamma}$ and $D^{(22)}_{N,\gamma}$ we compare a solution of an SDE with its respective Euler scheme, where both processes are defined with the same noise B. This allows us to do a pathwise comparison. We only detail the bound on $D^{(21)}_{N,\gamma}$, the bound on $D^{(22)}_{N,\gamma}$ can be obtained in the same way. Since $d_{CF,p}$ is an element of $\mathcal{D}_1$, we have:

$$\begin{aligned}\sup_{\theta\in\Theta}D^{(21)}_{N,\gamma}(\theta) &\le \frac{1}{N}\sum_{k=0}^{N-1}\sup_{\theta\in\Theta}|X^\theta_{k\gamma}-X^{\theta,\gamma}_{k\gamma}(B)|\\ &\le C_q\sum_{i=0}^{q}\frac{1}{N}\sum_{k=0}^{N-1}\sup_{\theta\in\Theta}|Y^\theta_{k\gamma+ih}-Y^{\theta,\gamma}_{k\gamma+ih}(B)|.\end{aligned}$$

Recall that $\alpha\ge 2$ in $\mathbf{I_s}$. Hence, an application of Jensen's inequality gives

$$\mathbb{E}\sup_{\theta\in\Theta}D^{(21)}_{N,\gamma}(\theta)^\alpha \le C_{q,\alpha}\sum_{i=0}^{q}\mathbb{E}\left[\sup_{\theta\in\Theta}\frac{1}{N}\sum_{k=0}^{N-1}|Y^\theta_{k\gamma+ih}-Y^{\theta,\gamma}_{k\gamma+ih}(B)|^\alpha\right]$$

Define

$$\mathcal{I}:=\sum_{i=0}^{q}\sup_{\theta\in\Theta}\frac{1}{N}\sum_{k=0}^{N-1}|Y^\theta_{k\gamma+ih}-Y^{\theta,\gamma}_{k\gamma+ih}(B)|^2.$$

We will first provide a bound on $\mathcal{I}$. Using (4.21), we have

$$\begin{aligned}&\sum_{i=0}^{q}|Y^\theta_{k\gamma+ih}-Y^{\theta,\gamma}_{k\gamma+ih}(B)|^2\\ &\quad\le C\sum_{i=0}^{q}\left(\gamma\int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma+\varepsilon_{k,i}}(1+|Y^\theta_s|^{2r})ds+|B_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}}-B_{(k\gamma+ih)_\gamma}|^2\right)\\ &\quad+C\sum_{i=0}^{q}\left(\sum_{j=0}^{\lfloor\frac{k\gamma+ih}{\gamma}\rfloor-1}\left(\gamma^2(1+|Y^{\theta,\gamma}_{j\gamma}(B)|^{2r})+\gamma^{-1}\int_0^\gamma|B_{j\gamma+t}-B_{j\gamma}|^2dt\right)\gamma e^{-\rho(\lfloor\frac{k\gamma+ih}{\gamma}\rfloor-j+1)}\right)\\ &=: C\sum_{i=0}^{q}\left(\mathcal{I}_{1,k}+\mathcal{I}_{2,k}+\mathcal{I}_{3,k}+\mathcal{I}_{4,k}\right),\end{aligned}\tag{4.22}$$

So we have

$$\mathcal{I} \le \sup_{\theta\in\Theta} \frac{1}{N}\sum_{k=0}^{N-1}\left(\mathcal{I}_{1,k}+\mathcal{I}_{2,k}+\mathcal{I}_{3,k}+\mathcal{I}_{4,k}\right).$$

Let us provide uniform bounds in θ on the sum over k of the terms $\mathcal{I}_{1,k}$, $\mathcal{I}_{2,k}$, $\mathcal{I}_{3,k}$, $\mathcal{I}_{4,k}$. First we have

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\mathcal{I}_{1,k} &\le \frac{1}{N}\int_{(ih)_\gamma}^{(N\gamma+ih)_\gamma}\gamma(1+|Y_s^\theta|^{2r})ds\\ &\le \frac{\gamma}{N}\int_{ih}^{N\gamma+ih}(1+|Y_s^\theta|^{2r})ds+\frac{\gamma}{N}\int_{(ih)_\gamma}^{ih}(1+|Y_s^\theta|^{2r})ds\\ &\le \frac{\gamma}{N}\int_{ih}^{N\gamma+ih}(1+|Y_s^\theta|^{2r})ds+2\gamma^2\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y_s^\theta|^{2r})ds\\ &\le \frac{\gamma^2}{N\gamma}\int_0^{N\gamma}(1+|Y_{s+ih}^\theta|^{2r})ds+2\gamma^2\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y_s^\theta|^{2r})ds. \end{aligned}\tag{4.23}$$

For $\mathcal{I}_{3,k}$, we write

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\mathcal{I}_{3,k} &\le C\frac{\gamma^2}{N}\sup_{\theta\in\Theta}\sum_{k=0}^{\lfloor\frac{N\gamma+ih}{\gamma}\rfloor-1}(1+|Y_{k\gamma}^{\theta,\gamma}(B)|^{2r})\\ &\le C\frac{\gamma}{N}\sup_{\theta\in\Theta}\int_0^{(N\gamma+ih)_\gamma-\gamma}(1+|Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r})\\ &\le C\frac{\gamma}{N}\sup_{\theta\in\Theta}\int_0^{N\gamma-\gamma}(1+|Y_{t_{\gamma+ih}}^{\theta,\gamma}(B)|^{2r})+C\mathbb{1}_{i\neq 0}\frac{\gamma}{N}\sup_{\theta\in\Theta}\int_0^{ih}(1+|Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r})\\ &\le C\frac{\gamma}{N}\sup_{\theta\in\Theta}\int_0^{N\gamma-\gamma}(1+|Y_{t_{\gamma+ih}}^{\theta,\gamma}(B)|^{2r})+C\mathbb{1}_{i\neq 0}\frac{\gamma}{N}\sup_{\theta\in\Theta}\frac{1}{ih}\int_0^{ih}(1+|Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r})\\ &\le C\frac{\gamma^2}{N\gamma}\sup_{\theta\in\Theta}\int_0^{N\gamma}(1+|Y_{t_{\gamma+ih}}^{\theta,\gamma}(B)|^{2r})+C\frac{\gamma^2}{N\gamma}\sup_{\theta\in\Theta}\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r}).\end{aligned}\tag{4.24}$$

For $\mathcal{I}_{4,k}$ we have

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\mathcal{I}_{4,k} &\le \frac{1}{N}\sum_{k=0}^{\lfloor\frac{N\gamma+ih}{\gamma}\rfloor-1}\gamma^{-1}\int_0^\gamma|B_{k\gamma+t}-B_{k\gamma}|^2dt\\ &\le \frac{\gamma^{-2}}{N}\int_0^{N\gamma+ih-\gamma}\left(\int_0^\gamma|B_{s_\gamma+t}-B_{s_\gamma}|^2dt\right)ds\\ &\le \frac{1}{N\gamma}\int_0^{ih}\left(\gamma^{-1}\int_0^\gamma|B_{s_\gamma+t}-B_{s_\gamma}|^2dt\right)ds+\frac{1}{N\gamma}\int_0^{N\gamma}\left(\gamma^{-1}\int_0^\gamma|B_{s_\gamma+ih+t}-B_{s_\gamma+ih}|^2dt\right)ds.\end{aligned}\tag{4.25}$$

Therefore, using (4.23), (4.24), (4.25) in (4.22), we get

$$\begin{aligned}\mathcal{I} \le C\sum_{i=0}^{q}\Bigg(&\sup_{\theta\in\Theta}\frac{\gamma^2}{N\gamma}\int_0^{N\gamma}(1+|Y_{s+ih}^\theta|^{2r})ds+2\sup_{\theta\in\Theta}\gamma^2\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y_s^\theta|^{2r})ds\\ &+C\frac{\gamma^2}{N\gamma}\sup_{\theta\in\Theta}\int_0^{N\gamma}(1+|Y_{t_{\gamma+ih}}^{\theta,\gamma}(B)|^{2r})+C\frac{\gamma^2}{N\gamma}\sup_{\theta\in\Theta}\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r})\\ &+\frac{1}{N\gamma}\int_0^{ih}\left(\gamma^{-1}\int_0^\gamma\sup_{\theta\in\Theta}|B_{s_\gamma+t}-B_{s_\gamma}|^2dt\right)ds+\frac{1}{N\gamma}\int_0^{N\gamma}\left(\gamma^{-1}\int_0^\gamma\sup_{\theta\in\Theta}|B_{s_\gamma+ih+t}-B_{s_\gamma+ih}|^2dt\right)ds\\ &+\frac{1}{N}\sum_{k=0}^{N-1}\sup_{\theta\in\Theta}|B_{(k\gamma+ih)_\gamma+\varepsilon_{k,i}}-B_{(k\gamma+ih)_\gamma}|^2\Bigg).\end{aligned}$$

Since $N\gamma \geq 1$ and $\varepsilon_{k,i} < \gamma$, Using Jensen's inequality ($\alpha/2 \geq 1$) and taking the expectation, we get by applying [14, Proposition 3.5] that for $\varepsilon \in (0, \alpha \min(\mathcal{H}))$,

$$\begin{aligned}
\mathbb{E}[\mathcal{I}^{\alpha/2}] \leq C_q \sum_{i=0}^{q} \Bigg(& \mathbb{E}\left[\sup_{\theta\in\Theta} \frac{1}{N\gamma}\int_0^{N\gamma} (1+|Y^\theta_{s+ih}|^{2r})ds\right]^{\alpha/2} + 2\gamma^2 \mathbb{E}\left[\sup_{\theta\in\Theta} \frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y^\theta_s|^{2r})ds\right]^{\alpha/2} \\
& + C\gamma^2\mathbb{E}\left[\frac{1}{N\gamma}\sup_{\theta\in\Theta}\int_0^{N\gamma}(1+|Y^{\theta,\gamma}_{t_{\gamma+ih}}(B)|^{2r})\right]^{\alpha/2} + C\gamma^2\mathbb{E}\left[\sup_{\theta\in\Theta}\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}(1+|Y^{\theta,\gamma}_{t_\gamma}(B)|^{2r})\right]^{\alpha/2} \\
& + \gamma^{\alpha\min(\mathcal{H})-\varepsilon} + \frac{1}{N\gamma}\int_0^{N\gamma}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}\,ds + \frac{1}{N}\sum_{k=0}^{N-1}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}\Bigg) \\
\leq C_q \sum_{i=0}^{q} \Bigg(& \mathbb{E}\left[\sup_{\theta\in\Theta} \frac{1}{N\gamma}\int_0^{N\gamma} |Y^\theta_{s+ih}|^{2r}ds\right]^{\alpha/2} + 2\gamma^2 \mathbb{E}\left[\sup_{\theta\in\Theta} \frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}|Y^\theta_s|^{2r}ds\right]^{\alpha/2} \\
& + C\gamma^2\mathbb{E}\left[\frac{1}{N\gamma}\sup_{\theta\in\Theta}\int_0^{N\gamma}|Y^{\theta,\gamma}_{t_{\gamma+ih}}(B)|^{2r}\right]^{\alpha/2} + C\gamma^2\mathbb{E}\left[\sup_{\theta\in\Theta}\frac{\mathbb{1}_{i\neq 0}}{ih}\int_0^{ih}|Y^{\theta,\gamma}_{t_\gamma}(B)|^{2r}\right]^{\alpha/2} \\
& + \gamma^\alpha + \gamma^{\alpha\min(\mathcal{H})-\varepsilon}\Bigg).
\end{aligned}$$

By Proposition B.1(iii), Proposition B.2(ii) and since $r \leq 1$, we have that the quantities

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\frac{1}{N\gamma}\int_0^{N\gamma}|Y^\theta_s|^{2r}ds\right]^{\alpha/2}, \mathbb{E}\left[\frac{1}{N\gamma}\sup_{\theta\in\Theta}\int_0^{N\gamma}|Y^{\theta,\gamma}_{t_\gamma}(B)|^{2r}\right]^{\alpha/2},$$

are finite uniformly in N and γ. One can check that the result still holds when the process is shifted by ih since the shifted process is still solution of an SDE that satisfies the necessary assumptions. Therefore, for $\varepsilon \in (0, \alpha \max(\mathcal{H}))$,

$$\mathbb{E}[\mathcal{I}^{\alpha/2}] \leq C\gamma^{\alpha\min(\mathcal{H})-\varepsilon}.$$

We conclude by observing that by Jensen's inequality

$$\begin{aligned}
\mathcal{I} &\geq \sum_{i=0}^{q}\left(\sup_{\theta\in\Theta}\frac{1}{N}\sum_{k=0}^{N-1}|Y^\theta_{k\gamma+ih} - Y^{\theta,\gamma}_{k\gamma+ih}(B)|\right)^2 \\
&\geq C\left(\sum_{i=0}^{q}\sup_{\theta\in\Theta}\frac{1}{N}\sum_{k=0}^{N-1}|Y^\theta_{k\gamma+ih} - Y^{\theta,\gamma}_{k\gamma+ih}(B)|\right)^2.
\end{aligned}$$

Hence

$$\mathbb{E}\sup_{\theta\in\Theta} D^{(21)}_{N,\gamma}(\theta)^\alpha \leq C\mathbb{E}[\mathcal{I}^{\alpha/2}] \leq C\gamma^{\alpha\min(\mathcal{H})-\varepsilon}. \tag{4.26}$$

Consider now $D^{(3)}_{N,\gamma}(\theta)$, which we recall is given by $D^{(3)}_{N,\gamma}(\theta) = d(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^\theta_{k\gamma}(\hat{B})}, \mu_\theta)$. Since μ_θ is also the stationary law of the process $X^\theta(\hat{B})$, we drop the dependence on $\hat{B}$ for the rest of the proof. This quantity is the hardest to handle, due to the fact that we wish to achieve a uniform bound in θ. We first start with the following decomposition:

$$\sup_{\theta\in\Theta} D^{(3)}_{N,\gamma}(\theta) \leq \sup_{\theta\in\Theta} D^{(31)}_{N,\gamma}(\theta) + \sup_{\theta\in\Theta} D^{(32)}_{N,\gamma}(\theta),$$

where, noticing that $\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X^\theta_{k\gamma}} = \frac{1}{T}\int_0^T \delta_{X^\theta_{t_\gamma}}\,dt$ for $T = N\gamma$,

$$D^{(31)}_{N,\gamma}(\theta) = d\left(\mu_\theta, \frac{1}{T}\int_0^T \delta_{X^\theta_t} dt\right)$$
$$D^{(32)}_{N,\gamma}(\theta) = d\left(\frac{1}{T}\int_0^T \delta_{X^\theta_t} dt, \frac{1}{T}\int_0^T \delta_{X^\theta_{t_\gamma}}\,dt\right).$$

Bound on $\sup_{\theta\in\Theta} D^{(32)}_{N,\gamma}(\theta)$. Similar arguments as before lead to

$$\mathbb{E}\sup_{\theta\in\Theta} D^{(32)}_{N,\gamma}(\theta)^\alpha \le C_{q,\alpha}\sum_{i=0}^{q}\frac{1}{T}\int_0^T \sup_{\theta\in\Theta}|Y^\theta_{t+ih} - Y^\theta_{t_\gamma+ih}|^\alpha\,dt.$$

