



HAL
open science

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

El Mehdi Haress, Alexandre Richard

► **To cite this version:**

El Mehdi Haress, Alexandre Richard. Estimation of several parameters in discretely-observed Stochastic Differential Equations with additive fractional noise. *Statistical Inference for Stochastic Processes*, In press. hal-04057186v3

HAL Id: hal-04057186

<https://universite-paris-saclay.hal.science/hal-04057186v3>

Submitted on 14 May 2024

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

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

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

El Mehdi Haress^{1*} and Alexandre Richard¹

¹Université Paris-Saclay, CentraleSupélec, MICS and CNRS FR-3487.

*Corresponding author(s). E-mail(s): el-mehdi.haress@centralesupelec.fr;
Contributing authors: alexandre.richard@centralesupelec.fr;

Abstract

We investigate the problem of joint statistical estimation of several parameters for a stochastic differential equation 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, which lies 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 permits to identify the parameters of the model. Under this assumption, we prove consistency results and derive a rate of convergence for the estimator. 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.

Keywords: fractional Brownian motion, parametric estimation, ergodicity

MSC Classification: 60H10 , 60G22 , 60G10 , 62F12 , 37M25

1 Introduction

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

$$Y_t = Y_0 + \int_0^t b_{\xi_0}(Y_s) ds + \sigma_0 B_t, \quad (1)$$

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 $b_\xi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$, 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. [Belfadli et al \(2011\)](#); [Hu and Nualart \(2010\)](#); [Prakasa Rao \(2010\)](#); [Tudor and Viens \(2007\)](#); [Hu et al \(2019\)](#)). 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 [Hu and Nualart \(2010\)](#); [Hu et al \(2019\)](#). In these works, the drift function 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 [Tudor and Viens \(2007\)](#). In addition, the process Y is observed in continuous time. In practical situations though, we only have access to discrete-time observations. Taking into account this constraint, two recent papers [Panloup et al \(2020\)](#); [Hu and Song \(2013\)](#) 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 [Hu and Song \(2013\)](#): while [Hu and Song \(2013\)](#) considers the fractional OU case, [Panloup et al \(2020\)](#) treats general drift functions which satisfy a coercivity assumption.

The diffusion coefficient σ_0 is usually estimated using the quadratic variations of Y , which is possible only when the process is either observed continuously or the step-size goes to zero (i.e. high frequency data), see [Xiao et al \(2011\)](#) and [Berzin et al \(2015\)](#). The Hurst parameter H_0 is also estimated using quadratic variations, see e.g. [Kubilius and Mishura \(2012\)](#), or by a direct access to discrete observations of a fractional Brownian motion path with a step-size that goes to zero as in [Gloter and Hoffmann \(2007\)](#).

When it comes to estimating all the parameters (ξ_0, σ_0, H_0) , we refer to [Brouste and Iacus \(2013\)](#) where the observations are assumed to be made continuously, and [Haress and Hu \(2021\)](#) 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 consider an ergodic setting that allows for (1) to have a stationary distribution for any $\theta_0 \in \Theta$. We work with the assumption that *the stationary distribution of Y identifies the parameters*, as initiated in [Panloup et al \(2020\)](#). However, as illustrated by the authors of [Haress and Hu \(2021\)](#), 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 [Haress and Hu \(2021\)](#), 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 to generalise the

approach presented in [Haress and Hu \(2021\)](#). 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 $(Y_{kh}^{\theta_0})_{k=0, \dots, n+q}$ and consider q linear transformations $\{\ell^i(Y_{kh}^{\theta_0}, \dots, Y_{kh+ih}^{\theta_0})\}_{k=0, \dots, n}$ where $i \in \llbracket 1, q \rrbracket$. Hence, we now have access to $q + 1$ paths, which we use to define the path of a higher-dimensional process X^{θ_0} that we call the augmented process associated to the SDE (1). With access to a path of X^{θ_0} , we construct the estimator of θ_0 by

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} d \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \mu_{\theta} \right), \quad (2)$$

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 obtain a rate of convergence.

In [Haress and Hu \(2021\)](#), the authors provided numerical evidence of the identifiability assumption (i.e the fact that the stationary distribution of Y and its increments identify the parameters). We prove here that in the setting of [Haress and Hu \(2021\)](#), i.e. of a fractional OU process, the aforementioned identifiability assumption holds. Also, as in [Panloup et al \(2020\)](#), we consider two variations of this assumption, a weak one which we will just call the identifiability assumption and a strong one. Moreover, to construct an estimator of the drift parameter ξ , the authors of [Panloup et al \(2020\)](#) 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 enough regularity so that its minimum is attained at the true parameter ξ_0 . Here, in view of estimating all the parameters, we will study the regularity of Y with respect to ξ, σ and H .

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. The fBm is known to be infinitely differentiable w.r.t its Hurst parameter (see [Koch and Neuenkirch \(2019\)](#)). In addition, other functionals of the fBm were considered. In [Jolis and Viles \(2010, 2007\)](#), the law of the integral w.r.t the fBm is proven to be continuous in H ; in [Richard \(2015\)](#), the Hölder continuity in H is obtained for generalised fractional Brownian fields; and in [Giordano et al \(2020\)](#), the law of stochastic heat and wave equations with additive fractional noise is proven to be continuous in H . Let us also mention that in [Richard and Talay \(2022\)](#), the law of functionals of fractional SDEs is proven to be Lipschitz continuous around its Markovian counterpart ($H = \frac{1}{2}$), including irregular functionals such as the law of the first hitting time (see also [Richard and Talay \(2017\)](#) 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 a separate paper [Haess and Richard \(2022\)](#).

In the formula (2), the stationary distribution μ_θ is generally unknown, except in some simple cases like for Ornstein-Uhlenbeck processes. This means that the estimator cannot be implemented. This problem can be solved by considering a numerical approximation of μ_θ via an Euler scheme $Y^{\theta,\gamma}$ of time-step $\gamma > 0$. Given $N + q$ simulated points of the form $(Y_{k\gamma}^{\theta,\gamma})_{k=0,\dots,N+q}$, we consider as before q linear transformations $\{\ell^i(Y_{k\gamma}^{\theta,\gamma}, \dots, Y_{k\gamma+i\gamma}^{\theta,\gamma})\}_{k=0,\dots,N}$, which we use to define a higher-dimensional process $X^{\theta,\gamma}$. We then define the estimator of θ_0 by

$$\hat{\theta}_{N,n,\gamma} = \operatorname{argmin}_{\theta \in \Theta} d \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kk}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta,\gamma}} \right). \quad (3)$$

We prove that $\hat{\theta}_{N,n,\gamma}$ is a strongly consistent estimator of θ_0 and obtain a rate of convergence.

Organisation of the paper.

In Section 2, we first detail the notations and some assumptions in Section 2.1. Then in Section 2.2 we explain how to approximate the invariant measure via an Euler scheme in order to implement the estimator (3). We present the main results for the estimators (2) and (3) in Section 2.3. In Section 2.4, we show that our results can be applied to the fractional Ornstein-Uhlenbeck process and to small perturbations of this process. In Section 3, we prove the strong consistency of the estimator (2) and its rate of convergence. In Section 4, we prove the strong consistency of the estimator (3) and its rate of convergence. In Section 5.1 and 5.2, we prove that the identifiability assumption holds in the case of a fractional Ornstein-Uhlenbeck process for the estimation of two parameters, and in Section 5.3, we exhibit a more general family of SDEs that verifies a stronger identifiability assumption for the estimation of one parameter. We also implement our method and run numerical simulations in Section 5.4. In the Appendix A, we recall some results from our companion paper [Haess and Richard \(2022\)](#). In Appendix B, we prove continuity and tightness results on Y and the solution of the Euler scheme associated to (1). Finally Appendix C is dedicated to the proof of Proposition 2.4.

2 A general procedure

We first give some general notations. Then we state the assumptions on the coefficients of (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 . For any given p , we will consider the p -Wasserstein distance, which is defined for every μ, ν in $\mathcal{M}_1(\mathbb{R}^d)$ as follows:

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

We denote by \mathcal{D}_p the set of distances dominated by the p -Wasserstein distance. As in Panloup et al (2020), we will also work with the distance $d_{CF,p} \in \mathcal{D}_1$ defined for $p > (\frac{d}{2} \vee 1)$ as

$$d_{CF,p}(\mathcal{L}(X), \mathcal{L}(Y)) = \left(\int_{\mathbb{R}^d} (\mathbb{E}[e^{i\langle \chi, X \rangle}] - \mathbb{E}[e^{i\langle \chi, Y \rangle}])^2 g_p(\chi) d\chi \right)^{1/2}, \quad (4)$$

where g_p is the integrable kernel given by

$$g_p(\chi) = c_p(1 + |\chi|^2)^{-p}, \quad (5)$$

and $c_p = (\int_{\mathbb{R}^d} (1 + |\chi|^2)^{-p} d\chi)^{-1}$ is a normalizing constant.

We denote by \mathbb{N}^* the set $\mathbb{N} \setminus \{0\}$ and by C a constant that can change from line to line and that does not depend on time and the parameters ξ, σ, H . When we want to make the dependence of C on some other parameter a explicit, we will write C_a .

The \mathbb{R}^d -fBm will be denoted by B , or by B^H if we need to emphasize on the Hurst parameter H of the process. Whenever we compare, on the same probability space, two fBm with different Hurst parameters B^{H_1} and B^{H_2} , it is assumed that they are built from the same Brownian motion W by the Mandelbrot-Van Ness formula:

$$B_t^{H_i} = \frac{1}{\Gamma(H_i + \frac{1}{2})} \int_{\mathbb{R}} \left((t-s)_+^{H_i - \frac{1}{2}} - (-s)_+^{H_i - \frac{1}{2}} \right) dW_s, \quad t \geq 0, \quad i = 1, 2. \quad (6)$$

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.

- A₀.** 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 .

A₁. $b \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Xi; \mathbb{R}^d)$ 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

$$\langle b_\xi(x) - b_\xi(y), x - y \rangle \leq -\beta|x - y|^2 \text{ and } |b_\xi(x) - b_\xi(y)| \leq K|x - y|. \quad (7)$$

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

$$|b_{\xi_1}(x) - b_{\xi_2}(x)| \leq c(1 + |x|^r). \quad (8)$$

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

$$Y_t^\theta = Y_0 + \int_0^t b_\xi(Y_s^\theta) ds + \sigma B_t, \quad (9)$$

where $Y_0 \in \mathbb{R}^d$ and B is an fBm of Hurst parameter H . Under **A₁**, [Hairer \(2005\)](#) (see also [Panloup et al, 2020](#), Remark 2.4) and the references therein) gives the existence and uniqueness of the invariant measure to (9). We denote by \bar{Y}^θ the unique stationary solution and by ν_θ its marginal distribution. For each $i \in \llbracket 1, q \rrbracket$, let ℓ^i be a linear transformation from $(\mathbb{R}^d)^{i+1}$ to \mathbb{R}^d .

Let us define the following processes for all $i \in \llbracket 1, q \rrbracket$:

$$\begin{aligned} Z_{\cdot}^{i,\theta} &= \ell^i(Y_{\cdot}^\theta, \dots, Y_{\cdot+ih}^\theta) \\ \bar{Z}_{\cdot}^{i,\theta} &= \ell^i(\bar{Y}_{\cdot}^\theta, \dots, \bar{Y}_{\cdot+ih}^\theta) \\ X_{\cdot}^\theta &= (Y_{\cdot}^\theta, Z_{\cdot}^{1,\theta}, \dots, Z_{\cdot}^{q,\theta}) \\ \bar{X}_{\cdot}^\theta &= (\bar{Y}_{\cdot}^\theta, \bar{Z}_{\cdot}^{1,\theta}, \dots, \bar{Z}_{\cdot}^{q,\theta}). \end{aligned} \quad (10)$$

$$\bar{X}_{\cdot}^\theta = (\bar{Y}_{\cdot}^\theta, \bar{Z}_{\cdot}^{1,\theta}, \dots, \bar{Z}_{\cdot}^{q,\theta}). \quad (11)$$

Typical linear transformations considered in applications (see the discussion in Section 2.4) will be the simple increments

$$\ell^i(U^{\theta_0}, \dots, U_{\cdot+ih}^{\theta_0}) = U_{\cdot+ih}^{\theta_0} - U_{\cdot}^{\theta_0}, \quad i = 1, \dots, q. \quad (12)$$

Observe that for all $\theta \in \Theta$ and $i \in \llbracket 1, q \rrbracket$, the processes $\bar{Z}^{i,\theta}$ and \bar{X}^θ are stationary. Denote by μ_θ the law of \bar{X}^θ . For simplicity, we will not write the parameter θ on the processes when θ is the true parameter θ_0 . The triangle inequality yields the following inequalities for all $\theta, \theta' \in \Theta$ and $p > 0$,

$$\begin{aligned} |X_{\cdot}^\theta|^p &\leq C_{p,q} \left(\sum_{i=0}^q |Y_{\cdot+ih}^\theta|^p \right) \\ |X_{\cdot}^\theta - X_{\cdot}^{\theta'}|^p &\leq C_{p,q} \left(\sum_{i=0}^q |Y_{\cdot+ih}^\theta - Y_{\cdot+ih}^{\theta'}|^p \right) \\ |X_{\cdot}^\theta - \bar{X}_{\cdot}^\theta|^p &\leq C_{p,q} \left(\sum_{i=0}^q |Y_{\cdot+ih}^\theta - \bar{Y}_{\cdot+ih}^\theta|^p \right), \end{aligned} \quad (13)$$

where $C_{p,q}$ is a constant that do not depend on θ or θ' . More precisely, $C_{q,p} \sim 2^p q^p L^p$, where L is the maximum of the Lipschitz constants of the applications ℓ^i for $i \in \llbracket 1, q \rrbracket$. This means that upper bounds on X will be obtained by 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 identifiability hypothesis reads as follows:

I_w. For any θ in Θ ,

$$\mu_\theta = \mu_{\theta_0} \iff \theta = \theta_0, \quad (14)$$

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

Remark 2.1. A similar assumption is considered in [Panloup et al \(2020\)](#) based on the stationary distribution of \bar{Y} : assume that $\nu_\theta = \nu_{\theta_0}$ iff $\theta = \theta_0$. Assumption **I_w** is weaker, in the sense that it is satisfied in situations where the assumption from [Panloup et al \(2020\)](#) is not (because we consider the process X^θ instead of Y^θ). Indeed, let $\theta, \theta_0 \in \Theta$ such that $\mu_\theta = \mu_{\theta_0}$. This implies $d_{CF,p}(\mu_\theta, \mu_{\theta_0}) = 0$. Using the definition of $d_{CF,p}$, we have

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

which implies that

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

Hence, if the assumption from [Panloup et al \(2020\)](#) holds, one gets $d_{CF,p}(\nu_\theta, \nu_{\theta_0}) = 0$, i.e. $\nu_\theta = \nu_{\theta_0}$, which then implies that $\theta = \theta_0$. To see that **I_w** is strictly weaker, we refer the reader to the example of the fractional OU process detailed in [Section 2.4](#): in dimension 1, the stationary measure is centred Gaussian and the variance, which depends on the drift coefficient, the diffusion coefficient and H , is not sufficient to identify all 3 parameters. However, considering the stationary measure of Y^θ and its increments permit to retrieve identifiability, see [Proposition 2.11](#).

