HAL
open science

A GENERAL DRIFT ESTIMATION PROCEDURE FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH ADDITIVE FRACTIONAL NOISE

Fabien Panloup, Samy Tindel, Maylis Varvenne

To cite this version:

Fabien Panloup, Samy Tindel, Maylis Varvenne. A GENERAL DRIFT ESTIMATION PROCEDURE FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH ADDITIVE FRACTIONAL NOISE. Electronic Journal of Statistics , 2020, 14 (1), pp.1075-1136. 10.1214/20-EJS1685 . hal-02077420v2

HAL Id: hal-02077420

https://hal.science/hal-02077420v2

Submitted on 26 Feb 2020

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

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

A GENERAL DRIFT ESTIMATION PROCEDURE FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH ADDITIVE FRACTIONAL NOISE

FABIEN PANLOUP, SAMY TINDEL, AND MAYLIS VARVENNE

Abstract. In this paper we consider the drift estimation problem for a general differential equation driven by an additive multidimensional fractional Brownian motion, under ergodic assumptions on the drift coefficient. Our estimation procedure is based on the identification of the invariant measure, and we provide consistency results as well as some information about the convergence rate. We also give some examples of coefficients for which the identifiability assumption for the invariant measure is satisfied.

Contents

1. Introduction 2
2. Main results 3
2.1. Notation 3
2.2. Assumptions 4
2.3. Statistical setting and construction of the estimator 6
2.4. Main consistency results 7
2.5. Rate of convergence 8
3. Preliminary Results 9
3.1. Ergodic properties of the SDE and of the Euler scheme 9
3.2. Continuity of $\vartheta \mapsto d(\nu_\vartheta, \nu_{\vartheta_0})$ 12
3.3. Further controls under $(\mathbf{C_s})$ 14
4. Proof of the consistency theorems 17
4.1. Uniform convergence of the contrast 17
4.2. A general convergence result 23
4.3. Proofs of the convergence theorems 24
5. Rate of convergence: proof of Theorem 2.13 24
5.1. L^q bound on $D_n^{(1)}$ 25
5.2. L^q bound on $D_{N,\gamma}^{(2)}$ 27
5.3. L^q bound on $D_{N,\gamma}^{(3)}$ 28
6. Identifiability assumption 30

Date: February 26, 2020.
2010 *Mathematics Subject Classification.* 62M09; 62F12.
Key words and phrases. fractional Brownian motion, parameter estimation, least square procedure, ergodicity.
S. Tindel is supported by the NSF grant DMS-1613163.

6.1. Case of a diffusion process 31
6.2. Fractional Brownian motion case 31
7. Numerical Discussions and Illustrations 35
Appendix A. Tightness and convergence of the occupation measures 40
A.1. Moment estimates 40
A.2. Proof of Proposition 3.3 45
Appendix B. Proof of Proposition 3.8 47
Acknowledgements 48
References 48

1. Introduction

Let B be a d-dimensional fractional Brownian motion with Hurst parameter $H \in (0,1)$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We recall that B is a centered Gaussian process. Its law is thus characterized by its covariance function, which is defined by

$$\mathbf{E}[B_t^i B_s^j] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right) \mathbf{1}_{\{0\}}(i-j), \qquad s,t \in \mathbb{R}.$$

The variance of the increments of B is then given by

$$\mathbf{E}\left[|B_t^i - B_s^i|^2\right] = |t-s|^{2H}, \qquad s,t \in \mathbb{R}, \quad i = 1,\dots,m, \tag{1.1}$$

and this implies that almost surely the fBm paths are γ-Hölder continuous for any $\gamma < H$.

In this article, we will consider the following $\mathbb{R}^d$-valued stochastic differential equation driven by B:

$$Y_t = y_0 + \int_0^t b_{\vartheta_0}(Y_s)\, ds + \sigma B_t, \qquad t \in [0,T]. \tag{1.2}$$

Here $y_0 \in \mathbb{R}^d$ is a given initial condition, $B = (B^1, \dots, B^d)$ is the aforementioned fractional Brownian motion (fBm), the unknown parameter ϑ_0 lies in a certain set Θ which will be specified later on, $\{b_\vartheta(\cdot),\, \vartheta \in \Theta\}$ is a known family of drift coefficients with $b_\vartheta(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, and σ is a $d \times d$-matrix which is supposed to be known. More precisely, we do not discuss here the problem of estimation of the diffusion parameter σ and of the Hurst index H (on this topic, see *e.g.* [8], [29] or [46]). For the sake of simplicity, we also assume throughout the paper that σ is invertible (on this topic, see Remark 2.4). Our aim is to get an accurate estimation of ϑ_0 according to some discrete observations of Y.

When $H = \frac{1}{2}$, *i.e.* when $(Y_t)_{t\geq 0}$ is a diffusion process, the drift estimation for solutions of stochastic differential equations is a widely investigated topic. Among many references devoted to parametric estimation procedures for discretely observed elliptic diffusions, let us quote the early contribution [30]. For adaptive and non parametric methods, one may refer to [13, 15, 27, 33, 38, 44] and to the references therein. In the fractional setting, the estimation problem for the drift term in equation (1.2) has also received a lot of attention in the recent past (see e.g [3, 28, 32, 34, 39, 45]). However, the following restrictions hold in all those contributions:

- The coefficient $b_{\vartheta_0}(Y_s)$ is of the form $\vartheta_0 b(Y_s)$ or even $\vartheta_0\, Y_s$ when Ornstein-Uhlenbeck processes are involved.
- The observation is either in continuous time or a discretized version of continuous observations.

• Rates of convergence of the estimators are not computed, a noticeable exception being the central limit theorems obtained in [28].

Let us also mention the nonparametric method put forward in the interesting recent paper [14]. The context in [14] is much more general than in the aforementioned references, but the estimation procedure is based on the observation of several paths of (1.2). This makes its practical implementation delicate.

With these considerations in mind, let us recall that the article [36] proposed an estimator valid for a wide class of functions b in (1.2), directly based on discrete observations of the process. This estimator is obtained through a least square procedure which is easily implemented. It is fairly general, but still exhibits some gaps that we aim at filling in this paper. Indeed, on the one hand [36] only focuses on the case $H > 1/2$. This obviously facilitates many of the stochastic calculus manipulations invoked in order to analyze the convergence of estimators. On the other hand [36] crucially assumes that the drift $b_\vartheta(\cdot)$ has a gradient shape. This is essential in order to identify the long-time behavior of certain stochastic integrals with respect to the underlying fBm (see Lemma 3.2 of [36] for details). Eventually [36] also required the following *identifiability* assumption:

$$\mathbf{E}\left[|b_{\vartheta_0}(\bar{Y}_0)|^2\right] = \mathbf{E}\left[|b_\vartheta(\bar{Y}_0)|^2\right] \quad \text{iff} \quad \vartheta = \vartheta_0. \tag{1.3}$$

Even though (1.3) can be considered as a quite standard hypothesis in an ergodic framework (see e.g [30] for a similar requirement), asking the class of models to be identifiable in this sense is mostly adapted to the drifts of the form $b_\vartheta(\cdot) = \vartheta h(\cdot)$. It may be restrictive or at least, difficult to check in practice.

In this paper, we thus aim at filling some of the aforementioned gaps. Our initial objective is to alleviate the identifiability condition. More precisely, instead of assuming that the model can be identified through the integral of one very specific function with respect to the invariant distribution, we will now just suppose that the model is identifiable by its invariant distribution, *i.e.* that two models are equal if and only if they have the same stationary distribution. We also get rid of the restriction on H and on the gradient shape of b_ϑ by considering an approach which essentially avoids to use stochastic calculus. In other words, our consistence and rate of convergence results will be mainly based on the study of the long-time behavior of distances between paths built with the same noise. In the additive noise case we are handling, those distances have the advantage to be regular. Eventually, even if our ergodic type assumptions are similar to those in [36] (and called $(\mathbf{C_s})$ in the current paper), we also obtain a (weak) consistence result under a weaker assumption (called $(\mathbf{C_w})$ below), where contractivity of the drift is only assumed out of a compact set.

2. Main results

In this section we will first give some general notation which will be used throughout the paper. Then we will specify our assumptions on the coefficients of (1.2) and describe the estimator we are considering. Eventually we give our almost sure convergence result as well as the convergence rate we have been able to obtain.

2.1. **Notation.** We consider the set of parameters Θ as a subset of $\mathbb{R}^q$ for $q \geq 1$. Let $f : \mathbb{R}^d \times \Theta \to \mathbb{R}$ be a $\mathcal{C}^{p_1,p_2}$ function, where p_1, p_2 are two integers greater than 1. Then for any $p \leq p_1$ and any tuple $(i_1, \dots i_p) \in \{1, \dots, d\}^p$, we set $\partial_x^{i_1 \dots i_p} f$ for $\frac{\partial^p f}{\partial x_{i_1} \dots \partial x_{i_p}}$. Analogously, for

$p \leq p_2$ we use the notation $\partial_\vartheta^{i_1 \dots i_p} f$ for $\frac{\partial^p f}{\partial \vartheta_{i_1} \dots \partial \vartheta_{i_p}}$ whenever $(i_1, \dots i_p) \in \{1, \dots, q\}^p$. Moreover, we will write $\partial_x f$ (resp. $\partial_\vartheta f$) for the Jacobi-matrices $(\partial_{x_1} f, \dots, \partial_{x_d} f)$ (resp. $(\partial_{\vartheta_1} f, \dots, \partial_{\vartheta_q} f)$).

Let $\mathcal{M}_1(\mathbb{R}^d)$ denotes the set of probability measures on $\mathbb{R}^d$. We say that d is a distance on $\mathcal{M}_1(\mathbb{R}^d)$ if it metrizes its usual topology, namely the weak convergence topology. Among those distances we will consider the p-Wasserstein distance, which is defined as follows: for every $\nu, \mu \in \mathcal{M}_1(\mathbb{R}^d)$, we introduce the set $\mathcal{C}(\nu, \mu)$ of couplings between ν and μ, that is

$$\mathcal{C}(\nu, \mu) = \{(X, Y);\, \mathcal{L}(X) = \nu,\, \mathcal{L}(Y) = \mu\}\,. \tag{2.1}$$

Then the p-Wasserstein distance is written as

$$\mathcal{W}_p(\nu, \mu) = \inf \left\{ \mathbf{E}[|X - Y|^p]^{\frac{1}{p}};\, (X, Y) \in \mathcal{C}(\nu, \mu) \right\}. \tag{2.2}$$

Remark 2.1. The distance $\mathcal{W}_1$ can also be represented as

$$\mathcal{W}_1(\nu, \mu) = \sup\{|\nu(h) - \mu(h)|;\ \|h\|_{\mathrm{Lip}} \leq 1\}. \tag{2.3}$$

We will denote by $\mathcal{D}_p$ the set of distances on $\mathcal{M}_1(\mathbb{R}^d)$ dominated by the p-Wasserstein distance for a given $p > 0$. Namely, we set

$$\mathcal{D}_p := \{\text{distances } d \text{ on } \mathcal{M}_1(\mathbb{R}^d);\, \exists\, c > 0 \text{ such that } \forall \nu, \mu \in \mathcal{M}_1(\mathbb{R}^d), d(\nu, \mu) \leq c\mathcal{W}_p(\nu, \mu)\} \tag{2.4}$$

In particular, a distance d that belongs to $\mathcal{D}_p$ induces a weaker topology than the p-Wasserstein distance. When necessary in some of the next results (or in the numerical experimentations), we will introduce specific distances which belong to $\mathcal{D}_p$.

Remark 2.2. The family of p-Wasserstein distances obviously provides examples of distances in $\mathcal{D}_p$. The Fortet-Mourier distance (see e.g [50, Chapter 6]), defined by

$$d_{\mathrm{FM}}(\nu, \mu) = \sup\left\{|\mathbf{E}[h(X)] - \mathbf{E}[h(Y)]|;\ \text{ where } X \sim \mu,\, Y \sim \nu,\ \text{ and } \|h\|_{\mathrm{Lip}} \leq 1,\, \|h\|_\infty \leq 1\right\},$$

is also easily seen to be an element of $\mathcal{D}_1$ thanks to (2.3). In this article we shall work with the distances $d_{\mathrm{CF},p}$ and d_s introduced below in Section 2.5, which are trivially proved to sit in $\mathcal{D}_1$ (due to relation (2.3)).

2.2. **Assumptions.** Before we proceed to a specific statement of our estimator, let us describe the assumptions under which we shall work. We start by a standard hypothesis on the parameter set Θ, which is supposed to be a compact set.

$(\mathbf{H_0})$: The set Θ is compactly embedded in $\mathbb{R}^q$ for a given $q \geq 1$.

Next we recall that our drift estimators rely on the invariant measure for the solution of equation (1.2). The existence and uniqueness of this invariant measure is usually obtained under some coercivity assumptions on the drift b. In the current paper we will distinguish between two notions of coercivity, respectively named weak and strong and denoted by $(\mathbf{C_w})$ and $(\mathbf{C_s})$. The weak assumption can be summarized as follows.

$(\mathbf{C_w})$: We have $b \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ and there exist constants $\alpha, \beta, C, L > 0$ and $r \in \mathbb{N}$ such that:

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

$$\langle b_\vartheta(x) - b_\vartheta(y),\, x - y\rangle \leq \beta - \alpha|x - y|^2 \quad \text{ and } \quad |b_\vartheta(x) - b_\vartheta(y)| \leq L|x - y| \tag{2.5}$$

(ii) For every $x \in \mathbb{R}^d$ and $\vartheta \in \Theta$ the following growth bound is satisfied:

$$|\partial_\vartheta b_\vartheta(x)| \leq C\left(1 + |x|^r\right). \tag{2.6}$$

The main part (2.5) of assumption $(\mathbf{C_w})$ states that the coefficient b is inward looking, except maybe on a compact set which is a neighborhood of 0. We now state the strong assumption $(\mathbf{C_s})$, which specifies that b should be inward looking everywhere, and can be expressed as a particular case of $(\mathbf{C_w})$.

$(\mathbf{C_s})$: Assumption $(\mathbf{C_w})$ holds with $\beta = 0$.

As mentioned above, Hypothesis $(\mathbf{C_w})$ combined with the invertibility of σ (and therefore $(\mathbf{C_s})$) classically involves (see *e.g.* [23]) the existence of a unique stationary solution for the solution of the following equation for any $\vartheta \in \Theta$:

$$dY_t^\vartheta = b_\vartheta(Y_t^\vartheta)\, dt + \sigma dB_t, \qquad t \in [0, T]. \tag{2.7}$$

Notice that the system (2.7) is identical to our original equation (1.2). However, let us notice that the fBm is unobserved. Moreover, the uniqueness of the stationary measure must be understood in a weak sense. Namely there exists a unique distribution $\mathbf{P}^\vartheta$ on $\mathcal{C}([0,\infty), \mathbb{R}^d)$ such that if $(\bar{Y}_t^\vartheta)$ denotes a process with distribution $\mathbf{P}^\vartheta$, then $(\bar{Y}_t^\vartheta)_{t\geq 0}$ is a stationary solution to (2.7), *i.e.* shift-invariant (when one considers its canonical version). We denote by ν_ϑ the distribution of $\bar{Y}^\vartheta$ at any instant $t \geq 0$, that is

$$\nu_\vartheta = \mathcal{L}(\bar{Y}_0^\vartheta). \tag{2.8}$$

Remark 2.3. Note that in this non-Markovian setting, ν_ϑ is not exactly the invariant distribution. More precisely, owing to [23], one can embed (2.7) into an infinite-dimensional Markovian structure which allows the construction of an adapted ergodic theory. An invariant distribution $\bar{\nu}_\vartheta$ is then defined on this enlarged structure. Without going into the details, one can just say that in this theory, the probability ν_ϑ can be retrieved as a marginal of the "true" invariant distribution. In the sequel, we will thus talk about *marginal invariant distribution* ν_ϑ.