We will show how to bound the quantity above for $i = 0$. The same arguments can be used for any value of i. Since Y^θ is a solution of (2.5), it follows by triangle inequality that

$$|Y^\theta_t - Y^\theta_{t_\gamma}| \le \int_{t_\gamma}^t |b_\xi(Y^\theta_s)|ds + |\sigma||B_t - B_{t_\gamma}|.$$

Moreover, using Jensen's inequality and integrating over t, we have

$$\frac{1}{T}\int_0^T |Y^\theta_t - Y^\theta_{t_\gamma}|^\alpha\,dt \le \gamma^{\alpha-1}\frac{1}{T}\int_0^T\int_{t_\gamma}^t |b_\xi(Y^\theta_s)|\,ds\,dt + |\sigma|^\alpha\frac{1}{T}\int_0^T |B_t - B_{t_\gamma}|^\alpha dt.$$

Integrating over t first, we get that

$$\frac{1}{T}\int_0^T |Y^\theta_t - Y^\theta_{t_\gamma}|^\alpha\,dt \le \gamma^{\alpha}\frac{1}{T}\int_0^T |b_\xi(Y^\theta_s)|ds + |\sigma|^\alpha\frac{1}{T}\int_0^T |B_t - B_{t_\gamma}|^\alpha dt.$$

The drift term above is bounded thanks to the sublinear growth of b_ξ given by (2.4) and the uniform bounds on the L^q moments of Y^θ_t given in Proposition B.1 (ii). As for the term $|B_t - B_{t_\gamma}|$, we have thanks to [14, Proposition 3.5] that for all $\varepsilon > 0$,

$$\mathbb{E}\left(\sup_{H\in\mathcal{H}}|B_t - B_{t_\gamma}|^\alpha\right) \le C\gamma^{\alpha\min(\mathcal{H})-\varepsilon}.$$

From here, it is readily checked that

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\frac{1}{T}\int_0^T |Y^\theta_t - Y^\theta_{t_\gamma}|^\alpha dt\right) \le C\gamma^{\alpha\min(\mathcal{H})-\varepsilon}.$$

Bound on $D^{(31)}_{N,\gamma}(\theta)$. The quantity $D^{(31)}_{N,\gamma}(\theta)$ can be handled the same way as $D^{(1)}_n$ in the proof of Lemma 3.5. Namely, we get that

$$\mathbb{E}D^{(31)}_{N,\gamma}(\theta)^\alpha \le C_\alpha\left(T^{-\alpha} + T^{-\frac{\alpha}{2}(2-\max(2\max(\mathcal{H}),1)}\right). \tag{4.27}$$

The hardest part is to obtain a bound on the supremum of $D^{(31)}_{N,\gamma}(\theta)$ over θ.

Bound on $\sup_{\theta\in\Theta} D^{(31)}_{N,\gamma}(\theta)$. Let $\varphi(\theta) = d(\mu_\theta, \frac{1}{T}\int_0^T \delta_{X^\theta_t} dt)$. In order to obtain a bound for the supremum over Θ, we discretise the parameter space Θ. Let $\varepsilon > 0$ and $\Theta^{(\varepsilon)} := \left\{\theta^{(\varepsilon)}_i \mid 1 \leqslant i \leqslant M_\varepsilon\right\}$ such that $\Theta \subset \bigcup_{i=1}^{M_\varepsilon} B\left(\theta^{(\varepsilon)}_i, \varepsilon\right)$ for some points $\theta^{(\varepsilon)}_i$ in Θ. Then, for any $\theta \in \Theta$,

$$\varphi(\theta) \leqslant |\varphi(\theta) - \varphi(\theta_\varepsilon)| + |\varphi(\theta_\varepsilon)|,$$

where $\theta_\varepsilon := \underset{\theta' \in \left\{\theta_i^{(\varepsilon)}\right\}}{\text{argmin}} |\theta' - \theta|$. Therefore

$$\varphi(\theta) \le |\varphi(\theta) - \varphi(\theta_\varepsilon)| + \max_{1 \le i \le M_\varepsilon} \left|\varphi\left(\theta_i^{(\varepsilon)}\right)\right|.$$

Using (4.27), we have

$$\mathbb{E} \sup_{\theta \in \Theta} \varphi(\theta)^\alpha \le C_\alpha \left(\mathbb{E} \sup_{\theta \in \Theta} |\varphi(\theta) - \varphi(\theta_\varepsilon)|^\alpha + M_\varepsilon \left(T^{-\alpha} + T^{-\frac{\alpha}{2}(2 - \max(2\max(\mathcal{H}), 1))}\right)\right).$$

Let us split the quantity $|\varphi(\theta) - \varphi(\theta_\varepsilon)|$ in two terms:

$$|\varphi(\theta) - \varphi(\theta_\varepsilon)| \le d(\mu_\theta, \mu_{\theta_\varepsilon}) + d\left(\frac{1}{T}\int_0^T \delta_{X_t^\theta} dt, \frac{1}{T}\int_0^T \delta_{X_t^{\theta_\varepsilon}} dt\right).$$

Since d belongs to $\mathcal{D}_2$, the second term in the right-hand side yields

$$d\left(\frac{1}{T}\int_0^T \delta_{X_t^\theta} dt, \frac{1}{T}\int_0^T \delta_{X_t^{\theta_\varepsilon}} dt\right) \le C_q \sum_{i=0}^q \frac{1}{T}\int_0^T |Y_{t+ih}^\theta - Y_{t+ih}^{\theta_\varepsilon}|^2 dt.$$

For $\varpi \in (0,1)$, Proposition B.3 yields that there exists a random variable $\mathbf{C}$ with finite moments such that

$$\frac{1}{T}\int_0^T |Y_t^\theta - Y_t^{\theta_\varepsilon}|^2 dt \le \mathbf{C}|\theta - \theta_\varepsilon|^{\frac{\varpi}{2}}.$$

This bound still holds if Y_t is replaced by Y_{t+ih} since

$$\frac{1}{T}\int_0^T |Y_{t+ih}^\theta - Y_{t+ih}^{\theta_\varepsilon}|^2 dt \le \frac{T+ih}{T}\frac{1}{T+ih}\int_0^{T+ih} |Y_t^\theta - Y_t^{\theta_\varepsilon}|^2 dt.$$

Overall, we get

$$d\left(\frac{1}{T}\int_0^T \delta_{X_t^\theta} dt, \frac{1}{T}\int_0^T \delta_{X_t^{\theta_\varepsilon}} dt\right) \le \mathbf{C}_q |\theta - \theta_\varepsilon|^{\frac{\varpi}{2}},$$

where $\mathbf{C}_q$ is a random variable that has a finite moments. By letting T go to infinity, we obtain a similar bound for $d(\mu_\theta, \mu_{\theta_\epsilon})$. It follows that

$$d\left(\frac{1}{T}\int_0^T \delta_{X_t^\theta} dt, \frac{1}{T}\int_0^T \delta_{X_t^{\theta_\varepsilon}} dt\right)^\alpha \le \mathbf{C}_q^\alpha |\theta - \theta_\varepsilon|^{\frac{\alpha\varpi}{2}}.$$

Hence we have obtained

$$\mathbb{E} \sup_{\theta \in \Theta} \varphi(\theta)^\alpha \le C_{\alpha,q,\varpi} \left(\varepsilon^{\frac{\alpha\varpi}{2}} + M_\varepsilon \left(T^{-\alpha} + T^{-\frac{\alpha}{2}(2 - \max(2\max(\mathcal{H}), 1))}\right)\right).$$

Choosing $M_\varepsilon \le \frac{C}{\varepsilon^d}$ and $\varepsilon = T^{-\chi}$ for some $\chi > 0$, we have that

$$\mathbb{E} \sup_{\theta \in \Theta} \varphi(\theta)^\alpha \le C_{\alpha,q,\varpi} \left(T^{-\chi\alpha\frac{\varpi}{2}} + T^{-\frac{\alpha}{2}(2 - \max(2\max(\mathcal{H}), 1) + \chi d}\right).$$

Finally we optimize over χ to get

$$\mathbb{E} \sup_{\theta \in \Theta} \varphi(\theta)^\alpha \le C_{\alpha,q,\varpi} T^{-\bar{\eta}},$$

for $\bar{\eta} = \frac{\varpi\alpha^2}{2(\alpha\varpi + 2d)}(2 - (2\max(\mathcal{H}) \vee 1))$. □

5 Fractional Ornstein-Uhlenbeck processes

We first study the identifiability assumption for the fractional Ornstein-Uhlenbeck (OU) process in Section 5.1, then a family of small perturbations of the fractional OU process in Section 5.2, and finally, in Section 5.3 we provide some numerical experiments to illustrate our main results.

5.1 Identifiability assumption

In this section, we provide an example of equation (1.1) for which the crucial assumption $\mathbf{I_s}$ is satisfied. More specifically, we consider the one-dimensional fractional Ornstein-Uhlenbeck process given by

$$\begin{aligned} &dU^\theta = -\xi U^\theta dt + \sigma dB \\ &U^\theta_0 = 0. \end{aligned} \tag{5.1}$$

For simplicity reasons, we consider that the linear transformations ℓ^i are increments of the form

$$\ell^1(U^\theta_\cdot, ..., U^\theta_{\cdot+ih}) = U^\theta_{\cdot+ih} - U^\theta_\cdot$$

We suppose here that θ is of dimension 2, i.e. only two parameters are unknown. We prove the following result.

Proposition 5.1. *Consider the fractional Ornstein-Uhlenbeck model defined by equation (5.1) and assume that one of the parameters ξ, σ or H is known. Let μ_θ denote the stationary measure of $(U^\theta_\cdot, U^\theta_{\cdot+h} - U^\theta_\cdot)$, then there exists $h_0 > 0$ such that for all $h < h_0$, we have*

$$\forall \theta_1, \theta_2 \in \Theta, \quad d(\mu_{\theta_1}, \mu_{\theta_2}) = 0 \quad \textit{iff} \quad \theta_1 = \theta_2.$$

The proof is given in Section 5.1.1 and is based on proving the injectivity of a specific function. This is stated in the following Lemma (the proof is given in Section 5.1.2 and Appendix D).

Lemma 5.2. *Assume one of the three cases $\theta = (\xi, H)$, $\theta = (\xi, \sigma)$ or $\theta = (H, \sigma)$, then the function f defined by*

$$f : \theta \to \begin{pmatrix} \sigma^2 H\Gamma(2H)\xi^{-2H} \\ \sigma^2\Gamma(2H+1)\frac{\sin(\pi H)}{\pi}\int_0^\infty \cos(hx)\frac{x^{1-2H}}{\xi^2+x^2}dx \end{pmatrix} \tag{5.2}$$

is one-to-one.

The computations done in this Section to prove the Proposition above can be generalized to the case when U^θ is a higher-dimensional fractional Ornstein-Uhlenbeck process with σ a multiple of the identity matrix.

5.1.1 Simplification of the problem: proof of Proposition 5.1

For the fractional Ornstein-Uhlenbeck process, the stationary measure is known to be Gaussian (see [5, Eq (2.2)]) and is given by

$$\mathcal{N}(0, \sigma^2 H\Gamma(2H)\xi^{-2H}). \tag{5.3}$$

Furthermore, the processes $\bar{U}^\theta_{\cdot+ih}$ are also Gaussian with the same law. The interaction of such processes is thus described by the covariance matrix, given by (see [5] for example):

$$\mathbb{E}(\bar{U}^\theta_t \bar{U}^\theta_{t+ih}) = \sigma^2 \frac{\Gamma(2H+1)\sin(\pi H)}{\pi}\int_0^\infty \cos(ihx)\frac{x^{1-2H}}{\xi^2+x^2}dx. \tag{5.4}$$

Now for θ_1, θ_2 in Θ, there is

$$d_{CF,2}(\mu_{\theta_1}, \mu_{\theta_2})^2 = \int_{\mathbb{R}^2}\left(\mathbb{E}e^{i\langle\chi,(\bar{U}^{\theta_1}_t, \bar{U}^{\theta_1}_{t+h} - \bar{U}^{\theta_1}_t)\rangle} - \mathbb{E}e^{i\langle\chi,(\bar{U}^{\theta_2}_t, \bar{U}^{\theta_2}_{t+h} - \bar{U}^{\theta_2}_t)\rangle}\right)^2 g(\chi)d\chi.$$

Since the process $(\bar{U}^{\theta}_{\cdot}, \bar{U}^{\theta}_{\cdot+h} - \bar{U}^{\theta}_{\cdot})$ is Gaussian and stationary, it comes:

$$d_{CF}(\mu_{\theta_1}, \mu_{\theta_2}) = 0 \quad \text{iff} \quad \begin{array}{c} \mathbb{E}(\bar{U}^{\theta_1}_0)^2 = \mathbb{E}(\bar{U}^{\theta_2}_0)^2 \\ \mathbb{E}\left(\bar{U}^{\theta_1}_0(\bar{U}^{\theta_1}_h - \bar{U}^{\theta_1}_0)\right) = \mathbb{E}\left(\bar{U}^{\theta_2}_0(\bar{U}^{\theta_2}_h - \bar{U}^{\theta_2}_0)\right) \end{array}$$

which now reads

$$d_{CF}(\mu_{\theta_1}, \mu_{\theta_2}) = 0 \text{ iff } \begin{array}{c} \mathbb{E}(\bar{U}^{\theta_1}_0)^2 = \mathbb{E}(\bar{U}^{\theta_2}_0)^2 \\ \mathbb{E}\left(\bar{U}^{\theta_1}_0\bar{U}^{\theta_1}_h\right) = \mathbb{E}\left(\bar{U}^{\theta_2}_0\bar{U}^{\theta_2}_h\right) \end{array}.$$

In view of (5.3) and (5.4), assumption $\mathbf{I_w}$ becomes equivalent to the injectivity of the function f defined in (5.2), which is therefore given by Lemma 5.2.

Remark 5.3. *In [13], the authors studied fractional OU processes and proposed a similar estimator for (H, ξ, σ) simultaneously. Similar to our case, for a strong consistency argument to hold, they are left to study the injectivity of:*

$$f : (\xi, H, \sigma) \to \begin{pmatrix} \sigma^2 H\Gamma(2H)\xi^{-2H} \\ \sigma^2\Gamma(2H+1)\frac{\sin(\pi H)}{\pi}\int_0^\infty \cos(hx)\frac{x^{1-2H}}{\xi^2+x^2}dx \\ \sigma^2\Gamma(2H+1)\frac{\sin(\pi H)}{\pi}\int_0^\infty \cos(2hx)\frac{x^{1-2H}}{\xi^2+x^2}dx \end{pmatrix} \text{ is injective.}$$

However, they did not prove the injectivity but provide numerical arguments that support their claim. Although, in our case, we deal with an easier problem (we estimate only two parameters), we manage to prove the injectivity of f.

5.1.2 Injectivity of f: proof of Lemma 5.2

When $\theta = (\xi, H)$ (only σ is known), the proof of the injectivity of f and therefore of the identifiability assumption (2.9) is done in Appendix D (due to the length and technical details of the proof). We know consider consider the cases $\theta = (H, \sigma)$ and $\theta = (\xi, \sigma)$ and prove the injectivity of f.