2.2 Approximation of the invariant measure μ_θ

To approximate μ_θ , we consider the Euler scheme of the stochastic process Y^θ , solution to [\(9\)](#). 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

$$\begin{aligned} Y_{(k+1)\gamma}^{\theta,\gamma} &= Y_{k\gamma}^{\theta,\gamma} + \gamma b_\xi(Y_{k\gamma}^{\theta,\gamma}) + \sigma(\widehat{B}_{(k+1)\gamma} - \widehat{B}_{k\gamma}) \\ Y_t^{\theta,\gamma} &= Y_{t_\gamma}^{\theta,\gamma} = Y_{k\gamma}^{\theta,\gamma} \text{ for } t \in [k\gamma, (k+1)\gamma), \end{aligned} \quad (15)$$

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 [\(9\)](#), 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 $Y^{\theta,\gamma}$ on \widehat{B} , we write $Y^{\theta,\gamma}(\widehat{B})$. We will say that $(\bar{Y}_t^{\theta,\gamma})_{t \geq 0}$ is a discrete stationary solution to (15) if it is a solution of (15) satisfying

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

By (Panloup et al, 2020, Proposition 3.4), there exists $\gamma_0 > 0$ such that for any $\gamma \in (0, \gamma_0]$ and $\theta \in \Theta$, (15) admits a unique stationary solution $\bar{Y}^{\theta,\gamma}$. As in Section 2.1, we define the augmented Euler scheme $X^{\theta,\gamma}$ by

$$X^{\theta,\gamma} = \left(Y^{\theta,\gamma}, \ell^1(Y^{\theta,\gamma}, Y_{+h}^{\theta,\gamma}), \dots, \ell^q(Y^{\theta,\gamma}, \dots, Y_{+qh}^{\theta,\gamma})\right). \quad (16)$$

Similarly, we write $X^{\theta,\gamma}(\widehat{B})$ 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

2.3 Main results

Assume that the solution Y is discretely observed at times $\{kh; k = 1, \dots, n+q\}$ for a fixed time step $h > 0$. Under Assumption **A**₁, we have 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*

$$d\left(\frac{1}{t} \int_0^t \delta_{X_s} ds, \mu_\theta\right) \xrightarrow[t \rightarrow +\infty]{} 0 \quad a.s.,$$

and

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

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, for some observations $X_0^{\theta_0}, \dots, X_{(n-1)h}^{\theta_0}$ and under the identifiability assumption **I**_w, the previous lemma justifies to use the estimator $\hat{\theta}_n$ defined in (2).

In practice, we want to implement the estimator $\hat{\theta}_{n,N,\gamma}$ defined in (3). The following result, coupled with Proposition 4.2(i) justifies the use of this estimator.

Proposition 2.4. *Let $(X_{k\gamma}^{\theta,\gamma})_{k \geq 0}$ be the augmented Euler scheme with time-step γ . Assume that **A**₀ and **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*

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

The proof is postponed to Appendix C.

The first result (Theorem 2.5) states the strong consistency of the estimator (2) under the assumptions $\mathbf{A}_0, \mathbf{A}_1, \mathbf{I}_w$ (see Section 3.3 for the proof).

Theorem 2.5. *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) is a strongly consistent estimator of θ_0 in the following sense:*

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

We also have strong consistency of the estimator (3) under the same assumptions.

Theorem 2.6. *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 . Assume that the exponent r in the sub-linear growth of b_ξ in (8) 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:*

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

We will also establish a rate of convergence of the estimators 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 for $\hat{\theta}_n$ and $\hat{\theta}_{n,N,\gamma}$, which will be proved in Section 3.4 and Section 4.3.

Theorem 2.7. *Assume that \mathbf{A}_0 and \mathbf{A}_1 hold, and that \mathbf{I}_s holds for $p > \frac{\alpha+d(q+1)}{2}$. Then $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$ a.s. and there exists a constant $C > 0$ such that for any $n \in \mathbb{N}^*$,*

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

Theorem 2.8. *Assume that \mathbf{A}_0 and \mathbf{A}_1 hold and that \mathbf{I}_s holds for $p > \frac{\alpha+d(q+1)}{2}$. Assume that the exponent r in the sub-linear growth of b_ξ in (8) satisfies $r \leq 1$. Then $\lim_{n \rightarrow \infty, N \rightarrow \infty, \gamma \rightarrow 0} \hat{\theta}_{n,N,\gamma} = \theta_0$ a.s. Moreover, for any $\varepsilon \in (0, \min(\mathcal{H}))$ and $\varpi \in (0, 1)$, there exists positive constants C, γ_0 such that for any $\gamma \in (0, \gamma_0]$ and $n, N \in \mathbb{N}$ satisfying $N\gamma \geq 1$, we have*

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

Remark 2.9. *We discuss here whether the above rate of convergence can be optimal. Assume first that $H \leq \frac{1}{2}$, $d = 1$ and $\alpha = 2$ (that is the case for the fractional Ornstein-Uhlenbeck (OU) process for example, see (Panloup et al, 2020, Lemma 6.2)). Then by*

Theorem 2.8, using that $N\gamma \leq N$, we have for $\varepsilon \in (0, \frac{1}{2})$

$$\left(\mathbb{E} \left| \hat{\theta}_{n,N,\gamma} - \theta_0 \right|^2 \right)^{\frac{1}{2}} \leq C \left(n^{-\frac{1}{2}} + \gamma^{\min(\mathcal{H})-\varepsilon} + (N\gamma)^{-\frac{1}{4}+\varepsilon} \right).$$

The term $n^{-\frac{1}{2}}$ corresponds to the convergence with respect to the sample size. It matches the CLT rate and generalises (Haress and Hu, 2021, Theorem 4.5) where, for the fractional OU process, the authors construct estimators of ξ, σ and H based on the invariant measure and prove a CLT with respect to the sample size. Moreover, taking $N = \gamma^{-(4\min(\mathcal{H})+1)}$, there is

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

The term $\gamma^{\min(\mathcal{H})-\varepsilon}$ corresponds to the convergence of the Euler scheme defined in (15) and is known to be optimal for strong errors. So the rate cannot be improved in this situation.

Beyond the fractional OU process, we generalise the rate obtained in (Panloup et al, 2020, Theorem 2.13) to an estimation of all the parameters. Finally, the CLT obtained in Haress and Hu (2021) holds for $H \in (0, \frac{3}{4})$ while here the rate we obtained is slower when $H > \frac{1}{2}$: for $\alpha = 2$, the first L^2 error term in the bound of Theorem 2.8 reads $n^{-1+\max(\mathcal{H})} \gg n^{-\frac{1}{2}}$.

Remark 2.10. The proofs of Theorem 2.7 and Theorem 2.8 rely on an upper bound of $\mathbb{E}[d(\mu_{\hat{\theta}_n}, \mu_{\theta_0})^\alpha]$. This involves bounding the quantities in the left-hand side of (13) with $p = \alpha$. In (13) the constant $C_{\alpha,q}$ is of order $q^\alpha L^\alpha$, where L is the biggest Lipschitz constant of the linear transformations. Hence the constant C in Theorem 2.7 depends on $(\ell^i)_{i=1,\dots,q}$ as $C \sim q^{\alpha+1} L^\alpha$. This can be useful in practice when choosing $(\ell^i)_{i=1,\dots,q}$.

2.4 Application to fractional Ornstein-Uhlenbeck-type processes

We first discuss the identifiability assumption for the fractional Ornstein-Uhlenbeck (OU) process, then for a family of small perturbations of the fractional OU process.

Identifiability assumption.

Consider the family of one-dimensional fractional OU processes given by

$$dU^\theta = -\xi U^\theta dt + \sigma dB, \quad U_0^\theta = 0. \quad (17)$$

It is known from the proof of (Hairer, 2005, Proposition 3.12) that the stationary measure of U^θ follows the Gaussian distribution

$$\mathcal{N}(0, \sigma^2 H \Gamma(2H) \xi^{-2H}). \quad (18)$$

In particular, this distribution alone does not permit to identify simultaneously ξ , σ and H . Hence the need to consider increments.

Here, ξ and σ are in compact subsets of $(0, \infty)$ and the linear transformation ℓ^1 takes the form of an increment

$$\ell^1(U^\theta, U_{\cdot+h}^\theta) = U_{\cdot+h}^\theta - U^\theta.$$

We suppose here that θ is of dimension 2, i.e. only two of the three parameters (ξ, σ, H) are unknown. In the following result, we establish that \mathbf{I}_w is verified.

Proposition 2.11. *Consider the fractional Ornstein-Uhlenbeck model defined by equation (17) and assume that one of the parameters ξ , σ or H is known. Let $p > 1$ and let μ_θ denote the stationary measure of $(U^\theta, U_{\cdot+h}^\theta - U^\theta)$. Then there exists $h_0 > 0$ such that for all $h \in (0, h_0)$, we have*

$$\forall \theta_1, \theta_2 \in \Theta, \quad d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2}) = 0 \quad \text{iff} \quad \theta_1 = \theta_2.$$

The proof is given in Section 5.1.

Strong identifiability assumption.

We show that Assumption \mathbf{I}_s holds for some specific examples of (9) and for the distance $d = d_{CF,p}$. Specifically, we consider a family $U^{\lambda,\theta}$ of real-valued processes defined by

$$dU_t^{\lambda,\theta} = \left(-\xi U_t^{\lambda,\theta} + \lambda b_\xi(U_t^{\lambda,\theta}) \right) dt + \sigma dB_t. \quad (19)$$

Under the assumption that the coefficient b_ξ is bounded altogether with its derivatives with respect to ξ and y , one can check that the drift term $b(\cdot) = -\xi \cdot + \lambda b_\xi(\cdot)$ satisfies \mathbf{A}_1 for λ small enough. Therefore, the equation has a unique invariant measure, which is denoted by μ_θ^λ . 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 (17). For the fractional OU process, we simply write μ_θ for the invariant measure. We make the following assumption on the parameters:

$\widetilde{\mathbf{A}}_0$. Assume that ξ , σ and H are one-dimensional parameters and that

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

We shall prove that $U^{\lambda,\theta}$ 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 $\theta \in [m_\Theta, M_\Theta]$.

The first lemma below states that \mathbf{I}_s is satisfied for $U^\theta := U^{0,\theta}$.

Lemma 2.12. *Let θ represent either ξ , σ or H . Assume that $\widetilde{\mathbf{A}_0}$ holds and if $\theta = H$, assume further that*

$$\xi > \sup_{H \in [m_{\mathcal{H}}, M_{\mathcal{H}}]} \exp\left(\frac{\Gamma(2H) + 2H\Gamma'(2H)}{2H\Gamma(2H)}\right) \text{ or } \xi < \inf_{H \in [m_{\mathcal{H}}, M_{\mathcal{H}}]} \exp\left(\frac{\Gamma(2H) + 2H\Gamma'(2H)}{2H\Gamma(2H)}\right). \quad (20)$$

Let $p \geq 1$, then for all $\theta_1, \theta_2 \in [m_{\Theta}, M_{\Theta}]$,

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

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

The previous lemma extends to the solution of Equation (19).

Proposition 2.13. *Let $U^{\lambda, \theta}$ be the process defined by (19) where θ is either ξ , σ or H , and let $p > 3/2$. Assume that $\widetilde{\mathbf{A}_0}$ holds and that b_{ξ} , $\partial_y b_{\xi}$, $\partial_{\xi} b_{\xi}$ are bounded. Moreover, if $\theta = \xi$, assume that $|\partial_{y, \xi}^2 b_{\xi}| \leq 1$ and if $\theta = H$, assume that (20) holds. Then there exists $\lambda_0 = \lambda_0(m_{\Theta}, M_{\Theta}, p) > 0$ and $c_{m_{\Theta}, M_{\Theta}, p} > 0$ such that for any $\lambda \in (0, \lambda_0)$ and any $\theta_1, \theta_2 \in [m_{\Theta}, M_{\Theta}]$,*

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

The proofs of Lemma 2.12 and Proposition 2.13 are given in Section 5.3.

3 Strong consistency and rate of convergence of the estimator $\hat{\theta}_n$

To prove the almost sure convergence, we will use (Panloup et al, 2020, 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 $\{L_v(\theta)\}_v$ indexed by a generic v which lies in a topological space and converges to a generic v_0 . In this Section, we always have $v \equiv n \in \mathbb{N}$, and so $\lim_{v \rightarrow v_0}$ is to be understood as $\lim_{n \rightarrow \infty}$. In Section 4, we will take $v \equiv (\gamma, n, N)$ with $\gamma \rightarrow 0$ and $n, N \rightarrow \infty$, and therefore $\lim_{v \rightarrow v_0}$ will be understood as $\lim_{n \rightarrow \infty, N \rightarrow \infty, \gamma \rightarrow 0}$.

Proposition 3.1 ((Panloup et al, 2020, Proposition 4.3)). *Let Θ be a compact set and $\{\theta \in \Theta \mapsto L_v(\theta)\}_v$ a family of non-negative stochastic processes. Assume that*

- (i) *Almost surely, $\lim_{v \rightarrow v_0} L_v(\theta) = L(\theta)$ uniformly in θ .*
- (ii) *$\theta \mapsto L(\theta)$ is deterministic and continuous in θ .*
- (iii) *For any v , the set $\text{argmin}\{L_v(\theta), \theta \in \Theta\}$ is non-empty.*

Let $\theta_v \in \text{argmin}_{\theta \in \Theta} L_v(\theta)$. If A is a limit point of $\{\theta_v\}_v$, then $A \in \text{argmin}_{\theta \in \Theta} L(\theta)$.

In this Section, we always have $L_v(\theta) = d(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}}, \mu_{\theta})$, with $v \equiv n$ and $v_0 \equiv \infty$.

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

First, we prove two lemmas that state the $L^p(\Omega)$ -continuity with respect to θ of the solution to (9), and the exponential convergence of the law of X^θ (defined in (10)) 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 $p > 0$. Let W be an \mathbb{R}^d -Brownian motion and for any $H \in (0, 1)$, denote by B^H the fBm with underlying noise W (i.e. as in (6)). There exists a constant $C_{T,p} > 0$ such that for any $\theta_1, \theta_2 \in \Theta$,*

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

where Y^{θ_1} (resp. Y^{θ_2}) is the solution to (9) with parameter θ_1 (resp. θ_2) and driving fBm B^{H_1} (resp. B^{H_2}), and both Y^{θ_1} and Y^{θ_2} start from the same initial condition.

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

- 1) $\theta_1 = (\xi, \sigma, H_1)$ and $\theta_2 = (\xi, \sigma, H_2)$
- 2) $\theta_1 = (\xi, \sigma_1, H)$ and $\theta_2 = (\xi, \sigma_2, H)$
- 3) $\theta_1 = (\xi_1, \sigma, H)$ and $\theta_2 = (\xi_2, \sigma, H)$.

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 $t \in [0, T]$,

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

Since b is K -Lipschitz, we get

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

By Jensen's inequality, we have

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

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 $(\mathbb{E}|B_t^{H_1} - B_t^{H_2}|^2)^{p/2}$. Using (Haess and Richard, 2022, Proposition 2.1), the fractional Brownian motion verifies

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

Therefore,

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

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

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

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

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

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 (Panloup et al, 2020, 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 (Panloup et al, 2020, Proposition 3.5) that $C_{T,p}$ does not depend on H or σ . \square

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

$$d(\mathcal{L}(X_t^\theta), \mu_\theta) \leq C e^{-\frac{1}{c}t}. \quad (21)$$

Proof. Since $d \in \mathcal{D}_p$, it comes:

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

Using \mathbf{A}_1 ,

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

It follows that $|Y_t^\theta - \bar{Y}_t^\theta|^2 \leq |Y_0^\theta - \bar{Y}_0^\theta|^2 e^{-2\beta t}$. Hence for $p \geq 2$,

$$\|Y_t^\theta - \bar{Y}_t^\theta\|_{L^p} \leq \|Y_0^\theta - \bar{Y}_0^\theta\|_{L^p} e^{-\beta t} \leq (\|Y_0^\theta\|_{L^p} + \|\bar{Y}_0^\theta\|_{L^p}) e^{-\beta t}. \quad (22)$$

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

$$\|\bar{Y}_0^\theta\|_{L^p} = \lim_{t \rightarrow \infty} \|Y_t^\theta\|_{L^p} \leq \sup_{t \geq 1} \sup_{\theta \in \Theta} \|Y_t^\theta\|_{L^p} < \infty.$$

This concludes the proof. \square

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 (13) 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 $\theta \mapsto d(\mu_\theta, \mu_{\theta_0})$. \square

3.2 Convergence of the contrast: proof of Lemma 2.2

Let $\theta = (\xi, \sigma, H) \in \Theta$. Recall that the Prokhorov distance is defined for any $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$ as

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ for any Borel set } A\},$$

where A^ε is the ε -neighbourhood of A . Convergence in law is equivalent to convergence with respect to the Prokhorov distance.