Remark 2.4. As mentioned before, the invertibility assumption on σ combined with $(\mathbf{C_w})$ ensures uniqueness of the invariant distribution. However, even though this hypothesis is of first importance under $(\mathbf{C_w})$ (in order to use irreducibility-type arguments), it could be entirely removed under $(\mathbf{C_s})$. Actually, in this case, the contraction assumption implies that two solutions of (2.7) driven by the same fBm but starting from different initial conditions come together at ∞, *a.s.* and in L^2, which classically involves uniqueness (see *e.g.* [9, Lemma 3(*ii*)] for details). This remark also holds for the Euler scheme (2.13) introduced in the next section but for sufficiently small step γ (see again [9, Lemma 3(*ii*)] for details).

As said previously, we shall obtain our drift estimators through the analysis of the marginal invariant distribution ν_ϑ defined by (2.8). If we want this strategy to be successful, it is natural to assume that ν_ϑ characterizes ϑ. We thus label this hypothesis as follows.

$(\mathbf{I_w})$: For all $\vartheta \in \Theta$, we have $\nu_\vartheta = \nu_{\vartheta_0}$ iff $\vartheta = \vartheta_0$.

It is worth noticing that if d denotes a distance on $\mathcal{M}_1(\mathbb{R}^d)$, then one can recast $(\mathbf{I_w})$ as:

$$d(\nu_\vartheta, \nu_{\vartheta_0}) = 0 \quad \text{iff} \quad \vartheta = \vartheta_0. \tag{2.9}$$

We shall use this characterization in order to construct the estimator $\hat{\vartheta}$ (see (2.12) below). Also notice that $(\mathbf{I_w})$ refers to a "weak" identifiability condition, which will be resorted to in order to derive the consistency of our estimator $\hat{\vartheta}$. In contrast, the following "strong" identifiability assumption $(\mathbf{I_s})$ defined for a given distance d on $\mathcal{M}_1(\mathbb{R}^d)$ will be used to get rates of convergence.

$(\mathbf{I_s})$ There exists a constant $C > 0$ and a parameter $\varsigma \in (0, 1]$ such that $\forall \vartheta \in \Theta$,

$$d(\nu_\vartheta, \nu_{\vartheta_0}) \geq C|\vartheta - \vartheta_0|^\varsigma. \tag{2.10}$$

Remark 2.5. We will construct a class of equations, basically obtained as perturbations of Langevin type equations, for which our assumptions $(\mathbf{I_w})$ and $(\mathbf{I_s})$ are satisfied. See Section 6 below.

2.3. **Statistical setting and construction of the estimator.** We wish to construct an estimator based on discrete observations. In this context, the simplest situation (which will mostly prevail in the paper) is to assume that the solution $(Y_t)_{t\geq 0}$ of (2.7) is discretely observed at some instants $\{t_k;\, 0\leq k\leq n\}$, with $t_{k+1}-t_k=\kappa$ for a given time step $\kappa>0$. Under $(\mathbf{C_w})$, it can be shown (see Proposition 3.3 below) that

$$\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}} \overset{n\to+\infty}{\Longrightarrow} \nu_{\vartheta_0} \quad a.s, \tag{2.11}$$

where $\Longrightarrow$ stands for the weak convergence of probability measures in $\mathbb{R}^d$. With this convergence in mind, the heart of our estimation method is then the following observation: under the identifiability assumption $(\mathbf{I_w})$, the most natural way to estimate ϑ_0 is to consider

$$\hat{\vartheta}=\underset{\vartheta\in\Theta}{\operatorname{argmin}}\; d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}},\nu_\vartheta\right), \tag{2.12}$$

where d is a given (arbitrary) distance on $\mathcal{M}_1(\mathbb{R}^d)$. However, in spite of the fact that our formula (2.12) is simple enough, it is also easily understood that ν_ϑ is far from being explicitly known (except in some very particular cases such as the Ornstein-Uhlenbeck process). In this paper, we propose to circumvent this difficulty by considering some estimators based on numerical approximations of ν_ϑ.

Specifically, the numerical approximations we will resort to are built through an Euler-type discretization of the stochastic process Y^ϑ solution to (2.7). Namely, let $(s_k)_{k\geq 0}$ be an increasing sequence of numbers such that $s_0=0$, and $\lim_{k\to\infty}s_k=+\infty$. The Euler-Maruyama scheme Z^ϑ is then recursively defined by $Z_0^\vartheta=z_0\in\mathbb{R}^d$ and:

$$\text{For all } k\geq 0,\quad Z^\vartheta_{s_{k+1}}=Z^\vartheta_{s_k}+(s_{k+1}-s_k)b_\vartheta(Z^\vartheta_{s_k})+\sigma\,(\tilde{B}_{s_{k+1}}-\tilde{B}_{s_k}), \tag{2.13}$$

where $\tilde{B}$ is a (simulated) m-dimensional fractional Brownian motion which is a priori different from the driving process B in equation (2.7) (since B is unobserved). When $s_k=k\gamma$ for a given $\gamma>0$, we say that the Euler scheme is a *constant step* sequence and we denote it by $Z^{\vartheta,\gamma}$. When $\gamma_k=s_k-s_{k-1}$ is a non-increasing sequence such that $\gamma_k\to 0$ as $k\to+\infty$, the Euler scheme will be called *decreasing step* Euler scheme. We will work with these two types of schemes in the sequel.

Remark 2.6. In practice it is natural to set $Z_0^\vartheta=Y_0$ where Y_0 is the first observation of the process $(Y_t)_{t\geq 0}$. Let us also remark that in the sequel, for notational sake, one usually denotes by B the fBm related to the Euler scheme Z^ϑ. However, the reader has to keep in mind that the fact that the fBms in (2.7) and in (2.13) are different. This certainly prevents us from any pathwise comparison between the observed process and the simulated one.

Remark 2.7. We refer to Section 7 for background on the simulation of the increments of the fBm.

Let us now give an explicit expression for the estimator we are considering in this article. We will focus on the constant step setting in (2.13) for sake of simplicity. Observe that under $(\mathbf{C_w})$,

it can be shown (see Proposition 3.4 below) that

$$\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{k\gamma}} \overset{n\to+\infty}{\Longrightarrow} \nu^{\gamma}_{\vartheta} \quad a.s. \tag{2.14}$$

where ν^{γ}_{ϑ} denotes the unique marginal invariant distribution of the Euler scheme Z^{ϑ}. By marginal, we mean again that $Z^{\vartheta,\gamma}$ can be endowed with a Feller infinite-dimensional Markov structure which admits a unique invariant distribution under $(\mathbf{C_w})$ (see [49] for details). The first marginal of this invariant distribution is ν^{γ}_{ϑ}. Similarly to what we proposed in (2.12), such a result suggests to define our estimator for ϑ as

$$\hat{\vartheta}_{N,n,\gamma} = \underset{\vartheta\in\Theta}{\operatorname{argmin}}\ d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}},\ \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{k\gamma}}\right), \tag{2.15}$$

where d is again a distance on $\mathcal{M}_1(\mathbb{R}^d)$. Note that in the decreasing step case analogous constructions may be carried out, and will be introduced later. Let us also remark that relation (2.15) only involves one Euler scheme path, which is relevant for numerical implementations.

We are now in a position to state our main results. We divide the presentation in two parts. In the next section, we focus on strong consistence results related to the family $\{\hat{\vartheta}_{N,n,\gamma};\ N\geq 1, n\geq 1, \gamma>0\}$ defined by (2.15), as well as its decreasing step counterpart. Then Section 2.5 is dedicated to the rate of convergence of the estimator $\hat{\vartheta}_{N,n,\gamma}$. In particular, this second part will involve concentration results related to the SDE and to its Euler discretization.

2.4. **Main consistency results.** We begin with a first result involving the weak assumption $(\mathbf{C_w})$, which requires to discretize the set Θ in the following sense. According to our hypothesis $(\mathbf{H_0})$, the set Θ is compact in $\mathbb{R}^q$. Therefore the Borel-Lebesgue property gives us the existence, for every $\varepsilon>0$, of $M_\varepsilon\in\mathbb{N}$ and $(\vartheta^{(\varepsilon)}_i)_{1\leqslant i\leqslant M_\varepsilon}\in\Theta^{M_\varepsilon}$ such that $\Theta\subset\bigcup_{i=1}^{M_\varepsilon}B(\vartheta^{(\varepsilon)}_i,\varepsilon)$. Thanks to this property, we define the following discretization for all $\varepsilon>0$ and $\vartheta\in\Theta$:

$$\vartheta^{(\varepsilon)} := \underset{\vartheta'\in\{\vartheta^{(\varepsilon)}_i\}}{\operatorname{argmin}}\ |\vartheta'-\vartheta|. \tag{2.16}$$

With this notation in hand, we can now state our first consistency theorem.

Theorem 2.8. *Assume* $(\mathbf{H_0})$, $(\mathbf{C_w})$ *and* $(\mathbf{I_w})$. *Let p be a strictly positive real number and consider a distance d on $\mathcal{M}_1(\mathbb{R}^d)$ which belongs to $\mathcal{D}_p$ (recall that $\mathcal{D}_p$ is defined by (2.4)). We consider the family $\{\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma};\ N\geq 1, n\geq 1, \gamma>0, \varepsilon>0\}$ defined by*

$$\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma} = \underset{\vartheta\in\Theta^{(\varepsilon)}}{\operatorname{argmin}}\ d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{s_k}}\right), \quad N,n\in\mathbb{N},\ \gamma>0,\ \varepsilon>0 \tag{2.17}$$

where $\Theta^{(\varepsilon)} := \{\vartheta^{(\varepsilon)}_i\ ;\ 1\leqslant i\leqslant M_\varepsilon\}$. Then $\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma}$ is a strong consistent estimator of ϑ_0. Specifically, we have

$$\lim_{\varepsilon\to 0}\lim_{\gamma\to 0}\lim_{N,n\to+\infty}\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma} = \vartheta_0 \quad a.s.$$

Let us remark that the discretization of Θ given by (2.16) is needed to get strong consistency, due to the fact that under $(\mathbf{C_w})$ we loose uniformity with respect to ϑ in some of our convergence results. For instance, $(\mathbf{C_w})$ only warranties the simple convergence of $d(\nu^{\gamma}_{\vartheta},\nu_{\vartheta})$ to 0 as $\gamma\to 0$ (see Proposition 3.4). The proof of Theorem 2.8 is achieved in Section 4.3.

We now turn to our main estimator defined in (2.15). The proof of the theorem below is detailed in Section 4.3.

Theorem 2.9. *Assume* $(\mathbf{H_0})$, $(\mathbf{C_s})$ *and* $(\mathbf{I_w})$. *As in Theorem 2.8, let* p *be a strictly positive real number and consider a distance* d *on* $\mathcal{M}_1(\mathbb{R}^d)$ *which belongs to* $\mathcal{D}_p$. *Then the family* $\{\hat{\vartheta}_{N,n,\gamma};\, N \geq 1, n \geq 1, \gamma > 0\}$ *defined by* (2.15) *is a strong consistent estimator of* ϑ_0 *in the following sense:*

$$\lim_{\gamma \to 0} \lim_{N,n \to +\infty} \hat{\vartheta}_{N,n,\gamma} = \vartheta_0 \quad a.s.$$

We close this section with a result concerning the approximation of invariant measures by an Euler scheme with decreasing time step. Namely we consider an approximation scheme denoted also by Z^{ϑ}, which is defined similarly to $Z^{\vartheta,\gamma}$ in (2.13) except for the fact that the sequence $(s_k)_{k\geq 0}$ satisfies:

$$s_{k+1} - s_k = \gamma_{k+1}, \quad k \geq 0, \tag{2.18}$$

where $(\gamma_k)_{k\geq 1}$ is a non-increasing sequence of positive numbers such that

$$\sum_{k\geq 1} \gamma_k = +\infty \quad \text{and} \quad \lim_{k\to+\infty} \gamma_k = 0. \tag{2.19}$$

The convergence theorem we obtain in the decreasing step case is the following. As for Theorem 2.9 , it is achieved under the strong coercivity assumption $(\mathbf{C_s})$. Its proof is developped in Section 4.3.

Theorem 2.10. *Assume* $(\mathbf{H_0})$, $(\mathbf{C_s})$ *and* $(\mathbf{I_w})$. *Let* $p \geq 2$ *and consider a distance* d *on* $\mathcal{M}_1(\mathbb{R}^d)$ *which belongs to* $\mathcal{D}_p$. *Let* $(\gamma_k)_{k\geq 1}$ *be a non-increasing sequence of positive numbers satisfying condition* (2.19) *and the technical condition* (4.3). *Denote by* Z^{ϑ} *the related Euler scheme given by* (2.13). *We consider an estimator* $(\hat{\vartheta}_{N,n})_{N,n}$ *defined as*

$$\hat{\vartheta}_{N,n} = \operatorname*{argmin}_{\vartheta\in\Theta} d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}}, \frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}}\right), \quad N \in \mathbb{N}. \tag{2.20}$$

Then, $(\hat{\vartheta}_{N,n})_{N,n}$ *is a strong consistent estimator of* ϑ_0, *namely:*

$$\lim_{N,n\to+\infty} \hat{\vartheta}_{N,n} = \vartheta_0 \quad a.s.$$

Remark 2.11. The technical condition (4.3) ($\sum_{k\geq 1} \gamma_k^{p'H+1} s_k^{-1} < +\infty$ for a given $p' \geq p$) is true in a very large setting. For instance, it can be checked that this is satisfied for any polynomial step sequence : $\gamma_k = \gamma k^{-\rho}$ with $\rho \in (0,1]$ and $\gamma \in \mathbb{R}_+^*$, but also for less decreasing sequences such as $\gamma_k = \gamma(\log k)^{-1}$. However, this is not true in full generality (the condition does not hold when $\gamma_k = (\log(\log k))^{-1}$ for instance).

2.5. **Rate of convergence.** Under our strong identifiability condition $(\mathbf{I_s})$, we will be able to get a rate of convergence for some of our estimators. In order to carry out this task, we shall assume that condition $(\mathbf{I_s})$ is verified for some specific distances on probability measures called respectively $d_{\mathrm{CF},p}$ and d_s. These distances are defined in the following way:

(*i*) Let X and Y be $\mathbb{R}^d$-valued random variables and $p > (\frac{d}{2} \vee 1)$. We consider the integrable kernel $g_p(\xi) := c_p(1+|\xi|^2)^{-p}$ where $c_p := \left(\int_{\mathbb{R}^d}(1+|\xi|^2)^{-p}d\xi\right)^{-1}$. Then the $d_{\mathrm{CF},p}$ distance between $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ is defined by:

$$d_{\mathrm{CF},p}(\mathcal{L}(X),\mathcal{L}(Y)) := \left(\int_{\mathbb{R}^d}\left(\mathbf{E}[e^{i\langle\xi,X\rangle}] - \mathbf{E}[e^{i\langle\xi,Y\rangle}]\right)^2 g_p(\xi)d\xi\right)^{1/2}. \tag{2.21}$$

(*ii*) Let $\{f_i\ ;\ i \geq 1\}$ be a family of $\mathcal{C}_b^1$, supposed to be dense in the space $\mathcal{C}_b^0$ of continuous and bounded functions and decreasing to 0 at ∞. Consider two probability measures ν and μ in $\mathcal{M}_1(\mathbb{R}^d)$. Then the distance d_s between ν and μ is defined as:

$$d_s(\nu,\mu) := \sum_{i=0}^{+\infty} 2^{-i}(|\nu(f_i) - \mu(f_i)| \wedge 1). \tag{2.22}$$

Remark 2.12. It is readily checked that both $d_{\mathrm{CF},p}$ and d_s metrize the convergence in law (their induced topology are in fact exactly the one induced by the convergence in law). The distance $d_{\mathrm{CF},p}$ is technically convenient for our purpose and close in spirit to the smooth Wasserstein distance invoked in the Stein method literature (see e.g [2]). The distance d_s is called weak-$\star$ distance in [50] and also used in [52] for filtering problems. Also notice that both $d_{\mathrm{CF},p}$ and d_s are elements of $\mathcal{D}_1$ where $\mathcal{D}_1$ is defined by (2.4).