The case $\theta = (H, \sigma)$. Let (a, b) in $Im(f)$, we will show that the equation

$$\begin{aligned} a &= \sigma^2 H\Gamma(2H)\xi^{-2H} \\ b &= \sigma^2\Gamma(2H+1)\frac{\sin(\pi H)}{\pi}\xi^{-2H}\int_0^\infty \cos(\xi hx)\frac{x^{1-2H}}{1+x^2}dx, \end{aligned} \tag{5.5}$$

has a unique solution in H, σ. First, notice that thanks to the first equation, we can write $\sigma^2 = \frac{a\xi^{2H}}{H\Gamma(2H)}$. Injecting this in the second equation we get:

$$b\pi = a\underbrace{\sin(\pi H)\int_0^\infty \cos(\xi hx)\frac{x^{1-2H}}{1+x^2}}_{:=g(H)}.$$

We will show that the function g is injective. For that, and since g is continuously differentiable, we will show that $g'(H) > 0$ for all $H \in \mathcal{H}$. We have:

$$\begin{aligned} g'(H) = \pi\cos(\pi H)&\int_0^\infty \cos(\xi hx)\frac{x^{1-2H}}{1+x^2}dx \\ - 2\sin(\pi H)&\int_0^1 cos(\xi hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx \\ - 2\sin(\pi H)&\int_1^\infty \cos(\xi hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

Let us assume first that $H > 1/2$**.** Let $\epsilon > 0$ and h small enough such that $\cos(\xi h x) \geq 1 - \epsilon$ for all $x \in (0,1)$ (which is possible since ξ lives in a compact), then:

$$g'(H) \geq \underbrace{-2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx}_{:=g_1(H)} - 2\epsilon\sin(\pi H)\int_0^1 \cos(\xi h x)\log(x)\frac{x^{1-2H}}{1+x^2}dx$$
$$\underbrace{-2\sin(\pi H)\int_1^\infty \cos(\xi h x)\log(x)\frac{x^{1-2H}}{1+x^2}dx}_{:=g_2(H)}$$
$$\underbrace{+\pi\cos(\pi H)\int_0^\infty \cos(\xi h x)\frac{x^{1-2H}}{1+x^2}dx}_{:=g_3(H)}.$$

Notice that since $H \in \mathcal{H}$, the absolute value of $2\epsilon\sin(\pi H)\int_0^1 \cos(\xi h x)\log(x)\frac{x^{1-2H}}{1+x^2}dx$ is bounded from above by a constant ϵC_1. Let us focus on the sum of g_1, g_2 and g_3. Notice that $g_1(H)$ is always positive. Bounding the term $\cos(\xi h x)$ in g_2 and g_3 by 1 (in absolute value), we get

$$\begin{aligned} g_1(H) + g_2(H) + g_3(H) > & -2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx \\ & - 2\sin(\pi H)\int_1^\infty \log(x)\frac{x^{1-2H}}{1+x^2}dx \\ & + \pi\cos(\pi H)\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

Observe that $H\Gamma(2H)\xi^{-2H} = 2H\Gamma(2H)\frac{\sin(\pi H)}{\pi}\xi^{-2H}\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx$. So by differentiating in H, the right-hand side above is actually equal to 0. It follows that there exists a constant C_2 such that

$$g_1(H) + g_2(H) + g_3(H) \geq C_2.$$

So overall, we have

$$g'(H) \geq C_2 - \epsilon C_1.$$

Choosing ϵ small enough (which means choosing h smaller than a theoretical h_0), we get

$$g'(H) > 0.$$

We now assume that $H < 1/2$**.** We have

$$\begin{aligned} g'(H) = & \pi\cos(\pi H)\int_0^1 \frac{x^{1-2H}}{1+x^2}dx + \pi\cos(\pi H)\int_0^1 (\cos(\xi h x) - 1)\frac{x^{1-2H}}{1+x^2}dx \\ & + \pi\cos(\pi H)\int_1^\infty \cos(\xi h x)\frac{x^{1-2H}}{1+x^2}dx \\ & - 2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx - 2\sin(\pi H)\int_0^1 (\cos(\xi h x) - 1)\log(x)\frac{x^{1-2H}}{1+x^2}dx \\ & - 2\sin(\pi H)\int_1^\infty \cos(\xi h x)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

It follows that

$$\begin{aligned} g'(H) > & \pi\cos(\pi H)\int_0^1 \frac{x^{1-2H}}{1+x^2}dx + \pi\cos(\pi H)\int_1^\infty \cos(\xi h x)\frac{x^{1-2H}}{1+x^2}dx \\ & - 2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx - 2\sin(\pi H)\int_1^\infty \cos(\xi h x)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

Recalling that $H\Gamma(2H)\xi^{-2H} = 2H\Gamma(2H)\frac{\sin(\pi H)}{\pi}\xi^{-2H}\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx$, we deduce that

$$\pi\cos(\pi H)\int_0^1 \frac{x^{1-2H}}{1+x^2}dx = -\pi\cos(\pi H)\int_1^\infty \frac{x^{1-2H}}{1+x^2}dx + \frac{\pi^2\cos(\pi H)}{\sin(\pi H)}$$
$$-2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx = 2\sin(\pi H)\int_1^\infty \log(x)\frac{x^{1-2H}}{1+x^2}dx - \frac{\pi^2\cos(\pi H)}{\sin(\pi H)},$$

where the second equation is obtained by differentiating with respect to H. Using this in our lower bound of g', we get

$$g'(H) > \pi\cos(\pi H)\int_1^\infty [\cos(\xi hx)-1]\frac{x^{1-2H}}{1+x^2}dx - 2\sin(\pi H)\int_1^\infty [\cos(\xi hx)-1]\log(x)\frac{x^{1-2H}}{1+x^2}dx.$$

Moving all the terms inside the integrals, we are left with

$$g'(H) > \int_1^\infty [1-\cos(\xi hx)][2\log(x)\sin(\pi H) - \pi\cos(\pi H)]\frac{x^{1-2H}}{1+x^2}dx.$$

We split the integral by the sign of $2\log(x)\sin(\pi H) - \pi\cos(\pi H)$ as follows:

$$g'(H) > \int_1^{e^{\frac{\pi\cos(\pi H)}{2\sin(\pi H)}}} [1-\cos(\xi hx)][2\log(x)\sin(\pi H) - \pi\cos(\pi H)]\frac{x^{1-2H}}{1+x^2}dx$$
$$+ \int_{e^{\frac{\pi\cos(\pi H)}{2\sin(\pi H)}}}^\infty [1-\cos(\xi hx)][2\log(x)\sin(\pi H) - \pi\cos(\pi H)]\frac{x^{1-2H}}{1+x^2}dx.$$

Since H lives in a compact set, we can choose h small enough so that the first term (the integral on a bounded interval) is as small as we want it to be. The second term is clearly positive. This is enough to conclude that there exists h_0 such that for $h < h_0$, we have

$$g'(H) > 0.$$

We have thus proved that f is one-to-one.

The case $\theta = (\xi, \sigma)$. As before, for (a, b) in the range of f, we need to show that (5.5) has a unique solution in σ, ξ. Notice that (5.5) is equivalent to

$$a = \sigma^2 H\Gamma(2H)\xi^{-2H}$$
$$b = a\frac{\sin(\pi H)}{\pi}\underbrace{\int_0^\infty \cos(\xi hx)\frac{x^{1-2H}}{1+x^2}dx}_{:=g(\xi)}.$$

Thus, it is enough to show that $g'(\xi) > 0$ for all ξ. We have

$$g'(\xi) = h\int_0^\infty \sin(\xi hx)\frac{x^{-2H}}{1+x^2}dx.$$

There is no problem differentiating with respect to ξ since using the Laplace transform one can show that $\int_0^\infty \frac{\sin(x)}{x^\alpha}dx = \frac{\Gamma(\alpha/2)\Gamma(1-\alpha/2)}{2\Gamma(\alpha)}$ for $\alpha \in (0,2)$. Let us now show that $g'(\xi) > 0$. There is

$$\begin{aligned} g'(\xi) &= h\left(\int_0^{1/h} \sin(\xi hx)\frac{x^{-2H}}{1+x^2}dx + \int_{1/h}^\infty \sin(\xi hx)\frac{x^{-2H}}{1+x^2}dx\right) \\ &> h\left(\int_0^{1/h} \xi h\frac{x^{-2H+1}}{1+x^2}dx - \int_0^{1/h}\frac{\xi^3h^3}{6}\frac{x^{-2H+3}}{1+x^2}dx - \int_{1/h}^\infty \frac{x^{-2H}}{1+x^2}dx\right) \\ &> h\left(\int_0^\infty \xi h\frac{x^{-2H+1}}{1+x^2}dx - \int_{1/h}^\infty \xi h\frac{x^{-2H+1}}{1+x^2}dx - \int_0^{1/h}\frac{\xi^3h^3}{6}\frac{x^{-2H+3}}{1+x^2}dx - \int_{1/h}^\infty \frac{x^{-2H}}{1+x^2}dx\right). \end{aligned}$$

The idea is to keep the first term since it is positive and bound from above the absolute value of the other terms. Since H and ξ both live in compact sets, we can find constants C_1, C_2, C_3 and C_4 such that

$$\begin{aligned} g'(\xi) &> h\Bigg(C_1 h - C_2 h (1/h)^{-2H} - C_3 h^3 (1/h)^{-2H+2} - C_4 (1/h)^{-2H-1} \Bigg) \\ &> h\Bigg(C_1 h - C_2 h^{2H+1} - C_3 h^{2H+1} - C_4 h^{2H+1} \Bigg) \\ &> h^2 \Bigg(C_1 - C_2 h^{2H} - C_3 h^{2H} - C_4 h^{2H} \Bigg). \end{aligned}$$

It follows that there exists $h_0 > 0$ such that for $h < h_0$, we have:

$$g'(\xi) > 0.$$

5.2 Strong Identifiability assumption for a small perturbation of the fractional OU process

In this section, we check assumption $\mathbf{I_s}$ for some specific examples of (2.5) and for the distance $d = d_{CF,p}$. Specifically, we shall consider a family $U^{\lambda,\theta}$ of real-valued processes defined by

$$dU_t^{\lambda,\theta} = \left(-\xi U_t^{\lambda,\theta} + \lambda b_\xi (U_t)^{\lambda,\theta} \right) dt + \sigma dB_t. \tag{5.6}$$

The quantity λ is a small enough parameter which is assumed to be known. The process $U^{\lambda,\theta}$ can be seen as a small perturbation of the fractional Ornstein-Uhlenbeck process, since $U^{0,\theta} = U^\theta$, where U^θ is the fractional OU process defined in (5.1). We also assume that ξ, σ and H are one dimensional parameters and:

$$\begin{aligned} \xi &\in [m_\Xi, M_\Xi], \text{ with } 0 < m_\Xi < M_\Xi < \infty \\ \sigma &\in [m_\Sigma, M_\Sigma], \text{ with } 0 < m_\Sigma < M_\Sigma < \infty \\ H &\in [m_\mathcal{H}, M_\mathcal{H}], \text{ with } 0 < m_\mathcal{H} < M_\mathcal{H} < 1. \end{aligned}$$

We shall prove that $Y^{\lambda,\theta}$ satisfies assumption $\mathbf{I_s}$ when only one parameter is unknown (so either $\theta = \xi$, $\theta = \sigma$ or $\theta = H$). When referring to θ, we will write our assumption above as $\theta \in [m_\Theta, M_\Theta]$.

Let us start with the process $U^\theta := Y^{0,\theta}$, that is the fractional OU itself. In this case, assumption $\mathbf{I_w}$ is satisfied as shown in the following lemma.

Lemma 5.4. *Consider the fractional Ornstein-Uhlenbeck process U^θ. We call μ_θ its stationary distribution, where θ represents either H, σ or ξ. Moreover, if $\theta = H$, assume that*

$$m_\Xi > \sup_{H \in [m_\mathcal{H}, M_\mathcal{H}]} e^{\frac{\Gamma'(2H+1)}{2\Gamma(2H+1)}} \text{ or } M_\Xi < \inf_{H \in [m_\mathcal{H}, M_\mathcal{H}]} e^{\frac{\Gamma'(2H+1)}{2\Gamma(2H+1)}}. \tag{5.7}$$

Then for all $\theta_1, \theta_2 \in [m_\Theta, M_\Theta]$,

$$d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2}) \geq c|\theta_1 - \theta_2|, \tag{5.8}$$

where c is a constant that does not depend on θ_1 or θ_2.

Proof. When $\theta = \xi$, this lemma has already been proved in [24, Lemma 6.2]. The fact that ξ is bounded away from 0 is crucial in their proof.

Let us now deal with the case $\theta = H$. We have already seen that $\mu_\theta = \mathcal{N}\left(0, \sigma^2 H \Gamma(2H) \xi^{-2H}\right)$. Taking into account the expression of $d_{CF,p}$ in (2.1), this yields:

$$d_{CF,p}^2(\mu_{\theta_1}, \mu_{\theta_2}) = \int_\mathbf{R} \left(\exp(-\frac{\sigma^2 H_1 \Gamma(2H_1)}{2\xi^{2H_1}} \eta^2) - \exp(-\frac{\sigma^2 H_2 \Gamma(2H_2)}{2\xi^{2H_2}} \eta^2) \right)^2 g_p(\eta) d\eta.$$

Therefore, it is sufficient to show that the derivative of $g(H) = \exp(-\frac{\sigma^2 H\Gamma(2H)}{2\xi^{2H}}\eta^2)$ is bounded away from 0.

$$g'(H) = \sigma^2\eta^2\frac{\xi^{-2H}}{2}\exp(-\frac{\sigma^2 H\Gamma(2H)}{2\xi^{2H}}\eta^2)\,(\Gamma'(2H+1) - \Gamma(2H+1)\log(\xi))\,.$$

Therefore, under (5.7), we either have $g'(H) < 0$ for all $H \in [m_{\mathcal{H}}, M_{\mathcal{H}}]$, or $g'(H) > 0$ for all $H \in [m_{\mathcal{H}}, M_{\mathcal{H}}]$.

A similar analysis can be done when $\theta = \sigma$. In this case, one needs to show that the derivative of $g(\sigma) = \exp(-\frac{\sigma^2 H\Gamma(2H)}{2\xi^{2H}}\eta^2)$ is bounded away from 0. Since there is

$$g'(\sigma) = -\frac{2\sigma H\Gamma(2H)}{2\xi^{2H}}\eta^2\exp(-\frac{\sigma^2 H\Gamma(2H)}{2\xi^{2H}}\eta^2),$$

and all the parameters live in compact sets that do not contain 0, it is clear that for all $\sigma \in [m_\Sigma, M_\Sigma]$, we have $g'(\sigma) < 0$. □

We now wish to extend Lemma 5.4 to the model given by equation (5.6). We will prove the following proposition.