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 Prokhorov distance d_P as follows (see (Gibbs and Su, 2002, Theorem 2)):

$$\begin{aligned} & d\left(\frac{1}{t} \int_0^t \delta_{X_s^\theta} ds, \mu_\theta\right) \\ & \leq C_p \sup_{t \geq 0} \left(\max\left(\frac{1}{t} \int_0^t |X_s^\theta|^2 ds, \mathbb{E}|\bar{X}_t^\theta|^2\right) + 1 \right) d_P\left(\frac{1}{t} \int_0^t \delta_{X_s^\theta} ds, \mu_\theta\right). \end{aligned} \quad (23)$$

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

$$\max\left(\frac{1}{t} \int_0^t |X_s^\theta|^2 ds, \mathbb{E}|\bar{X}_t^\theta|^2\right) \leq C_q \sum_{i=0}^q \max\left(\frac{1}{t} \int_0^t |Y_{s+ih}^\theta|^2 ds, \mathbb{E}|\bar{Y}_{t+ih}^\theta|^2\right). \quad (24)$$

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 (Panloup et al, 2020, 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 $\{\pi_t^\theta = \frac{1}{t} \int_0^t \delta_{X_{s+}^\theta} ds\}_{t \geq 0}$. The criterion from (Billingsley, 1999, 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 $t_0 \in [0, T]$,

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

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

$$\forall T > 0, \exists \delta > 0, r > 0, \rho > 0 \text{ s.t. } \forall t_0 \in [0, T],$$

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \int_0^t \sup_{u \in [t_0, t_0+\delta]} |X_{s+u}^\theta - X_{s+t_0}^\theta|^r ds \leq C_{r,T} \delta^{1+\rho} \quad \text{a.s.} \quad (25)$$

For $T, r, \delta > 0$, by definition of X^θ and (13), there is

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

By (Panloup et al, 2020, 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 (25).

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 probability measure γ . As in Appendix A.2 of (Panloup et al, 2020, Proposition 3.3), we get that γ is stationary. Let us now prove that γ is the law of \bar{X}^θ . A process $x_t = (y_t, z_t^1, \dots, z_t^q)$ has the law of X^θ if

$$\begin{aligned}
& y. - y_0 - \int_0^\cdot b_\xi(y_u) du \text{ has the law of } \sigma B \text{ where } B \text{ has Hurst parameter } H; \\
& z^i - \ell^i \left(\int_0^\cdot b_\xi(y_u) du, \dots, \int_0^{\cdot+ih} b_\xi(y_u) du \right) \text{ has the law of } \sigma \ell^i(B., \dots, B_{\cdot+ih}) \\
& \text{for all } i \in \llbracket 1, q \rrbracket.
\end{aligned}$$

Let us define

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

and

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

Hence we have to prove that $\gamma \circ G^{-1}$ is the law of $\mathbf{B}.$. This follows the same lines as the end of Appendix A.2 of Panloup et al (2020) and we omit the details.

The same analysis presented in this Section still holds if we replace $\frac{1}{t} \int_0^t |X_s^\theta|^p ds$ by $\frac{1}{n} \sum_{k=0}^{n-1} |X_{kh}^\theta|^p$. This is mostly due to the fact that in Proposition B.1, we also proved that the moments $\frac{1}{n} \sum_{k=0}^{n-1} |X_{kh}^\theta|^p$ are finite uniformly in n , and therefore the right-hand side in (24) is finite even when the integral is replaced by a discrete sum.

3.3 Proof of Theorem 2.5

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

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

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 **I_w** is reduced to $\{\theta_0\}$.

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

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

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

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

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

3.4 Proof of Theorem 2.7

Since **I_s** implies **I_w** and $d_{CF,p} \in \mathcal{D}_1 \subset \mathcal{D}_2$, we can apply Theorem 2.5 to obtain the strong consistency. For the rest of this section, d always refer to the distance $d_{CF,p}$. We recall that $X = X^{\theta_0}$ denotes the observed process with the true parameter θ_0 . In view of the strong identifiability assumption **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 Panloup et al (2020), with adaptations due to the estimation of σ and H . It is based on the following decomposition: since $\hat{\theta}_n$ minimizes the function $\theta \mapsto d(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}}, \mu_\theta)$, we have

$$d(\mu_{\hat{\theta}_n}, \mu_{\theta_0}) \leq d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}}, \mu_{\theta_0}\right) + d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}}, \mu_{\hat{\theta}_n}\right)$$

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

$L^\alpha(\Omega)$ bound on $D_n^{(1)}$.

Following the proof of (Panloup et al, 2020, 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 $p > \frac{\alpha+d(q+1)}{2}$. There exists a positive constant $C_{\alpha,q}$ such that for any $n \in \mathbb{N}$,

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

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 $D_n^{(1)} \leq D_n^{(11)} + D_n^{(12)}$ where

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

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 .

As in the proof of (Panloup et al, 2020, Lemma 5.3), the bound on $D_n^{(11)}$ relies on the pathwise exponential convergence of $Y_t - \bar{Y}_t$ towards 0, which implies the same exponential convergence of $X_t - \bar{X}_t$. For the Lipschitz function $f_\chi(x) = e^{i\langle \chi, x \rangle}$, following the aforementioned proof leads to

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E} f_\chi(X_{kh}) - \mu_{\theta_0}(f_\chi) \right| \leq \frac{1}{n} \|f_\chi\|_{Lip} \sum_{k=0}^{n-1} \mathbb{E} [|X_{kh} - \bar{X}_{kh}|] \leq \frac{C_q}{n} \|f_\chi\|_{Lip} \quad (26)$$

and then

$$\mathbb{E}[|D_n^{(11)}|^\alpha] \leq \frac{C_{\alpha,q}}{n^\alpha}. \quad (27)$$

Let us now bound $D_n^{(12)}$. This is a concentration result inspired by (Varvenne, 2019, Theorem 2.3). Namely, apply (Panloup et al, 2020, Lemma 5.5) to $\frac{1}{n} \sum_{k=0}^{n-1} f_\chi(X_{kh})$ to get

$$\mathbb{E} \left[\left| \frac{1}{n} \sum_{k=0}^{n-1} f_\chi(X_{kh}) - \mathbb{E}(f_\chi(X_{kh})) \right|^\alpha \right] \leq C_\alpha \|f_\chi\|_{Lip}^\alpha n^{-\frac{\alpha}{2}(2-\max(2H,1))},$$

for some positive constant C_α that depends on α . Using the definition of $d_{CF,p}$ and Jensen's inequality, one gets

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

As $\|f_\chi\|_{Lip} \leq |\chi|$, there is $\mathbb{E}[|D_n^{(12)}|^\alpha] \leq C_\alpha n^{-\frac{\alpha}{2}(2-\max(2H,1))} \int_{\mathbb{R}^{d(q+1)}} |\chi|^\alpha g_p(\chi) d\chi$, the integral being finite since $p > \frac{\alpha+d(q+1)}{2}$. This bound with (27) yield the result. \square

4 Strong consistency and rate of convergence of the estimator $\hat{\theta}_{n,N,\gamma}$

In this section, the results on the estimator $\hat{\theta}_{n,N,\gamma}$ are proven: first, $\hat{\theta}_{n,N,\gamma}$ is shown to be strongly consistent (Theorem 2.6), then the rate presented in Theorem 2.8 is established under the strong identifiability assumption \mathbf{I}_s .

4.1 Proof of Theorem 2.6 and Theorem 2.8

First, using intermediary results that we shall prove in Section 4.2 and Section 4.3, we provide a proof of the strong consistency and a rate of convergence for the estimator $\hat{\theta}_{n,N,\gamma}$ defined in (3).

Proof of Theorem 2.6. We use again Proposition 3.1 with

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

this time with $v = (n, N, \gamma)$. We will prove in Section 4.2 that the contrast $L_v(\theta)$ converges uniformly as $(n, N, \gamma) \rightarrow (\infty, \infty, 0)$ 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. \square

Proof of Theorem 2.8. Since \mathbf{I}_s implies \mathbf{I}_w and $d_{CF,p} \in \mathcal{D}_1 \subset \mathcal{D}_2$, we can apply Theorem 2.6 to obtain the strong consistency. To prove the convergence above, we proceed similarly to 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 the triangle inequality to get

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

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

$$\begin{aligned} d\left(\mu_{\theta_0}, \mu_{\hat{\theta}_{n,N,\gamma}}\right) &\leq d\left(\mu_{\theta_0}, \frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}\right) + d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0,\gamma}}\right) \\ &\quad + \sup_{\theta \in \Theta} d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta,\gamma}}, \mu_{\theta}\right). \end{aligned} \quad (28)$$

To allow pathwise comparison, let us define the following processes. For any $\theta \in \Theta$, define $Y^{\theta,\gamma}(B)$, an Euler scheme of Y^θ defined with the same fBm B . Namely, $Y^{\theta,\gamma}(B)$ is defined by (15) where \hat{B} is replaced by B . As in Section 2.1, define $X^{\theta,\gamma}(B)$ by

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

We also define $Y(\hat{B})$ which is the solution to (9) with the fBm \hat{B} , and similarly we define $X(\hat{B})$. Now, we can do pathwise comparison between X^θ and $X^{\theta,\gamma}(B)$, and between $X^{\theta,\gamma}$ and $X^\theta(\hat{B})$.

Bounding the second term in (28).

Split the second term in the right-hand side of (28) as follows

$$\begin{aligned} d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0,\gamma}}\right) &\leq d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \mu_{\theta_0}\right) + d\left(\mu_{\theta_0}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0,\gamma}}\right) \\ &\quad + d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0,\gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0,\gamma}}\right). \end{aligned} \quad (29)$$

Furthermore, split the last term above as

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

Moreover,

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

$$\begin{aligned}
&\leq 2d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0, \gamma(B)}} \right) + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}}, \mu_{\theta_0} \right) \\
&\quad + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})} \right) + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})}, \mu_{\theta_0} \right).
\end{aligned}$$

Injecting the above bound into (29), we get

$$\begin{aligned}
&d \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{L_{k\gamma}^{\theta_0}} \right) \\
&\leq d \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \mu_{\theta_0} \right) + 2d \left(\mu_{\theta_0}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}} \right) + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})}, \mu_{\theta_0} \right) \\
&\quad + 2d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0, \gamma(B)}} \right) + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})} \right).
\end{aligned} \tag{30}$$

Bounding the third term in (28).

Split the third term in (28) as follows

$$\begin{aligned}
&\sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta, \gamma}}, \mu_{\theta} \right) \\
&\leq \sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})} \right) + \sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})}, \mu_{\theta} \right).
\end{aligned} \tag{31}$$

Final bound on $d(\mu_{\theta_0}, \mu_{\widehat{\theta}_{n, N, \gamma}})$.

Using (30) and (31) in (28), we get

$$\begin{aligned}
&d(\mu_{\theta_0}, \mu_{\widehat{\theta}_{n, N, \gamma}}) \\
&\leq 2d \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{kh}^{\theta_0}}, \mu_{\theta_0} \right) + 2d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}}, \mu_{\theta_0} \right) + d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0}(\widehat{B})}, \mu_{\theta_0} \right) \\
&\quad + \sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta}(\widehat{B})}, \mu_{\theta} \right) \\
&\quad + 2 \sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_0, \gamma(B)}} \right) + 2 \sup_{\theta \in \Theta} d \left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta}(\widehat{B})} \right).
\end{aligned} \tag{32}$$

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, one thus gets

$$d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_{kh}^{\theta_0}},\mu_{\theta_0}\right)\leq C_{\alpha,q}\left(n^{-\alpha}+n^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}\right)\quad (33)$$

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta_0}},\mu_{\theta_0}\right)\leq C_{\alpha,q}\left(N^{-\alpha}+N^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}\right)\quad (34)$$

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta_0}(\hat{B})},\mu_{\theta_0}\right)\leq C_{\alpha,q}\left(N^{-\alpha}+N^{-\frac{\alpha}{2}(2-2\max(\mathcal{H})\vee 1)}\right).\quad (35)$$

Remark 4.1. For the term $d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta_0}(\hat{B})},\mu_{\theta_0}\right)$, notice that μ_{θ_0} is also the law of $\bar{X}^{\theta_0,\hat{B}}$, the stationary augmented process associated to (1) with the fBm \hat{B} instead of B , so (26) 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 $t \rightarrow \infty$ by Proposition 3.3.

Let us define

$$\begin{aligned} D_{N,\gamma}^{(21)}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}(B)}\right) \\ D_{N,\gamma}^{(22)}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta,\gamma}},\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta}(\hat{B})}\right) \\ D_{N,\gamma}^{(3)}(\theta) &:= d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{X_{k\gamma}^{\theta}(\hat{B})},\mu_{\theta}\right). \end{aligned}\quad (36)$$

In Section 4.3, we show how to bound the moments of $\sup_{\theta\in\Theta}D_{N,\gamma}^{(21)}(\theta)$, $\sup_{\theta\in\Theta}D_{N,\gamma}^{(22)}(\theta)$ and $\sup_{\theta\in\Theta}D_{N,\gamma}^{(3)}(\theta)$. 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 $N \geq 1$ and $\gamma \leq \gamma_0$ with $N\gamma \geq 1$, the following bounds hold:

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(21)}(\theta)\right|^\alpha\leq C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}\quad (37)$$

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(22)}(\theta)\right|^\alpha\leq C_{\alpha,\varepsilon}\gamma^{\alpha\min(\mathcal{H})-\varepsilon}\quad (38)$$

$$\mathbb{E}\sup_{\theta\in\Theta}\left|D_{N,\gamma}^{(3)}(\theta)\right|^\alpha\leq C_{\alpha,\varepsilon,\varpi}\left(\gamma^{\alpha\min(\mathcal{H})-\varepsilon}+(N\gamma)^{-\bar{\eta}}\right),\quad (39)$$

with $\bar{\eta} = \frac{\varpi\alpha^2}{2(\alpha\varpi+2d)}(2 - (2\max(\mathcal{H}) \vee 1))$. Injecting the bounds (33), (34), (35), (37), (38) and (39) into the decomposition (32) concludes the proof. \square

4.2 Proof of the uniform convergence of the contrast

In this section, we obtain 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 $d(\mu_{\theta_0}, \mu_{\theta})$, that is used in the proof of Theorem 2.6. First we prove that almost surely, there is convergence as $(n, N, \gamma) \rightarrow (\infty, \infty, 0)$ for each fixed θ . We have already proven 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 2.4, $d(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta, \gamma}}, \mu_{\theta}^{\gamma})$ converges to 0 as $N \rightarrow \infty$. Finally, we prove in Proposition 4.2(i) that $d(\mu_{\theta}, \mu_{\theta}^{\gamma})$ converges to 0 as $\gamma \rightarrow 0$. Therefore

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

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

Proposition 4.2. *Let $0 < p \leq 2$ and $d \in \mathcal{D}_p$. Under the assumptions **A₀**, **A₁** and **I_w**, there exists $\gamma_0 > 0$ such that for all $\gamma \in (0, \gamma_0]$, the following assertions hold true.*

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

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

Proof of (i).