With the distances $d_{\mathrm{CF},p}$ and d_s in hand, our main result about rates of convergence is the following:

Theorem 2.13. *Assume* $(\mathbf{H_0})$, $(\mathbf{C_s})$ *and* $(\mathbf{I_s})$ *hold true, where* (2.10) *in hypothesis* $(\mathbf{I_s})$ *is considered for* $d = d_s$ *or* $d = d_{\mathrm{CF},p}$ *with* $p > (q+d)/2$ *and* $\varsigma = 2/q$ *for a given* $q \geq 2$. *Let* $\hat{\vartheta}_{N,n,\gamma}$ *be the estimator given by* (2.15). *Then, we get the following rate of convergence: there exists* $C_q > 0$ *such that*

$$\mathbf{E}\left[|\hat{\vartheta}_{N,n,\gamma} - \vartheta_0|^2\right] \leq C_q\left(n^{-\frac{q}{2}(2-(2H\vee 1))} + \gamma^{qH} + T^{-\tilde{\eta}}\right) \tag{2.23}$$

with $\tilde{\eta} := \frac{q^2}{2(q+d)}(2-(2H\vee 1))$ *and* $T := N\gamma$.

Remark 2.14. This non-asymptotical bound theoretically enables to calibrate the "free parameters" γ and N in terms of the number of observations n, which is fixed by the statistical setting. For instance, when $\varsigma = 1$ (*i.e.* when $q = 2$), the first term is of order $n^{-(2-(2H\vee 1))}$ and hence, in order to to preserve this rate order, we have to fix $\gamma \leq n^{-\frac{1-(H\vee\frac{1}{2})}{H}}$ and $N \geq n^{\frac{4+2d}{4}}\gamma^{-1}$. More precisely, for these choices of parameters, the quadratic error induced by this estimator (when $\varsigma = 1$) is of order $n^{-\frac{1}{2}}$ if $H < 1/2$ and $n^{H-\frac{1}{2}}$ if $H > 1/2$. The constant ς, which appears in Assumption $(\mathbf{I_s})$, comes from the fact that, the bounds are first established on the distances between the invariant distributions $\nu_{\hat{\vartheta}_{N,n,\gamma}}$ and ν_{ϑ_0}. Nevertheless, except some particular settings such as the Ornstein-Uhlenbeck process, this exponent ς is unfortunately difficult to compute in some general settings. Finally, let us remark that L^p-bounds can be easily deduced from the proof for any $p \geq 2$. However, since they do not modify significantly the results, we chose here to only state the quadratic one.

3. Preliminary Results

In this section we label some basic results about equation (2.7) and its invariant measure for further use. We first recall some ergodic properties of stochastic differential equations driven by a fBm, then we study the continuity of the invariant measure ν_ϑ with respect to the parameter ϑ. Under the strong coercivity assumption $(\mathbf{C_s})$, we quantify the distance between the empirical measures respectively related to the process Y^ϑ and its Euler approximation Z^ϑ. Eventually we give some convergence results for the quantities involved in the right hand side of (2.17).

3.1. Ergodic properties of the SDE and of the Euler scheme.

In this section we review several ergodic properties for equation (2.7). These properties are at the heart of our estimation procedure.

3.1.1. *Convergence of $\mathcal{L}(Y_t)$.* We start by giving the basic convergence in law towards the invariant distribution for our processes Y^ϑ.

Proposition 3.1. *Assume* $(\mathbf{H_0})$ *and* $(\mathbf{C_w})$ *and consider the family of processes* $\{Y^\vartheta;\, \vartheta \in \Theta\}$ *defined by* (2.7)*. Then the following properties hold true:*

(i) Existence and uniqueness hold for the invariant distribution related to the dynamical system (2.7)*. Furthermore, having in mind the notations introduced in* (2.8)*, for all* $\varepsilon > 0$ *there exists a constant* $C_\varepsilon > 0$ *independent of* $\vartheta \in \Theta$ *such that*

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(Y_t^\vartheta\right), \nu_\vartheta\right) \leqslant C_\varepsilon t^{-(\alpha_H - \varepsilon)}, \tag{3.1}$$

with an exponent α_H *given by*

$$\alpha_H = \begin{cases} \frac{1}{8} & \text{if} \quad H \in (\frac{1}{4}, 1) \setminus \left\{\frac{1}{2}\right\} \\ H(1-2H) & \text{if} \quad H \in (0, \frac{1}{4}). \end{cases}$$

(ii) For any $p > 0$ *and for any distance* $d \in \mathcal{D}_p$*, the following upper bound holds uniformly in* $\vartheta \in \Theta$*:*

$$d\left(\mathcal{L}(Y_t^\vartheta), \nu_\vartheta\right) \leq C\, t^{-\frac{\alpha_H}{4p}}, \tag{3.2}$$

for a strictly positive constant C*. In particular, for any* $p > 0$*, we have*

$$\sup_{t \geq 0} \mathbf{E}[|Y_t|^p] < +\infty. \tag{3.3}$$

Proof. We prove the two statements of our proposition separately.

Proof of item (i). The only difference between our claim and [23, Theorems 1.2 and 1.3] is the uniformity with respect to $\vartheta \in \Theta$ in the convergence in total variation result. However, following carefully the proof of [23], it can be shown that the constants therein do not depend on ϑ if Hypothesis $(\mathbf{C_w})$ is satisfied. Therefore the constant C_ε in (3.1) is uniform in ϑ.

Proof of item (ii). Relation (3.3) is proved in Proposition A.1 of the Appendix. In order to prove (3.2), consider a couple (X_1, X_2) of random variables such that $X_1 \sim Y_t^\vartheta$ and X_2 has distribution ν_ϑ. By Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbf{E}[|X_1 - X_2|^p]^{\frac{1}{p}} &= \mathbf{E}[|X_1 - X_2|^p\, \mathbf{1}_{(X_1 \neq X_2)}]^{\frac{1}{p}} \\ &\leq \left(\mathbf{E}[|X_1|^{2p}]^{\frac{1}{2p}} + \mathbf{E}[|X_2|^{2p}]^{\frac{1}{2p}}\right) \mathbf{P}(X_1 \neq X_2)^{\frac{1}{2p}}. \end{aligned} \tag{3.4}$$

We now bound separately the two terms on the right hand side of (3.4). If we denote by $\mathcal{C}(Y_t^\vartheta, \nu_\vartheta)$ the set of couplings between $\mathcal{L}(Y_t^\vartheta)$ and ν_ϑ (defined as in (2.1)), then we have

$$d_{\mathrm{TV}}\left(\mathcal{L}\left(Y_t^\vartheta\right), \nu_\vartheta\right) = \inf\left\{\mathbf{P}(\tilde{X}_1 \neq \tilde{X}_2);\, (\tilde{X}_1, \tilde{X}_2) \in \mathcal{C}(Y_t^\vartheta, \nu_\vartheta)\right\}.$$

Therefore, owing to (3.1) we can choose a coupling $(X_1, X_2) \in \mathcal{C}(Y_t^\vartheta, \nu_\vartheta)$ and a constant $C > 0$ such that

$$\mathbf{P}(X_1 \neq X_2) \leq C t^{-(\alpha_H - \varepsilon)}. \tag{3.5}$$

In addition, according to (3.3) and a uniform integrability argument, we easily get the following inequality for the coupling (X_1, X_2) we have chosen in (3.5):

$$\mathbf{E}[|X_1|^{2p}]^{\frac{1}{2p}} + \mathbf{E}[|X_2|^{2p}]^{\frac{1}{2p}} \leq 2 \sup_{\vartheta \in \Theta} \sup_{t \geqslant 0} \mathbf{E}[|Y_t^\vartheta|^{2p}]^{\frac{1}{2p}} < +\infty. \tag{3.6}$$

We plug (3.5) and (3.6), applied to $\varepsilon = \alpha_H/2$, into (3.4). Going back to the definition (2.2) of the distance $\mathcal{W}_p$, we obtain that there exists a strictly positive constant C such that for any $t \geq 0$ and any $\vartheta \in \Theta$ we have

$$\mathcal{W}_p(\mathcal{L}(Y_t^\vartheta), \nu_\vartheta) \leq C t^{-\frac{\alpha_H}{4p}}.$$

The result (3.2) follows. □

Next we observe that whenever $(\mathbf{C_s})$ is fulfilled, the polynomial convergence in (3.2) can be replaced by an exponential rate. This is summarized in the following proposition.

Proposition 3.2. *Let $\{Y^\vartheta; \vartheta \in \Theta\}$ be the family of processes defined by (2.7). Suppose that Hypothesis $(\mathbf{H_0})$ and Hypothesis $(\mathbf{C_s})$ are met. Let d be a distance in $\mathcal{D}_p$. Then, we have*

$$d(\mathcal{L}(Y_t^\vartheta), \nu_\vartheta) \leq c_1 e^{-c_2 t},$$

with $c_2 = \alpha/2$ where α is the constant featured in equation (2.5), and where $c_1 = c_1(H, \alpha)$.

Proof. Let $\bar{Y}^\vartheta$ be the stationary solution of equation (2.7). One can easily show, by means of the same arguments as in [21], that

$$\mathbf{E}\left[|Y_t^\vartheta - \bar{Y}_t^\vartheta|^p\right]^{1/p} \leq c_1 e^{-c_2 t}. \tag{3.7}$$

The result follows trivially. □

3.1.2. *Ergodic Theorems for the SDE.* We now summarize the limit theorems obtained for equation (2.7) which will be relevant for our purposes, with a special emphasis on the occupation measure $\delta_{Y_s^\vartheta}$.

Proposition 3.3. *Let Y^ϑ be the unique solution of (2.7) obtained for a parameter $\vartheta \in \Theta$. Assume $(\mathbf{H_0})$ and $(\mathbf{C_w})$ hold true and let ν_ϑ be the measure defined by (2.8). Then for all $\vartheta \in \Theta$, for any $p > 0$ and for any distance $d \in \mathcal{D}_p$, we have*

(i) The distance between ν_ϑ and the normalized occupation measure of Y^ϑ converges to 0 as $t \to \infty$. That is

$$\lim_{t\to\infty} d\left(\frac{1}{t}\int_0^t \delta_{Y_s^\vartheta}\, ds,\, \nu_\vartheta\right) = 0. \tag{3.8}$$

In particular, we have an almost sure uniform bound for the p-th powers of Y_s^ϑ:

$$\sup_{\vartheta\in\Theta}\sup_{t\geq 1} \frac{1}{t}\int_0^t |Y_s^\vartheta|^p ds < \infty \quad \text{a.s.} \tag{3.9}$$

(ii) Some discrete versions of (3.8) and (3.9) are also available. Specifically, let $\eta > 0$ and set $t_k = k\eta$ for $k \geq 0$. Then for any $\vartheta \in \Theta$ we have

$$\lim_{n\to\infty} d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}^\vartheta}, \nu_\vartheta\right) = 0.$$

In particular, the following uniform bound holds true:

$$\sup_{\vartheta\in\Theta}\sup_{n\geq 1} \frac{1}{n}\sum_{k=0}^{n-1} |Y_{t_k}^\vartheta|^p < \infty \quad \text{a.s.} \tag{3.10}$$

Proof. Relations (3.9) and (3.10) are proved in Proposition A.1. As far as the identification of the limit is concerned, the proof follows the lines of [9] and is detailed in Section A.2. □

3.1.3. *Ergodic Theorems for the Euler scheme.* Recall that we denote by $(Z_{k\gamma}^{\vartheta,\gamma})_{k\geq 0}$ the Euler scheme with step γ related to $(Y_t^{\vartheta})_{t\geq 0}$, as defined in (2.13). This section focuses on the asymptotic behavior of $Z_{k\gamma}^{\vartheta,\gamma}$ as $k\to\infty$ and $\gamma\to 0$.

Proposition 3.4. *Let Y^{ϑ} be the unique solution of (2.7) and consider the Euler scheme $(Z_{k\gamma}^{\vartheta,\gamma})_{k\geq 0}$ with step γ related to $(Y_t^{\vartheta})_{t\geq 0}$. Assume $(\mathbf{H_0})$ and $(\mathbf{C_w})$ hold true. Then for all $\vartheta\in\Theta$, for any $p>0$ and for any distance $d\in\mathcal{D}_p$, we have*

(i) There exists $\gamma_0>0$ and a unique family of measures $(\nu_{\vartheta}^{\gamma})_{\gamma\leq\gamma_0}$ such that for all $\vartheta\in\Theta$ and $\gamma\in(0,\gamma_0]$, we have

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

In particular, we get the following uniform bound for the p-th powers of $Z_{k\gamma}^{\vartheta,\gamma}$:

$$\sup_{\vartheta\in\Theta}\sup_{N\geq 1}\frac{1}{N}\sum_{k=0}^{N-1}|Z_{k\gamma}^{\vartheta,\gamma}|^p<\infty \quad a.s. \tag{3.11}$$

(ii) The invariant measure ν_{ϑ}^{γ} for the Euler scheme converges to the invariant measure ν_{ϑ} of Y^{ϑ} as the mesh of the partition goes to 0:

$$\lim_{\gamma\to 0} d\left(\nu_{\vartheta}^{\gamma},\nu_{\vartheta}\right)=0.$$

Proof. The weak convergence of $\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta,\gamma}}$ to ν_{ϑ}^{γ} as $n\to+\infty$, as well as the convergence of ν_{ϑ}^{γ} to ν_{ϑ} as $\gamma\to 0$ are consequences of [10, Theorem 1]. The extension to distances d dominated by $\mathcal{W}^p$ follows from Proposition A.2. More precisely for (ii), we can deduce from Proposition A.2 and from Fatou's lemma that for any $M>0$, for any $\gamma\in(0,\gamma_0]$,

$$\nu_{\vartheta}^{\gamma}(|x|^p\wedge M)\leq\liminf_{N\to+\infty}\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{E}[|Z_{k\gamma}^{\vartheta,\gamma}|^p]\leq C$$

where C is a positive constant independent of M and γ. Hence, taking limits as M goes to ∞, this yields

$$\sup_{\gamma\in(0,\gamma_0]}\nu_{\vartheta}^{\gamma}(|x|^p)<+\infty,$$

for any $p>0$. Then one can conclude as in the proof of Proposition 3.1 (see relation (3.4)), where the distance $d_{\mathrm{TV}}(\nu_{\vartheta}^{\gamma},\nu_{\vartheta})$ is upper bounded thanks to [10, Theorem 1]. □

3.2. **Continuity of $\vartheta\mapsto d(\nu_{\vartheta},\nu_{\vartheta_0})$.** The convergence of our estimator $\hat{\vartheta}_{N,n,\gamma}$ defined by (2.15) depends crucially on continuity properties of the family $\{\nu_{\vartheta};\,\vartheta\in\Theta\}$. To this aim, we first prove a basic result on the continuity of the map $\vartheta\mapsto Y_t^{\vartheta}$.