Proposition 5.5. *Let $Y^{\lambda,\theta}$ be the process defined by (5.6) where θ is either H, σ or ξ, and consider $p > 3/2$. We assume $\theta \in [m_\Theta, M_\Theta]$ as in Lemma 5.4 and $\lambda \in (0, \lambda_0)$ with a small enough $\lambda_0 = \lambda_0(m_\Theta, M_\Theta, p)$. Also, assume without loss of generality that b_ξ, $\partial_y b_\xi$, $\partial_\xi b_\xi$, $\partial^2_{y,\xi} b_\xi$ are all bounded by 1. Moreover, if $\theta = H$, assume that (5.7) holds. Then the following lower bound holds true for any $\theta_1, \theta_2 \in [m_\Theta, M_\Theta]$:*

$$d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2}) \geq c_{m_\Theta, M_\Theta, p}|\theta_1 - \theta_2|.$$

Remark 5.6. *In Proposition 5.5, the assumption that $\partial_\xi b_\xi$ and $\partial^2_{\xi,y} b_\xi$ are both bounded by 1 is only needed when $\theta = \xi$, that is, the case already handled by [24, Proposition 6.4].*

Proof. Again, the case $\theta = \xi$ has already been handled in [24, Proposition 6.4]. Our proof for the general case will be very similar. More specifically, we decompose $d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2})$ as

$$d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2}) \geq I_3^{1/2} - \left(I_2^{1/2} + I_{11}^{1/2} + I_{12}^{1/2}\right), \tag{5.9}$$

where

$$\begin{aligned}
I_{1j} &= \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta\bar{U}_t^{\lambda,\theta_j})] - \mathbb{E}[\exp(i\eta U_t^{\lambda,\theta_j})]\right)^2 g_p(\eta)d\eta, \quad j = 1, 2,\\
I_3 &= \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta U_t^{0,\theta_1})] - \mathbb{E}[\exp(i\eta U_t^{0,\theta_2})]\right)^2 g_p(\eta)d\eta,\\
I_2 &= \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta U_t^{\lambda,\theta_1})] - \mathbb{E}[\exp(i\eta U_t^{0,\theta_1}) - \mathbb{E}[\exp(i\eta U_t^{\lambda,\theta_2})] + \mathbb{E}[\exp(i\eta U_t^{0,\theta_2})]\right)^2 g_p(\eta)d\eta.
\end{aligned}$$

In our definition above, t is an arbitrary large time to be determined later. Our goal now is to bound I_3 from below and bound I_2 and I_{1j} from above.

Lower bound for I_3. We bound I_3 from below as follows:

$$\begin{aligned}
I_3 \geq &\int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta\bar{U}_t^{0,\theta_1})] - \mathbb{E}[\exp(i\eta\bar{U}_t^{0,\theta_2})]\right)^2 g_p(\eta)d\eta\\
&- \left(\int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta\bar{U}_t^{0,\theta_1})] - \mathbb{E}[\exp(i\eta U_t^{0,\theta_1})]\right)^2 g_p(\eta)d\eta\right.\\
&\left.+ \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta\bar{U}_t^{0,\theta_2})] - \mathbb{E}[\exp(i\eta U_t^{0,\theta_2})]\right)^2 g_p(\eta)d\eta\right)
\end{aligned}$$

Now, by Lemma 5.4, there exists a constant c_1 such that the first term is bounded from below by $c_1|\theta_1-\theta_2|^2$. In view of Proposition 3.3, the other terms are upper bounded by Ce^{-ct}. Choosing t large enough, we can thus bound I_3 from below by

$$I_3 \geq \frac{c_1}{2}|\theta_1-\theta_2|^2.$$

Upper bound for I_{1j}. The term I_{1j} also represents a distance between the solution of (5.6) and its stationary version, so we want to use Proposition 3.3 again. For that, we need the drift $-\xi.+\lambda b_\xi(.)$ to verify assumption $\mathbf{A_1}$. It was already checked in [24] that this is the case when b_ξ, $\partial_\lambda b_\xi$ are bounded by 1, and λ is small enough ($\lambda < m_\Theta(1-\epsilon)$ for some $\epsilon>0$). So we clearly have $I_{1j}\leq Ce^{-ct}$. Setting t large enough we get that

$$I_{1j} \leq \frac{c_1}{16}|\theta_1-\theta_2|^2.$$

Upper bound for I_2. We refer here to the upper bound for I_2 from the proof of Proposition 6.4 in [24]. Namely, it was shown that (see equation (6.17) in [24]) that

$$I_2 \leq C\mathbb{E}\left[\left(|U_t^{0,\theta_2}-U_t^{0,\theta_1}|+|\Delta_R(U_t)|\right)\left(|U_t^{\lambda,\theta_2}-U_t^{0,\theta_2}|+|\Delta_R(U_t)|\right)+|\Delta_R(U_t)|\right],$$

where $\Delta_R(U_t)$ are the rectangular increments defined by

$$\Delta_R(Y_t)=U_t^{\lambda,\theta_1}-U_t^{0,\theta_1}-U_t^{\lambda,\theta_2}+U_t^{0,\theta_2}.$$

Notice that when $\theta=H$ or $\theta=\sigma$, we have $\Delta_R(U_t)=0$. So we get

$$I_2 \leq \lambda^2 C\mathbb{E}\left(|U_t^{0,\theta_2}-U_t^{0,\theta_1}|^2\|\partial_\lambda U^\theta\|_\infty^2\right).$$

It was shown in [24] that $\|\partial_\lambda U^\theta\|_\infty^2 \leq c_{m_\Xi,M_\Xi,\epsilon}$ when b_ξ and $\partial_\lambda b_\xi$ are both bounded by 1 and $\lambda \leq m_\Xi(1-\epsilon)$. We end up with

$$I_2 \leq C_{m_\Xi,M_\Xi,\epsilon}\lambda^2\mathbb{E}|U_t^{0,\theta_2}-U_t^{0,\theta_1}|^2.$$

Now if $\theta=H$, the increments of the stationary Ornstein-Uhlenbeck process are shown to be bounded in [14, Lemma A.1]. Via a similar analysis for the non-stationary Ornstein-Uhlenbeck process, one can show that

$$\mathbb{E}\left|U_t^{0,\theta_2}-U_t^{0,\theta_1}\right|^2 \leq C|\theta_2-\theta_1|^2,$$

where C does not depend on t. When $\theta=\sigma$, we have

$$\begin{aligned}\mathbb{E}\left|U_t^{0,\theta_2}-U_t^{0,\theta_1}\right|^2 &= \mathbb{E}\left(\int_0^t(\theta_2-\theta_1)e^{-t+u}dB_u\right)^2\\ &\leq C(\theta_2-\theta_1)^2.\end{aligned}$$

So, finally, our bound on I_2 becomes

$$I_2 \leq C_{m_\Xi,M_\Xi,\epsilon}\lambda^2|\theta_2-\theta_1|^2.$$

Finally, we choose λ small enough so that:

$$I_2 \leq \frac{c_1}{16}|\theta_1-\theta_2|^2.$$

To finish the proof, it remains to combine the bounds we obtained for I_3,I_2 and I_{1j} into (5.9). □

5.3 Numerical results

In this section, we provide numerical examples to illustrate our main results. We only deal with the one-dimensional Ornstein-Uhlenbeck model defined in (5.1) that starts from 0, as we think it already raises numerous questions about the numerical implementation. We explain at the end how one might extend our approach to more general SDEs of the form (1.1).

Simulated data. The fractional Ornstein-Uhlenbeck process can not be simulated exactly. Therefore, we have chosen a discretization procedure thanks to a simple first order Euler scheme with very small time-step $\underline{\gamma}$ (namely $\underline{\gamma} = 10^{-3}$) in order to get a good approximation of the process.

Let us recall that in the setting of equation (1.1), the simple Euler scheme converges strongly to the true SDE. The same is tue for the augmented process X^{θ_0}, since we have:

$$|X^{\theta_0}_{.} - X^{\theta_0,\underline{\gamma}}_{.}| \leq C \sum_{i=0}^{q} |Y^{\theta_0}_{.+ih} - Y^{\theta_0,\underline{\gamma}}_{.+ih}|,$$

and the right hand side is bounded from above thanks to [24, Proposition 3.7 (i)]. Furthermore, taking the expectation in the previous result leads to a marginal control of the L^2-distance between the Euler scheme and the true SDE (with same fBm) of order $\underline{\gamma}^H$ (independently of the horizon). This confirms that our approximation of the observations is reasonable when H is not too small.

Let us also recall that the fractional Brownian motion can be simulated through the Davies-Harte method. Therefore, up to the approximation of the true SDE, we now assume that we are given a sequence $(Y_{k\underline{\gamma}})_{k\geq 0}$, where $(Y_t)_{t\geq 0}$ is a solution to (1.1) with a given θ_0. Then we create from this path a subsequence of observations $(X_{t_k})_{k=1}^n$ as defined in (2.6). Here we take $q = 2$ and consider the linear transformation to be the simple increments:

$$\ell^i(Y^{\theta_0}_{.}, ..., Y^{\theta_0}_{.+ih}) = Y^{\theta_0}_{.+ih} - Y^{\theta_0}_{.} \quad i = 1, 2.$$

Furthermore, we consider the time-steps t_k to be of the form $t_k = k\gamma$, which means in particular that we assume γ to be of the form $k_0\underline{\gamma}$ with $k_0 \in \mathbb{N}_*$ (namely $k_0 = 100$). Recall that the observations $(X^{\theta_0}_{t_k})_{k=1}^n$ contain three observed paths (2.6).

For the rest of this section, we will use the following terminology:

- One-dimensional case: This is when we only use the first path of X^{θ_0} (i.e Y^{θ_0}) as observations. This means that we are only interested in estimating one parameter (either the drift, the diffusion or the hurst parameter) and we assume the other two known in advance. There are thus three choices to consider.
- Two-dimensional case: This when we want to estimate two parameters and therefore take the first two paths of X^{θ_0} as observations. There are also three choices to consider.
- Three-dimensional case: This when we want to estimate all the parameters and therefore consider all the paths included in X^{θ_0}. Here, there is only one setting to consider.

Computation of the distance between the empirical measures. The theoretical construction of an estimator like (2.10) involves in practice the computation of the distance d between the empirical measures of the observed process and the stationary distribution, for a distance $d \in \mathcal{D}_p$. We describe how to compute this kind of distance.

Whenever d is the Wasserstein distance, an explicit computation of the distance is possible if the observed process is 1-dimensional. However, as we explained in Section 1, using the observations of Y^{θ_0} only allows us to estimate one parameter. If we want to estimate more, we need to add increments of the process into the observations. This gives birth to a augmented process X^{θ_0} that lives in a higher dimension. Unfortunately, in higher dimensions, the computation of the Wasserstein distance requires approximation/optimization methods that are highly expensive complexity-wise and are out of the scope of this paper. In this context and as in [24], it seems to be numerically simpler to work with an approximation of the distance $d_{CF,p}$ (defined in (2.1)), which is also used for our analysis of the rate of convergence. Such an approximation can be obtained by a standard discretization of the integral which appears in (2.1).

Minimization of the distance with respect to θ. Eventually the implementation of our estimation procedure relies on an optimization problem in order to compute the argmin in (2.10). More specifically, in the Ornstein-Uhlenbeck case, we already have an expression of the stationary distribution (5.3). Furthermore, we also know how to express the covariance between the process and its increments (5.4). Since the stationary distribution is Gaussian $\mu_\theta \sim N(0, \Sigma_\theta)$, we thus have all the information we need to simulate it.

In this case, we want to minimize:

$$F : \theta \to d(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{t_k}^{\theta_0}}, \mu_\theta). \tag{5.10}$$

In the one-dimensional case, the computation of F is quite fast, so we simply use the Python library *scipy.optimize* to minimize F.

In higher dimensions, when we use more than one path of X, we adapt the same technique as [24] described mainly in Equation (7.6). Taking $d = d_{CF,p}$, the idea is to write the functional F as:

$$F(\theta) = d_{CF,p}(\mu, \mu_\theta) = \mathbb{E}[|\mu(f_\Phi) - \mu_\theta(f_\Phi)|^2]. \tag{5.11}$$

where $\mu = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{X_{t_k}^{\theta_0}}$, $f_\phi(x) = e^{i\langle x,\phi\rangle}$ and Φ is random variable that has g_p as density (see (2.2)).

Writing F this way allows us to perform a stochastic gradient descent algorithm. In fact, the gradient ∇F is formally obtained as:

$$\nabla F = \mathbb{E}\Lambda(\theta, \Phi), \tag{5.12}$$

where

$$\begin{aligned}
\Lambda(\theta, \phi) &= \partial_\theta \left(|\mu(f_\phi) - \mu_\theta(f_\phi)|^2\right) \\
&= 2\left(\frac{1}{n}\sum_{k=0}^{n-1}\cos(\langle\phi, X_{t_k}^{\theta_0}\rangle) - e^{-\frac{1}{2}\phi^T\Sigma_\theta\phi}\right)\nabla\left(e^{-\frac{1}{2}\phi^T\Sigma_\theta\phi}\right) \\
&= -\left(\frac{1}{n}\sum_{k=0}^{n-1}\cos(\langle\phi, X_{t_k}^{\theta_0}\rangle) - e^{-\frac{1}{2}\phi^T\Sigma_\theta\phi}\right)\left(e^{-\frac{1}{2}\phi^T\Sigma_\theta\phi}\right)\nabla\left(\phi^T\Sigma_\theta\phi\right).
\end{aligned}$$

Thus, our gradient algorithm reads:

$$\theta_{n+1} = \theta_n - \eta_n\Lambda(\theta_n, \Phi_{n+1}), \tag{5.13}$$

where $(\eta_n)_n$ is a sequence of positive steps and $(\Xi_n)_n$ is a sequence of i.i.d random variables with common distribution g_p.

Simulation of the variable Φ. To perform a stochastic gradient descent, we want to replace the expression of the gradient above by simply $\Lambda(\theta, \Phi)$ where we recall that Φ has density

$$g_p(\phi) = c_p(1 + |\phi|^2)^{-p}.$$

Since g_p has a spherical form, Φ can be simulated using the spherical coordinates and the inversion method. More specifically, the simulation of Φ can be reduced to simulating independent one-dimensional variables that either have a uniform distribution or if not, can be simulated through the inverse method.

The two-dimensional case (i.e when we use the first two paths of X^{θ_0} as observations) is described in [24, Section 7], basically Φ can be simulated as $(R\cos(\Theta), R\sin(\Theta))$ where R has density $q_p(r) = \frac{2(p-1)r}{(1+r^2)^p}$ and Θ has a uniform distribution on $[0, 2\pi]$. The reader can check that their approach can be generalized to any dimension. For instance, in the three-dimensional case, Φ can be simulated as $[R\cos(U)\sin(V), R\sin(U)\sin(V), R\cos(V)]$ where R has density $q_p^1(r) = c\frac{r^2}{(1+r^2)^p}$, V has density $q^2(v) = c\sin(v)$ on $[0, \pi]$ and U has uniform density on $[0, 2\pi]$.

Numerical Illustrations. Let us now turn to some numerical tests. We consider the process Y given by (1.1) where the drift $b_\xi(.) = -\xi.$ (i.e the OU process). We assume θ to sit in a compact interval. The assumptions $\mathbf{A_0}$ and $\mathbf{A_1}$ are clearly satisfied, where $\mathbf{I_w}$ follows from Proposition 5.1. Moreover, Lemma 5.4 proves that $\mathbf{I_s}$ is satisfied when we are only interested in estimating one parameter. Using the strategy described before, we get a discretely observed path of X with the following parameters:

$$\begin{aligned} \theta_0 = (\xi_0, H_0, \sigma_0) &= (2, 0.5, 0.7) \\ \underline{\gamma} &= 10^{-3} \\ \gamma &= 10^{-1} \\ n &= 10000 \\ q &= 3. \end{aligned}$$

We start with the one-dimensional case (Figure 1). In this case, we can use the Wasserstein distance as its implementation is quite fast. We compute the functional F defined in (5.10) and minimize it using the Python function *minimize* (from the *scipy.optimize* library). We compute the minimum of F over many trials, which allows us to plot statistics like the mean and the variance.