By the triangle inequality,

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

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

$$\begin{aligned} d(\mu_{\theta}, \mu_{\theta}^{\gamma}) &\leq \mathcal{W}_2(\mu_{\theta}, \mathcal{L}(X_{N\gamma}^{\theta})) + \mathcal{W}_2(\mu_{\theta}^{\gamma}, \mathcal{L}(X_{N\gamma}^{\theta, \gamma})) + \mathcal{W}_2(\mathcal{L}(X_{N\gamma}^{\theta, \gamma}), \mathcal{L}(X_{N\gamma}^{\theta})) \\ &=: W^{(1)} + W^{(2)} + W^{(3)}. \end{aligned} \tag{40}$$

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

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

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

$$\begin{aligned} W^{(2)} &= \mathcal{W}_2(\mu_\theta^\gamma, \mathcal{L}(X_{N\gamma}^{\theta,\gamma})) \leq \left(\mathbb{E} |\bar{X}_{N\gamma}^{\theta,\gamma} - X_{N\gamma}^{\theta,\gamma}|^2 \right)^{\frac{1}{2}} \\ &\leq C_q \sum_{i=0}^q \left(\mathbb{E} |\bar{Y}_{N\gamma+ih}^{\theta,\gamma} - Y_{N\gamma+ih}^{\theta,\gamma}|^2 \right)^{\frac{1}{2}}. \end{aligned} \quad (41)$$

By (Panloup et al, 2020, Equation (4.2)), we have for any $k \in \mathbb{N}$,

$$\left| \bar{Y}_{k\gamma}^{\theta,\gamma} - Y_{k\gamma}^{\theta,\gamma} \right|^2 \leq (1 - 2\gamma\beta + \gamma^2 K^2)^k \left| \bar{Y}_0^{\theta,\gamma} - Y_0^{\theta,\gamma} \right|^2. \quad (42)$$

Furthermore, for any $i \in \llbracket 0, q \rrbracket$, 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 (42) holds for all the terms in (41). 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 $N \rightarrow \infty$. Now for the last term in (40), by definition of the Wasserstein distance, we have

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

In (Panloup et al, 2020, 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 $\phi_j(z) = \gamma^3 |b_\xi(z)|^2 + \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt$. 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

$$|Y_{m\gamma}^\theta - Y_{m\gamma}^{\theta,\gamma}(B)|^2 \leq C \sum_{j=0}^{m-1} \left(\gamma^3 (1 + |Y_{j\gamma}^{\theta,\gamma}(B)|^{2r}) + \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt \right) e^{-\rho\gamma(m-j+1)}. \quad (43)$$

Now for $i \in \llbracket 0, q \rrbracket$ and $k \in \mathbb{N}$, since the process $Y^{\theta, \gamma}$ is constant over intervals of size γ , recalling the notation $t_\gamma = \gamma \lfloor \frac{t}{\gamma} \rfloor$, we can always write

$$Y_{k\gamma+ih}^\theta - Y_{k\gamma+ih}^{\theta, \gamma}(B) = \left(Y_{(k\gamma+ih)_\gamma + \varepsilon_{k,i}}^\theta - Y_{(k\gamma+ih)_\gamma}^\theta \right) + \left(Y_{(k\gamma+ih)_\gamma}^\theta - Y_{(k\gamma+ih)_\gamma}^{\theta, \gamma}(B) \right), \quad (44)$$

where

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

For the first term in (44), using the sub-linear growth of b , we write

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

It follows from Jensen's inequality that

$$\begin{aligned} & |Y_{j(k\gamma+ih)_\gamma + \varepsilon_{k,i}}^\theta - Y_{(k\gamma+ih)_\gamma}^\theta|^2 \\ & \leq C \left(\varepsilon_{k,i} \int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma + \varepsilon_{k,i}} (1 + |Y_s^\theta|^{2r}) ds + |B_{(k\gamma+ih)_\gamma + \varepsilon_{k,i}} - B_{(k\gamma+ih)_\gamma}|^2 \right). \end{aligned} \quad (45)$$

The second term in (44) can be bounded using (43) with $m \equiv \frac{(k\gamma+ih)_\gamma}{\gamma}$. Combining this and (45) in (44), we get that for any $k \in \mathbb{N}$,

$$\begin{aligned} & \sum_{i=0}^q |Y_{k\gamma+ih}^\theta - Y_{k\gamma+ih}^{\theta, \gamma}(B)|^2 \\ & \leq C \sum_{i=0}^q \gamma \int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_\gamma + \varepsilon_{k,i}} (1 + |Y_s^\theta|^{2r}) ds + |B_{(k\gamma+ih)_\gamma + \varepsilon_{k,i}} - B_{(k\gamma+ih)_\gamma}|^2 \\ & \quad + C \sum_{i=0}^q \sum_{j=0}^{\lfloor \frac{k\gamma+ih}{\gamma} \rfloor - 1} \left(\gamma^2 (1 + |Y_{j\gamma}^{\theta, \gamma}(B)|^{2r}) + \gamma^{-1} \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt \right) \gamma e^{-\rho(\lfloor \frac{k\gamma+ih}{\gamma} \rfloor - j + 1)}. \end{aligned} \quad (46)$$

Taking the expectation, using $\limsup_{\substack{n \rightarrow \infty \\ \gamma \rightarrow 0}} \gamma \sum_{j=0}^n e^{-\rho(n-j+1)} < +\infty$ and $r \leq 1$, we get

$$\begin{aligned} & \sup_{\theta \in \Theta, \gamma \in (0, \gamma_0)} \limsup_{k \rightarrow \infty} \gamma^{-2 \max(\mathcal{H})} \sum_{i=0}^q \mathbb{E} \left| Y_{N\gamma+ih}^\theta - Y_{N\gamma+ih}^{\theta, \gamma}(B) \right|^2 \\ & \leq C \left(1 + \sup_{\theta \in \Theta, \gamma \in (0, \gamma_0)} \limsup_{k \rightarrow \infty} \mathbb{E} |Y_{k\gamma}^{\theta, \gamma}(B)|^{2r} + \sup_{\theta \in \Theta} \limsup_{t \rightarrow \infty} \mathbb{E} |Y_t^\theta|^{2r} \right). \end{aligned}$$

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 concludes 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 $\{\theta \mapsto d(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta, \gamma}}, \mu_{\theta}^{\gamma}); N \geq 1; \theta \in \Theta\}$ is equicontinuous for a fixed $\gamma \in (0, \gamma_0]$. For some θ_1 and θ_2 in Θ , there is

$$\begin{aligned} & \left| d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_1, \gamma}}, \mu_{\theta_1}^{\gamma}\right) - d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_2, \gamma}}, \mu_{\theta_2}^{\gamma}\right) \right| \\ & \leq d(\mu_{\theta_1}^{\gamma}, \mu_{\theta_2}^{\gamma}) + d\left(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_1, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_2, \gamma}}\right). \end{aligned}$$

Decompose the second term to get

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

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

$$\frac{1}{N} \sum_{k=0}^{N-1} |Y_{k\gamma}^{\theta_1, \gamma} - Y_{k\gamma}^{\theta_2, \gamma}|^2 \leq \mathbf{C} (1 \wedge |\theta_1 - \theta_2|^{\varpi}).$$

These results still hold when replacing $Y_{k\gamma}^{\theta, \gamma}$ by $Y_{k\gamma+ih}^{\theta, \gamma}$, since we compare two piecewise constant processes. Thus $d(\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_1, \gamma}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^{\theta_2, \gamma}})$ goes to 0 as $|\theta_1 - \theta_2| \rightarrow 0$ uniformly in N . The same goes for $d(\mu_{\theta_1}^{\gamma}, \mu_{\theta_2}^{\gamma})$ by taking the limit $N \rightarrow \infty$. This concludes the proof of the equicontinuity and therefore the proof of (ii). \square

4.3 Proof of the bounds (37), (38) and (39)

We prove here the bounds (37), (38) and (39) on $D_{N, \gamma}^{(21)}$, $D_{N, \gamma}^{(22)}$ and $D_{N, \gamma}^{(3)}$ that were defined in (36). In this section, d always refer to the distance $d_{CF, p}$.

Proposition 4.3. *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 (8) satisfies $r \leq 1$. For any $\varepsilon \in (0, \alpha \min(\mathcal{H}))$ and any $\varpi \in (0, 1)$, there exist constants $C_{\alpha, \varepsilon} > 0$ and*

$C_{\alpha,\varepsilon,\varpi} > 0$ such that for all $\gamma \in (0, \gamma_0]$ and $N \geq 1$ satisfying $N\gamma \geq 1$, the inequalities (37), (38) and (39) hold.

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

$$\begin{aligned} \sup_{\theta \in \Theta} D_{N,\gamma}^{(21)}(\theta) &\leq \frac{1}{N} \sum_{k=0}^{N-1} \sup_{\theta \in \Theta} |X_{k\gamma}^\theta - X_{k\gamma}^{\theta,\gamma}(B)| \\ &\leq C_q \sum_{i=0}^q \frac{1}{N} \sum_{k=0}^{N-1} \sup_{\theta \in \Theta} |Y_{k\gamma+ih}^\theta - Y_{k\gamma+ih}^{\theta,\gamma}(B)|. \end{aligned}$$

Recall that $\alpha \geq 2$ in **I_s**. Hence, an application of Jensen's inequality gives

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

Define

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

We will first provide a bound on \mathcal{I} . Using (46),

$$\begin{aligned} &\sum_{i=0}^q |Y_{k\gamma+ih}^\theta - Y_{k\gamma+ih}^{\theta,\gamma}(B)|^2 \\ &\leq C \sum_{i=0}^q \left(\gamma \int_{(k\gamma+ih)_\gamma}^{(k\gamma+ih)_{\gamma+\varepsilon_{k,i}}} (1 + |Y_s^\theta|^{2r}) ds + |B_{(k\gamma+ih)_{\gamma+\varepsilon_{k,i}}} - B_{(k\gamma+ih)_\gamma}|^2 \right) \\ &\quad + C \sum_{i=0}^q \left(\sum_{j=0}^{\lfloor \frac{k\gamma+ih}{\gamma} \rfloor - 1} \left(\gamma^2 (1 + |Y_{j\gamma}^{\theta,\gamma}(B)|^{2r}) + \gamma^{-1} \int_0^\gamma |B_{j\gamma+t} - B_{j\gamma}|^2 dt \right) \gamma e^{-\rho(\lfloor \frac{k\gamma+ih}{\gamma} \rfloor - j + 1)} \right) \\ &=: C \sum_{i=0}^q (\mathcal{I}_{1,k}(i) + \mathcal{I}_{2,k}(i) + \mathcal{I}_{3,k}(i) + \mathcal{I}_{4,k}(i)). \end{aligned} \tag{47}$$

Hence there is $\mathcal{I} \leq C_q \sup_{\theta \in \Theta} \sum_{i=0}^q \frac{1}{N} \sum_{k=0}^{N-1} (\mathcal{I}_{1,k}(i) + \mathcal{I}_{2,k}(i) + \mathcal{I}_{3,k}(i) + \mathcal{I}_{4,k}(i))$. Let us provide uniform bounds in θ on the sum over k of the terms $\mathcal{I}_{1,k}(i)$, $\mathcal{I}_{2,k}(i)$, $\mathcal{I}_{3,k}(i)$, $\mathcal{I}_{4,k}(i)$. First we have

$$\frac{1}{N} \sum_{k=0}^{N-1} \mathcal{I}_{1,k}(i) \leq \frac{1}{N} \int_{(ih)_\gamma}^{(N\gamma+ih)_\gamma} \gamma (1 + |Y_s^\theta|^{2r}) ds$$

$$\begin{aligned}
&\leq \frac{\gamma}{N} \int_{ih}^{N\gamma+ih} (1 + |Y_s^\theta|^{2r}) ds + \frac{\gamma}{N} \int_{(ih)_\gamma}^{ih} (1 + |Y_s^\theta|^{2r}) ds \\
&\leq \frac{\gamma^2}{N\gamma} \int_0^{N\gamma} (1 + |Y_{s+ih}^\theta|^{2r}) ds + 2\gamma^2 \mathbf{1}_{i \neq 0} \left(\frac{1}{ih} \int_0^{ih} (1 + |Y_s^\theta|^{2r}) ds \right).
\end{aligned} \tag{48}$$

For $\mathcal{I}_{3,k}(i)$, write

$$\begin{aligned}
\frac{1}{N} \sum_{k=0}^{N-1} \mathcal{I}_{3,k}(i) &\leq C \frac{\gamma^2}{N} \sup_{\theta \in \Theta} \sum_{k=0}^{\lfloor \frac{N\gamma+ih}{\gamma} \rfloor - 1} (1 + |Y_{k\gamma}^{\theta,\gamma}(B)|^{2r}) \\
&\leq C \frac{\gamma}{N} \sup_{\theta \in \Theta} \int_0^{(N\gamma+ih)_\gamma - \gamma} (1 + |Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r}) dt \\
&\leq C \frac{\gamma}{N} \sup_{\theta \in \Theta} \int_0^{N\gamma - \gamma} (1 + |Y_{t_\gamma+ih}^{\theta,\gamma}(B)|^{2r}) dt + C \mathbf{1}_{i \neq 0} \frac{\gamma}{N} \sup_{\theta \in \Theta} \int_0^{ih} (1 + |Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r}) dt \\
&\leq C \frac{\gamma^2}{N\gamma} \sup_{\theta \in \Theta} \int_0^{N\gamma} (1 + |Y_{t_\gamma+ih}^{\theta,\gamma}(B)|^{2r}) dt + C \frac{\gamma^2}{N\gamma} \mathbf{1}_{i \neq 0} \sup_{\theta \in \Theta} \frac{1}{ih} \int_0^{ih} (1 + |Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r}) dt.
\end{aligned} \tag{49}$$

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

$$\begin{aligned}
\frac{1}{N} \sum_{k=0}^{N-1} \mathcal{I}_{4,k}(i) &\leq \frac{1}{N} \sum_{k=0}^{\lfloor \frac{N\gamma+ih}{\gamma} \rfloor - 1} \gamma^{-1} \int_0^\gamma |B_{k\gamma+t} - B_{k\gamma}|^2 dt \\
&\leq \frac{\gamma^{-2}}{N} \int_0^{N\gamma+ih-\gamma} \left(\int_0^\gamma |B_{s_\gamma+t} - B_{s_\gamma}|^2 dt \right) ds \\
&\leq \frac{1}{N\gamma} \int_0^{ih} \left(\gamma^{-1} \int_0^\gamma |B_{s_\gamma+t} - B_{s_\gamma}|^2 dt \right) ds + \frac{1}{N\gamma} \int_0^{N\gamma} \left(\gamma^{-1} \int_0^\gamma |B_{s_\gamma+ih+t} - B_{s_\gamma+ih}|^2 dt \right) ds.
\end{aligned} \tag{50}$$

Therefore, using (48), (49), (50) in (47), it comes

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

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 (Haress and Richard, 2022, Proposition 3.5) that for $\varepsilon \in (0, \alpha \min(\mathcal{H}))$,

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

It follows that

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

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

$$\mathbb{E} \left[\sup_{\theta \in \Theta} \frac{1}{N\gamma} \int_0^{N\gamma} |Y_s^\theta|^{2r} ds \right]^{\alpha/2} \quad \text{and} \quad \mathbb{E} \left[\frac{1}{N\gamma} \sup_{\theta \in \Theta} \int_0^{N\gamma} |Y_{t_\gamma}^{\theta,\gamma}(B)|^{2r} dt \right]^{\alpha/2}$$

are bounded 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

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

Hence

$$\mathbb{E} \sup_{\theta \in \Theta} D_{N,\gamma}^{(21)}(\theta)^\alpha \leq C \mathbb{E}[\mathcal{I}^{\alpha/2}] \leq C \gamma^{\alpha \min(\mathcal{H}) - \varepsilon}. \quad (51)$$

Consider now $D_{N,\gamma}^{(3)}(\theta)$, which was defined in (36). 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. We first start with the following decomposition:

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

where, noticing that $\frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma}^\theta} = \frac{1}{T} \int_0^T \delta_{X_{t\gamma}^\theta} dt$ for $T = N\gamma$,

$$\begin{aligned} D_{N,\gamma}^{(31)}(\theta) &= d \left(\mu_\theta, \frac{1}{T} \int_0^T \delta_{X_t^\theta} dt \right) \\ D_{N,\gamma}^{(32)}(\theta) &= d \left(\frac{1}{T} \int_0^T \delta_{X_t^\theta} dt, \frac{1}{T} \int_0^T \delta_{X_{t\gamma}^\theta} dt \right). \end{aligned}$$

Bound on $\sup_{\theta \in \Theta} D_{N,\gamma}^{(32)}(\theta)$.

Similar arguments as before lead to

$$\mathbb{E} \sup_{\theta \in \Theta} D_{N,\gamma}^{(32)}(\theta)^\alpha \leq C_{q,\alpha} \sum_{i=0}^q \frac{1}{T} \int_0^T \sup_{\theta \in \Theta} |Y_{t+ih}^\theta - Y_{t\gamma+ih}^\theta|^\alpha dt.$$

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 (9), it follows from using Jensen's inequality and integrating over t that

$$\frac{1}{T} \int_0^T |Y_t^\theta - Y_{t\gamma}^\theta|^\alpha dt \leq 2^{\alpha-1} \gamma^{\alpha-1} \frac{1}{T} \int_0^T \int_{t\gamma}^t |b_\xi(Y_s^\theta)|^\alpha ds dt + 2^{\alpha-1} |\sigma|^\alpha \frac{1}{T} \int_0^T |B_t - B_{t\gamma}|^\alpha dt.$$

By Fubini's theorem, we get that

$$\frac{1}{T} \int_0^T |Y_t^\theta - Y_{t_\gamma}^\theta|^\alpha dt \leq 2^{\alpha-1} \gamma^\alpha \frac{1}{T} \int_0^T |b_\xi(Y_s^\theta)|^\alpha ds + 2^{\alpha-1} |\sigma|^\alpha \frac{1}{T} \int_0^T |B_t - B_{t_\gamma}|^\alpha dt.$$

The drift term above is bounded thanks to the sublinear growth of b_ξ given by (8) and the uniform bounds on the L^q moments of Y_t^θ given in Proposition B.1 (ii). As for the term $|B_t - B_{t_\gamma}|$, we have thanks to (Haress and Richard, 2022, Proposition 3.5) that for all $\varepsilon > 0$,

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

Hence, the two previous inequalities yield

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

Bound on $D_{N,\gamma}^{(31)}(\theta)$.