Proposition 3.5. *Let ϑ_1 and ϑ_2 be elements of Θ, and consider the respective solutions $(Y_t^{\vartheta_1})_{t\geqslant 0}$ and $(Y_t^{\vartheta_2})_{t\geqslant 0}$ of equation (2.7). Assume hypothesis $(\mathbf{H_0})$ and $(\mathbf{C_w})$ are satisfied. Then for any $p>0$ and $T>0$, there exists $C_{T,p}>0$ independent of ϑ_1 and ϑ_2 such that*

$$\|Y_t^{\vartheta_1}-Y_t^{\vartheta_2}\|_{L^p(\Omega)}\leqslant C_{T,p}|\vartheta_1-\vartheta_2|. \tag{3.12}$$

Proof. By monotonicity of the norms in $L^p(\Omega)$, it is enough to consider the case $p \geq 2$. Furthermore, it is readily seen from (2.7) that we have

$$Y_t^{\vartheta_1} - Y_t^{\vartheta_2} = \int_0^t (b_{\vartheta_1}(Y_s^{\vartheta_1}) - b_{\vartheta_2}(Y_s^{\vartheta_2}))ds. \tag{3.13}$$

Starting from (3.13), we easily get the following identity for the square of $Y_t^{\vartheta_1} - Y_t^{\vartheta_2}$:

$$\frac{d}{dt}|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^2 = 2\langle Y_t^{\vartheta_1} - Y_t^{\vartheta_2}, b_{\vartheta_1}(Y_t^{\vartheta_1}) - b_{\vartheta_2}(Y_t^{\vartheta_2})\rangle. \tag{3.14}$$

We now invoke the fact that b is Lipschitz continuous under $(\mathbf{C_w})$ plus inequality (2.6) on $\partial_\vartheta b_\vartheta$ in order to get

$$\begin{aligned}\frac{d}{dt}|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^2 &= 2\langle Y_t^{\vartheta_1} - Y_t^{\vartheta_2}, b_{\vartheta_1}(Y_t^{\vartheta_1}) - b_{\vartheta_1}(Y_t^{\vartheta_2})\rangle + 2\langle Y_t^{\vartheta_1} - Y_t^{\vartheta_2}, b_{\vartheta_1}(Y_t^{\vartheta_2}) - b_{\vartheta_2}(Y_t^{\vartheta_2})\rangle \\ &\leqslant c_1 \left|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}\right|^2 + 2c_2|\vartheta_1 - \vartheta_2| \left(1 + \left|Y_t^{\vartheta_2}\right|^r\right) \left|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}\right|,\end{aligned}$$

where c_1 and c_2 are two strictly positive constants. Now apply the elementary inequality $|ab| \leqslant \frac{1}{2}(|a|^2 + |b|^2)$ with $a = |\vartheta_1 - \vartheta_2|(1 + |Y_t^{\vartheta_2}|^r)$ and $b = |Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|$. We deduce the existence of a constant $c > 0$ such that

$$\frac{d}{dt}|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^2 \leqslant c\left(\left|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}\right|^2 + |\vartheta_1 - \vartheta_2|^2 \left(1 + |Y_t^{\vartheta_2}|^{2r}\right)\right). \tag{3.15}$$

With relation (3.15) in hand, a standard application of Gronwall's Lemma yields

$$|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^2 \leqslant c|\vartheta_1 - \vartheta_2|^2 \int_0^t e^{c(t-s)} \left(1 + |Y_s^{\vartheta_2}|^{2r}\right) ds. \tag{3.16}$$

Let us now get some information about a generic p-th power of $Y_t^{\vartheta_1} - Y_t^{\vartheta_2}$ for $p \geq 2$. To this aim, we resort to Jensen's inequality in relation (3.16). This gives the existence of a constant $c(T,p)$ such that for any $t \in [0,T]$ we have

$$|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^p \leqslant c(T,p)|\vartheta_1 - \vartheta_2|^p \int_0^t \left(1 + |Y_s^{\vartheta_2}|^{2r}\right)^{\frac{p}{2}} ds.$$

Taking the expectation, we finally get

$$\mathbf{E}\left[|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}|^p\right]^{1/p} \leqslant \tilde{c}(T,p)|\vartheta_1 - \vartheta_2| \left(\int_0^t \mathbf{E}\left[1 + |Y_s^{\vartheta_2}|^{pr}\right] ds\right)^{1/p},$$

where $\tilde{c}(T,p)$ is another finite constant. Hence our result (3.12) follows from the bound (3.3). □

We now state the announced continuity property for the family $\{\nu_\vartheta;\, \vartheta \in \Theta\}$.

Proposition 3.6. *Let $\{Y^\vartheta;\, \vartheta \in \Theta\}$ be the family of processes defined by (2.7). Assume $(\mathbf{H_0})$ and $(\mathbf{C_w})$ hold true and consider the family $\{\nu_\vartheta;\, \vartheta \in \Theta\}$ of invariant measures given by Proposition 3.1. Let $p > 0$ and pick any distance $d \in \mathcal{D}_p$, where we recall that $\mathcal{D}_p$ is defined by (2.4). Then the map $\vartheta \mapsto d(\nu_\vartheta, \nu_{\vartheta_0})$ is continuous on Θ.*

Proof. Owing to the very definition (2.4) of $\mathcal{D}_p$, it is enough to prove the result for $d = \mathcal{W}_p$ and for an arbitrary $p \geq 1$. Next we apply the triangle inequality and the fact that $\mathcal{W}_p$ is defined in (2.2) by an infimum over all couplings. This yields the following inequality, valid for any $t \geq 0$:

$$\mathcal{W}_p(\nu_{\vartheta_1}, \nu_{\vartheta_2}) \leq 2 \sup_{\vartheta \in \Theta} \mathcal{W}_p(\mathcal{L}(Y_t^\vartheta), \nu_\vartheta) + \|Y_t^{\vartheta_1} - Y_t^{\vartheta_2}\|_{L^p(\Omega)}. \tag{3.17}$$

We now bound the two terms in the right hand side of (3.17). In order to handle the term $\mathcal{W}_p(\mathcal{L}(Y_t^\vartheta),\nu_\vartheta)$, we consider a small parameter $\varepsilon>0$. By Proposition 3.1(ii), there exists t_0 large enough such that

$$2\sup_{\vartheta\in\Theta}\mathcal{W}_p(\mathcal{L}(Y_{t_0}^\vartheta),\nu_\vartheta)\le\frac{\varepsilon}{2}. \tag{3.18}$$

We will fix this value of t_0 in the right hand side of (3.17). Then the difference $Y_{t_0}^{\vartheta_1}-Y_{t_0}^{\vartheta_2}$ is handled thanks to Proposition 3.5. Namely consider $\delta>0$ such that (with the notations of Proposition 3.5) we have $C_{t_0}\delta\le\frac{\varepsilon}{2}$. We get that for all $(\vartheta_1,\vartheta_2)\in\Theta^2$ such that $|\vartheta_1-\vartheta_2|\le\delta$, we have

$$\|Y_t^{\vartheta_1}-Y_t^{\vartheta_2}\|_{L^p(\Omega)}\le\frac{\varepsilon}{2} \tag{3.19}$$

We conclude by plugging (3.18) and (3.19) into (3.17). This yields

$$\mathcal{W}_p(\nu_{\vartheta_1},\nu_{\vartheta_2})\le\varepsilon, \tag{3.20}$$

for all $(\vartheta_1,\vartheta_2)\in\Theta^2$ such that $|\vartheta_1-\vartheta_2|\le\delta$. The continuity of $\vartheta\mapsto\mathcal{W}_p(\nu_\vartheta,\nu_{\vartheta_0})$ on Θ follows. □

3.3. **Further controls under $(\mathbf{C_s})$.** Up to now we have derived properties of the system (2.7) under the weak coercive assumption $(\mathbf{C_w})$. In this section, we focus on possible additional bounds one can obtain under the stronger hypothesis $(\mathbf{C_s})$. We will first see how $(\mathbf{C_s})$ guarantees a uniform control on the distance between the Euler scheme and the SDE, for a general decreasing sequence of time steps. Then we will show that $(\mathbf{C_s})$ ensures a some additional uniform continuity in ϑ for the occupation measures of Y^ϑ.

We consider here Euler type approximations in continuous time, with time steps γ_n satisfying (2.19). In order to define this Euler approximation $(Z_t^\vartheta)_{t\ge0}$, we set $s_0=0$ and $s_n=\sum_{i=1}^n\gamma_i$ for all $n\ge1$. Then for any $n\ge0$, the process $(Z_t^\vartheta)_{t\ge0}$ is given recursively by

$$Z_{s_n+t}^\vartheta=Z_{s_n}^\vartheta+tb_\vartheta(Z_{s_n}^\vartheta)+\sigma(B_{s_n+t}-B_{s_n}),\quad t\in[0,s_{n+1}-s_n]. \tag{3.21}$$

Notice that the fractional Brownian motion B in (3.21) is the same as the fBm driving equation (2.7). The control we get on Z^ϑ is summarized in the following proposition.

Proposition 3.7. *Let Y^ϑ be the solution of equation (2.7), and consider the continuous-time Euler scheme $(Z_t^\vartheta)_{t\ge0}$ with a time steps sequence $(\gamma_n)_{n\ge1}$ defined by (3.21). We assume that $(\mathbf{C_s})$ holds. Then the following assertions hold true.*

(i) For any $p\ge2$, there exist some positive constants ρ and C such that for any $n\ge1$ we have

$$|Y_{s_n}^\vartheta-Z_{s_n}^\vartheta|^p\le e^{-\rho s_n}|Y_0^\vartheta-Z_0^\vartheta|^p+C\sum_{k=0}^{n-1}\phi_{k,p}(Z_{s_k}^\vartheta)e^{-\rho(s_n-s_{k+1})}, \tag{3.22}$$

where the function $\phi_{k,p}$ is defined, for any $k\ge0$, by

$$\phi_{k,p}(z)=\gamma_{k+1}^{p+1}|b_\vartheta(z)|^p+\int_0^{\gamma_{k+1}}|B_{s_k+t}-B_{s_k}|^pdt. \tag{3.23}$$

(ii) Assume in addition that $\gamma_n\to0$ as $n\to+\infty$. Then for any $p\ge2$, there exists $n_0\in\mathbb{N}$ and some positive constants ρ and C such that for any $n\ge n_0$ we have

$$|Y_{s_n}^\vartheta-Z_{s_n}^\vartheta|^p\le e^{-\rho(s_n-s_{n_0})}|Y_{s_{n_0}}^\vartheta-Z_{s_{n_0}}^\vartheta|^p+C\sum_{k=n_0}^{n-1}\phi_{k,p}(Y_{s_k})e^{-\rho(s_n-s_{k+1})}. \tag{3.24}$$

Proof. Let $n \geq 0$ and consider the dynamics of $Y^\vartheta - Z^\vartheta$ on $[s_n, s_{n+1})$. That is, set $\varepsilon_t := Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t}$ for $t \in [0, \gamma_{n+1})$. Then ε_t verifies the relation

$$\varepsilon_t = Y^\vartheta_{s_n} - Z^\vartheta_{s_n} + \int_{s_n}^{s_n+t} \left(b_\vartheta(Y^\vartheta_s) - b_\vartheta(Z^\vartheta_{s_n}) \right) ds. \tag{3.25}$$

Starting from this equation, we divide the proof in several steps.

Step 1: Contracting bound for ε_t. Consider a parameter $\eta > 0$. We wish to use the coercivity assumption $(\mathbf{C_s})(i)$ in order to get an upper-bound on the following derivative:

$$\left(e^{\eta t} |\varepsilon_t|^p \right)' = e^{\eta t} \left(p|\varepsilon_t|^{p-2} \langle \varepsilon_t, \varepsilon_t' \rangle + \eta |\varepsilon_t|^p \right). \tag{3.26}$$

To this aim, observe that thanks to (3.25) the quantity $\langle \varepsilon_t, \varepsilon_t' \rangle$ can be expressed as:

$$\begin{aligned} \langle \varepsilon_t, \varepsilon_t' \rangle &= \langle Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t},\, b_\vartheta(Y^\vartheta_{s_n+t}) - b_\vartheta(Z^\vartheta_{s_n}) \rangle \\ &= \langle Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t},\, b_\vartheta(Y^\vartheta_{s_n+t}) - b_\vartheta(Z^\vartheta_{s_n+t}) \rangle + \langle Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t},\, b_\vartheta(Z^\vartheta_{s_n+t}) - b_\vartheta(Z^\vartheta_{s_n}) \rangle. \end{aligned}$$

Then we invoke $(\mathbf{C_s})(i)$ and the elementary inequality $ab \leq a^2 + b^2$, valid for all $a, b \geq 0$. We obtain

$$\begin{aligned} \langle \varepsilon_t, \varepsilon_t' \rangle &\leqslant -\alpha |Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t}|^2 + \frac{\alpha}{2} |Y^\vartheta_{s_n+t} - Z^\vartheta_{s_n+t}|^2 + \frac{2}{\alpha} |b_\vartheta(Z^\vartheta_{s_n+t}) - b_\vartheta(Z^\vartheta_{s_n})|^2 \\ &\leqslant -\frac{\alpha}{2} |\varepsilon_t|^2 + \frac{2L^2}{\alpha} |Z^\vartheta_{s_n+t} - Z^\vartheta_{s_n}|^2, \end{aligned} \tag{3.27}$$

where the second inequality is due to relation (2.5) and the definition (3.25) of ε_t. We now plug relation (3.21) into this inequality in order to get

$$\langle \varepsilon_t, \varepsilon_t' \rangle \leqslant -\frac{\alpha}{2} |\varepsilon_t|^2 + \frac{2L^2}{\alpha} \left| t b(Z^\vartheta_{s_n}) + \sigma(B_{s_n+t} - B_{s_n}) \right|^2,$$

from which we easily end up with

$$p|\varepsilon_t|^{p-2} \langle \varepsilon_t, \varepsilon_t' \rangle \leqslant -\frac{p\alpha}{2} |\varepsilon_t|^p + \frac{4pL^2}{\alpha} |\varepsilon_t|^{p-2} \left(t^2 |b_\vartheta(Z^\vartheta_{s_n})|^2 + |\sigma|^2 |B_{s_n+t} - B_{s_n}|^2 \right). \tag{3.28}$$

Eventually we apply Young's inequality with parameters $\bar{p} = p/(p-2)$ and $\bar{q} = p/2$ and some appropriate weights to (3.28). This yields the existence of a constant $C = C_{p,L}$ such that

$$p|\varepsilon_t|^{p-2} \langle \varepsilon_t, \varepsilon_t' \rangle \leqslant -\frac{p\alpha}{4} |\varepsilon_t|^p + C \left(t^p |b_\vartheta(Z^\vartheta_{s_n})|^p + |B_{s_n+t} - B_{s_n}|^p \right). \tag{3.29}$$

We are now ready to give some information about expressions of the form $e^{\eta t} |\varepsilon_t|^p$. Namely we set $\eta = p\alpha/4$, then we apply identity (3.26) and inequality (3.29). This easily yields

$$\left(e^{\eta t} |\varepsilon_t|^p \right)' \leqslant e^{\eta t}\ C(t^p |b_\vartheta(Z^\vartheta_{s_n})|^p + |B_{s_n+t} - B_{s_n}|^p). \tag{3.30}$$

Step 2: Inductive procedure. Let us integrate (3.30) on the interval $[0, \gamma_{n+1}]$, where we recall that $\gamma_{n+1} = s_{n+1} - s_n$. With the definition (3.25) in mind, this gives

$$|Y^\vartheta_{s_{n+1}} - Z^\vartheta_{s_{n+1}}|^p \leqslant e^{-\eta\gamma_{n+1}} |Y^\vartheta_{s_n} - Z^\vartheta_{s_n}|^p + C \int_0^{\gamma_{n+1}} e^{-\eta(\gamma_{n+1}-t)} (t^p |b_\vartheta(Z^\vartheta_{s_n})|^p + |B_{s_n+t} - B_{s_n}|^p) dt,$$

which yields

$$|Y^\vartheta_{s_{n+1}} - Z^\vartheta_{s_{n+1}}|^p \leqslant e^{-\eta\gamma_{n+1}} |Y^\vartheta_{s_n} - Z^\vartheta_{s_n}|^p + C \left(\gamma_{n+1}^{p+1} |b_\vartheta(Z^\vartheta_{s_n})|^p + \int_0^{\gamma_{n+1}} |B_{s_n+t} - B_{s_n}|^p) dt \right). \tag{3.31}$$

Thus, setting $\phi_{n,p}(x) = \left(\gamma_{n+1}^{p+1}|b_\vartheta(x)|^p + \int_0^{\gamma_{n+1}} |B_{s_n+t} - B_{s_n}|^p)dt\right)$, an elementary induction procedure yields the following relation for every $n \geq 1$:

$$|Y_{s_n}^\vartheta - Z_{s_n}^\vartheta|^p \leq |Y_0^\vartheta - Z_0^\vartheta|^p e^{-\eta s_n} + \sum_{k=0}^{n-1} \phi_{k,p}(Z_{s_k}^\vartheta) e^{-\eta(s_n - s_{k+1})}.$$

This proves our claim (3.22).