We then move on to estimating two parameters, assuming the last one is known (Figure 2). In this case, we implement the stochastic descent described above. We let our gradient descent run until it reaches 1000 iterations. Since the parameters ξ, H and σ are not comparable, we decided to plot the loss function

$$\mathrm{Loss}(\theta) = \frac{1}{3}\sum_{i=1}^{3} \frac{|\theta^i - \theta_0^i|}{|\theta^{0,i} - \theta_0^i|},$$

where θ^0 denotes the initial point in our algorithm and θ^i is the i-th coordinate of θ.

Finally, we estimate all the parameters. In all experiments, we take $\theta^0 = [\xi_0 = 2, H_0 = 0.7, \sigma_0 = 0.5]$. And in the last one, we test another initial point $\theta_2^0 = [1, 0.5, 0.6]$.

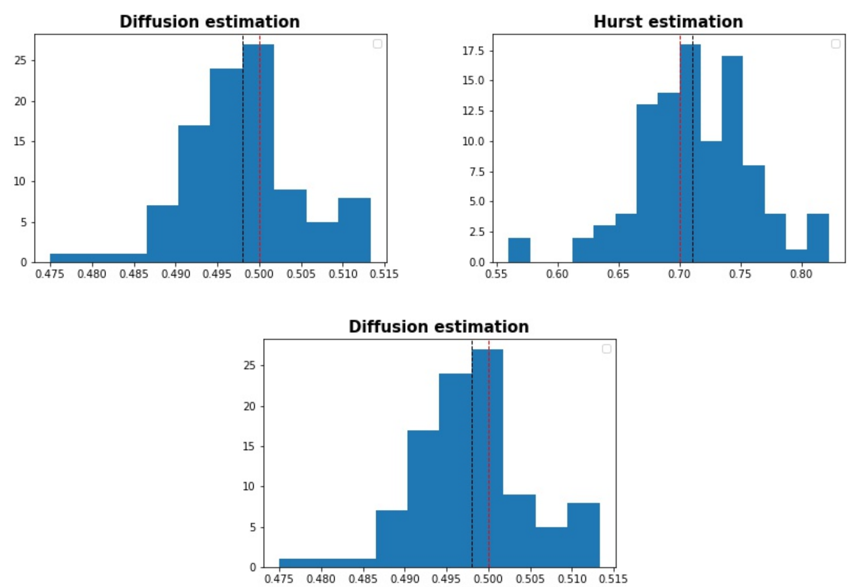

Figure 1: Histograms for the estimation of each parameter separately. The true parameters are highlighted in *red* and the empirical mean of the estimators in *black*. Left: estimation of σ, the empirical variance is $\sim 10^{-5}$. Right: estimation of H, the empirical variance is $\sim 10^{-3}$. Bottom: estimation of ξ, the empirical variance is $\sim 10^{-3}$.

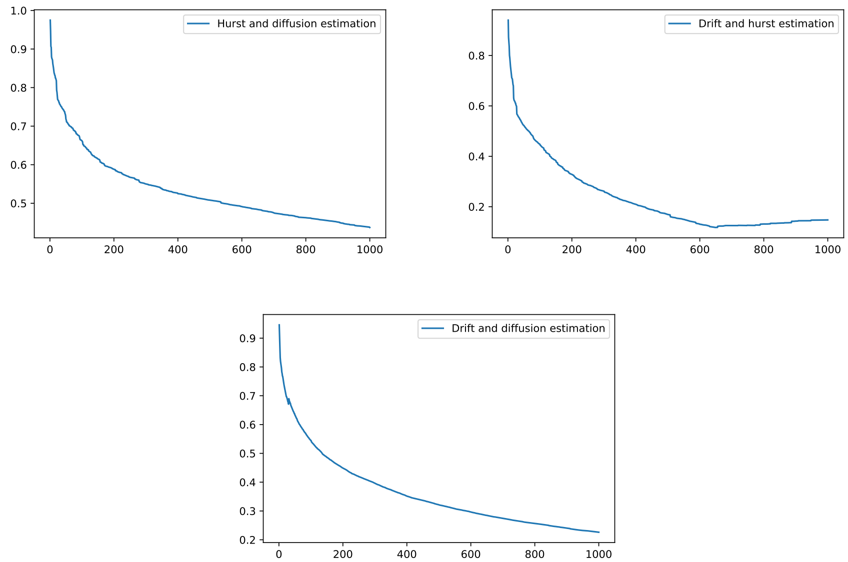

Figure 2: The evolution of the loss function with respect to the number of iterations in the gradient descent, the size of the sample is fixed to 10000. The true parameters are $\xi = 2,\ H = 0.7,\ \sigma = 0.5$. Left: the estimated parameters are $H = 0.678,\ \sigma = 0.57$. Right: the estimated parameters are $\xi = 1.82,\ H = 0.72$. Bottom: the estimated parameters are $\xi = 1.838, \sigma = 0.529$.

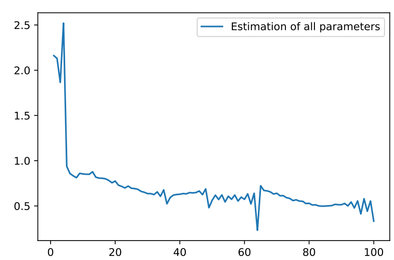

Figure 3: The evolution of the loss function with respect to the number of iterations in the gradient descent, the size of the sample is fixed to 10000. The true parameters are $\xi = 2,\ H = 0.7,\ \sigma = 0.5$. The estimated parameters are $\xi = 1.5,\ H = 0.8,\ \sigma = 0.4$.

Discussion. In the one-dimensional case (see Figure 1), we manage to get reasonably accurate estimators of the parameters. While the drift estimator seems to suffer from a slight bias, the mean and variance of all the estimators over the hundred trials are satisfying. Notice that we always have $\frac{variance}{mean} \lesssim 0.01$.

In the two-dimensional case (Figure 2), we used a mini-batch procedure. That is, in equation (5.13), we replace the random simulated term $\Lambda(\theta_n, \Phi_{n+1})$ by an average over $m = 100$ simulations. The mini-batch procedure is known to reduce the randomness of the algorithm. The results displayed in Figure 2 show that our approach yields a good convergence of the estimate to the true parameter θ. However, our simulations also reveal that the gradient gets flat near the true parameter. Therefore the algorithm moves very slowly after a large number of iterations. This is the reason why we decided to stop the algorithm after 1000 iterations. Also, we noticed that the partial derivatives $\Lambda(\theta, \Phi)^i$ have different magnitudes and decided to adapt our steps η_n accordingly.

In the three-dimensional case (Figure 3), the gradient gets flat very soon and the algorithm moves very slowly after a few iterations. Here, we stopped the algorithm after 100 iterations as the three-dimensional case is quite expensive complexity-wise.

Overall, these attempts to combine our statistical procedure with gradient descent algorithms would certainly need to be enriched with deeper considerations. So this task is deferred to a subsequent paper for the sake of conciseness.

Beyond the OU model. When the stationary distribution is unknown, one can approximate $\mu_\theta(f_\xi)$ in (5.11) by $\frac{1}{N}\sum_{k=0}^{N-1} f(X^{\theta,\gamma}_{k\bar\gamma})$ for some large N and small $\bar\gamma$ as we explained in Section 4. In this case, we can write the gradient Λ as in [24, Eq (7.6)]:

$$\begin{aligned}\Lambda(\theta,\phi) = 2(\mu_\theta-\mu)\,\cos(\langle\phi,.\rangle)\,\rho_\theta(-\sin\langle\phi,.\rangle)\\ +\,2(\mu_\theta-\mu)\,\sin(\langle\phi,.\rangle)\,\rho_\theta(\cos\langle\phi,.\rangle),\end{aligned} \tag{5.14}$$

where for any function $g:\mathbb{R}\to\mathbb{R}$, each component of $\rho_\theta(g(\langle\phi,.\rangle))$ reads:

$$\rho_\theta(g(\langle\phi,.\rangle))^i = \frac{1}{N} g(\langle\phi, X^{\theta,\gamma}_{k\bar\gamma}\rangle)\langle\phi,\partial_{\theta^i}X^{\theta,\gamma}_{k\bar\gamma}\rangle.$$

Therefore, the question is how to simulate paths of the process $\partial_{\theta^i}X^{\theta,\gamma}_{.}$. In [24] the authors handle the case when θ^i is the drift parameter ξ and explain how the process can be simulated recursively as

$$\partial_\xi Y^{\theta,\gamma}_{(k+1)\bar\gamma} = \partial_\xi Y^{\theta,\gamma}_{k\bar\gamma} + \bar\gamma\left(\partial_\xi b_\xi(Y^{\theta,\gamma}_{k\bar\gamma}) + \nabla_z b_\xi(Y^{\theta,\gamma}_{k\bar\gamma})\partial_\xi Y^{\theta,\gamma}_{k\bar\gamma}\right).$$

The same technique can be used when θ^i is the diffusion parameter σ:

$$\partial_\sigma Y^{\theta,\gamma}_{(k+1)\bar\gamma} = \partial_\sigma Y^{\theta,\gamma}_{k\bar\gamma} + \bar\gamma\nabla_z b_\xi(Yk\bar\gamma^{\theta,\gamma})\partial_\sigma Y^{\theta,\gamma}_{k\bar\gamma} + \left(B_{(k+1)\bar\gamma} - B_{k\bar\gamma}\right).$$

Finally, in order to compute $\partial_H Z_.$ in the same way, one needs to compute $\partial_H B_.$ which is not an obvious task. For instance, using the representation of the fBm on an interval [23, Equation 5.8]

$$B_t = \int_0^t K_H(t,s)dW_s,$$

one cannot simply differentiate the kernel K_H with respect to H to get $\partial_H B_.$. In [20], it is shown that for all $t\geq 0$, B_t is almost surely infinitely differentiable with respect to H. But since we consider ergodic increments, we need a result that states: almost surely, for all $t\geq 0$, B_t is infinitely differentiable with respect to H. So maybe in this case, one should look into derivative-free methods for optimization (e.g [8]), where one can perform a gradient descent without having to compute the gradient. Or maybe the $d_{CF,p}$ distance is not appropriate and we should consider other distances. This opens new potential problems when considering models beyond the fractional Ornstein-Uhlenbeck, which we leave for future investigation.

A Regularity in the Hurst parameter

In this section, we borrow some results from our companion paper [14, Section 4 and Section 5] that state the regularity in the Hurst parameters of continuous and discrete ergodic means. Recall that the fractional OU process is defined by (5.1), and let us denote by $\bar U^\theta$ the stationary fractional OU process.

Lemma A.1. *Let $\mathcal{H}$ be a compact subset of $(0,1)$. For each $H\in\mathcal{H}$, let Y^H be the solution of (2.5) with the drift b satisfying $\mathbf{A_1}$. Let $\varpi\in(0,1)$ and $p\geq 1$. There exists a random variable $\mathbf{C}$, independent of ξ and σ, with a finite moment of order p such that almost surely, for any $t\geq 0$ and any $\theta_1=(\xi,H_1,\sigma),\theta_2=(\xi,H_2,\sigma)\in\mathcal{H}$,*

$$\frac{1}{t+1}\int_0^{t+1}\left|Y^{\theta_1}_s - Y^{\theta_2}_s\right|^2 ds \leq \mathbf{C}|H_1-H_2|^\varpi.$$

Proof. For any $\theta \in \Theta$, the process $\sigma^{-1}Y^{\theta}_{\cdot}$ is solution to the SDE

$$\sigma^{-1}Y_t^{\theta} = \sigma^{-1}Y_0 + \int_0^t \tilde{b}_{\xi}(\sigma^{-1}Y_s^{\theta})ds + B_t,$$

with $\tilde{b}_{\xi}(x) = \sigma^{-1}b_{\xi}(\sigma\cdot)$. We have $\tilde{b}_{\xi} \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Xi, \mathbb{R}^d)$ and since σ lives in the compact set Σ, on can check that $\tilde{b}_{\xi}$ still satisfies (2.3) and (2.4). Looking through the proof of [14, Theorem 4.5], a comparison with the stationary OU process $\bar{U}$ gives

$$\frac{1}{t+1}\int_0^{t+1} \left|\sigma^{-1}(Y_s^{\theta_1} - Y_s^{\theta_2})\right|^2 ds \leq C|\bar{U}_0^{(1,H_1,Id)} - \bar{U}_0^{(1,H_2,Id)}|^2 + \frac{1}{t+1}\int_0^{t+1} |\bar{U}_s^{(1,H_1,Id)} - \bar{U}_s^{(1,H_2,Id)}|^2 ds.$$

We can now apply [14, Proposition 4.2] with $t' = t = 0$ and [14, Proposition 4.4] with $H' = K = K'$ and $t' = t$ to get that there exists a random variable $\mathbf{C}_1$ (independent of ξ and σ) with a finite moment of order p such that

$$\frac{1}{t+1}\int_0^{t+1} \left|\sigma^{-1}(Y_s^{\theta_1} - Y_s^{\theta_2})\right|^2 ds \leq \mathbf{C}_1|H_1 - H_2|^{\varpi}.$$

Since $\left|\sigma^{-1}(Y_s^{\theta_1} - Y_s^{\theta_2})\right| \geq |\sigma^{-1}|\left|(Y_s^{\theta_1} - Y_s^{\theta_2})\right|$, dividing by $|\sigma^{-1}|$ and taking the supremum over Σ, we get the desired result by setting $\mathbf{C} = |\sigma^{-1}|^{-1}\mathbf{C}_1$. ☐

Lemma A.2. *Let $\mathcal{H}$ be a compact subset of $(0,1)$, $\varpi \in (0,1)$, and $p \geq 1$. There exists $\gamma_0 > 0$ such that for $\gamma \in (0, \gamma_0)$, there exists a random variable $\mathbf{C}$ with a finite moment of order p such that almost surely, for all $t, t' \geq 0$ and all $H_1, H_2 \in \mathcal{H}$,*

$$\frac{1}{t+1}\int_0^{t+1} |\bar{U}_{s_\gamma}^{(1,H_1,Id)} - \bar{U}_{s_\gamma}^{(1,H_2,Id)}|^2 ds \leq \mathbf{C}|H_1 - H_2|^{\varpi},$$

where s_γ denotes the leftmost point in a time-discretisation of step γ.