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

$$\mathbb{E} D_{N,\gamma}^{(31)}(\theta)^\alpha \leq C_\alpha \left(T^{-\alpha} + T^{-\frac{\alpha}{2}(2 - \max(2 \max(\mathcal{H}), 1))} \right). \quad (52)$$

Bound on $\sup_{\theta \in \Theta} D_{N,\gamma}^{(31)}(\theta)$.

Let $\varphi(\theta) = d(\mu_\theta, \frac{1}{T} \int_0^T \delta_{X_t^\theta} dt)$. Let $\varepsilon > 0$ and $\Theta^{(\varepsilon)} := \{\theta_i^{(\varepsilon)} \mid 1 \leq i \leq M_\varepsilon\}$ such that $\Theta \subset \bigcup_{i=1}^{M_\varepsilon} B(\theta_i^{(\varepsilon)}, \varepsilon)$ for some points $\theta_i^{(\varepsilon)}$ in Θ . Then, as in (Panloup et al, 2020, Eq. (5.27)-(5.28)) and using (52), one gets

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

where $\theta_\varepsilon := \operatorname{argmin}_{\theta' \in \{\theta_i^{(\varepsilon)}\}} |\theta' - \theta|$. Now $|\varphi(\theta) - \varphi(\theta_\varepsilon)| \leq d(\mu_\theta, \mu_{\theta_\varepsilon}) + d\left(\frac{1}{T} \int_0^T \delta_{X_t^\theta} dt, \frac{1}{T} \int_0^T \delta_{X_t^{\theta_\varepsilon}} dt\right)$.

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

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

For $\varpi \in (0, 1)$, Proposition B.3 gives the existence of a random variable \mathbf{C} with finite moments such that

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

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

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

Overall, we get

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

Hence we have obtained that

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

Choosing $M_\varepsilon \leq \frac{C}{\varepsilon^d}$ and $\varepsilon = T^{-\chi}$ for some $\chi > 0$, it comes that

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

Finally optimize over χ to get $\mathbb{E} \sup_{\theta \in \Theta} \varphi(\theta)^\alpha \leq C_{\alpha, q, \varpi} T^{-\bar{\eta}}$, for $\bar{\eta} = \frac{\varpi \alpha^2}{2(\alpha \varpi + 2d)} (2 - (2 \max(\mathcal{H}) \vee 1))$. \square

5 Application to fractional Ornstein-Uhlenbeck processes

In this section, we prove the results of Section 2.4 and provide numerical experiments to illustrate the convergence of the estimators in the case of fractional Ornstein-Uhlenbeck processes. We first prove the identifiability assumption for the fractional Ornstein-Uhlenbeck (OU) process in Section 5.1 and Section 5.2, then for a family of small perturbations of the fractional OU process in Section 5.3 and finally, in Section 5.4, we provide numerical results.

5.1 Identifiability assumption: proof of Proposition 2.11

The proof of Proposition 2.11 is based on the injectivity of a specific function, as stated in the following Lemma (the proof is given in Section 5.2).

Lemma 5.1. *Assume one of the three cases $\theta = (\xi, H)$, $\theta = (\xi, \sigma)$ or $\theta = (\sigma, H)$, then there exists $h_0 > 0$ such that for $h \in (0, h_0)$, the function f defined by*

$$f : \theta \mapsto \left(\begin{array}{c} \sigma^2 H \Gamma(2H) \xi^{-2H} \\ \sigma^2 \Gamma(2H + 1) \frac{\sin(\pi H)}{\pi} \int_0^\infty \cos(hx) \frac{x^{1-2H}}{\xi^2 + x^2} dx \end{array} \right) \quad (53)$$

is one-to-one.

For the fractional OU process, recall that the stationary measure follows the Gaussian distribution given by (18). Furthermore, the processes $\bar{U}_{\cdot+ih}^\theta$ are also Gaussian with the same law. The correlation between these processes is given by (see (Cheridito et al, 2003, Eq (2.2))):

$$\mathbb{E}(\bar{U}_t^\theta \bar{U}_{t+ih}^\theta) = \sigma^2 \frac{\Gamma(2H + 1) \sin(\pi H)}{\pi} \int_0^\infty \cos(ihx) \frac{x^{1-2H}}{\xi^2 + x^2} dx. \quad (54)$$

Now for θ_1, θ_2 in Θ , there is

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

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

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

which thus reads

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

In view of (18) and (54), assumption \mathbf{I}_w becomes equivalent to the injectivity of the function f defined in (53), which is therefore given by Lemma 5.1.

Remark 5.2. *In Haress and Hu (2021), the authors studied fractional OU processes and proposed a similar estimator for (ξ, σ, H) simultaneously. Similarly to our case, for a consistency argument to hold, they are left to study the injectivity of*

$$f : (\xi, \sigma, H) \mapsto \left(\begin{array}{c} \sigma^2 H \Gamma(2H) \xi^{-2H} \\ \sigma^2 \Gamma(2H + 1) \frac{\sin(\pi H)}{\pi} \int_0^\infty \cos(hx) \frac{x^{1-2H}}{\xi^2 + x^2} dx \\ \sigma^2 \Gamma(2H + 1) \frac{\sin(\pi H)}{\pi} \int_0^\infty \cos(2hx) \frac{x^{1-2H}}{\xi^2 + x^2} dx \end{array} \right). \quad (55)$$

The injectivity was not proven but numerical arguments were provided to support this claim.

5.2 Injectivity of f : proof of Lemma 5.1

The case $\theta = (\sigma, H)$.

Let (a, b) be in the range of f . We will show that the equation

$$\begin{aligned} a &= \sigma^2 H \Gamma(2H) \xi^{-2H} \\ b &= \sigma^2 \Gamma(2H + 1) \frac{\sin(\pi H)}{\pi} \xi^{-2H} \int_0^\infty \cos(\xi h x) \frac{x^{1-2H}}{1+x^2} dx, \end{aligned} \quad (56)$$

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

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

We will show that the function g is injective. Since g is continuously differentiable, it suffices to show that $g'(H) > 0$ for all $H \in \mathcal{H}$. We have

$$g'(H) = \pi \cos(\pi H) \int_0^\infty \cos(\xi h x) \frac{x^{1-2H}}{1+x^2} dx - 2 \sin(\pi H) \int_0^\infty \cos(\xi h x) \log(x) \frac{x^{1-2H}}{1+x^2} dx. \quad (57)$$

By (18) and (54), we have

$$\mathbb{E}(\bar{U}_t^\theta \bar{U}_{t+\theta}^\theta) = \sigma^2 \frac{\Gamma(2H + 1) \sin(\pi H)}{\pi} \int_0^\infty \frac{x^{1-2H}}{\xi^2 + x^2} dx = \mathbb{E}(\bar{U}_t^\theta)^2 = \sigma^2 H \Gamma(2H) \xi^{-2H}.$$

Using $\Gamma(2H + 1) = 2H \Gamma(2H)$ and the change of variables $y = x/\xi$ yields

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

By differentiating with respect to H , we get

$$\pi \cos(\pi H) \int_0^\infty \frac{x^{1-2H}}{1+x^2} dx - 2 \sin(\pi H) \int_0^\infty \log(x) \frac{x^{1-2H}}{1+x^2} dx = 0.$$

Subtracting this term to $g'(H)$ in (57), we get

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

Let $\beta_H = e^{\frac{\pi \cos(\pi H)}{2 \sin(\pi H)}}$. Then

$$g'(H) = \int_0^{\beta_H} (1 - \cos(\xi h x)) (2 \sin(\pi H) \log(x) - \pi \cos(\pi H)) \frac{x^{1-2H}}{1+x^2} dx$$

$$+ \int_{\beta_H}^{\infty} (1 - \cos(\xi hx))(2 \sin(\pi H) \log(x) - \pi \cos(\pi H)) \frac{x^{1-2H}}{1+x^2} dx.$$

For $x \in (0, \beta_H]$, $2 \sin(\pi H) \log(x) - \pi \cos(\pi H) \leq 0$. In addition, using $1 - \cos(x) \leq \frac{x^2}{2}$ in the first integral and the change of variables $y = hx$ in the second integral, we get

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

For $x \geq \beta_H h$, we have $2 \sin(\pi H) \log(x) - 2 \sin(\pi H) \log(h) - \pi \cos(\pi H) \geq 0$. Assuming $h \leq 1$, it thus follows that

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

Since $H \in \mathcal{H}$ and $\xi \in \Xi$, we deduce that there exists $C_1, C_2, C_3 > 0$ such that

$$g'(H) \geq C_1 h^{2H} |\log(h)| + C_2 h^{2H} - C_3 h^2.$$

Therefore, there exists $C > 0$ and $h_0 > 0$ such that for $h \in (0, h_0)$, we have

$$g'(H) \geq C h^{2H} |\log(h)| > 0. \tag{58}$$

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

The case $\theta = (\xi, H)$.

Let (a, b) be in the range of f . We prove that the following equation has a unique solution in (ξ, H) :

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

which is equivalent to solving

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

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

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

We will show that for all possible values of a , g_a is a bijection and therefore there exists a unique H such that $g_a(H) = \frac{\pi b}{2a}$. For this H , ξ is then uniquely determined by the equality $\xi = \left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}}$.

We plan to differentiate g_a . For $H > 1/2$, the derivative in the H variable of the function $x \mapsto \cos \left(\left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}} hx \right) \frac{x^{1-2H}}{1+x^2}$ is integrable and we get

$$\begin{aligned} g'_a(H) &= \pi \cos(\pi H) \int_0^\infty \cos \left(\left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}} hx \right) \frac{x^{1-2H}}{1+x^2} dx \\ &\quad - \sin(\pi H) \left[H \mapsto \left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}} \right]' (H) h \int_0^\infty x \sin \left(\left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}} hx \right) \frac{x^{1-2H}}{1+x^2} dx \\ &\quad - 2 \sin(\pi H) \int_0^\infty \cos \left(\left(\frac{a}{H\Gamma(2H)} \right)^{-\frac{1}{2H}} hx \right) \log(x) \frac{x^{1-2H}}{1+x^2} dx \\ &=: g'_{a,1}(H) + g'_{a,2}(H) + g'_{a,3}(H). \end{aligned} \tag{59}$$

Unfortunately when $H \leq 1/2$, the integral that appears in $g'_{a,2}(H)$ is not defined in Lebesgue's sense. However, we have for any $A > 1$ that

$$\int_1^A \sin(Cx) \frac{x^{2-2H}}{1+x^2} dx = \int_1^A \sin(Cx) x^{-2H} dx - \int_1^A \sin(Cx) \frac{x^{-2H}}{(1+x^2)} dx.$$

The first integral in the right-hand side converges in Riemann's sense as $A \rightarrow \infty$ and the second one converges as a classical Lebesgue's integral. Thus we get that $g'_{a,1}(H) + g'_{a,2}(H) + g'_{a,3}(H)$ is well-defined even for $H \in (0, 1/2]$, and then that the equality (59) also holds for $H \leq 1/2$.

Now notice that $g'_{a,1}(H) + g'_{a,3}(H)$ is exactly the term $g'(H)$ handled in the previous case $\theta = (\sigma, H)$ with $\xi \equiv \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}}$. We have shown in (58) that there exists $C_1 > 0$ and $h_0 > 0$ such that for $h \in (0, h_0)$,

$$g'_{a,1}(H) + g'_{a,3}(H) > C_1 h^{2H} |\log(h)|. \quad (60)$$

We now prove an upper bound on the absolute value of $g'_{a,2}(H)$. Using the change of variable $y = hx$, we have

$$|g'_{a,2}(H)| \quad (61)$$

$$\leq \sin(\pi H) \left| \left[H \mapsto \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} \right]' (H) \right| h^{2H} \left| \int_0^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) \frac{x^{2-2H}}{h^2 + x^2} dx \right|. \quad (62)$$

Let us show that the integral $\mathcal{J} = \left| \int_0^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) \frac{x^{2-2H}}{h^2 + x^2} dx \right|$ is bounded uniformly in $h \in (0, h_0)$. Using that $|\sin(x)| \leq x$ for $x \geq 0$ we have for $\alpha_H = 2\pi \left(\frac{a}{H\Gamma(2H)}\right)^{\frac{1}{2H}}$ that

$$\begin{aligned} \mathcal{J} &\leq \int_0^{\alpha_H} \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} \frac{x^{3-2H}}{h^2 + x^2} dx + \left| \int_{\alpha_H}^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) \frac{x^{2-2H}}{h^2 + x^2} dx \right| \\ &\leq \int_0^{\alpha_H} \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x^{1-2H} dx \\ &\quad + \left| \int_{\alpha_H}^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) \left(\frac{x^{2-2H}}{h^2 + x^2} - \frac{x^{2-2H}}{x^2} \right) dx \right| \\ &\quad + \left| \int_{\alpha_H}^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) x^{-2H} dx \right|. \end{aligned}$$

Hence bounding the sine function by 1 in the second integral and using the change of variables $y = \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x$ in the third, we get

$$\begin{aligned} \mathcal{J} &\leq \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} \frac{\alpha_H^{2-2H}}{2-2H} + h^2 \left| \int_{\alpha_H}^\infty \sin \left(\left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} x \right) \frac{x^{-2H}}{h^2 + x^2} dx \right| \\ &\quad + \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{2H-1}{2H}} \left| \int_{2\pi}^\infty \frac{\sin(x)}{x^{2H}} dx \right| \end{aligned}$$

$$\begin{aligned}
&\leq \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} \frac{\alpha_H^{2-2H}}{2-2H} + h^2 \int_{\alpha_H}^{\infty} x^{-2-2H} dx + \left(\frac{a}{H\Gamma(2H)}\right)^{\frac{1-2H}{2H}} \left| \int_{2\pi}^{\infty} \frac{\sin(x)}{x^{2H}} dx \right| \\
&\leq \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}} \frac{1}{2-2H} + h^2 \frac{\alpha_H^{-1-2H}}{1+2H} + \left(\frac{a}{H\Gamma(2H)}\right)^{\frac{1-2H}{2H}} \left| \int_{2\pi}^{\infty} \frac{\sin(x)}{x^{2H}} dx \right|.
\end{aligned}$$

Writing $\int_{2\pi}^{\infty} \frac{\sin(x)}{x^{2H}} dx$ as the sum of positive terms

$$\begin{aligned}
\int_{2\pi}^{\infty} \sin(x)x^{-2H} dx &= \sum_{k=0}^{\infty} \int_{2k\pi}^{(2k+1)\pi} \frac{\sin(x)}{x^{2H}} dx + \int_{(2k+1)\pi}^{(2k+2)\pi} \frac{\sin(x)}{x^{2H}} dx \\
&= \sum_{k=0}^{\infty} \int_{2k\pi}^{(2k+1)\pi} \sin(x) \left(\frac{1}{x^{2H}} - \frac{1}{(x+\pi)^{2H}} \right) dx \quad (63) \\
&\leq \sum_{k=1}^{\infty} \frac{\pi}{(2k\pi)^{2H}} \left(1 - \frac{1}{(1+k^{-1})^{2H}} \right),
\end{aligned}$$

we get that the last sum can be bounded uniformly for $H \in \mathcal{H}$. Thus, \mathcal{J} can be bounded uniformly for $H \in \mathcal{H}$ by a constant $C_2 > 0$. From (61), we thus get

$$|g'_{a,2}(H)| \leq C_2 h^{2H} \sin(\pi H) \left| [H \mapsto \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}}]'(H) \right|.$$

Since the mapping $H \mapsto \left(\frac{a}{H\Gamma(2H)}\right)^{-\frac{1}{2H}}$ is smooth on $(0, +\infty)$ and \mathcal{H} is a compact subset of $(0, 1)$, we deduce that there exists a constant $\tilde{C}_2 > 0$ such that

$$|g'_{a,2}(H)| \leq \tilde{C}_2 h^{2H}.$$

Combining this with (60), we conclude that for any $h \in (0, h_0)$,

$$g'_{a,1}(H) + g'_{a,2}(H) + g'_{a,3}(H) \geq C_1 h^{2H} |\log(h)| - \tilde{C}_2 h^{2H}.$$

Hence, there exists $C, h_1 > 0$ such that for any $h \in (0, h_1)$, we have

$$g'_{a,1}(H) + g'_{a,2}(H) + g'_{a,3}(H) \geq C h^{2H} |\log(h)| > 0.$$

This proves that g_a is a bijection.