Step 3: Proof of (3.24). In order to obtain our second statement (3.24), one needs to go back to inequality (3.31). Then observe that hypothesis (2.5) yields

$$|b(Z_{s_n}^\vartheta)|^p \leq 2^{p-1}\left(|b(Y_{s_n}^\vartheta)|^p + L^p|Y_{s_n}^\vartheta - Z_{s_n}^\vartheta|^p\right),$$

which leads for any p to the existence of a constant C such that

$$|Y_{s_{n+1}}^\vartheta - Z_{s_{n+1}}^\vartheta|^p \leqslant \left(e^{-\eta\gamma_{n+1}} + C\gamma_{n+1}^{p+1}\right)|Y_{s_n}^\vartheta - Z_{s_n}^\vartheta|^p + C\phi_{n,p}(Y_{s_n}). \tag{3.32}$$

Since $\lim_{n\to\infty}\gamma_n = 0$, one checks easily that there exists a n_0 such that for any $n \geq n_0$ we have the following inequality:

$$e^{-\eta\gamma_{n+1}} + C\gamma_{n+1}^{p+1} \leq e^{-\frac{\eta}{2}\gamma_{n+1}}.$$

Plugging this information into (3.32), we end up with

$$|Y_{s_{n+1}}^\vartheta - Z_{s_{n+1}}^\vartheta|^p \leqslant e^{-\frac{\eta}{2}\gamma_{n+1}}|Y_{s_n}^\vartheta - Z_{s_n}^\vartheta|^p + C\phi_{k,p}(Y_{s_n}).$$

Our assertion (3.24) then follows by an induction procedure exactly as for Step 2. □

We now give a continuity results (with respect to the parameter ϑ) for some occupation measures related to our processes of interest. The proofs are postponed to Appendix B.

Proposition 3.8. *As in Proposition 3.6, let Y^ϑ be the solution of equation (2.7) and consider the Euler scheme $Z^{\vartheta,\gamma}$ defined by (2.13). Also consider $p > 0$ and $d \in \mathcal{D}_p$. We assume that $(\mathbf{C_s})$ holds true. Then,*

(i) The occupation measures of the process Y^ϑ are Lipschitz with respect to ϑ, that is there exists a positive random variable C_p such that for all $t \geq 1$:

$$d\left(\frac{1}{t}\int_0^t \delta_{Y_s^{\vartheta_1}} ds, \frac{1}{t}\int_0^t \delta_{Y_s^{\vartheta_2}} ds\right) \leq C_p|\vartheta_1 - \vartheta_2|. \tag{3.33}$$

(ii) The occupation measures of the Euler approximation $Z^{\vartheta,\gamma}$ are also Lipschitz with respect to ϑ. Namely there exists $\gamma_0 > 0$ such that: for any $\gamma \in (0, \gamma_0]$, there exists a positive random variable $C_p(\gamma)$ such that for all $N \geq 1$

$$d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta_1,\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta_2,\gamma}}\right) \leq C_p(\gamma)|\vartheta_1 - \vartheta_2|. \tag{3.34}$$

Remark 3.9. In the sequel we will analyze several quantities like (3.34), where we compare two discrete random measures on $\mathbb{R}^d$ $\nu_1 = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$ and $\nu_2 = \frac{1}{N}\sum_{i=1}^N \delta_{Y_i}$ with $X_i = X_i(\omega)$ and $Y_i = Y_i(\omega)$. In this context we will always upper bound quantities of the form $d(\nu_1, \nu_2)$ for a distance d in $\mathcal{D}_p$. To this aim, resorting a trivial coupling between ν_1 and ν_2, it is enough to prove an almost sure upper bound on

$$\frac{1}{N}\sum_{i=1}^N |X_i - Y_i|^p.$$

We will adopt this strategy throughout the paper, the typical outcome being an a.s. bound on $d(\nu_1, \nu_2)$. Straightforward extensions to a continuous time setting allow to handle quantities of the form (3.33).

4. Proof of the consistency theorems

The aim of this section is to achieve the proof of Theorems 2.8, 2.9 and 2.10. We first establish a general asymptotic result for a family of contrasts in Sections 4.1 and 4.2. Then we will combine this general proposition with our preliminary results of Section 3 in order to prove our main claims.

4.1. Uniform convergence of the contrast.

In this section we state some uniform convergence results for the *contrast*, i.e for the function involved in the definition of estimators such as (2.15). We should notice at this point that our uniform convergence results hold only under the assumption $(\mathbf{C_s})$. In case of a constant time step Euler scheme, we get the following result.

Proposition 4.1. *We consider the same setting as in Proposition 3.8. In particular, we assume that* $(\mathbf{C_s})$ *holds true for the coefficients of equation* (2.7)*. Then the following assertions hold true.*

(i) *The invariant measure* ν^γ *of the Euler scheme converges uniformly to the invariant measure of* Y^ϑ*. Namely*

$$\limsup_{\gamma\to 0} \gamma^{-H} \sup_{\vartheta\in\Theta} d(\nu_\vartheta, \nu_\vartheta^\gamma) < +\infty.$$

In particular, we have

$$\lim_{\gamma\to 0} \sup_{\vartheta\in\Theta} d(\nu_\vartheta, \nu_\vartheta^\gamma) = 0.$$

(ii) *The occupation measure of the Euler scheme converges to the invariant measure* ν_ϑ^γ *as the number of steps goes to* ∞*, that is:*

$$\lim_{N\to+\infty} \sup_{\vartheta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta_{Z_{s_k}^{\vartheta,\gamma}}, \nu_\vartheta^\gamma\right) = 0.$$

(iii) *We have*

$$\lim_{\gamma\to 0} \lim_{N,n\to+\infty} \sup_{\vartheta\in\Theta} \left| d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1} \delta_{Z_{s_k}^{\vartheta,\gamma}}\right) - d(\nu_{\vartheta_0}, \nu_\vartheta)\right| = 0.$$

Proof. We prove the three items separately.

Proof of (i). For sake of simplicity, we only detail the proof for $p = 2$. The extension to a general p does not generate particular difficulties and can be done as in Proposition 3.7. We start from the following inequality:

$$d\left(\nu_\vartheta^\gamma, \nu_\vartheta\right) \leqslant d\left(\mathcal{L}\left(Z_{N\gamma}^{\vartheta,\gamma}\right), \mathcal{L}\left(Y_{N\gamma}^{\vartheta}\right)\right) + d\left(\nu_\vartheta^\gamma, \mathcal{L}\left(Z_{N\gamma}^{\vartheta,\gamma}\right)\right) + d\left(\mathcal{L}\left(Y_{N\gamma}^{\vartheta}\right), \nu_\vartheta\right). \tag{4.1}$$

Let us consider the three terms of the right-hand side of (4.1) successively. First, without loss of generality, we can assume that $Y_0^\vartheta = Z_0^{\vartheta,\gamma}$. Furthermore, we have that $Z_{n\gamma}^{\vartheta,\gamma} = Z_{n\gamma}^{\vartheta}$ for all $n \geq 1$, where Z^ϑ is defined by (3.21). Then, by Proposition 3.7 (i) applied with $\gamma_n = \gamma$, we have

$$\mathbf{E}[|Y_{n\gamma}^\vartheta - Z_{n\gamma}^{\vartheta,\gamma}|^2] \leq C \sum_{k=0}^{n-1} \mathbf{E}[\phi_{k,2}(Z_{k\gamma}^{\vartheta,\gamma})] e^{-\rho\gamma(n-1-k)},$$

where we recall that $\phi_{k,p}$ is defined by (3.23). Using that b_ϑ is sublinear (uniformly in ϑ) and the fact that the increments of B satisfy relation (1.1), one obtains

$$\mathbf{E}[|Y^\vartheta_{n\gamma} - Z^{\vartheta,\gamma}_{n\gamma}|^2] \le C \sum_{k=0}^{n-1} \left(\gamma^3 \left(1 + \mathbf{E}[|Z^{\vartheta,\gamma}_{k\gamma}|^2]\right) + \gamma^{2H+1}\right) e^{-\rho\gamma(n-1-k)}.$$

It follows easily that

$$\sup_{\vartheta\in\Theta,\gamma\in(0,\gamma_0]} \limsup_{n\to+\infty} \gamma^{-2H}\mathbf{E}[|Y^\vartheta_{n\gamma} - Z^{\vartheta,\gamma}_{n\gamma}|^2] \le 1 + \sup_{\vartheta\in\Theta,\gamma\in(0,\gamma_0]} \limsup_{n\to+\infty} \mathbf{E}[|Z^{\vartheta,\gamma}_{k\gamma}|^2].$$

Moreover, owing to Proposition A.2, there exists γ_0 such that

$$\limsup_{n\to+\infty} \mathbf{E}[|Z^{\vartheta,\gamma}_{k\gamma}|^2] < \infty.$$

Summarizing, we have obtained

$$\sup_{\vartheta\in\Theta,\gamma\in(0,\gamma_0]} \limsup_{n\to+\infty} \gamma^{-2H}\mathbf{E}[|Y^\vartheta_{n\gamma} - Z^{\vartheta,\gamma}_{n\gamma}|^2] < \infty.$$

We now consider the term $d(\nu^\gamma_\vartheta, \mathcal{L}(Z^{\vartheta,\gamma}_{N\gamma}))$ in the right hand side of (4.1). Using that ν^γ_ϑ is invariant, we remark that

$$d\left(\nu^\gamma_\vartheta, \mathcal{L}\left(Z^{\vartheta,\gamma}_{N\gamma}\right)\right) \le \mathbf{E}[|Z^{\vartheta,\gamma}_{N\gamma} - \bar{Z}^{\vartheta,\gamma}_{N\gamma}|^2]^{\frac{1}{2}}$$

where $\bar{Z}^{\vartheta,\gamma}$ denotes a stationary Euler scheme built with the same noise process as for for $Z^{\vartheta,\gamma}$. Thus, thanks to the fact that

$$Z^{\vartheta,\gamma}_{k\gamma} - \bar{Z}^{\vartheta,\gamma}_{k\gamma} = Z^{\vartheta,\gamma}_{(k-1)\gamma} - \bar{Z}^{\vartheta,\gamma}_{(k-1)\gamma} + \gamma\left(b(Z^{\vartheta,\gamma}_{(k-1)\gamma}) - b(\bar{Z}^{\vartheta,\gamma}_{(k-1)\gamma})\right)$$

a straightforward induction under assumption $(\mathbf{C_s})$, similar to (3.27), leads to

$$\left|Z^{\vartheta,\gamma}_{k\gamma} - \bar{Z}^{\vartheta,\gamma}_{k\gamma}\right|^2 \leqslant (1 - 2\gamma\alpha + \gamma^2 L^2)^k \left|Z^{\vartheta,\gamma}_0 - \bar{Z}^{\vartheta,\gamma}_0\right|^2. \tag{4.2}$$

We choose $\gamma_0 = \alpha/L^2$ in such a way that $2\alpha - \gamma L^2 \ge \alpha$ for any $\gamma \in (0, \gamma_0]$. In addition, recall that $Z^{\vartheta,\gamma}_0 = z_0$. Then, by possibly picking a smaller value of γ_0, we deduce from Proposition A.2 that

$$\limsup_{N\to+\infty} \sup_{\vartheta\in\Theta} d\left(\nu^\gamma_\vartheta, \mathcal{L}\left(Z^{\vartheta,\gamma}_{N\gamma}\right)\right) = 0.$$

Eventually, the last term in the right hand side of (4.1) tends to 0 uniformly in ϑ as $N \to +\infty$ by Proposition 3.1. This concludes the proof of (i).

Proof of (ii). By Proposition 3.4, the convergence of $d(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{k\gamma}}, \nu^\gamma_\vartheta)$ to 0 is true for the simple convergence. In order to extend this result to a uniform convergence in ϑ, we use Proposition 3.8(ii) to obtain that the family $\{\vartheta \mapsto d(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{k\gamma}}, \nu^\gamma_\vartheta);\ N \ge 1, \vartheta \in \Theta\}$ is equicontinuous for a fixed $\gamma \in (0, \gamma_0]$. Actually, for some given ϑ_1 and ϑ_2,

$$\left|d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta_1,\gamma}_{k\gamma}}, \nu^\gamma_{\vartheta_1}\right) - d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta_2,\gamma}_{k\gamma}}, \nu^\gamma_{\vartheta_2}\right)\right| \le d(\nu^\gamma_{\vartheta_1}, \nu^\gamma_{\vartheta_2}) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta_1,\gamma}_{k\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta_2,\gamma}_{k\gamma}}\right).$$

The second term goes to 0 as $\vartheta_1 - \vartheta_2 \to 0$ by Proposition 3.8(ii). This is also the case for the first one by letting N go to ∞ in Proposition 3.8(ii) (for instance).

Proof of (iii). This is a simple consequence of the two previous statements and of Proposition 3.3(ii). □

We now generalize Proposition 4.1 to the case of a decreasing time step for the Euler scheme (2.13).