Proof. Apply [14, Proposition 5.1], with $t' = t$ and $H' = K' = K$ to get that

$$\frac{1}{t+1}\int_0^{t+1} |\bar{U}_{s_\gamma}^{(1,H_1,Id)} - \bar{U}_{s_\gamma}^{(1,H_2,Id)}|^2 ds \leq \mathbf{C}|H_1 - H_2|^{\varpi} + C\mathbb{E}|\bar{U}_0^{(1,H_1,Id)} - \bar{U}_0^{(1,H_2,Id)}|^2.$$

Now apply [14, Proposition 4.2] with $t = t' = 0$ to get the desired result. ☐

Lemma A.3. *Let $\mathcal{H}$ be a compact subset of $(0,1)$. Let $\varpi \in (0,1)$ and $p \geq 1$. There exists $\gamma_0 > 0$ such that for $\gamma \in (0, \gamma_0]$, there exists a random variable $\mathbf{C}$ with a finite moment of order p such that almost surely, for any $N \in \mathbb{N}^*$ and any $\theta = (\xi, H_1, \sigma)$, $\theta_2 = (\xi, H_2, \sigma) \in \Theta$,*

$$\frac{1}{N}\sum_{k=1}^N \left|Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma}\right|^2 \leq \mathbf{C}|H_1 - H_2|^{\varpi}.$$

Proof. For any $\theta \in \Theta$, the process $\sigma^{-1}Y^{\theta,\gamma}_{\cdot}$ is solution to the SDE

$$\sigma^{-1}Y_t^{\theta,\gamma} = \sigma^{-1}Y_0 + \int_0^t \tilde{b}_{\xi}(\sigma^{-1}Y_{s_\gamma}^{\theta,\gamma})ds + B_t^H,$$

with $\tilde{b}_{\xi}(x) = \sigma^{-1}b_{\xi}(\sigma\cdot)$. We have $\tilde{b}_{\xi} \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Xi, \mathbb{R}^d)$ and since σ lives in the compact set Σ, one can check that $\tilde{b}_{\xi}$ still satisfies (2.3) and (2.4). Looking through the proof of [14, Eq (5.5)], a comparison with the stationary OU process $\bar{U}$ gives

$$\frac{1}{N}\sum_{k=0}^N \left|\sigma^{-1}(Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma})\right|^2 \leq C\Bigg(\frac{1}{N}\sum_{k=0}^N \left|\bar{U}_{j\gamma}^{(1,H_1,Id)} - \bar{U}_{j\gamma}^{(1,H_2,Id)}\right|^2 + \left|\bar{U}_0^{(1,H_1,1)} - \bar{U}_0^{(1,H_2,Id)}\right|^2 + \frac{1}{N\gamma}\int_0^{N\gamma} \left|U_s^{(1,H_1,Id)} - U_s^{(1,H_2,Id)}\right|^2 ds\Bigg).$$

The regularity of the second term in the right-hand side is given by [14, Proposition 4.2] and the regularity of the third term is given by [14, Theorem 4.5]. To bound the first term, we apply Lemma A.2. To conclude the proof, we notice that $|\sigma^{-1}(Y^{\theta_1,\gamma}_{k\gamma} - Y^{\theta_2,\gamma}_{k\gamma})| \geq |\sigma^{-1}||(Y^{\theta_1}_{k\gamma} - Y^{\theta_2}_{k\gamma})|$, divide by $|\sigma^{-1}|$ and take the supremum over Σ. □

B Continuity and Tightness results

In Proposition B.1 and Proposition B.2, we prove that the solutions Y^θ and $Y^{\theta,\gamma}$ to (2.5) and (4.1) and their ergodic means have finite moments uniformly in time and θ. Finally, in Proposition B.3, we state a result on the the regularity of the ergodic means in θ.

Proposition B.1. *Assume $\mathbf{A_0}$ and $\mathbf{A_1}$. Let Y^θ_t be the unique solution of (2.5). Let $p > 1$. Then the following inequalities hold true:*

(i) $\sup_{t\geq 0}\sup_{\theta\in\Theta} \mathbb{E}\left[|Y^\theta_t|^p\right] < \infty.$

(ii) $\mathbb{E}\left(\sup_{t\geq 0}\sup_{\theta\in\Theta} \frac{1}{t}\int_0^t |Y^\theta_s|^2 ds\right)^p < \infty.$

(iii) $\mathbb{E}\left(\sup_{t\geq 0}\sup_{\theta\in\Theta} \frac{1}{n}\sum_{k=0}^{n-1} |Y^\theta_{kh}|^2 ds\right)^p < \infty.$

Proof. Throughout the proof, C will denote a constant that do not depend on θ or t and that may change from line to line. Observe that when the supremum is taken only over ξ, the proof is already done in [24, Proposition A.1]. The proofs of all three items are based on a comparison with fractional OU processes defined in (5.1).

For the proof of (i), by [12, p 725], a comparison with the stationary fractional OU process $U^{(1,H,\sigma)}$ yields that there exist constants $c_1, c_2 > 0$ independent of ξ such that,

$$|Y_t - \bar{U}^{(1,H,\sigma)}_t|^p \leq e^{-2c_1 t}|Y_0|^p + c_2\int_0^t e^{-2c_2(t-s)}(1 + |\bar{U}^{(1,H,\sigma)}_s|^p)ds.$$

Moreover, since $U^{(1,H,\sigma)}$ is a Gaussian process, for any $t \geq 1$, we have $\mathbb{E}|\bar{U}^{(1,H,\sigma)}_t|^p \lesssim (\mathbb{E}|\bar{U}^{(1,H,\sigma)}_t|^2)^{p/2}$. By (5.3), we know that $\mathbb{E}|\bar{U}^{(1,H,\sigma)}_t|^2 = \sigma^2 H\Gamma(2H)$. Therefore

$$\sup_{t\geq 0}\sup_{\theta\in\Theta}\mathbb{E}|Y^\theta_t|^p \leq C(1 + \sup_{t\geq 0}\sup_{\theta\in\Theta}\mathbb{E}|\bar{U}^{(1,H,\sigma)}_t|^p) < \infty.$$

For the proof of (ii), we follow the steps of the proof of Proposition A.1 in [24] (see equation (A.6) and what follows), to get that for all $t \geq 0$,

$$\begin{aligned}\frac{1}{t}\int_0^t \sup_{\theta\in\Theta}|Y^\theta_s|^2 ds &\leq C\frac{1}{t}\int_0^t \sup_{\theta=(1,H,\sigma)\in\Theta}|U^\theta_s|^2 ds\\ &\leq C\frac{1}{t}\int_0^t \sup_{\theta=(1,H,\sigma)\in\Theta}|\sigma||U^{(1,H,Id)}_s|^2 ds.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{1}{t}\int_0^t \sup_{\theta}|Y^\theta_s|^2 ds &\leq C\frac{1}{t}\int_0^t \sup_{H\in\mathcal{H}}|U^{(1,H,Id)}_s|^2 ds\\ &\leq C\left(\sup_{H\in\mathcal{H}}\frac{1}{t}\int_0^t |U^{(1,H,Id)}_s - U^{(1,1/2,Id)}_s|^2 ds + \frac{1}{t}\int_0^t |U^{(1,1/2,Id)}_s|^2 ds\right).\end{aligned}$$

Moreover, by Lemma A.1 (applied for $Y^\theta \equiv U^{(1,H,Id)}$), for any $\beta \in (0,1)$, there exists a random variable $\mathbf{C}$ with a finite moment of order p such that for any $t \geq 1$

$$\frac{1}{t}\int_0^t \sup_\theta |Y_s^\theta|^2 ds \leq C\left(\mathbf{C}\sup_{H\in\mathcal{H}}|H-\frac{1}{2}|^\beta ds + \frac{1}{t}\int_0^t |\bar{U}_s^{(1,1/2,Id)}|^2 ds + 1\right),$$

We also know by [22, Section 1.3.2.2] that $\frac{1}{t}\int_0^t |\bar{U}_s^{(1,1/2,Id)}|^2 ds$ converges to $\mathbb{E}|\bar{U}_0^{(1,1/2,Id)}|^2 = \sigma^2 H\Gamma(2H)$ as t goes to ∞. It follows that

$$\mathbb{E}\left(\sup_{t\geq 0}\frac{1}{t}\int_0^t \sup_\theta |Y_s^\theta|^2 ds\right)^p < \infty.$$

The proof of (iii) can be done in the exact same way by transcribing all the integrals to discrete sums and using Lemma A.2. □

Proposition B.2. *Assume* $\mathbf{A_0}$ *and* $\mathbf{A_1}$*. Let* $Y^{\theta,\gamma}_{\cdot}$ *be the unique solution of* (4.1)*. Then there exists* $\gamma_0 > 0$ *such that for any* $p > 1$ *we have*

(i) $\sup_{\theta\in\Theta,\gamma\in(0,\gamma_0)} \limsup_{N\to\infty} \mathbb{E}\left|Y_{N\gamma}^{\theta,\gamma}\right|^p < \infty.$

(ii) For $\gamma \in (0,\gamma_0]$, $\mathbb{E}\left(\sup_{\theta\in\Theta}\sup_{N\geq 1}\frac{1}{N}\sum_{k=0}^{N-1}|Y_{k\gamma}^{\theta,\gamma}|^2 ds\right)^p < \infty.$

Proof. Note that the same results are proven in [24, Proposition A.4] when Θ only represents the range of the parameter ξ. With this in mind, as in Proposition B.1, the proof of (i) is based on comparisons with the discrete Ornstein-Uhlenbeck process, which has finite moments uniformly in θ. The proof (ii) is the same as the proof of (ii) in Proposition B.1 and is based on a comparison with the discrete OU process and Lemma A.3. □

Proposition B.3. *Let* θ_1 *and* θ_2 *in* Θ*, consider the respective solutions* $\{Y_t^{\theta_1}\}_t$ *and* $\{Y_t^{\theta_2}\}_t$ *of* (2.5)*. Assume hypothesis* $\mathbf{A_0}$ *and* $\mathbf{A_1}$ *are satisfied. And assume that the exponent* r *in the sub-linear growth of* b_ξ *in* (2.4) *satisfies* $r \leq 2$*. Let* $p \geq 1$*, and* $\varpi \in (0,1)$*, there exists a positive random variable* $\mathbf{C}$ *that has a finite p-moment, such that almost surely for all* $\theta_1, \theta_2 \in \Theta$ *and for all* $t \geq 1$*,*

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1} - Y_s^{\theta_2}|^2 ds \leq \mathbf{C}\left(1 \wedge |\theta_1-\theta_2|^\varpi\right). \tag{B.1}$$

Furthermore, there exists γ_0 *such that for* $\gamma \in (0,\gamma_0]$ *similar results hold for the occupation measures of the Euler approximation* M^θ*, that is almost surely, for any* $\theta_1, \theta_2 \in \Theta$ *and any* $N \geq 1$*,*

$$\frac{1}{N}\sum_0^N |Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma}|^2 ds \leq \mathbf{C}\left(1 \wedge |\theta_1-\theta_2|^{\frac{\varpi}{2}}\right). \tag{B.2}$$

Proof. In the proof, we denote by C a constant independent of time and θ that may change from line to line. Similarly, $\mathbf{C}$ will denote a positive random variable that has a finite p-moment and that do not depend on θ and may change from line to line.

We will focus on the proof of (B.1). The proof of (B.2) can be obtained using the same tools, plus some discrete computations and the discrete analogue of the results we borrow from [24] and [14]. Up to introducing pivot terms, we can consider three different cases,

$$\begin{aligned}
&1)\ \theta_1 = (\xi_1, H, \sigma), \theta_2 = (\xi_2, H, \sigma)\\
&2)\ \theta_1 = (\xi, H_1, \sigma),\ \theta_2 = (\xi, H_2, \sigma)\\
&3)\ \theta_1 = (\xi, H, \sigma_1),\ \theta_2 = (\xi, H, \sigma_2).
\end{aligned}$$

In the first case, we have by [24, Equation 5.32] that

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1}-Y_s^{\theta_2}|^2 ds \le C|\xi_1-\xi_2|^2\left(\sup_{\theta\in\Theta}\frac{1}{t}\int_0^t |Y_s^{\theta}|^{2r} ds\right),$$

where r is the exponent in the sub-linear growth assumption on b_ξ. Since $r\le 1$, we have

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1}-Y_s^{\theta_2}|^2 ds \le C|\xi_1-\xi_2|^2\left(1+\sup_{\theta\in\Theta}\frac{1}{t}\int_0^t |Y_s^{\theta}|^{2} ds\right),$$

It follows from the uniform bound on the moments of Y_t^θ in Proposition B.1(ii) that there exists a random variable $\mathbf{C}$ with finite moment of order p such that

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1}-Y_s^{\theta_2}|^2 ds \le \mathbf{C}|\xi_1-\xi_2|^2. \tag{B.3}$$

In the second case, let $\varpi\in(0,1)$. By Lemma A.1, we get

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1}-Y_s^{\theta_2}|^2 ds \le \mathbf{C}|H_1-H_2|^{\varpi}. \tag{B.4}$$

In the third case, the idea here is to compare the process Y with the fractional OU processes $U^{(1,H,\sigma_1)}$ and $U^{(1,H,\sigma_2)}$ defined by (5.1). For $s\ge 1$, we have

$$\begin{aligned}
&\frac{\partial}{\partial s}|Y_s^{\theta_1}-Y_s^{\theta_2}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)|^2\\
&\quad = 2\langle Y_s^{\theta_1}-Y_s^{\theta_2}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right), b(Y_s^{\theta_1})-b(Y_s^{\theta_2})+\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)\rangle\\
&\quad \le -c_1|Y_s^{\theta_1}-Y_s^{\theta_2}|^2-c_2|U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2+c_3|Y_s^{\theta_1}-Y_s^{\theta_2}||U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|,
\end{aligned}$$

where the last inequality follows from $\mathbf{A_1}$. Next, we apply Young's inequality to get

$$\begin{aligned}
&\frac{\partial}{\partial s}|Y_s^{\theta_1}-Y_s^{\theta_2}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)|^2\\
&\quad \le -c_1|Y_s^{\theta_1}-Y_s^{\theta_2}|^2-c_2|U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2+c_3|U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2\\
&\quad \le -c_1|Y_s^{\theta_1}-Y_s^{\theta_2}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)|^2+c_2|U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2.
\end{aligned}$$

We can now apply Grönwall's lemma to get

$$|Y_s^{\theta}-Y_s^{\theta}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)|^2 \le C\int_0^s e^{-(s-u)}|U_u^{(1,H,\sigma_1)}-U_u^{(1,H,\sigma_2)}|^2 du.$$

Jensen's inequality yields that:

$$|Y_s^{\theta_1}-Y_s^{\theta_2}-\left(U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}\right)|^2 \le C\int_0^s e^{-(s-u)}|U_u^{(1,H,\sigma_1)}-U_u^{(1,H,\sigma_2)}|^2 du,$$

and therefore

$$|Y_s^{\theta_1}-Y_s^{\theta_2}|^2 \le C\left(|U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2+\int_0^s e^{-(s-u)}|U_u^{(1,H,\sigma_1)}-U_u^{(1,H,\sigma_2)}|^2 du\right).$$

Then, using Fubini's theorem, it comes that

$$\begin{aligned}
\frac{1}{t}\int_0^t |Y_s^{\theta_1}-Y_s^{\theta_2}|^2 ds &\le \frac{C}{t}\int_0^t |U_u^{(1,H,\sigma_1)}-U_u^{(1,H,\sigma_2)}|^2\int_u^t \mathbf{1}_{[0,s]}e^{-(s-u)} ds du\\
&\quad +\frac{1}{t}\int_0^t |U_s^{(1,H,\sigma_1)}-U_s^{(1,H,\sigma_2)}|^2 ds\\
&\le \frac{C}{t}\int_0^t |U_u^{(1,H,\sigma_1)}-U_u^{(1,H,\sigma_2)}|^2 du.
\end{aligned}$$

Now, observe that

$$U_s^{(1,H,\sigma_1)} - U_s^{(1,H,\sigma_2)} = (\sigma_1 - \sigma_2) U_s^{(1,H,Id)}.$$

Since $U_s^{(1,H,Id)}$ has finite moments uniformly in θ (recall (5.3) and that $U^{(1,H,Id)}$ is a Gaussian process), it follows that

$$\frac{1}{t}\int_0^t |Y_s^{\theta_1} - Y_s^{\theta_2}|^2 ds \leq \mathbf{C}|\sigma_1 - \sigma_2|^2.$$

where $\mathbf{C}$ has a finite moment of order p.