The case $\theta = (\xi, \sigma)$.

As before, for (a, b) in the range of f , we need to show that (56) has a unique solution in ξ, σ , for a given H . Notice that (56) is equivalent to

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

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

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

Let C be a constant that may depend only on Θ and may change from line to line. We decompose $\tilde{g}'(\xi)$ as

$$\tilde{g}'(\xi) = -h \int_0^1 \sin(\xi hx) \frac{x^{2-2H}}{1+x^2} dx - h \int_1^\infty \sin(\xi hx) \frac{x^{2-2H}}{1+x^2} dx.$$

Using $|\sin(x)| \leq x$ in the first integral, we get

$$\begin{aligned} \tilde{g}'(\xi) &\leq Ch^2 - h \int_1^\infty \sin(\xi hx) \frac{x^{2-2H}}{1+x^2} dx \\ &= Ch^2 + h \int_1^\infty \sin(\xi hx) \frac{x^{-2H}}{1+x^2} dx - h \int_1^\infty \sin(\xi hx) x^{-2H} dx. \end{aligned}$$

Since ξ is in a compact, we use in the first integral that $|\sin(\xi hx)| \leq Chx$. As for the second integral, we use the change of variables $y = \xi hx$ to get

$$\begin{aligned} \tilde{g}'(\xi) &\leq Ch^2 + Ch^2 \int_1^\infty \frac{x^{1-2H}}{1+x^2} dx - \xi^{2H-1} h^{2H} \int_{\xi h}^\infty \sin(x) x^{-2H} dx \\ &\leq Ch^2 + \xi^{2H-1} h^{2H} \int_0^{\xi h} \sin(x) x^{-2H} dx - \xi^{2H-1} h^{2H} \int_0^\infty \sin(x) x^{-2H} dx. \end{aligned}$$

Using the inequality $|\sin(x)| \leq x$ and the fact that ξ is in a compact, we have $|\xi^{2H-1} h^{2H} \int_0^{\xi h} \sin(x) x^{-2H} dx| \leq Ch^2$. As for the last term, we write $\int_0^\infty \sin(x) x^{-2H} dx = \int_0^{2\pi} \sin(x) x^{-2H} dx + \int_{2\pi}^\infty \sin(x) x^{-2H} dx$. The second term is positive by (63), therefore

$$\begin{aligned} \int_0^\infty \sin(x) x^{-2H} dx &\geq \int_0^{2\pi} \sin(x) x^{-2H} dx = \int_0^\pi \sin(x) x^{-2H} dx + \int_\pi^{2\pi} \sin(x) x^{-2H} dx \\ &= \int_0^\pi \sin(x) \left(\frac{1}{x^{2H}} - \frac{1}{(x+\pi)^{2H}} \right) dx > 0. \end{aligned}$$

Since the last integral is continuous in H , it follows that $\inf_{\theta \in \Theta} \xi^{2H-1} \int_0^\infty \sin(x) x^{-2H} dx \geq c > 0$ and we get

$$\tilde{g}'(\xi) \leq Ch^2 - ch^{2H}.$$

It follows that there exists $h_0 > 0$ such that for any $h < h_0$, we have $g'(\xi) < 0$.

5.3 Strong identifiability assumption: proof of Lemma 2.12 and Lemma 2.13

In this section, we prove Lemma 2.12 and Lemma 2.13, that is the strong identifiability assumption for the fractional Ornstein-Uhlenbeck process and small perturbations of this process.

Proof of Lemma 2.12. The condition $p \geq 1$ ensures that $d_{CF,p}$ is well-defined in dimension $d = 1$. When $\theta = \xi$, this lemma was proved in (Panloup et al, 2020, Lemma 6.2).

Let us deal with the case $\theta = H$. We have already seen that $\mu_\theta = \mathcal{N}(0, \sigma^2 H \Gamma(2H) \xi^{-2H})$. Taking into account the expression of $d_{CF,p}$ in (4) yields

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

Let $g(H, \eta) = \exp(-\frac{\sigma^2 H \Gamma(2H)}{2\xi^{2H}} \eta^2)$, then we have

$$d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2})^2 = \int_{\mathbb{R}} \left(\int_{H_1}^{H_2} \partial_H g(H, \eta) dH \right)^2 g_p(\eta) d\eta. \quad (64)$$

We will show that $\partial_H g(H, \eta)$ is bounded away from 0. We have

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

Under (20), we have $|\Gamma(2H) + 2H\Gamma'(2H) - 2H\Gamma(2H) \log(\xi)| > 0$ for all $H \in \mathcal{H}$. Hence, there exists two positive constants c, C that depend only on Θ such that, we have

$$|\partial_H g(H, \eta)| \geq C\eta^2 \exp(-c\eta^2) \text{ for all } H \in [m_{\mathcal{H}}, M_{\mathcal{H}}].$$

Using this in (64), it follows that

$$d_{CF,p}(\mu_{\theta_1}, \mu_{\theta_2})^2 \geq C^2 |H_1 - H_2|^2 \int_{\mathbb{R}} \eta^2 \exp(-2c\eta^2) g_p(\eta) d\eta.$$

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

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

and all the parameters live in compact sets that do not contain 0, there exists positive constants \tilde{C}, \tilde{c} such that all $\sigma \in [m_{\Sigma}, M_{\Sigma}]$, we have $g'(\sigma) < -\tilde{C}\eta^2 \exp(-\tilde{c}\eta^2)$. Hence we can conclude as in the previous case. \square

Proof of Lemma 2.13. The case $\theta = \xi$ was considered in (Panloup et al, 2020, Proposition 6.4) under the same assumptions. Our proof for $\theta = H$ or $\theta = \sigma$ will be very similar. More specifically, we decompose $d_{CF,p}(\mu_{\theta_1}^\lambda, \mu_{\theta_2}^\lambda)$ as

$$d_{CF,p}(\mu_{\theta_1}^\lambda, \mu_{\theta_2}^\lambda) \geq I_3^{1/2} - \left(I_2^{1/2} + I_{11}^{1/2} + I_{12}^{1/2} \right), \quad (65)$$

where

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

In the above definition of I_{1j} , I_2 and I_3 , t is an arbitrary large time to be determined later. Our goal 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:

$$\begin{aligned} I_3 &\geq \frac{1}{3} \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta \bar{U}_t^{0, \theta_1})] - \mathbb{E}[\exp(i\eta \bar{U}_t^{0, \theta_2})] \right)^2 g_p(\eta) d\eta \\ &\quad - \left(\int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta \bar{U}_t^{0, \theta_1})] - \mathbb{E}[\exp(i\eta U_t^{0, \theta_1})] \right)^2 g_p(\eta) d\eta \right. \\ &\quad \left. + \int_{\mathbb{R}} \left(\mathbb{E}[\exp(i\eta \bar{U}_t^{0, \theta_2})] - \mathbb{E}[\exp(i\eta U_t^{0, \theta_2})] \right)^2 g_p(\eta) d\eta \right) \end{aligned}$$

Now, by Lemma 2.12, 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 bounded by Ce^{-ct} . Choosing t large enough, we can thus bound I_3 from below by

$$I_3 \geq \frac{c_1}{6} |\theta_1 - \theta_2|^2.$$

Upper bound for I_{1j} .

The term I_{1j} also represents a distance between the solution of (19) and its stationary version. Under the assumption that b_ξ , $\partial_\xi b_\xi$ and $\partial_y b_\xi$ are bounded and λ is small enough, the drift $-\xi + \lambda b_\xi(\cdot)$ satisfies assumption **A1**. Therefore by Proposition 3.3 we 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 .

It was shown in (Panloup et al, 2020, Equation (6.17)) under $p > 3/2$ 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

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

But $\Delta_R(U_t) = 0$ when $\theta = H$ or $\theta = \sigma$. So $I_2 \leq \lambda^2 C \mathbb{E}(|U_t^{0,\theta_2} - U_t^{0,\theta_1}|^2 \|\partial_\lambda U^{\lambda,\theta}\|_\infty^2)$. It was also proved in (Panloup et al, 2020, equation (6.18) and thereafter) that when b_ξ and $\partial_y b_\xi$ are both bounded and $\lambda \leq m_\Xi(1 - \epsilon)$, there is $\|\partial_\lambda U^{\lambda,\theta}\|_\infty^2 \leq c_{m_\Xi, M_\Xi, \epsilon}$. Hence we deduce that

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

Now if $\theta = H$, we get from the same computation as from the stationary case (see (Haress and Richard, 2022, Lemma A.1)) 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 $\theta = \sigma$,

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

Thus our bound on I_2 becomes $I_2 \leq C_{m_\Xi, M_\Xi, \epsilon} \lambda^2 |\theta_2 - \theta_1|^2$. Finally, choose λ small enough so that

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

To finish the proof, combine the bounds obtained for I_{1j} , I_2 and I_3 into (65). \square

5.4 Numerical results

In this section, we provide numerical examples to illustrate the main results of this paper. We only deal with the 1d OU model defined in (17) that starts from 0, as 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).

Simulated data.

The fractional OU process cannot be simulated exactly. Therefore, we have chosen to approximate it by the Euler scheme with very small time-step \underline{h} (namely $\underline{h} = 10^{-3}$). Recall that the L^2 -distance between the true SDE and the Euler scheme is of order \underline{h}^H when both are defined with the same fBm. This result holds independently of the time horizon when the drift is contractive, see e.g. (Panloup et al, 2020, Proposition 3.7 (i)). Recall also that the fBm 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 $(U_{kh})_{k \geq 0}$, where $(U_t)_{t \geq 0}$ is a solution to (17) with a given θ_0 . Then we create from this path a subsequence of augmented observations $(X_{t_k})_{k=1, \dots, n}$ as defined in (10). Here we consider the linear transformation to be the simple increments as in (12). Furthermore, we consider the time-steps t_k to be of the form $t_k = kh$, which means in particular that we assume h to be of the form $k_0 \underline{h}$ with $k_0 \in \mathbb{N}_*$ (namely $k_0 = 100$). Moreover, to compare the estimators (2) and (3), for $\gamma = 0.1$, we simulate N of the Euler approximation $U^{\theta, \gamma}$ defined in (15) with $b_\theta(x) = -\theta x$. Then we create from this path a subsequence of augmented Euler approximations $(X_{k\gamma}^{\theta, \gamma})_{k=1, \dots, N}$ as defined in (16) taking again the linear transformations as simple increments. For the rest of this section, we will use the following terminology:

- One-dimensional case: This is when we only use the first component of X^{θ_0} (i.e. U^{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 are known. There are thus three choices to consider.
- Two-dimensional case: This is when we want to estimate two parameters and therefore take the first two components of X^{θ_0} as observations. There are also three choices to consider.
- Three-dimensional case: This is when we want to estimate all the parameters and therefore consider all the components included in X^{θ_0} .
- In the simulations, we shall refer to the $\hat{\theta}_n$ as the oracle and $\hat{\theta}_{n, N, \gamma}$ as the estimate.

Computation of the distance between the empirical measures.

In practice, to implement the estimator (2), one needs to compute the distance $d \in \mathcal{D}_p$ between the average of Dirac measures and the stationary distribution. If the observed process is \mathbb{R} -valued, and d is given by the Wasserstein distance, an explicit computation is possible. However, as we explained in the introduction, using the observations of U^{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. Unfortunately, the computation of the Wasserstein distance in higher dimension requires approximation/optimization methods that are highly expensive in terms of complexity and are not discussed in this paper. In this context and as in Panloup et al (2020), it is simple to consider an approximation of the distance $d_{CF, p}$ defined in (4), which we also worked with to obtain the rate of convergence. More specifically, we want a discretisation technique for the integral that appears in (4).

Minimization of the distance with respect to θ .

To implement the estimators, we see the problem of computing the argmin in (2) as an optimization problem. More specifically, in the Ornstein-Uhlenbeck case, we already have an expression of the stationary distribution (18). Furthermore, we also know how to express the covariance between the process and its increments (54). Since the stationary distribution is $\mu_\theta \sim N(0, \Sigma_\theta)$, we have all the information that is needed to simulate it. In this case, to compute $\hat{\theta}_n$ defined in (2) we want to minimise

$$F : \theta \mapsto d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{t_k}^{\theta_0}}, \mu_\theta\right). \quad (66)$$

We adapt the technique described in (Panloup et al, 2020, Equations (7.5)-(7.6)). Taking $d = d_{CF,p}$, the idea is to write the functional F as

$$F(\theta) = d_{CF,p}(\mu, \mu_\theta) = \mathbb{E}[|\mu(f_\Phi) - \mu_\theta(f_\Phi)|^2]. \quad (67)$$

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 (5)). Writing F like this allows to perform a gradient descent algorithm. In fact, an approximation of the gradient ∇F is formally obtained as

$$\widehat{\nabla} F := \frac{1}{R} \sum_{r=1}^R \Lambda(\theta, \Phi^r),$$

where $(\Phi^r)_{r=1, \dots, R}$ is a sequence of i.i.d random variables with law g_p and

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

Hence the gradient algorithm reads

$$\forall t \in \{0, \dots, T\}, \theta_{t+1} = \theta_t - \eta_t \frac{1}{R} \sum_{r=1}^R \Lambda(\theta_t, \Phi_{t+1}^r), \quad (68)$$

where $(\eta_t)_t$ is a sequence of positive steps and at each gradient step t , $(\Phi_t^r)_{t,r}$ is a sequence of i.i.d random variables with law g_p . Moreover, to compute $\hat{\theta}_{n,N,\gamma}$ defined

in (3), we want to minimise

$$F : \theta \mapsto d\left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_{t_k}^{\theta,0}}, \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{t_k}^{\theta,\gamma}}\right).$$

In this case, write $F(\theta) = \mathbb{E}[|\mu(f_\Phi) - \frac{1}{N} \sum_{k=0}^{N-1} e^{i\langle X_{k\gamma}^{\theta,\gamma}, \Phi \rangle}|^2]$. The gradient Λ can be written as in (Panloup et al, 2020, Eq (7.6)):

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

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

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

Therefore, the question is how to simulate paths of the process $\partial_{\theta^i} X_{k\gamma}^{\theta,\gamma}$. In Panloup et al (2020) the authors handle the case when θ^i is the drift parameter ξ and explain how the process can be simulated recursively as

$$\partial_\xi U_{(k+1)\gamma}^{\theta,\gamma} = \partial_\xi U_{k\gamma}^{\theta,\gamma} + \gamma \left(\partial_\xi b_\xi(U_{k\gamma}^{\theta,\gamma}) + \nabla b_\xi(U_{k\gamma}^{\theta,\gamma}) \partial_\xi U_{k\gamma}^{\theta,\gamma} \right).$$

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

$$\partial_\sigma U_{(k+1)\gamma}^{\theta,\gamma} = \partial_\sigma U_{k\gamma}^{\theta,\gamma} + \gamma \nabla b_\xi(U_{k\gamma}^{\theta,\gamma}) \partial_\sigma U_{k\gamma}^{\theta,\gamma} + (B_{(k+1)\gamma} - B_{k\gamma}).$$

Finally, in order to compute $(\partial_H U_{k\gamma})_{k=0,\dots,N}$ in the same way, one needs to compute $\partial_H B$, which is not an obvious task. For instance, using the Mandelbrot-Van Ness representation (6), one cannot simply differentiate the integrand with respect to H to get $\partial_H B$. In Koch and Neuenkirch (2019), 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 . In Haress and Richard (2022), it is shown that the solution to (9) is $(1 - \varepsilon)$ -Hölder continuous in H . So in this case, one simply approximates the gradient of $\theta \mapsto F(\theta)$ by a finite difference with step $\delta = 0.1$. Denoting this approximation by $\widehat{\Lambda}(\theta, \Phi)$, the gradient algorithm reads as in (68) with $\widehat{\Lambda}(\theta_t, \Phi_t^r)$ in place of $\Lambda(\theta_t, \Phi_t^r)$.

Simulation of the variable Φ .

Since g_p has a spherical form, Φ can be simulated using the spherical coordinates and the inverse transform sampling method in any dimension (see e.g. (Panloup et al, 2020, Section 7) for $d = 2$).