Proposition 4.2. *We consider the same setting as in Proposition 3.8. In particular, we assume that* $(\mathbf{C_s})$ *holds true for the coefficients of equation* (2.7)*. Let* $p \geq 2$ *and consider* $d \in \mathcal{D}_p$*. Let* $\{s_k;\, k \geq 0\}$ *be the sequence of time steps defined by* (2.18)*, which is assumed to verify*

$$\exists\, p' \geq p \text{ such that } \sum_{k=1}^{+\infty} \frac{\gamma_{k+1}^{p'H+1}}{s_k} < +\infty. \tag{4.3}$$

Then we have

$$\lim_{N,n\to+\infty} \sup_{\vartheta\in\Theta} \left| d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}}, \frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}} \right) - d(\nu_{\vartheta_0}, \nu_{\vartheta}) \right| = 0. \tag{4.4}$$

Proof. For notational convenience, the proof will be detailed for the continuous-time Euler approximation $(Z^{\vartheta}_t)_{t\geq 0}$ defined by (3.21), with step sequence $(\gamma_n)_{n\geq 1}$. An application of the triangular inequality allows us to bound the left hand side of (4.4) as follows:

$$\left| d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}}, \frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}} \right) - d(\nu_{\vartheta_0}, \nu_{\vartheta}) \right| \leqslant A_{1,n} + A_{2,N}(\vartheta). \tag{4.5}$$

with

$$A_{1,n} := d\left(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}}, \nu_{\vartheta_0} \right), \qquad A_{2,N}(\vartheta) := d\left(\frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}}, \nu_{\vartheta} \right).$$

Our claim can thus be reduced to prove that

$$\lim_{n\to+\infty} A_{1,n} = 0, \quad \text{and} \quad \lim_{N\to+\infty} \sup_{\vartheta\in\Theta} A_{2,N}(\vartheta) = 0.$$

Furthermore, the fact that $\lim_{n\to+\infty} A_{1,n} = 0$ is a direct consequence of Proposition 3.3-(ii). We thus focus on the asymptotic behavior of $A_{2,N}$ in the remainder of the proof.

In order to bound $A_{2,N}$, let us set $\underline{s} = \max\{s_k, s_k \leq s\}$. Then we observe that

$$\frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}} = \frac{1}{s_N}\int_0^{s_N} \delta_{Z^{\vartheta}_{\underline{s}}} ds.$$

Therefore we can split $A_{2,N}$ into

$$A_{2,N}(\vartheta) \leqslant A_{21,N}(\vartheta) + A_{22,N}(\vartheta) + A_{23,N}(\vartheta) \tag{4.6}$$

where the quantities $A_{2j,N}(\vartheta)$ for $j = 1, 2, 3$ are defined as follows:

$$A_{21,N}(\vartheta) := d\left(\frac{1}{s_N}\int_0^{s_N} \delta_{Z^{\vartheta}_{\underline{s}}} ds, \frac{1}{s_N}\int_0^{s_N} \delta_{Y^{\vartheta}_{\underline{s}}} ds \right),$$

$$A_{22,N}(\vartheta) := d\left(\frac{1}{s_N}\int_0^{s_N} \delta_{Y^{\vartheta}_{\underline{s}}} ds, \frac{1}{s_N}\int_0^{s_N} \delta_{Y^{\vartheta}_{s}} ds \right), \tag{4.7}$$

$$A_{23,N}(\vartheta) := d\left(\frac{1}{s_N}\int_0^{s_N} \delta_{Y^{\vartheta}_{s}} ds, \nu_{\vartheta} \right). \tag{4.8}$$

We will now treat those three terms separately.

The term $A_{23,N}$ is easily handled by applying Proposition 3.3 (i) (simple convergence) and Proposition 4.1 (i) (equicontinuity). We thus get

$$\lim_{N\to+\infty} \sup_{\vartheta\in\Theta} A_{23,N}(\vartheta) = 0.$$

Now, let us prove that $\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} A_{22,N}(\vartheta)=0$. Since $d\in\mathcal{D}_p\subset\mathcal{D}_{p'}$, we can assume without loss of generality that $d=\mathcal{W}_{p'}$. We invoke the strategy outlined in Remark 3.9. This means that we are reduced to prove the following limit :

$$\sup_{\vartheta\in\Theta}\frac{1}{s_N}\int_0^{s_N}|Y_t^\vartheta-Y_{\underline{t}}^\vartheta|^{p'}dt\xrightarrow{N\to+\infty}0\quad a.s. \tag{4.9}$$

To this aim, we first note that

$$|Y_t^\vartheta-Y_{\underline{t}}^\vartheta|\le\int_{\underline{t}}^t|b_\vartheta(Y_s^\vartheta)|ds+\|\sigma\||B_t-B_{\underline{t}}|\,, \tag{4.10}$$

and thus there exists a constant $c_{p',\sigma}$ such that

$$\frac{1}{s_N}\int_0^{s_N}|Y_t^\vartheta-Y_{\underline{t}}^\vartheta|^{p'}dt\le c_{p,\sigma}\left[\frac{1}{s_N}\int_0^{s_N}\left|\int_{\underline{t}}^t|b_\vartheta(Y_s^\vartheta)|ds\right|^{p'}dt+\frac{1}{s_N}\int_0^{s_N}|B_t-B_{\underline{t}}|^{p'}dt\right]. \tag{4.11}$$

Next we upper bound the first term in the right hand side of (4.11) invoking successively Jensen's inequality and Fubini's theorem. We obtain that

$$\begin{aligned}\int_0^{s_N}\left|\int_{\underline{t}}^t|b_\vartheta(Y_s^\vartheta)|ds\right|^{p'}dt &\le \int_0^{s_N}|t-\underline{t}|^{p'-1}\left(\int_{\underline{t}}^t|b_\vartheta(Y_s^\vartheta)|^{p'}ds\right)dt\\ &= \int_0^{s_N}|b_\vartheta(Y_s^\vartheta)|^{p'}\left(\int_s^{\bar{s}}|t-\underline{s}|^{p'-1}dt\right)ds\,,\end{aligned}$$

where we have introduced the additional notation $\bar{s}=\min\{s_k,s_k>s\}$. Taking into account the fact that $|t-\underline{s}|\le\gamma_{k+1}$ for any $s,t\in[s_k,s_{k+1})$, we end up with

$$\int_0^{s_N}\left|\int_{\underline{t}}^t|b_\vartheta(Y_s^\vartheta)|ds\right|^{p'}dt\le\sum_{k=0}^{N-1}\gamma_{k+1}^{p'}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds. \tag{4.12}$$

Furthermore, since b_ϑ is uniformly sublinear in ϑ and owing to (3.9) we know that

$$C_1:=\sup_{\vartheta\in\Theta}\sup_{N\ge1}\frac{1}{s_N}\int_0^{s_N}|b_\vartheta(Y_s^\vartheta)|^{p'}ds=\sup_{\vartheta\in\Theta}\sup_{N\ge1}\frac{1}{s_N}\sum_{k=0}^{N-1}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds<+\infty\quad a.s.$$

Moreover, $\lim_{k\to+\infty}\gamma_{k+1}^{p'}=0$. Hence, for all $\varepsilon>0$ there exists $k_0\ge0$ such that for all $k\ge k_0$, we have $\gamma_{k+1}^{p'}<\varepsilon$. One can thus deduce that

$$\begin{aligned}&\sup_{\vartheta\in\Theta}\frac{1}{s_N}\sum_{k=0}^{N-1}\gamma_{k+1}^{p'}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds\\ &\qquad\leqslant\sup_{\vartheta\in\Theta}\frac{1}{s_N}\sum_{k=0}^{k_0-1}\gamma_{k+1}^{p'}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds+\sup_{\vartheta\in\Theta}\frac{1}{s_N}\sum_{k=k_0}^{N-1}\gamma_{k+1}^{p'}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds\\ &\qquad\leqslant C_1\left(\frac{\gamma_1^{p'}s_{k_0}}{s_N}+\varepsilon\right),\end{aligned} \tag{4.13}$$

from which it is easily seen that

$$\sup_{\vartheta\in\Theta}\frac{1}{s_N}\sum_{k=0}^{N-1}\gamma_{k+1}^{p'}\int_{s_k}^{s_{k+1}}|b_\vartheta(Y_s^\vartheta)|^{p'}ds\xrightarrow{N\to+\infty}0\quad a.s. \tag{4.14}$$

We can conclude that the first term in the right hand side of (4.11) vanishes as $N\to+\infty$ due to our identity (4.12).

In order to prove that the second term in the right hand side of (4.11) converges to 0, we invoke Kronecker's lemma (in its continuous version, see [17, Theorem 2.1]). We get that it is sufficient to prove that:

$$\int_0^{+\infty} \frac{|B_t - B_{\underline{t}}|^{p'}}{1+t} dt < +\infty \quad a.s.$$

However, due to the fact that $(t - \underline{t}) \leqslant \gamma_{k+1}$ if $t \in [s_k, s_{k+1})$, we have,

$$\mathbf{E}\left[\int_0^{+\infty} \frac{|B_t - B_{\underline{t}}|^{p'}}{1+t} dt\right] \leq \int_0^{+\infty} \frac{1}{1+t}(t-\underline{t})^{p'H} dt \leq c \sum_{k=1}^{+\infty} \frac{\gamma_{k+1}^{p'H+1}}{s_k} < +\infty$$

where c is a positive constant and where the last inequality stems from hypothesis (4.3). Hence, Kronecker's lemma yields

$$\frac{1}{s_N} \int_0^{s_N} |B_t - B_{\underline{t}}|^{p'} ds \xrightarrow{N \to +\infty} 0 \quad a.s. \tag{4.15}$$

Now inequality (4.11) combined with (4.14) and (4.15) easily yields (4.9). We conclude that $\lim_{N\to+\infty} \sup_{\vartheta \in \Theta} A_{22,N}(\vartheta) = 0$.

Going back to our decomposition (4.6), we still have to prove that $\lim_{N\to+\infty} \sup_{\vartheta\in\Theta} A_{21,N}(\vartheta) = 0$. To this end, let us write $A_{21,N}(\vartheta)$ in its discrete form:

$$A_{21,N}(\vartheta) = d\left(\frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}}, \frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}\delta_{Y^{\vartheta}_{s_k}}\right). \tag{4.16}$$

Then invoking Remark 3.9 once more, we are reduced to prove

$$\frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}|Y^{\vartheta}_{s_k} - Z^{\vartheta}_{s_k}|^{p'} \xrightarrow{N\to+\infty} 0 \quad a.s. \tag{4.17}$$

In order to achieve (4.17) consider the integer n_0 given by Proposition 3.7 (ii). For $k \leqslant n_0$ we trivially bound $|Y^{\vartheta}_{s_k} - Z^{\vartheta}_{s_k}|$ using (3.22). Since (3.21) asserts that $\sup_{\vartheta\in\Theta, k\in\{1,\dots,n_0\}} |Z^{\vartheta}_{s_k}| < +\infty$, we get that

$$\sup_{\vartheta\in\Theta, k\in\{1,\dots,n_0\}} |Y^{\vartheta}_{s_k} - Z^{\vartheta}_{s_k}| =: C(\omega) < +\infty \quad a.s. \tag{4.18}$$

We now bound the right hand side of (4.17) by means of (4.18) whenever $k \leqslant n_0$ and invoking (3.24) when $k \geqslant n_0 + 1$. This gives

$$\begin{aligned}\frac{1}{s_N}\sum_{k=0}^{N-1} \gamma_{k+1}|Y^{\vartheta}_{s_k} - Z^{\vartheta}_{s_k}|^{p'} &\leq C(\omega)^{p'} \frac{1}{s_N}\left(\sum_{k=0}^{n_0}\gamma_{k+1} + \sum_{k=n_0+1}^{N-1}\gamma_{k+1}e^{-\rho(s_k - s_{n_0})}\right) \\ &\quad + \frac{1}{s_N}\sum_{k=n_0+1}^{N-1}\gamma_{k+1}\sum_{\ell=n_0}^{k-1}\phi_{\ell,p'}(Y^{\vartheta}_{s_\ell})e^{-\rho(s_k - s_{\ell+1})}.\end{aligned} \tag{4.19}$$

The first term in the right-hand side of (4.19) is clearly evanescent as $N \to +\infty$. Let us focus on the second one. By a Fubini type transformation, we get

$$\begin{aligned}\sum_{k=n_0+1}^{N-1}\gamma_{k+1}\sum_{\ell=n_0}^{k-1}\phi_{\ell,p'}(Y^{\vartheta}_{s_\ell})e^{-\rho(s_k - s_{\ell+1})} &= \sum_{\ell=n_0}^{N-2}\phi_{\ell,p'}(Y^{\vartheta}_{s_\ell})\sum_{k=\ell+1}^{N-1}\gamma_{k+1}e^{-\rho(s_k - s_{\ell+1})} \\ &\leq \sum_{\ell=0}^{N-2}\phi_{\ell,p'}(Y^{\vartheta}_{s_\ell})\sum_{k=\ell+1}^{N-1}\gamma_{k}e^{-\rho(s_k - s_{\ell+1})}\end{aligned}$$

where the last inequality is due to the fact that (γ_k) is non increasing. Since $x \mapsto e^{-\rho x}$ is a non-increasing function, we remark that for $\ell \in \{0, \dots, N-2\}$,

$$\sum_{k=\ell+1}^{N-1} \gamma_k e^{-\rho(s_k - s_{\ell+1})} \leq e^{\rho s_{\ell+1}} \int_{s_\ell}^{s_{N-1}} e^{-\rho x} dx = \frac{1}{\rho}\left(e^{\rho\gamma_{\ell+1}} - e^{-\rho(s_{N-1}-s_{\ell+1})}\right) \leq C_\rho.$$

Thus, in order to see that the right hand side of (4.19) vanishes as $N \to +\infty$, it remains to show that

$$\lim_{N\to+\infty} \sup_{\vartheta\in\Theta} \frac{1}{s_N} \sum_{\ell=0}^{N-1} \phi_{\ell,p'}(Y^\vartheta_{s_\ell}) = 0 \tag{4.20}$$

where we recall that $\phi_{\ell,p'}$ is defined by (3.23) and thus

$$\frac{1}{s_N}\sum_{\ell=0}^{N-1} \phi_{\ell,p'}(Y^\vartheta_{s_\ell}) = \frac{1}{s_N}\sum_{\ell=0}^{N-1} \gamma_{\ell+1}^{p'+1} |b_\vartheta(Y^\vartheta_{s_\ell})|^{p'} + \frac{1}{s_N}\sum_{\ell=0}^{N-1}\int_0^{\gamma_{\ell+1}} |B_{s_\ell+t} - B_{s_\ell}|^{p'} dt. \tag{4.21}$$

Let us begin by the term involving b_ϑ in (4.21). First notice that

$$\mathcal{W}_{p'}\left(\frac{1}{s_N}\int_0^{s_N} \delta_{Y^\vartheta_{\underline{s}}} ds, \nu_\vartheta\right) \leqslant A_{22,N}(\vartheta) + A_{23,N}(\vartheta),$$

where $A_{22,N}$ and $A_{23,N}$ are respectively defined by (4.7) and (4.8) with $d = \mathcal{W}_{p'}$. Furthermore, we have seen that $\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} A_{22,N}(\vartheta) = 0$ and $\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} A_{23,N}(\vartheta) = 0$. We immediately deduce that

$$\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} \mathcal{W}_{p'}\left(\frac{1}{s_N}\int_0^{s_N} \delta_{Y^\vartheta_{\underline{s}}} ds, \nu_\vartheta\right) = 0.$$

Moreover, by Proposition A.1 for instance, we have $\sup_{\vartheta\in\Theta}\nu_\vartheta(|\cdot|^{p'}) < \infty$. We thus deduce that

$$\sup_{\vartheta\in\Theta}\sup_{N\geqslant 1} \frac{1}{s_N}\sum_{\ell=0}^{N-1} \gamma_{\ell+1}|Y^\vartheta_{s_\ell}|^{p'} < +\infty \quad a.s.$$