□

C Proof of Proposition 4.1

The proof follows the same steps as the proof of Lemma 2.2. Let $\theta = (\xi, H, \sigma) \in \Theta$. We will first prove that almost surely, the random measure $\frac{1}{t}\int_0^t \delta_{X_s^{\theta,\gamma}} ds$ converges in law to μ_θ^γ as $t \to \infty$. This implies that $\frac{1}{t}\int_0^t \delta_{X_s^{\theta,\gamma}} ds$ converges to μ_θ in the Prokhorov distance. To extend this result to distances d in $\mathcal{D}_2$ (i.e dominated by the 2-Wasserstein distance), we use the fact that the 2-Wasserstein distance is dominated by the Prokorov distance d_P as follows (see [9, Theorem 2]):

$$d(\frac{1}{t}\int_0^t \delta_{X_s^{\theta,\gamma}} ds, \mu_\theta) \leq C_p \sup_{t\geq 0}\left(\max(\frac{1}{t}\int_0^t |X_s^{\theta,\gamma}|^2 ds \vee \mathbb{E}|X_t^{\bar{\theta},\gamma}|^2) + 1\right) d_P(\frac{1}{t}\int_0^t \delta_{X_s^{\theta,\gamma}} ds, \mu_\theta)$$

By definition of the process $X^{\theta,\gamma}$, we have that

$$\left(\frac{1}{t}\int_0^t |X_s^{\theta,\gamma}|^2 ds \vee \mathbb{E}|X_t^{\bar{\theta},\gamma}|^2\right) \leq C_q \sum_{i=0}^{q}\left(\frac{1}{t}\int_0^t |Y_{s+ih}^{\theta,\gamma}|^2 ds \vee \mathbb{E}|\bar{Y}_{s+ih}^{\theta,\gamma}|^2\right) \tag{C.1}$$

Therefore, we conclude thanks to Proposition B.2 that in the present case, the convergence in law is equivalent to the convergence for the 2-Wasserstein distance. Similarly to Section 3.2, we consider a family of probability measures on the set of càdlàg functions for which the identification of the limit will be easier, namely $\{\pi_t = \frac{1}{N}\sum_{k=0}^{N-1} \delta_{X_{k\gamma+.}^{\theta,\gamma}}\}_{t\geq 0}$. We first prove that the family is tight and then identify the limit as the discretely stationary law of the augmented process $\bar{X}^{\theta,\gamma}$.

For tightness, we have to prove the following two points (see *e.g* [3, Theorem 13.2]

1. $\forall T > 0, (\mu_T^{(N)})$ defined by

$$\mu_T^N = \frac{1}{N}\sum_{k=0}^{N-1} \delta_{\{\sup_{t\in[0,T]}|X_{k\gamma+t}^{\theta,\gamma}|\}}$$

   is a.s. a tight sequence.

2. For every $\eta > 0$,

$$\limsup_{\delta\to 0}\limsup_{n\to+\infty} \frac{1}{N}\sum_{k=0}^{N-1} \delta_{\{\omega'_T(X_{k\gamma+.}^{\theta,\gamma},\delta)\geq\eta\}} = 0 \quad \text{a.s.}$$

   with

$$w'_T(x,\delta) = \inf_{\{t_i\}}\left\{\max_{i\leq r}\sup_{s,t\in[t_i,t_{i+1})}|x_t - x_s|\right\}$$

   where the infimum extends over finite sets $\{t_i\}$ satisfying:

$$0 = t_0 < t_1 < \cdots < t_r = T \quad \text{and} \quad \inf_{i\leq r}(t_i - t_{i-1}) \geq \delta.$$

Since the process has only jumps at times $n\gamma$ with $n \in \mathbb{N}, \omega'_T\left(X^{\theta,\gamma}, \delta\right) = 0$ when $\delta < \gamma$. It follows that the second point is obvious. Then, let us prove the first point. By definition of $\widehat{X}$, for $k \geq N$, we get

$$|X^{\theta,\gamma}_{k\gamma}|^2 \leq C_q \sum_{i=0}^{q} |Y^{\theta,\gamma}_{k\gamma+ih}|^2.$$

In the proof of [24, Proposition 2], we have

$$\sup_{t\in[0,T]} |Y^{\theta,\gamma}_{k\gamma+t}|^2 = \sup_{k\in\{N,\dots,N+\lfloor T/\gamma\rfloor\}} |\widehat{Y}_{k\gamma}|^2 \leq |Y^{\theta,\gamma}_{N\gamma}|^2 + C\left(1 + \sum_{l=N}^{N+\lfloor T/\gamma\rfloor-1} |\hat{B}_{(l+1)\gamma} - \hat{B}_{l\gamma}|^2\right).$$

Thus, if $V(x) = |x|^2$, we deduce that

$$\begin{aligned}
\mu^N_T(V) &\leq C_q\left(\frac{1}{n}\sum_{k=0}^{n} |Y^{\theta,\gamma}_{k\gamma}|^2 + 1 + \frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=k}^{k+\lfloor T/\gamma\rfloor-1} |\hat{B}_{(l+1)\gamma} - \hat{B}_{l\gamma}|^2\right)\\
&\leq C_q\left(\frac{1}{n}\sum_{k=0}^{n} |Y^{\theta,\gamma}_{k\gamma}|^2 + 1 + \lfloor\frac{T}{\gamma}\rfloor \sup_{n\geq 1}\frac{1}{n}\sum_{k=0}^{n+\lfloor T/\gamma\rfloor-1} |\hat{B}_{(k+1)\gamma} - \hat{B}_{k\gamma}|^2\right).
\end{aligned}$$

Using Proposition B.2 and [24, Eq (15)], we conclude that $\sup_{N\geq 1}\mu^{(N)}_T(V) < +\infty$ a.s, which implies that $(\mu^{(N)}_T)_{N\geq 1}$ is a.s. tight on $\mathbb{R}^d$ (see *e.g.* [7, Proposition 2.1.6]).

Now, let $(t_n)_{n\geq 1}$, be an increasing sequence going to $+\infty$ and $\{\frac{1}{t_n}\sum_{k=0}^{t_n-1}\delta_{X^{\theta,\gamma}_{k\gamma+\cdot}}\}_{n\geq 1}$ be a (pathwise) sequence with limiting distribution ρ. We first show that ρ is the law of a stationary process. Let $M \geq 1$, $(u_1,\dots,u_M) \in \mathbb{R}^M$ and $f : \mathbb{R}^{M(q+1))} \mapsto \mathbb{R}$, then for all $T > 1$

$$\begin{aligned}
&\frac{1}{t_n}\sum_{k=0}^{t_n-1}\left(f\left(X^{\theta,\gamma}_{u_1+k\gamma},\dots,X^{\theta,\gamma}_{u_M+k\gamma}\right) - f\left(X^{\theta,\gamma}_{u_M+k\gamma+T\gamma},\dots,X^{\theta,\gamma}_{u_M+k\gamma+T\gamma}\right)\right)\\
&\quad= \frac{1}{t_n}\left(\sum_{k=0}^{T-1}(f\left(X^{\theta,\gamma}_{u_1+k\gamma},\dots,X^{\theta,\gamma}_{u_M+k\gamma}\right) - \sum_{k=t_n}^{t_n+T-1}(f\left(X^{\theta,\gamma}_{u_1+k\gamma},\dots,X^{\theta,\gamma}_{u_M+k\gamma}\right)\right).
\end{aligned}$$

The last term converges to 0 when $t_n \to \infty$ a.s since f is bounded. Therefore, ρ is the law of a stationary process. Let us now prove that ρ is the law of $\bar{X}^{\theta,\gamma}$.

A process $x_t = (y_t, z^1_t, .., z^q_t)$ has the law of $\bar{X}^{\theta,\gamma}$ if $x_t = x_{k\gamma}$ for $t \in [k\gamma,(k+1)\gamma]$, and

$y_\cdot - y_0 - \int_0^{\cdot_\gamma} b_\xi(y_u)du$ has the law of a $\sigma B_{\cdot_\gamma}$ where B has Hurst parameter H;

$z^i_\cdot - \ell^i\left(\int_0^{\cdot_\gamma} b_\xi(y_u)du,\dots,\int_0^{(\cdot+ih)_\gamma} b_\xi(y_u)du\right)$ has the law of $\sigma\ell^i(B_{\cdot_\gamma},\dots,B_{(\cdot+ih)_\gamma})$ for all $i \in [\![1,q]\!]$,

where for all $t \geq 0$, $t_\gamma = \gamma\lfloor t/\gamma\rfloor$. Let us define

$$G_\gamma(x_\cdot) = \begin{pmatrix} y_\cdot - y_0 - \int_0^{\cdot_\gamma} b_\xi(y_u)du \\ z^1_\cdot - \ell^1\left(\int_0^{\cdot_\gamma} b_\xi(y_u)du, \int_0^{(\cdot+h)_\gamma} b_\xi(y_u)du\right) \\ \cdot \\ \cdot \\ \cdot \\ z^q_\cdot - \left(\int_0^{\cdot_\gamma} b_\xi(y_u)du,\dots,\int_0^{(\cdot+qh)_\gamma} b_\xi(y_u)du\right) \end{pmatrix},$$

and

$$\mathbf{B}_{\cdot} = \left(\sigma B_{\cdot_\gamma}, \ldots, \sigma \ell^q (B_{\cdot_\gamma}, \ldots, B_{(\cdot+qh)_\gamma})\right).$$

In other words, we have to prove that

$$\gamma \circ G_\gamma^{-1} \text{ is the law of } \mathbf{B}_{\cdot},$$

Since G_γ is continuous for the u.s.c topology, we have

$$\gamma \circ G_\gamma^{-1} = \lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} \delta_{G_\gamma(X^{\theta,\gamma}_{s+\cdot})} ds.$$

Let $T > 0$ and $F : \mathcal{C}([0,T], \mathbb{R}^{d(q+1)}) \mapsto \mathbb{R}$ be a bounded measurable function. We want to show that

$$\lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} F\left(G_\gamma(X^{\theta,\gamma}_{s+\cdot})\right) ds = \mathbb{E}F(\mathbf{B}_{\cdot}).$$

It is sufficient to check the convergence for the finite dimensional distributions. For any $N \geq 1$, $\{u_1, \ldots, u_N\} \in \mathbb{R}^N$ and measurable and bounded $f : \mathbb{R}^{d(q+1)N} \mapsto \mathbb{R}$, we want to show that

$$\lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} f(G_\gamma(X^{\theta,\gamma}_{s+u_1}), \ldots, G(X_{s+u_N}))ds = \mathbb{E}f\left(\mathbf{B}_{u_1}, \ldots, \mathbf{B}_{u_N}\right).$$

By construction, we have

$$G_\gamma(X^{\theta,\gamma}_{s+\cdot}) = \begin{pmatrix} \sigma(B_{(s+\cdot)_\gamma} - B_{s_\gamma}) \\ \sigma\ell^1(B_{(s+\cdot)_\gamma} - B_{s_\gamma}, B_{(s+h+\cdot)_\gamma} - B_{s_\gamma}) \\ \cdot \\ \cdot \\ \cdot \\ \sigma\ell^q(B_{(s+\cdot)_\gamma} - B_{s_\gamma}, \ldots, B_{(s+qh+\cdot)_\gamma} - B_{s_\gamma}) \end{pmatrix}.$$

Therefore, we can write

$$\begin{aligned} f(G_\gamma(X^{\theta,\gamma}_{s+u_1}), \ldots, G_\gamma(X^{\theta,\gamma}_{s+u_N})) &= \tilde{f}\left(\{B_{(s+u_1+ih)_\gamma} - B_{s_\gamma}\}_{i=0,\ldots,q}, \ldots, \{B_{(s+u_N+ih)_\gamma} - B_{s_\gamma}\}_{i=0,\ldots,q}\right) \\ f\left(\mathbf{B}_{u_1}, \ldots, \mathbf{B}_{u_N}\right) &= \tilde{f}\left(\{B_{(u_1+ih)_\gamma}\}_{i=0,\ldots,q}, \ldots, \{B_{(u_N+ih)_\gamma}\}_{i=0,\ldots,q}\right), \end{aligned}$$

where $\tilde{f} = f \circ \lambda$ for some linear transformation λ, so $\tilde{f}$ is still a bounded measurable function. By the ergodicity of the increments of the fractional Brownian motion ([6, Eq 5]), we have

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} f(G_\gamma(X^{\theta,\gamma}_{s+u_1}), \ldots, G_\gamma(X^{\theta,\gamma}_{s+u_N}))ds \\ &= \lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} \tilde{f}\left(\{B_{(s+u_1+ih)_\gamma} - B_s\}_{i=0,\ldots,q}, \ldots, \{B_{(s+u_N+ih)_\gamma} - B_{s_\gamma}\}_{i=0,\ldots,q}\right) ds \\ &= \mathbb{E}\tilde{f}\left(\{B_{(u_1+ih)_\gamma}\}_{i=0,\ldots,q}, \ldots, \{B_{(u_N+ih)_\gamma}\}_{i=0,\ldots,q}\right) = \mathbb{E}f(\mathbf{B}_{u_1}, \ldots, \mathbf{B}_{u_N}). \end{aligned}$$

Hence, $\rho \circ G_\gamma^{-1}$ has the law of $\mathbf{B}$.

D Proof of Lemma 5.2 when $\theta = (\xi, H)$

D.1 The case $H > 1/2$

We will prove that $f : \Theta \times \mathcal{H} \to \mathrm{Im}(f)$ is bijective for h smaller than a theoretical h_0. Let (a, b) be a vector belonging to $Im(f)$. We will show that the following equation has a unique solution in (θ, H):

$$\begin{aligned} a &= H\Gamma(2H)\xi^{-2H} \\ b &= 2H\Gamma(2H)\xi^{-2H} \frac{\sin(\pi H)}{\pi} \int_0^\infty \cos(\xi h x) \frac{x^{1-2H}}{1+x^2} dx \end{aligned}$$

which is equivalent to solving:

$$\xi = (\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}$$
$$b = 2a\frac{\sin(\pi H)}{\pi}\int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\frac{x^{1-2H}}{1+x^2}dx,$$

For the rest of this section, we will focus on the function

$$g_a(H) = \sin(\pi H)\int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\frac{x^{1-2H}}{1+x^2}dx,$$

and we will show that for all possible values of a, g_a is bijective and therefore there exists a unique H such that $g_a(H) = \frac{\pi b}{2a}$, which implies the uniqueness of ξ by the equation $\theta = (\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}$.