Numerical illustrations.

Recall that we consider the process U given by (17) and we assume θ to be in a compact interval. The assumptions \mathbf{A}_0 and \mathbf{A}_1 are clearly satisfied, where \mathbf{I}_w follows from Proposition 2.11. Moreover, Lemma 2.12 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 and an Euler approximation $X^{\theta,\gamma}$. We set the following parameters:

$$\begin{aligned}\theta_0 &= (\xi_0, \sigma_0, H_0) = (2, 0.5, 0.7) \\ \underline{h} &= 10^{-3}, \quad q = 3, \quad h = 10^{-1} \\ n &= 1000, \quad N = 10000, \quad \gamma = 10^{-1} \\ R &= 100, \quad T = 100, \quad p = 2.\end{aligned}$$

Let us start with the one-dimensional case (Figure 1). We perform the gradient descents described above over 100 realisations of the observations $(X_{kh}^{\theta_0})_{k=1,\dots,n}$ and plot a histogram highlighting the empirical mean obtained and the empirical variance. More precisely, we denote the empirical variance when estimating ξ_0, σ_0 and H_0 respectively by $\text{Var}_\xi, \text{Var}_\sigma$ and Var_H . We use the same algorithm in the two-dimensional case (Figures 2, 3 and 4) and the three-dimensional case (Figure 5) and plot histograms for all the parameters we are interested in estimating. Take $\theta^0 = (\xi^0 = 1, \sigma^0 = 1, H^0 = 0.5)$ as the initial point in the gradient descents.

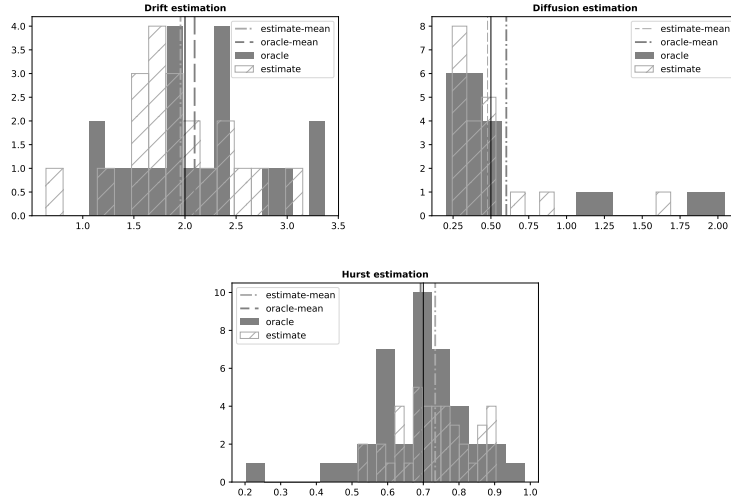


Fig. 1 Histograms for the estimation of each parameter separately. The filled vertical lines are for the true parameters. The dash lines represent the empirical mean of $\hat{\theta}_n$ and the dash-dotted lines the empirical mean of $\hat{\theta}_{n,N,\gamma}$. Left: $\text{Var}_\xi(\hat{\theta}_n) \sim 0.1$, $\text{Var}_\xi(\hat{\theta}_{n,N,\gamma}) \sim 0.1$. Right: $\text{Var}_\sigma(\hat{\theta}_n) \sim 0.01$, $\text{Var}_\sigma(\hat{\theta}_{n,N,\gamma}) \sim 0.01$. Bottom: $\text{Var}_H(\hat{\theta}_n) \sim 0.01$, $\text{Var}_H(\hat{\theta}_{n,N,\gamma}) \sim 0.01$.

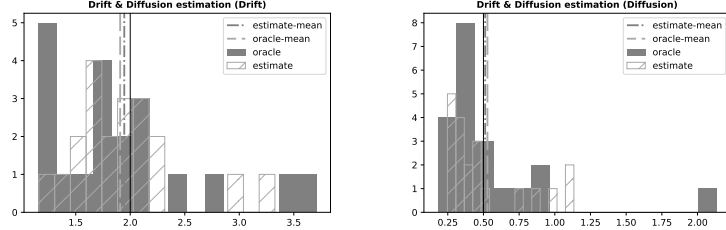


Fig. 2 Histograms for the estimation of the drift and the diffusion. The filled vertical lines are for the true parameters. The dash lines represent the empirical mean of $\hat{\theta}_n$ and the dash-dotted lines the empirical mean of $\hat{\theta}_{n,N,\gamma}$. Left: $\text{Var}_\xi(\hat{\theta}_n) \sim 0.1$, $\text{Var}_\xi(\hat{\theta}_{n,N,\gamma}) \sim 0.1$. Right: $\text{Var}_\sigma(\hat{\theta}_n) \sim 0.01$, $\text{Var}_\sigma(\hat{\theta}_{n,N,\gamma}) \sim 0.01$.

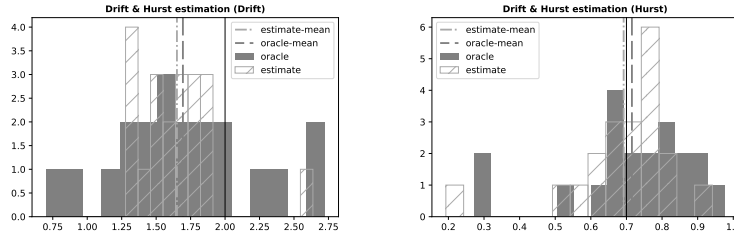


Fig. 3 Histograms for the estimation of the drift and the Hurst parameter. The filled vertical lines are for the true parameters. The dash lines represent the empirical mean of $\hat{\theta}_n$ and the dash-dotted lines the empirical mean of $\hat{\theta}_{n,N,\gamma}$. Left: $\text{Var}_\xi(\hat{\theta}_n) \sim 0.1$, $\text{Var}_\xi(\hat{\theta}_{n,N,\gamma}) \sim 0.1$. Right: $\text{Var}_H(\hat{\theta}_n) \sim 0.01$, $\text{Var}_H(\hat{\theta}_{n,N,\gamma}) \sim 0.01$.

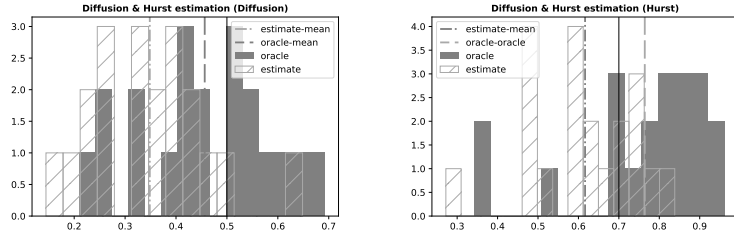


Fig. 4 Histograms for the estimation of the diffusion and the Hurst parameter. The filled vertical lines are for the true parameters. The dash lines represent the empirical mean of $\hat{\theta}_n$ and the dash-dotted lines the empirical mean of $\hat{\theta}_{n,N,\gamma}$. Left: $\text{Var}_\sigma(\hat{\theta}_n) \sim 0.01$, $\text{Var}_\sigma(\hat{\theta}_{n,N,\gamma}) \sim 0.01$. Right: $\text{Var}_H(\hat{\theta}_n) \sim 0.01$, $\text{Var}_H(\hat{\theta}_{n,N,\gamma}) \sim 0.01$.

Discussion.

In the $1d$ case, we get accurate estimators of the parameters (see Figure 1). In the $2d$ case, one observes a decrease in the accuracy of the estimates (see the left histogram on Figure 3 and Figure 4 for instance), which is due to a higher bias, as the variances of the estimators stay the same as in the $1d$ case. In the $3d$ case, one observes an

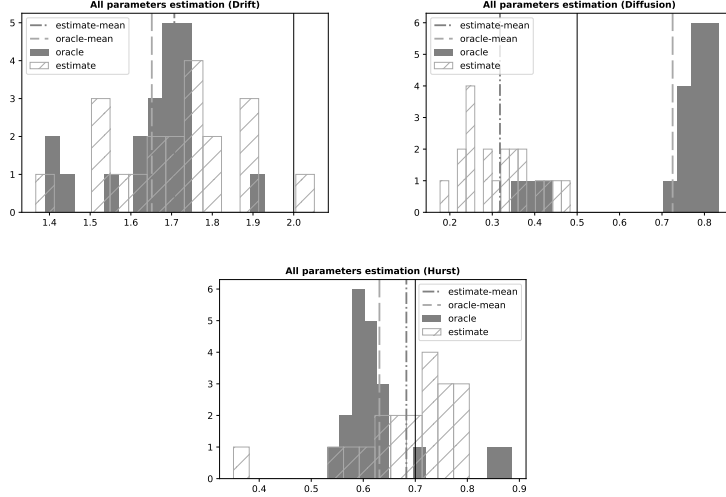


Fig. 5 Histograms for the estimation of all the parameters. The filled vertical lines are for the true parameters. The dash lines represent the empirical mean of $\hat{\theta}_n$ and the dash-dotted lines the empirical mean of $\hat{\theta}_{n,N,\gamma}$. Left: $\text{Var}_\xi(\hat{\theta}_n) \sim 0.1$, $\text{Var}_\xi(\hat{\theta}_{n,N,\gamma}) \sim 0.1$. Right: $\text{Var}_\sigma(\hat{\theta}_n) \sim 0.01$, $\text{Var}_\sigma(\hat{\theta}_{n,N,\gamma}) \sim 0.01$. Bottom: $\text{Var}_H(\hat{\theta}_n) \sim 0.01$, $\text{Var}_H(\hat{\theta}_{n,N,\gamma}) \sim 0.01$.

increase in both the bias and the variances of the estimators (specially in the estimation of the drift). Furthermore, observe that for $N = 10000$ and $\gamma = 0.1$, the estimator $\hat{\theta}_{n,N,\gamma}$ behaves qualitatively like $\hat{\theta}_n$. While $\hat{\theta}_{n,N,\gamma}$ is slow to compute, this ensures that the error generated by replacing $\theta \mapsto \mu_\theta$ by its approximation is not significant. Finally, note that in practice in the $3d$ case, the oracle $\hat{\theta}_n$ is slow to compute. In fact, computing $\theta \mapsto \mu_\theta$ requires computing the integrals appearing in the right-hand side of (55) which are slow to converge. Moreover, the error when approximating these integrals can be significant as shown in Figure 4. Thus, even when the invariant measure is known, it can be more efficient to compute $\hat{\theta}_{n,N,\gamma}$ than the oracle in higher dimensions. In general, further exploration and in-depth analysis would be necessary to enhance the integration of our statistical procedure with gradient descent algorithms. In particular, going beyond the fractional Ornstein-Uhlenbeck model requires looking into derivative-free methods to approximate $\nabla\Lambda(\theta, \phi)$ (e.g. Flaxman et al (2005)) and faster methods to compute the approximation of $\theta \mapsto \mu_\theta$. This aspect remains open for future investigation.

Declarations

Funding

This work is supported by the Labex Mathématique Hadamard and the SIMALIN project ANR-19-CE40-0016 and the SDAIM project ANR-22-CE40-0015 from the French National Research Agency.

Conflict of interest/Competing interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval

All authors declare to comply with the Journal Ethics.

Consent to participate

All the authors consent to participate in this research.

Appendix A Regularity in the Hurst parameter

In this section, we recall and adapt some results from our companion paper (Haress and Richard, 2022, Sections 4 and 5) that state the regularity in the Hurst parameter of continuous and discrete ergodic means. Recall that the fractional OU process is defined by (17), and let us denote by $\bar{U}^{(1,\sigma,H)}$ the stationary fractional OU process with drift $\xi = 1$, diffusion matrix σ and Hurst parameter H .

In the whole Appendix, let \mathcal{H} be a compact subset of $(0, 1)$, Ξ be a compact subset of \mathbb{R}^m , Σ a compact subset of the space of $d \times d$ invertible matrices and denote $\Theta = \Xi \times \Sigma \times \mathcal{H}$.

Lemma A.1. *Let $\varpi \in (0, 1)$ and $p \geq 1$. Let W be an \mathbb{R}^d -Brownian motion and for any $H \in (0, 1)$, denote by B^H the fBm with underlying noise W (i.e. as in (6)). There exists a random variable \mathbf{C} with a finite moment of order p such that almost surely, for any $t \geq 0$, any $\theta_1 = (\xi, \sigma, H_1) \in \Theta$ and $\theta_2 = (\xi, \sigma, H_2) \in \Theta$,*

$$\frac{1}{t+1} \int_0^{t+1} |Y_s^{\theta_1} - Y_s^{\theta_2}|^2 ds \leq \mathbf{C} |H_1 - H_2|^\varpi,$$

where Y^{θ_1} (resp. Y^{θ_2}) is the solution to (9) with parameter θ_1 (resp. θ_2), a drift b_ξ satisfying A₁ and driving fBm B^{H_1} (resp. B^{H_2}), and Y^{θ_1} and Y^{θ_2} start from the same initial condition.

Proof. For $i = 1, 2$, the process $\sigma^{-1}Y^{\theta_i}$ is solution to the SDE

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

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 , \tilde{b}_ξ still satisfies (7) and (8). We choose the stationary fractional OU $\bar{U}^{(1,\text{Id},H_1)}$ with the same noise B^{H_1} as Y^{θ_1} (similarly for $\bar{U}_s^{(1,\text{Id},H_2)}$). As in the proof of (Haress and Richard, 2022, Theorem 4.5), a comparison between $Y_s^{\theta_1} - Y_s^{\theta_2}$ and

$\bar{U}^{(1,\text{Id},H_1)} - \bar{U}^{(1,\text{Id},H_2)}$ gives

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

We can now apply (Haress and Richard, 2022, Proposition 4.2) with $t' = t = 0$ and (Haress and Richard, 2022, 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

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

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

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}_γ with a finite moment of order p such that almost surely, for all $t, t' \geq 0$ and all $H_1, H_2 \in \mathcal{H}$,*

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

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

Proof. Apply (Haress and Richard, 2022, Proposition 5.1), with $t' = t$ and $H' = K' = K$ to get that

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

Now apply (Haress and Richard, 2022, Proposition 4.2) with $t = t' = 0$ to get the desired result. \square

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}_γ with a finite moment of order p such that almost surely, for any $N \in \mathbb{N}^*$, any $\theta = (\xi, \sigma, H_1)$ and any $\theta_2 = (\xi, \sigma, H_2) \in \Theta$,*

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

where $Y^{\theta_1, \gamma}$ and $Y^{\theta_2, \gamma}$ are Euler schemes (15) with the same initial condition and driven by fBm with the same underlying noise (see (6)).

Proof. For any $\theta \in \Theta$, the process $\sigma^{-1}Y^{\theta,\gamma}$ 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}_ξ still satisfies (7) and (8). As in the proof of (Haress and Richard, 2022, Eq. (5.5)), a comparison with the stationary fractional OU process \bar{U} gives

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

The regularity of the second term in the right-hand side is given by (Haress and Richard, 2022, Proposition 4.2) and the regularity of the third term is given by (Haress and Richard, 2022, Theorem 4.5). To bound the first term, we apply Lemma A.2. To conclude the proof, we notice that $|\sigma^{-1}(Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma})| \geq |\sigma^{-1}| |(Y_{k\gamma}^{\theta_1} - Y_{k\gamma}^{\theta_2})|$, divide by $|\sigma^{-1}|$ and take the supremum over Σ . \square

Appendix 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 (9) and (15) and their ergodic means have finite moments uniformly in time and θ . Finally, in Proposition B.3, we state a result on the regularity of the ergodic means in θ .

Proposition B.1. *Assume \mathbf{A}_0 and \mathbf{A}_1 . Let Y^θ be the unique solution of (9). Let $p > 1$. Then the following inequalities hold true:*

- (i) $\sup_{t \geq 0} \sup_{\theta \in \Theta} \mathbb{E} |Y_t^\theta|^p < \infty$.
- (ii) $\mathbb{E} \left(\sup_{t \geq 0} \sup_{\theta \in \Theta} \frac{1}{t} \int_0^t |Y_s^\theta|^2 ds \right)^p < \infty$.
- (iii) $\mathbb{E} \left(\sup_{n \in \mathbb{N}^*} \sup_{\theta \in \Theta} \frac{1}{n} \sum_{k=0}^{n-1} |Y_{kh}^\theta|^2 \right)^p < \infty$.