Furthermore, since b_ϑ is uniformly sublinear in ϑ and $\lim_{\ell\to+\infty}\gamma_{\ell+1}^{p'} = 0$, it easily follows along the same lines as for (4.14) that:

$$\sup_{\vartheta\in\Theta}\frac{1}{s_N}\sum_{\ell=0}^{N-1}\gamma_{\ell+1}^{p'+1}|b_\vartheta(Y_{s_\ell})|^{p'} \xrightarrow{N\to+\infty} 0. \tag{4.22}$$

We now turn to the term in (4.21) involving the fBm, for which we use the classical discrete Kronecker lemma. To this aim, we remark that

$$\mathbf{E}\left[\sum_{k=1}^{+\infty}\frac{1}{s_k}\int_0^{\gamma_{k+1}} |B_{s_k+t} - B_{s_k}|^{p'} dt\right] = \frac{1}{p'H+1}\sum_{k=1}^{+\infty}\frac{\gamma_{k+1}^{p'H+1}}{s_k} < +\infty, \tag{4.23}$$

where the last inequality stems from assumption (4.3). From (4.23), it is easily seen that

$$\sum_{\ell=0}^{+\infty}\frac{1}{s_\ell}\int_0^{\gamma_{\ell+1}} |B_{s_\ell+t} - B_{s_\ell}|^{p'} dt < +\infty \quad a.s.$$

We are thus in position to apply Kronecker's lemma to the sequence $(\int_0^{\gamma_{\ell+1}} |B_{s_\ell+t} - B_{s_\ell}|^{p'} dt)_{\ell\geqslant 0}$. This yields

$$\frac{1}{s_N}\sum_{\ell=0}^{N-1}\int_0^{\gamma_{\ell+1}} |B_{s_\ell+t} - B_{s_\ell}|^{p'} dt \xrightarrow{N\to+\infty} 0. \tag{4.24}$$

Let us summarize our computations so far. Gathering (4.22), (4.24) and (4.21), we have obtained relation (4.20). This easily implies that (4.17) holds true and hence $A_{21,N}(\vartheta)$ defined by (4.16) satisfies

$$\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} A_{21,N}(\vartheta)=0.$$

Since similar results have also been shown for $A_{22,N}(\vartheta)$ and $A_{23,N}(\vartheta)$, relation (4.6) gives

$$\lim_{N\to+\infty}\sup_{\vartheta\in\Theta} A_{2,N}(\vartheta)=0.$$

Eventually, plugging this information into (4.5) yields our claim (4.4). □

4.2. **A general convergence result.** The consistence of our estimators will rely on the following general proposition about convergence of minimizers for a sequence of random functions. Observe that our sequences below are indexed by a generic r which sits in an unspecified set. This simplifies the subsequent applications of the proposition to our indices N, n, γ in the remainder of the section.

Proposition 4.3. *Let Θ be a compact set and $(\vartheta\mapsto L_r(\vartheta))_r$ denote a family of non-negative random functions. Assume that:*

(1) With probability one, $\lim_r L_r(\vartheta)=L(\vartheta)$ uniformly in $\vartheta\in\Theta$.
(2) The function $\vartheta\mapsto L(\vartheta)$ is non-random and continuous on Θ.
(3) For any r, the set $\operatorname{argmin}\{L_r(\vartheta),\vartheta\in\Theta\}$ is nonempty.

Then, for a fixed r, let $\hat\vartheta_r\in\operatorname{argmin}\{L_r(\vartheta),\vartheta\in\Theta\}$. Let $\mathcal{A}$ denote the limit points of $(\hat\vartheta_r)_r$. Then we have

$$\mathcal{A}\subset\operatorname{argmin}\{L(\vartheta),\vartheta\in\Theta\}.$$

In particular, if L attains its minimum for a unique $\vartheta^\star$, then $\lim_r\hat\vartheta_r=\vartheta^\star$.

Proof. Let $\vartheta^\star$ be an element of $\operatorname{argmin}\{L(\vartheta),\vartheta\in\Theta\}$. We consider a generic element $\vartheta^\infty\in\mathcal{A}$ and its related convergent subsequence $(\hat\vartheta_{r_n})_{n\geqslant 0}$. Then we can upper bound $L(\vartheta^\infty)$ as follows:

$$L(\vartheta^\infty)\leq L_{r_n}(\hat\vartheta_{r_n})+|L(\vartheta^\infty)-L_{r_n}(\hat\vartheta_{r_n})|. \tag{4.25}$$

We now bound the two terms in the right hand side of (4.25). On the one hand, by definition of $\hat\vartheta_r$,

$$L_{r_n}(\hat\vartheta_{r_n})\leq L_{r_n}(\vartheta^\star).$$

Hence, thanks to the fact that $\lim_r L_r(\vartheta)=L(\vartheta)$ for all $\vartheta\in\Theta$, we get

$$\limsup_{n\to+\infty} L_{r_n}(\hat\vartheta_{r_n})\leq L(\vartheta^\star). \tag{4.26}$$

On the other hand, we also have

$$|L(\vartheta^\infty)-L_{r_n}(\hat\vartheta_{r_n})|\leq|L(\vartheta^\infty)-L(\hat\vartheta_{r_n})|+\sup_{\vartheta\in\Theta}|L(\vartheta)-L_{r_n}(\vartheta)|. \tag{4.27}$$

Therefore we can invoke the continuity of L to bound the first term in the right hand side of (4.27), plus the uniform convergence of L_r to L in order to handle the second term. This yealds

$$\limsup_{n\to+\infty}|L(\vartheta^\infty)-L_{r_n}(\hat{\vartheta_{r_n}})|=0 \tag{4.28}$$

Plugging (4.28) and (4.26) into (4.25), we obtain that $L(\vartheta^\infty)\leq L(\vartheta^\star)$ and thus ϑ^∞ belongs to the set $\operatorname{argmin}\{L(\vartheta),\vartheta\in\Theta\}$. This finishes the proof. □

4.3. **Proofs of the convergence theorems.** With all our preliminary considerations in hand, we are now ready to prove the main convergence results for our estimators. This is briefly outlined below.

Proof of Theorem 2.9. Recall that the family $\{\hat{\vartheta}_{N,n,\gamma},\ (N,n,\gamma) \in \mathbb{N}^2 \times \mathbb{R}_+^*\}$ is defined by (2.15). We wish to apply Proposition 4.3 with $r = (N,n,\gamma) \in \mathbb{N}^2 \times \mathbb{R}_+^*$. We set $L_{N,n,\gamma}(\vartheta) = d(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{s_k}})$. By Proposition 4.1, we have uniformly in $\vartheta \in \Theta$,

$$\lim_{\gamma\to 0}\lim_{N,n\to+\infty} L_{N,n,\gamma}(\vartheta) = d(\nu_{\vartheta_0},\nu_\vartheta) =: L(\vartheta). \tag{4.29}$$

In addition, owing to Proposition 3.6 and Assumption $(\mathbf{I_w})$, L is continuous and ϑ_0 is the unique minimum of L. We have thus checked that the hypothesis of Proposition 4.3 are fulfilled, from which Theorem 2.9 is easily deduced. □

Proof of Theorem 2.10 . The proof goes along the same lines as for Theorem 2.9. Namely we apply Proposition 4.3 to the sequence $\{\hat{\vartheta}_{N,n},\ (N,n) \in \mathbb{N}^2\}$ defined by (2.20). To this aim, we set

$$L_{N,n}(\vartheta) = d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\gamma_{k+1}\delta_{Z^{\vartheta}_{s_k}}\right).$$

Then according to Proposition 4.2, the sequence $(L_{N,n})_{N,n}$ converges uniformly to L defined by (4.29) when $N,n \to +\infty$. Furthermore, the continuity of L follows as in the proof of Theorem 2.9. Our claim is thus easily deduced. □

Proof of Theorem 2.8 . We still wish to apply Proposition 4.3 to the family $\{\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma},\ (N,n,\gamma) \in \mathbb{N}^2 \times \mathbb{R}_+^*\}$ defined by (2.17). However, since we only assume $(\mathbf{C_w})$ instead of $(\mathbf{C_s})$, one is only able to obtain simple convergence properties on Θ. In order to circumvent this problem, we have restricted our analysis to the discretized parameter set $\Theta^{(\varepsilon)}$ introduced in (2.17). For a given $\varepsilon > 0$, $\Theta^{(\varepsilon)}$ is finite and hence, one deduces from Propositions 3.3(ii) and 3.4, that

$$\lim_{\gamma\to 0}\lim_{N,n\to+\infty}\sup_{\vartheta\in\Theta^{(\varepsilon)}} |L_{N,n,\gamma}(\vartheta) - L(\vartheta)| = 0,$$

where L is defined by (4.29). Now, denote by $\mathcal{A}^{(\varepsilon)}$ the set of limit points of $(\hat{\vartheta}^{(\varepsilon)}_{N,n,\gamma})_{N,n,\gamma}$. From Proposition 4.3, one deduces that

$$\mathcal{A}^{(\varepsilon)} \subset \operatorname{argmin}\{L(\vartheta), \vartheta \in \Theta^{(\varepsilon)}\}.$$

Furthermore, L is a continuous function such that $L(\vartheta_0) = 0$. Thus, since $\operatorname{dist}(\vartheta_0, \Theta^{(\varepsilon)}) \to 0$ as $\varepsilon \to 0$, one deduces that $\min_{\vartheta\in\Theta^{(\varepsilon)}} L(\vartheta) \to 0$ as $\varepsilon \to 0$. Owing to $(\mathbf{I_w})$, this implies that any sequence $(\vartheta^{(\varepsilon)})_\varepsilon$ of $\mathcal{A}^{(\varepsilon)}$ converges to ϑ_0. This concludes the proof. □

5. Rate of convergence: proof of Theorem 2.13

All along this section, we assume $(\mathbf{C_s})$ and $(\mathbf{I_s})$. Our aim is to bound the quantity $\mathbf{E}[|\hat{\vartheta}_{N,n,\gamma} - \vartheta_0|^2]$ where $\hat{\vartheta}_{N,n,\gamma}$ is defined by (2.15). Owing to $(\mathbf{I_s})$, we are reduced to study

$$\mathbf{E}\left[d\left(\nu_{\vartheta_0}, \nu_{\hat{\vartheta}_{N,n,\gamma}}\right)^q\right] \tag{5.1}$$

where $q := 2/\varsigma$ and $\varsigma \in (0,1]$ is given in $(\mathbf{I_s})$. Our strategy of proof is based on the following decomposition

Lemma 5.1. *Let $\hat{\vartheta}_{N,n,\gamma}$ be the estimator defined by (2.15) and recall that ν_ϑ is defined by (2.8) for all $\vartheta \in \Theta$. Then $d\left(\nu_{\vartheta_0}, \nu_{\hat{\vartheta}_{N,n,\gamma}}\right)$ can be decomposed as*

$$d\left(\nu_{\vartheta_0}, \nu_{\hat{\vartheta}_{N,n,\gamma}}\right) \leq 2D_n^{(1)} + 2\sup_{\vartheta\in\Theta} D_{N,\gamma}^{(2)}(\vartheta) + 2\sup_{\vartheta\in\Theta} D_{N,\gamma}^{(3)}(\vartheta) \tag{5.2}$$

where $D_n^{(1)}$, $D_{N,\gamma}^{(2)}(\vartheta)$, and $D_{N,\gamma}^{(3)}(\vartheta)$ are respectively by

$$D_n^{(1)} := d\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}\right), \tag{5.3}$$

$$D_{N,\gamma}^{(2)}(\vartheta) := d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta,\gamma}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Y_{k\gamma}^{\vartheta}}\right), \tag{5.4}$$

$$D_{N,\gamma}^{(3)}(\vartheta) := d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Y_{k\gamma}^{\vartheta}}, \nu_\vartheta\right). \tag{5.5}$$

Proof. Let us write $\hat{\vartheta}$ for $\hat{\vartheta}_{N,n,\gamma}$ throughout the proof in order to ease notations. We first apply the triangular inequality, which yields

$$d\left(\nu_{\vartheta_0}, \nu_{\hat{\vartheta}}\right) \leq d\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\hat{\vartheta},\gamma}}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\hat{\vartheta},\gamma}}, \nu_{\hat{\vartheta}}\right)$$

where $(Y_t)_{t\geq 0}$ is the observation process given by (1.2). Next, we invoke the fact that $\hat{\vartheta}$ minimizes the quantity $d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta,\gamma}}\right)$ in Θ, which gives

$$d\left(\nu_{\vartheta_0}, \nu_{\hat{\vartheta}}\right) \leq d\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta_0,\gamma}}\right) + \sup_{\vartheta\in\Theta} d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta,\gamma}}, \nu_\vartheta\right). \tag{5.6}$$

We further split the second term in the right hand side of (5.6) as follows:

$$\begin{aligned} &d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta_0,\gamma}}\right) \\ &\quad\leq d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \nu_{\vartheta_0}\right) + d\left(\nu_{\vartheta_0}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Y_{k\gamma}^{\vartheta_0}}\right) + d\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta_{Y_{k\gamma}^{\vartheta_0}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z_{k\gamma}^{\vartheta_0,\gamma}}\right), \end{aligned} \tag{5.7}$$

and resort to a similar decomposition for the third term in the right hand side of (5.6). It is then readily checked that plugging (5.7) into (5.6) we end up with our claim (5.2). □

In the remainder of the section, we shall handle the L^q-moments of $D_n^{(1)}$, $\sup_{\vartheta\in\Theta}|D_{N,\gamma}^{(2)}(\vartheta)|$ and $\sup_{\vartheta\in\Theta}|D_{N,\gamma}^{(3)}(\vartheta)|$ separately, respectively in Sections 5.1, 5.2 and 5.3.

5.1. L^q bound on $D_n^{(1)}$.

We start this section by giving a notation concerning expectations of empirical measures.

Notation 5.2. *Let Y be the solution of equation (2.7) and $t \geq 0$. As previously, δ_{Y_t} denotes the Dirac measure at Y_t, considered as a random measure. Then $\mathbf{E}[\delta_{Y_t}]$ is the deterministic measure such that for all continuous and bounded $f : \mathbb{R}^d \to \mathbb{R}$ we have*

$$\mathbf{E}[\delta_{Y_t}](f) = \mathbf{E}[f(Y_t)].$$

With this notation in mind, we can now deliver our L^q estimate for $D_n^{(1)}$.