Now let us differentiate g_a. For all $H > 1/2$

$$\begin{aligned} g_a'(H) = &\underbrace{\pi\cos(\pi H)\int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\frac{x^{1-2H}}{1+x^2}dx}_{:=g'_{a,1}(H)} \\ &\underbrace{-\sin(\pi H)[(\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}]'h\int_0^\infty x\sin((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\frac{x^{1-2H}}{1+x^2}dx}_{g'_{a,2}(H)} \\ &\underbrace{-2\sin(\pi H)\int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx}_{g'_{a,3}(H)}. \end{aligned} \tag{D.1}$$

Remark D.1. *In the expression above of g_a', we see that we need $H > 1/2$ for the integral $\int_0^\infty x\sin((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\frac{x^{1-2H}}{1+x^2}dx$ to be absolutely convergent. For the case $H < 1/2$, we need to check the convergence of $\int_0^\infty \frac{sin(Cx)}{x^\alpha}$ for $1 < \alpha < 1$ if we want to differentiate g_a.*

We will show that for any $H > 1/2$, $g_a'(H) > 0$ which proves that g_a is bijective since it is continuous. Let us first handle the term $g'_{a,2}$. Notice that $H \to (\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}$ is $C^1(\mathcal{H}, \mathbf{R})$ and therefore is bounded and has a bounded derivative. So we can bound $g'_{a,2}$ in absolute value by a $C_1 h$.

Let us now handle the term $g'_{a,3}$, we first split the integral between $[0,1]$ and $[1,\infty)$.

$$\begin{aligned} g'_{a,3}(H) = &-2\sin(\pi H)\int_0^1 \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx \\ &- 2\sin(\pi H)\int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

Notice that for h small enough, the first term in $g'_{a,3}$ is positive. In fact, we will rely on this term to compensate all the negative terms in our lower-bound of g_a'. For all ϵ, there exists h_1 such that for $h \leq h_1$, we have:

$$\begin{aligned} g'_{a,3}(H) \geq &-2\sin(\pi H)\int_0^1 (1-\epsilon)\log(x)\frac{x^{1-2H}}{1+x^2}dx \\ &- 2\sin(\pi H)\int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

Since $H \in \mathcal{H}$, a compact set, we can further lower bound $g'_{a,3}$ by

$$\begin{aligned} g'_{a,3}(H) \geq &-\epsilon C_2 - 2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx \\ &- 2\sin(\pi H)\int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx. \end{aligned}$$

For the moment, the lower-bound we have on g_a' is composed of the following terms

$$\begin{aligned} g_a'(H) \geq & -hC_1 - \epsilon C_2 + g_{a,1}' \\ & \underbrace{-2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx}_{=:g_{31}'(H)} \\ & \underbrace{-2\sin(\pi H)\int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx}_{=:g_{a,32}'(H)}. \end{aligned}$$

The idea is to show that the sum of the terms $g_{31}', g_{a,32}'$ and $g_{a,1}'$ is positive so that we can choose h_0 such that the sum of all terms stays positive. For all $H \in \mathcal{H}$ with $H > 1/2$, we have

$$|-\sin(\pi H)\int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}hx)\log(x)\frac{x^{1-2H}}{1+x^2}dx| < \sin(\pi H)\int_1^\infty \log(x)\frac{x^{1-2H}}{1+x^2}dx,$$

and

$$|\pi\cos(\pi H)\int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}hx)\frac{x^{1-2H}}{1+x^2}dx| < -\pi\cos(\pi H)\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx.$$

Therefore, the sum of the terms $g_{31}', g_{a,32}'$ and $g_{a,1}'$ is strictly lower-bounded by

$$\begin{aligned} g_{31}'(H) + g_{a,32}'(H) + g_{a,1}'(H) > & -hC_1 - \epsilon C_2 + -2\sin(\pi H)\int_0^1 \log(x)\frac{x^{1-2H}}{1+x^2}dx \\ & -2\sin(\pi H)\int_1^\infty \log(x)\frac{x^{1-2H}}{1+x^2}dx \\ & +\pi\cos(\pi H)\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx. \end{aligned} \tag{D.2}$$

Recall that $H\Gamma(2H)\xi^{-2H} = 2H\Gamma(2H)\frac{\sin(\pi H)}{\pi}\xi^{-2H}\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx$. Thus we have $\int_0^\infty \frac{x^{1-2H}}{1+x^2}dx = \frac{\pi}{2\sin(\pi H)}$ and $\int_0^\infty -2\log(x)\frac{x^{1-2H}}{1+x^2}dx = \frac{-\pi^2\cos(\pi H)}{2(\sin(\pi H))^2}$ (by differentiating). Therefore, the lower bound in (D.2) is actually equal to 0.

In conclusion, we have shown that for all $H \in \mathcal{H}$ satisfying $H > 1/2$,

$$g_a'(H) \geq -hC_1 - \epsilon C_2 + C_3, \tag{D.3}$$

where $C_1, C_2, C_3 > 0$. Choose h and ϵ small enough such that $-hC_1 - \epsilon C_2 + C_3 > 0$ to conclude the proof.

D.2 The case $H < 1/2$

Using the Laplace transform, one can show that $\int_0^\infty \frac{\sin(x)}{x^\alpha}dx = \frac{\Gamma(\alpha/2)\Gamma(1-\alpha/2)}{2\Gamma(\alpha)}$ for $\alpha \in (0,2)$. So we can now differentiate g_a even for $H < 1/2$.

The arguments used to obtain the lower-bound (D.1) on g_a' are still valid even when $H < 1/2$. The difference here is that the term "$\cos(\pi H)$" in $g_{a,1}'$ is positive and therefore we want to bounde $g_{a,1}'$ from below.

For $H \in \mathcal{H}$ with $H < 1/2$, we write as in Section D.1:

$$g_a'(H) = \underbrace{\pi \cos(\pi H) \int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) \frac{x^{1-2H}}{1+x^2} dx}_{=: g_{a,1}'(H)}$$

$$\underbrace{- \sin(\pi H)[(\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}}]' h \int_0^\infty x \sin((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) \frac{x^{1-2H}}{1+x^2} dx}_{=: g_{a,2}'(H)}$$

$$\underbrace{-2 \sin(\pi H) \int_0^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) \log(x) \frac{x^{1-2H}}{1+x^2} dx}_{=: g_{a,3}'(H)}.$$

As in Section D.1, we have $g_{a,2}' \geq -hC$. But now, we need different bounds on $g_{a,1}'$ and $g_{a,3}'$. We split the two integrals in $g_{a,1}'$ and $g_{a,3}'$ between $(0,1)$ and $(1,\infty)$. Notice that on $(0,1)$, the term $\cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)$ goes to 1 as $h \to 0$ uniformly in x. So for $\varepsilon \in (0,1)$, there exists h_1 such that for $h \leq h_1$, we have

$$\begin{aligned} g_{a,1}'(H) + g_{a,2}'(H) + g_{a,3}'(H) \geq -hC - \varepsilon C + \pi \cos(\pi H) \int_0^1 \frac{x^{1-2H}}{1+x^2} dx - 2\sin(\pi H) \int_0^1 \log(x) \frac{x^{1-2H}}{1+x^2} dx \\ + \pi \cos(\pi H) \int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) \frac{x^{1-2H}}{1+x^2} dx \\ - 2\sin(\pi H) \int_1^\infty \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) \log(x) \frac{x^{1-2H}}{1+x^2} dx. \end{aligned}$$

Recall that $\int_0^\infty \frac{x^{1-2H}}{1+x^2} dx = \frac{\pi}{2\sin(\pi H)}$ and $\int_0^\infty -2\log(x) \frac{x^{1-2H}}{1+x^2} dx = \frac{-\pi^2 \cos(\pi H)}{2(\sin(\pi H))^2}$ (by differentiating). Therefore, we can simplify the previous bound as

$$\begin{aligned} g_{a,1}'(H) + g_{a,2}'(H) + g_{a,3}'(H) \geq -hC - \varepsilon C + \pi \cos(\pi H) \int_1^\infty [\cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) - 1] \frac{x^{1-2H}}{1+x^2} dx \\ - 2\sin(\pi H) \int_1^\infty [\cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx) - 1] \log(x) \frac{x^{1-2H}}{1+x^2} dx. \end{aligned}$$

Moving all the terms inside the integrals, we are left with:

$$\begin{aligned} g_{a,1}'(H) + g_{a,2}'(H) + g_{a,3}'(H) \geq -hC - \varepsilon C + \int_1^\infty [1 - \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)] \\ \times [2\log(x)\sin(\pi H) - \pi\cos(\pi H)] \frac{x^{1-2H}}{1+x^2} dx. \end{aligned}$$

We split the integral by the sign of $2\log(x)\sin(\pi H) - \pi\cos(\pi H)$ as follows:

$$\begin{aligned} & g_{a,1}'(H) + g_{a,2}'(H) + g_{a,3}'(H) \\ & \geq -hC - \varepsilon C + \int_1^{e^{\frac{\pi\cos(\pi H)}{2\sin(\pi H)}}} [1 - \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)][2\log(x)\sin(\pi H) - \pi\cos(\pi H)] \frac{x^{1-2H}}{1+x^2} dx \\ & \quad + \int_{e^{\frac{\pi\cos(\pi H)}{2\sin(\pi H)}}}^\infty [1 - \cos((\frac{a}{H\Gamma(2H)})^{-\frac{1}{2H}} hx)][2\log(x)\sin(\pi H) - \pi\cos(\pi H)] \frac{x^{1-2H}}{1+x^2} dx. \end{aligned}$$

Since H lives in a compact set, we can choose h small enough so that the first integral (the integral on a bounded interval) is as small as we want it to be. The second term is clearly positive.

Altogether, we have shown that for all $\epsilon > 0$, there exists h_2 such that for all $h < h_2$, for all $H \in \mathcal{H}$ satisfying $H < 1/2$, we have

$$g_a'(H) > -hC_1 - \epsilon C_2 + C_3, \tag{D.4}$$

where $C_1, C_2, C_3 > 0$. This proves the existence of h_0, such that for all $h \leq h_0$, we have $g_a'(H) > 0$, and thus concludes on the injectivity of f when $H < 1/2$.

References

[1] Rachid Belfadli, Khalifa Es-Sebaiy, and Youssef Ouknine. Parameter estimation for fractional Ornstein-Uhlenbeck processes: non-ergodic case. *Preprint arXiv:1102.5491*, 2011.

[2] Corinne Berzin, Alain Latour, and José R. León. Variance estimator for fractional diffusions with variance and drift depending on time. *Electron. J. Stat.*, 9(1):926–1016, 2015.

[3] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[4] Alexandre Brouste and Stefano M. Iacus. Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Comput. Statist.*, 28(4):1529–1547, 2013.

[5] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab.*, 8:no. 3, 14, 2003.

[6] Serge Cohen and Fabien Panloup. Approximation of stationary solutions of Gaussian driven stochastic differential equations. *Stochastic Process. Appl.*, 121(12):2776–2801, 2011.

[7] Marie Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

[8] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394, 2005.

[9] Alison L. Gibbs and Francis E. Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[10] Luca M. Giordano, Maria Jolis, and Lluís Quer-Sardanyons. SPDEs with fractional noise in space: continuity in law with respect to the Hurst index. *Bernoulli*, 26(1):352–386, 2020.

[11] Arnaud Gloter and Marc Hoffmann. Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.*, 35(5):1947–1974, 2007.

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

[13] El Mehdi Haress and Yaozhong Hu. Estimation of all parameters in the fractional Ornstein-Uhlenbeck model under discrete observations. *Stat. Inference Stoch. Process.*, 24(2):327–351, 2021.

[14] El Mehdi Haress and Alexandre Richard. Long time Hurst regularity of fractional SDEs and their ergodic means. *Preprint arXiv:2206.06648*, 2022.

[15] Yaozhong Hu and David Nualart. Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist. Probab. Lett.*, 80(11-12):1030–1038, 2010.

[16] Yaozhong Hu, David Nualart, and Hongjuan Zhou. Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter. *Stat. Inference Stoch. Process.*, 22(1):111–142, 2019.

[17] Yaozhong Hu and Jian Song. Parameter estimation for fractional Ornstein-Uhlenbeck processes with discrete observations. In *Malliavin calculus and stochastic analysis*, volume 34 of *Springer Proc. Math. Stat.*, pages 427–442. Springer, New York, 2013.

[18] Maria Jolis and Noèlia Viles. Continuity with respect to the Hurst parameter of the laws of the multiple fractional integrals. *Stochastic Process. Appl.*, 117(9):1189–1207, 2007.

[19] Maria Jolis and Noèlia Viles. Continuity in the Hurst parameter of the law of the Wiener integral with respect to the fractional Brownian motion. *Statist. Probab. Lett.*, 80(7-8):566–572, 2010.

[20] Stefan Koch and Andreas Neuenkirch. The Mandelbrot–Van Ness fractional Brownian motion is infinitely differentiable with respect to its Hurst parameter. *Discrete Contin. Dyn. Syst. Ser. B*, 24(8):3865–3880, 2019.

[21] K. Kubilius and Y. Mishura. The rate of convergence of Hurst index estimate for the stochastic differential equation. *Stochastic Process. Appl.*, 122(11):3718–3739, 2012.

[22] Kestutis Kubilius, Yuliya Mishura, and Kostiantyn Ralchenko. *Parameter estimation in fractional diffusion models*, volume 8 of *Bocconi & Springer Series*. Bocconi University Press, [Milan]; Springer, Cham, 2017.

[23] David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[24] Fabien Panloup, Samy Tindel, and Maylis Varvenne. A general drift estimation procedure for stochastic differential equations with additive fractional noise. *Electron. J. Stat.*, 14(1):1075–1136, 2020.

[25] B. L. S. Prakasa Rao. *Statistical inference for fractional diffusion processes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2010.

[26] Alexandre Richard. A fractional Brownian field indexed by L^2 and a varying Hurst parameter. *Stochastic Process. Appl.*, 125(4):1394–1425, 2015.

[27] Alexandre Richard and Denis Talay. Noise sensitivity of functionals of fractional Brownian motion driven stochastic differential equations: results and perspectives. In *Modern problems of stochastic analysis and statistics*, volume 208 of *Springer Proc. Math. Stat.*, pages 219–235. Springer, Cham, 2017.

[28] Alexandre Richard and Denis Talay. Lipschitz continuity in the Hurst parameter of functionals of Stochastic Differential Equations driven by fractional Brownian motion. *Preprint arXiv:1605.03475v4*, 2022.

[29] Ciprian A. Tudor and Frederi G. Viens. Statistical aspects of the fractional stochastic calculus. *Ann. Statist.*, 35(3):1183–1212, 2007.

[30] Maylis Varvenne. Concentration inequalities for stochastic differential equations with additive fractional noise. *Electron. J. Probab.*, 24:Paper No. 124, 22, 2019.

[31] Weilin Xiao, Weiguo Zhang, and Weidong Xu. Parameter estimation for fractional Ornstein-Uhlenbeck processes at discrete observation. *Appl. Math. Model.*, 35(9):4196–4207, 2011.