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 (Panloup et al, 2020, Proposition A.1). The proofs of all three items are based on a comparison with fractional OU processes (17).

For the proof of (i), by (Hairer, 2005, p.725), a comparison with the stationary fractional OU process $\bar{U}^{(1,\sigma,H)}$ yields that there exist constants $c_1, c_2 > 0$ independent of ξ such that,

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

Moreover, since $U^{(1,\sigma,H)}$ is a Gaussian process, for any $t \geq 1$, we have $\mathbb{E}|\bar{U}_t^{(1,\sigma,H)}|^p \lesssim (\mathbb{E}|\bar{U}_t^{(1,\sigma,H)}|^2)^{p/2}$. By (18), we know that $\mathbb{E}|\bar{U}_t^{(1,\sigma,H)}|^2 = \sigma^2 H \Gamma(2H)$. Therefore

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

For the proof of (ii), we follow the steps of the proof of Proposition A.1 in Panloup et al (2020) (see equation (A.6) and what follows), to get that for all $t > 0$,

$$\sup_{\theta \in \Theta} \frac{1}{t} \int_0^t |Y_s^\theta|^2 ds \leq C \sup_{\theta \in \Theta} \frac{1}{t} \int_0^t |U_s^\theta|^2 ds \leq C \sup_{\theta=(1,\sigma,H) \in \Theta} \frac{1}{t} \int_0^t |\sigma| |U_s^{(1,\text{Id},H)}|^2 ds.$$

It follows that

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

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

$$\frac{1}{t} \int_0^t \sup_{\theta \in \Theta} |Y_s^\theta|^2 ds \leq C \left(\mathbf{C} \sup_{H \in \mathcal{H}} |H - \frac{1}{2}|^\varpi + \frac{1}{t} \int_0^t |U_s^{(1,\text{Id},1/2)}|^2 ds \right).$$

The ergodicity of $U^{(1,\text{Id},1/2)}$ implies that $\frac{1}{t} \int_0^t |U_s^{(1,\text{Id},1/2)}|^2 ds$ converges as $t \rightarrow \infty$. It follows that

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

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. \square

Proposition B.2. *Assume \mathbf{A}_0 and \mathbf{A}_1 . Let $Y^{\theta,\gamma}$ be the Euler scheme (15). Then there exists $\gamma_0 > 0$ such that for any $p > 1$, there is*

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

Proof. Note that the same results are proven in (Panloup et al, 2020, 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. \square

Proposition B.3. *Let the assumptions \mathbf{A}_0 and \mathbf{A}_1 hold. Assume also that the exponent r in the sub-linear growth of b_ξ in (8) satisfies $r \leq 1$. Let $p \geq 1$ and $\varpi \in (0, 1)$, then there exists a positive random variable \mathbf{C} that has a finite moment of order p , such that almost surely for all $\theta_1, \theta_2 \in \Theta$ and for all $t \geq 1$,*

$$\frac{1}{t} \int_0^t |Y_s^{\theta_1} - Y_s^{\theta_2}|^2 ds \leq \mathbf{C} |\theta_1 - \theta_2|^\varpi, \quad (\text{B1})$$

where Y^{θ_1} and Y^{θ_2} are solutions to (9) with the same initial condition and driven by an fBm with the same underlying noise (see (6)). Furthermore, there exists γ_0 such that for any $\gamma \in (0, \gamma_0]$, there exists a positive random variable \mathbf{C}_γ that has a finite moment of order p , such that almost surely, for any $\theta_1, \theta_2 \in \Theta$ and any $N \geq 1$,

$$\frac{1}{N} \sum_{k=0}^N |Y_{k\gamma}^{\theta_1, \gamma} - Y_{k\gamma}^{\theta_2, \gamma}|^2 ds \leq \mathbf{C}_\gamma |\theta_1 - \theta_2|^\varpi. \quad (\text{B2})$$

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 moment of order p , that does not depend on θ and may change from line to line. Let us first focus on the proof of (B1). Up to introducing pivot terms, we can consider three different cases:

- (1) $\theta_1 = (\xi_1, \sigma, H), \theta_2 = (\xi_2, \sigma, H)$
- (2) $\theta_1 = (\xi, \sigma, H_1), \theta_2 = (\xi, \sigma, H_2)$
- (3) $\theta_1 = (\xi, \sigma_1, H), \theta_2 = (\xi, \sigma_2, H).$

In the first case, we have by (Panloup et al, 2020, Eq. (5.32)) that

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

where r is the exponent in the sub-linear growth assumption on b_ε . Since $r \leq 1$,

$$\frac{1}{t} \int_0^t |Y_s^{\theta_1} - Y_s^{\theta_2}|^2 ds \leq C |\xi_1 - \xi_2|^2 \left(1 + \sup_{\theta \in \Theta} \frac{1}{t} \int_0^t |Y_s^\theta|^2 ds \right). \quad (\text{B3})$$

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

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

The second case (2) is directly the result of Lemma A.1.

As for the third case (3), the idea is to compare the process Y with the fractional OU processes $U^{(1,\sigma_1,H)}$ and $U^{(1,\sigma_2,H)}$ defined by (17) with the same initial condition and the same driving fBm. For $s \geq 1$, it comes

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

where the last inequality follows from A₁. Next, apply Young's inequality to get

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

We can now apply Grönwall's lemma to get

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

Jensen's inequality yields that

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

Then, using Fubini's theorem, it comes that

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

$$\leq \frac{C}{t} \int_0^t |U_u^{(1,\sigma_1,H)} - U_u^{(1,\sigma_2,H)}|^2 du. \quad (\text{B4})$$

Now, observe that $U_s^{(1,\sigma_1,H)} - U_s^{(1,\sigma_2,H)} = (\sigma_1 - \sigma_2)U_s^{(1,\text{Id},H)}$. Since $U_s^{(1,\text{Id},H)}$ has finite moments uniformly in θ (recall (18) and that $U^{(1,\text{Id},H)}$ 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 . This concludes the proof of (B1).

The proof of (B2) is obtained using discrete analogues of the previous arguments. More precisely, in the first case (1), similarly to (B3), we have from (Panloup et al, 2020, Proposition 3.8 (ii)) that

$$\frac{1}{N} \sum_{k=0}^N |Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma}|^2 \leq C |\xi_1 - \xi_2|^2 \left(1 + \sup_{\theta \in \Theta} \frac{1}{N} \sum_{k=0}^N |Y_{k\gamma}^{\theta,\gamma}|^2 \right).$$

While the dependence of the RHS on σ and H in Panloup et al (2020) is not explicit, one can show that the upper bound they obtain in the continuous setting (i.e. (Panloup et al, 2020, Eq. (5.32))) still holds if the integrals are replaced by discrete sums. Then conclude using the uniform bound on the moments of $Y^{\theta,\gamma}$ in Proposition B.2(ii).

The second case (2) is directly the result of Lemma A.3. In the third case (3), similarly to (B4), via a comparison with the discrete fractional OU process $U^{\theta,\gamma}$ (that solves (15) with $b_\xi(\cdot) = -\xi \cdot$), one can show that

$$\frac{1}{N} \sum_{k=0}^N |Y_{k\gamma}^{\theta_1,\gamma} - Y_{k\gamma}^{\theta_2,\gamma}|^2 \leq \frac{C}{N} \sum_{k=0}^N |U_{k\gamma}^{(1,\sigma_1,H),\gamma} - U_{k\gamma}^{(1,\sigma_2,H),\gamma}|^2.$$

Then we use the linearity of $U^{(1,\sigma,H),\gamma}$ in σ to conclude. \square

Appendix C Proof of Proposition 2.4

The proof follows the same steps as the proof of Lemma 2.2. Let $\theta = (\xi, \sigma, H) \in \Theta$. 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 $t \rightarrow \infty$. 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 Prokhorov distance d_P as in (23):

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

$$\times d_P \left(\frac{1}{t} \int_0^t \delta_{X_s^{\theta,\gamma}} ds, \mu_\theta \right).$$

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

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

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 $\{\pi_N = \frac{1}{N} \sum_{k=0}^{N-1} \delta_{X_{k\gamma+}}^{\theta,\gamma}\}_{N \geq 0}$. We first prove that the family is tight and then identify the limit as the stationary law of the augmented process $\bar{X}^{\theta,\gamma}$. Tightness in $D(\mathbb{R}_+, \mathbb{R}^d)$, the space of functions that are right-continuous and have limits from the left is equivalent to tightness in $D([0, T], \mathbb{R}^d)$ for every $T > 0$. Thus by (Billingsley, 1999, Theorem 13.2), tightness is equivalent to the following two points that must hold for any $T > 0$:

- (i) $\lim_{a \rightarrow \infty} \limsup_{N \rightarrow \infty} \pi_N \left(x : \sup_{t \in [0, T]} |x_t| \geq a \right) = 0$.
- (ii) Denote $w'_T(x, \delta) = \inf_{\{t_i\}} \left\{ \max_{i \leq r} \sup_{s, t \in [t_i, t_{i+1})} |x_t - x_s| \right\}$. Then for any $\eta > 0$, almost surely,

$$\limsup_{\delta \rightarrow 0} \limsup_{N \rightarrow +\infty} \pi_N (x : w'_T(x, \delta) \geq \eta) = \limsup_{\delta \rightarrow 0} \limsup_{N \rightarrow +\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{1}_{\{w'_T(X_{k\gamma+}^{\theta,\gamma}, \delta) \geq \eta\}} = 0,$$

where the infimum runs over finite sets $\{t_i\}_{i=1, \dots, r}$, $r \in \mathbb{N}^*$, satisfying

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

Since the process has only jumps at times $n\gamma$ with $n \in \mathbb{N}$, $w'_T(X^{\theta,\gamma}, \delta) = 0$ when $\delta < \gamma$, which implies that the second condition (ii) holds.

The first condition (i) is equivalent to tightness in the space of probability measures on \mathbb{R} of the sequence $(\mu_T^N)_{N \in \mathbb{N}^*}$ defined by

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

Recall that by definition of X we have $|X_{k\gamma}^{\theta,\gamma}|^2 \leq C_q \sum_{i=0}^q |Y_{k\gamma+ih}^{\theta,\gamma}|^2 \leq C_q \sup_{i \in [0, q]} |Y_{k\gamma+ih}^{\theta,\gamma}|^2$. Hence, for $V(x) = |x|^2$ and

$$\tilde{\mu}_T^N = \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\{\sup_{t \in [0, T+qh]} |Y_{k\gamma+t}^{\theta,\gamma}|\}},$$

we deduce that $\mu_T^N(V) \leq C_q \tilde{\mu}_T^N(V)$. From the last equation in the proof of (Cohen and Panloup, 2011, Proposition 2), one has $\sup_{N \geq 1} \tilde{\mu}_T^N(V) < +\infty$ almost surely, which implies that $(\mu_T^N)_{N \geq 1}$ is a.s. tight on \mathbb{R} (see *e.g.* (Duflo, 1997, Proposition 2.1.6)).

Now let $(t_n)_{n \geq 1}$ be an increasing sequence going to $+\infty$ and $\{\frac{1}{t_n} \sum_{k=0}^{t_n-1} \delta_{X_{k\gamma}^{\theta, \gamma}}\}_{n \geq 1}$ be a (pathwise) sequence with limiting distribution ρ . As in Appendix A.2 of (Panloup et al, 2020, Proposition 3.3), we get that γ is stationary. Let us now prove that ρ is the law of $\bar{X}^{\theta, \gamma}$.

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

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

$$z_{\cdot}^i - \ell^i \left(\int_0^{\cdot \gamma} b_{\xi}(y_u) du, \dots, \int_0^{(\cdot + ih)\gamma} b_{\xi}(y_u) du \right) \text{ has the law of } \sigma \ell^i(B_{\cdot \gamma}, \dots, B_{(\cdot + ih)\gamma})$$

for all $i \in \llbracket 1, q \rrbracket$,

where for all $t \geq 0$, $t_{\gamma} = \gamma \lfloor t/\gamma \rfloor$. Now one can proceed as in the end of Section 3.2 to deduce that ρ is the law of $\bar{X}^{\theta, \gamma}$.

References

- Belfadli R, Es-Sebaiy K, Ouknine Y (2011) Parameter estimation for fractional Ornstein-Uhlenbeck processes: non-ergodic case. Preprint [arXiv:11025491](https://arxiv.org/abs/11025491)
- Berzin C, Latour A, León JR (2015) Variance estimator for fractional diffusions with variance and drift depending on time. Electron J Stat 9(1):926–1016
- Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication
- Brouste A, Iacus SM (2013) Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. Comput Statist 28(4):1529–1547
- Cheridito P, Kawaguchi H, Maejima M (2003) Fractional Ornstein-Uhlenbeck processes. Electron J Probab 8:no. 3, 14
- Cohen S, Panloup F (2011) Approximation of stationary solutions of Gaussian driven stochastic differential equations. Stochastic Process Appl 121(12):2776–2801
- Duflo M (1997) Random iterative models, Applications of Mathematics (New York), vol 34. Springer-Verlag, Berlin, translated from the 1990 French original by Stephen S. Wilson and revised by the author

- Flaxman AD, Kalai AT, McMahan BH (2005) Online convex optimization in the bandit setting: gradient descent without a gradient. In: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, pp 385–394
- Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. *International statistical review* 70(3):419–435
- Giordano LM, Jolis M, Quer-Sardanyons L (2020) SPDEs with fractional noise in space: continuity in law with respect to the Hurst index. *Bernoulli* 26(1):352–386
- Gloter A, Hoffmann M (2007) Estimation of the Hurst parameter from discrete noisy data. *Ann Statist* 35(5):1947–1974
- Hairer M (2005) Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann Probab* 33(2):703–758.
- Haress EM, Hu Y (2021) Estimation of all parameters in the fractional Ornstein-Uhlenbeck model under discrete observations. *Stat Inference Stoch Process* 24(2):327–351
- Haress EM, Richard A (2022) Long time Hurst regularity of fractional SDEs and their ergodic means. Preprint [arXiv:2206.06648](https://arxiv.org/abs/2206.06648)
- Hu Y, Nualart D (2010) Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist Probab Lett* 80(11-12):1030–1038
- Hu Y, Song J (2013) Parameter estimation for fractional Ornstein-Uhlenbeck processes with discrete observations. In: *Malliavin calculus and stochastic analysis*, Springer Proc. Math. Stat., vol 34. Springer, New York, p 427–442
- Hu Y, Nualart D, Zhou H (2019) Parameter estimation for fractional Ornstein-Uhlenbeck processes of general Hurst parameter. *Stat Inference Stoch Process* 22(1):111–142
- Jolis M, Viles N (2007) Continuity with respect to the Hurst parameter of the laws of the multiple fractional integrals. *Stochastic Process Appl* 117(9):1189–1207.
- Jolis M, Viles N (2010) 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
- Koch S, Neuenkirch A (2019) 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.
- Kubilius K, Mishura Y (2012) The rate of convergence of Hurst index estimate for the stochastic differential equation. *Stochastic Process Appl* 122(11):3718–3739

- Panloup F, Tindel S, Varvenne M (2020) A general drift estimation procedure for stochastic differential equations with additive fractional noise. *Electron J Stat* 14(1):1075–1136
- Prakasa Rao BLS (2010) *Statistical inference for fractional diffusion processes*. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester
- Richard A (2015) A fractional Brownian field indexed by L^2 and a varying Hurst parameter. *Stochastic Process Appl* 125(4):1394–1425
- Richard A, Talay D (2017) Noise sensitivity of functionals of fractional Brownian motion driven stochastic differential equations: results and perspectives. In: *Modern problems of stochastic analysis and statistics*, Springer Proc. Math. Stat., vol 208. Springer, Cham, p 219–235,
- Richard A, Talay D (2022) Lipschitz continuity in the Hurst parameter of functionals of Stochastic Differential Equations driven by fractional Brownian motion. Preprint [arXiv:160503475v4](https://arxiv.org/abs/160503475v4)
- Tudor CA, Viens FG (2007) Statistical aspects of the fractional stochastic calculus. *Ann Statist* 35(3):1183–1212
- Varvenne M (2019) Concentration inequalities for stochastic differential equations with additive fractional noise. *Electron J Probab* 24:Paper No. 124, 22
- Xiao W, Zhang W, Xu W (2011) Parameter estimation for fractional Ornstein-Uhlenbeck processes at discrete observation. *Appl Math Model* 35(9):4196–4207