Lemma 5.3. *Let $D_n^{(1)}$ be the random variable defined by (5.3). Then, whenever d is given by $d_{\mathrm{CF},p}$ or d_s, defined respectively by (2.21) with $p > (q+d)/2$ and (2.22), we have:*

$$\mathbf{E}\left[|D_n^{(1)}|^q\right] \leqslant C_q\left(n^{-q} + n^{-\frac{q}{2}(2-(2H\vee 1))}\right). \tag{5.8}$$

Proof. We decompose $D_n^{(1)}$ as follows:

$$D_n^{(1)} \leq d\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[\delta_{Y_{t_k}}]\right) + d\left(\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[\delta_{Y_{t_k}}], \frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}\right) =: D_n^{(11)} + D_n^{(12)}. \tag{5.9}$$

For the term $D_n^{(11)}$, we can use the contractivity assumption $(\mathbf{C_s})$ on the drift b which implies that two solutions of the SDE (1.2) with different initial conditions converge exponentially pathwise to each other as $t \to +\infty$ (see e.g. [21]). More specifically, we have already seen in (3.7) that the arguments of [21] entail $\|Y_t - \bar{Y}_t\|_{L^p(\Omega} \leq c_1 e^{-c_2 t}$ for two positive constants c_1, c_2, where we recall that $\bar{Y}$ designates the stationary solution of (2.7). Hence for a Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ we easily get the existence of a constant $C > 0$ such that

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[f(Y_{t_k})] - \nu_{\vartheta_0}(f)\right| \leq \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[|f(Y_{t_k}) - f(\bar{Y}_{t_k})|] \leq \frac{C}{n}\|f\|_{\mathrm{Lip}}. \tag{5.10}$$

It remains to take into account the distance d. Recall that we only consider the two distances $d_{\mathrm{CF},p}$ and d_s defined in Subsection 2.5. We thus easily deduce from (5.10) and the definitions of $d_{\mathrm{CF},p}$ and d_s that there exists a positive constant $\tilde{C}$ such that

$$D_n^{(11)} \leq \max\left\{d_{\mathrm{CF},p}\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[\delta_{Y_{t_k}}]\right), d_s\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[\delta_{Y_{t_k}}]\right)\right\} \leq \frac{\tilde{C}}{n}. \tag{5.11}$$

The term $D_n^{(12)}$ is handled in Proposition 5.4 below and specifically in relation (5.12). Therefore, plugging (5.11) and (5.12) into (5.9), relation (5.8) is proved. □

Proposition 5.4. *Let d be one of the two distances d_s and $d_{\mathrm{CF},p}$ with $p > (q+d)/2$. Then,*

$$\mathbf{E}\left[d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{n}\sum_{k=0}^{n-1}\mathbf{E}[\delta_{Y_{t_k}}]\right)^q\right] \leq C_q n^{-\frac{q}{2}(2-(2H\vee 1))}. \tag{5.12}$$

The proof of the proposition is based on the following lemma:

Lemma 5.5. *Recall that $(Y_t)_{t\geq 0}$ is given by (1.2). Then for all $q \geqslant 1$ and for all Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\mathbf{E}\left[\left|\frac{1}{n}\sum_{k=0}^{n-1} f(Y_{t_k}) - \mathbf{E}[f(Y_{t_k})]\right|^q\right] \leqslant C_q\|f\|_{\mathrm{Lip}}^q n^{-\frac{q}{2}(2-(2H\vee 1))}. \tag{5.13}$$

Proof. We invoke a concentration result for large time borrowed from [48, Theorem 2.3]. This result asserts that: there exists $C > 0$ such that for all Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ and for all $r \geqslant 0$,

$$\mathbf{P}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(Y_{t_k}) - \mathbf{E}[f(Y_{t_k})] \geq r\right) \leq C\exp(-C\|f\|_{\mathrm{Lip}}^{-2} r^2 n^{2-(2H\vee 1)}). \tag{5.14}$$

Therefore, one can check that (5.13) holds true by plugging inequality (5.14) into the classical formula

$$\mathbf{E}[X^q] = \int_0^{+\infty} q x^{q-1} \mathbf{P}(X \geq x) dx,$$

which is valid for any positive random variable X. □

Proof of Proposition 5.4. We will only give details about $d_{\mathrm{CF},p}$ since d_s can be treated exactly along the same lines. Furthermore, since our parameter q is greater than 2, by using Jensen inequality and the linearity of $\mathbf{E}$ into the definition (2.21) of $d_{\mathrm{CF},p}$, we get:

$$\mathbf{E}\left[d_{\mathrm{CF},p}\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}[\delta_{Y_{t_k}}] \right)^q \right] \leqslant \int_{\mathbb{R}^d} \mathbf{E}\left[\left| \frac{1}{n}\sum_{k=0}^{n-1} f_\xi(Y_{t_k}) - \mathbf{E}[f_\xi(Y_{t_k})] \right|^q \right] g_p(\xi) d\xi,$$

where $f_\xi(x) = e^{i\langle \xi, x\rangle}$. Since $\|f_\xi\|_{\mathrm{Lip}} \leq |\xi|$, we thus deduce from Lemma 5.5 that

$$\mathbf{E}\left[d_{\mathrm{CF},p}\left(\nu_{\vartheta_0}, \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}[\delta_{Y_{t_k}}] \right)^q \right] \leqslant C_q n^{-\frac{q}{2}(2-(2H\vee 1))} \int_{\mathbb{R}^d} |\xi|^q g_p(\xi) d\xi.$$

The integral in the last inequality is finite owing to the fact that we chose $p > (q+d)/2$. Our claim thus follows. □

5.2. L^q **bound on** $D^{(2)}_{N,\gamma}$**.** Our aim in this section is to get an equivalent of relation (5.8) for the term $D^{(2)}_{N,\gamma}$ defined by (5.4), namely

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |D^{(2)}_{N,\gamma}(\vartheta)|^q \right] \leq C\gamma^{qH}. \tag{5.15}$$

where the distance d in the definition of $D^{(2)}_{N,\gamma}$ is either d_s or $d_{\mathrm{CF},p}$. To this end, resorting to the fact that $d_{\mathrm{CF},p}$ and d_s are both elements of $\mathcal{D}_1$, the quantity (5.4) can be upper-bounded as follows:

$$\sup_{\vartheta\in\Theta} D^{(2)}_{N,\gamma}(\vartheta) \leq \frac{1}{N}\sum_{k=0}^{N-1} \sup_{\vartheta\in\Theta} |Y^\vartheta_{k\gamma} - Z^{\vartheta,\gamma}_{k\gamma}|.$$

We thus deduce that

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |D^{(2)}_{N,\gamma}(\vartheta)|^q \right] \leq \mathbf{E}\left[\left(\frac{1}{N}\sum_{k=0}^{N-1} \sup_{\vartheta\in\Theta} |Y^\vartheta_{k\gamma} - Z^{\vartheta,\gamma}_{k\gamma}| \right)^q \right].$$

Recall that $q \geq 2$. Hence a direct application of Jensen's inequality gives

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |D^{(2)}_{N,\gamma}(\vartheta)|^q \right] \leq \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{E}\left[\sup_{\vartheta\in\Theta} |Y^\vartheta_{k\gamma} - Z^{\vartheta,\gamma}_{k\gamma}|^q \right]. \tag{5.16}$$

Now according to Proposition 3.7 (i) and Proposition A.2 (ii) (see also the proof of Proposition 4.1 (i)), it is readily checked that

$$\sup_{k\geq 0} \mathbf{E}\left[\sup_{\vartheta\in\Theta} |Y^\vartheta_{k\gamma} - Z^{\vartheta,\gamma}_{k\gamma}|^q \right] \leqslant C\gamma^{qH}.$$

Gathering this information with (5.16), inequality (5.15) is easily deduced.

5.3. L^q **bound on** $D^{(3)}_{N,\gamma}$**.** The quantity (5.5) is the hardest to treat among the terms in our decomposition (5.2), due to the fact that we wish to achieve a uniform bound in ϑ. We summarize our analysis in the following lemma.

Lemma 5.6. *Let $D^{(3)}_{N,\gamma}$ be the random variable defined by (5.5), and assume that d is either d_s or $d_{\mathrm{CF},p}$ with $p > (q+d)/2$. Then, we have*

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} D^{(3)}_{N,\gamma}(\vartheta)^q\right] \leq C_q\left(\gamma^{qH} + T^{-\tilde{\eta}}\right) \tag{5.17}$$

with $\tilde{\eta} := \frac{q^2}{2(q+d)}(2-(2H\vee 1))$ and $T = N\gamma$.

Proof. We will further decompose the term $D^{(3)}_{N,\gamma}$ and then divide our analysis in several steps. First, let us introduce some notations: denote by T the quantity $N\gamma$ and for all $t\in[0,T]$, set $\underline{t} := \inf\{k\gamma \mid k\gamma \leqslant t < (k+1)\gamma\}$ as we did in the proof of Proposition 4.2. With this notations in hand, we have $D^{(3)}_{N,\gamma}(\vartheta) = d(\nu_\vartheta, \frac{1}{T}\int_0^T \delta_{Y^\vartheta_{\underline{t}}}dt)$ from which we deduce the following decomposition:

$$\sup_{\vartheta\in\Theta} D^{(3)}_{N,\gamma}(\vartheta) \leq \sup_{\vartheta\in\Theta} D^{(31)}_{N,\gamma}(\vartheta) + \sup_{\vartheta\in\Theta} D^{(32)}_{N,\gamma}(\vartheta) \tag{5.18}$$

where

$$D^{(31)}_{N,\gamma}(\vartheta) = d\left(\nu_\vartheta, \frac{1}{T}\int_0^T \delta_{Y^\vartheta_t}dt\right) \quad \text{and} \quad D^{(32)}_{N,\gamma}(\vartheta) = d\left(\frac{1}{T}\int_0^T \delta_{Y^\vartheta_t}dt, \frac{1}{T}\int_0^T \delta_{Y^\vartheta_{\underline{t}}}dt\right).$$

We will now handle those two terms separately:

Step 1: Bound on $D^{(32)}_{N,\gamma}$. As in Section 5.2, since $d\in\mathcal{D}_1$ whenever $d=d_s$ or $d=d_{\mathrm{CF},p}$, we have

$$\sup_{\vartheta\in\Theta} D^{(32)}_{N,\gamma}(\vartheta) \leq \frac{1}{T}\int_0^T \sup_{\vartheta\in\Theta} |Y^\vartheta_t - Y^\vartheta_{\underline{t}}|dt. \tag{5.19}$$

We now proceed as in Section 5.2 in order to get the equivalent of (5.16) thanks to Jensen's inequality. We get

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |D^{(32)}_{N,\gamma}(\vartheta)|^q\right] \leq \frac{1}{T}\int_0^T \mathbf{E}\left[\sup_{\vartheta\in\Theta} |Y^\vartheta_t - Y^\vartheta_{\underline{t}}|^q\right]dt. \tag{5.20}$$

In order to bound the right hand side of (5.20), we start by recalling the bound (4.10) for $Y^\vartheta_t - Y^\vartheta_{\underline{t}}$:

$$|Y^\vartheta_t - Y^\vartheta_{\underline{t}}| \leq \int_{\underline{t}}^t |b_\vartheta(Y^\vartheta_s)|ds + \|\sigma\||B_t - B_{\underline{t}}|.$$

The drift term above is now bounded thanks to the sublinear growth of b_ϑ given by (2.5) and the uniform bound on the L^q moments of Y^ϑ_t given by Proposition A.1. As far as the term $|B_t - B_{\underline{t}}|$ is concerned , we obviously have thanks to (1.1) and the fact that $|t-\underline{t}|\leq\gamma$:

$$\mathbf{E}\left[|B_t - B_{\underline{t}}|^q\right] \leq C_H\gamma^{qH}.$$

From here, it is readily checked that

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |Y^\vartheta_t - Y^\vartheta_{\underline{t}}|^q\right] \leq C_H\ \gamma^{qH}.$$

Plugging this information into (5.20) we end up with

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} |D^{(32)}_{N,\gamma}(\vartheta)|^q\right] \leq C_q\gamma^{qH}. \tag{5.21}$$

Step 2: Bound on $D^{(31)}_{N,\gamma}(\vartheta)$ for a fixed ϑ. For a fixed value of $\vartheta \in \Theta$, the term $D^{(31)}_{N,\gamma}(\vartheta)$ will be handled similarly to Section 5.1. Namely, along the same lines as for relation (5.9) we write

$$D^{(31)}_{N,\gamma}(\vartheta) \leq d\left(\nu_\vartheta, \frac{1}{T}\int_0^T \mathbf{E}[\delta_{Y_t^\vartheta}]dt\right) + d\left(\frac{1}{T}\int_0^T \mathbf{E}[\delta_{Y_t^\vartheta}]dt, \frac{1}{T}\int_0^T \delta_{Y_t^\vartheta}dt\right). \tag{5.22}$$

Then the first term in the right hand side of (5.22) is handled exactly as (5.10) in Section 5.1, which yields

$$\left|\frac{1}{T}\int_0^T \mathbf{E}[f(Y_t^\vartheta)]dt - \nu_\vartheta(f)\right| \leqslant \frac{C}{T}\|f\|_{\mathrm{Lip}}. \tag{5.23}$$

The second term in the right hand side of (5.22) can be upper bounded thanks to a continuous time version of Lemma 5.5 (also based on [48, Theorem 2.3] and left to the reader for the sake of conciseness). We get

$$\mathbf{E}\left[\left|\frac{1}{T}\int_0^T \left(f(Y_t^\vartheta) - \mathbf{E}[f(Y_t^\vartheta)]\right)dt\right|^q\right] \leqslant C_q\|f\|^q_{\mathrm{Lip}} T^{-\frac{q}{2}(2-(2H\vee 1))}. \tag{5.24}$$

Therefore putting together (5.23) and (5.24) and arguing as in Section 5.1, we get that for $d = d_{\mathrm{CF},p}$ with $p > (q+d)/2$ or $d = d_s$ and for any $\vartheta \in \Theta$,

$$\mathbf{E}\left[|D^{(31)}_{N,\gamma}(\vartheta)|^q\right] = \mathbf{E}\left[d\left(\nu_\vartheta, \frac{1}{T}\int_0^T \delta_{Y_t^\vartheta}dt\right)^q\right] \leqslant C_q\left(T^{-q} + T^{-\frac{q}{2}(2-(2H\vee 1))}\right) \tag{5.25}$$

where C_q is a positive constant which does not depend on ϑ.

Step 3: Bound on $\sup_{\vartheta\in\Theta} D^{(31)}_{N,\gamma}(\vartheta)$. In order to gor from (5.25) to a bound for the supremum over Θ, we proceed to a discretization of the parameter space Θ as in Section 2.4. Towards this aim, we will use the following notation: for any $\vartheta \in \Theta$, we set

$$\varphi(\vartheta) := d\left(\nu_\vartheta, \frac{1}{T}\int_0^T \delta_{Y_t^\vartheta}dt\right). \tag{5.26}$$

Let $\varepsilon > 0$ and recall that $\Theta^{(\varepsilon)} := \{\vartheta_i^{(\varepsilon)} \mid 1 \leqslant i \leqslant M_\varepsilon\}$ is defined at the beginning of Subsection 2.4 in such a way that $\Theta \subset \bigcup_{i=1}^{M_\varepsilon} B(\vartheta_i^{(\varepsilon)}, \varepsilon)$. Then, for any $\vartheta \in \Theta$,

$$\varphi(\vartheta) \leqslant |\varphi(\vartheta) - \varphi(\vartheta^{(\varepsilon)})| + |\varphi(\vartheta^{(\varepsilon)})|$$

where $\vartheta^{(\varepsilon)}$ is defined by (2.16). Therefore

$$\varphi(\vartheta) \leq |\varphi(\vartheta) - \varphi(\vartheta^{(\varepsilon)})| + \max_{1\leq i\leq M_\varepsilon} |\varphi(\vartheta_i^{(\varepsilon)})|$$

and finally

$$\begin{aligned}\mathbf{E}\left[\sup_{\vartheta\in\Theta}\varphi(\vartheta)^q\right] &\leq c_q\mathbf{E}\left[\sup_{\vartheta\in\Theta}|\varphi(\vartheta) - \varphi(\vartheta^{(\varepsilon)})|^q\right] + c_q\mathbf{E}\left[\max_{1\leq i\leq M_\varepsilon}|\varphi(\vartheta_i^{(\varepsilon)})|^q\right] \\ &\leq c_q\mathbf{E}\left[\sup_{\vartheta\in\Theta}|\varphi(\vartheta) - \varphi(\vartheta^{(\varepsilon)})|^q\right] + c_q\sum_{i=1}^{M_\varepsilon}\mathbf{E}\left[|\varphi(\vartheta_i^{(\varepsilon)})|^q\right].\end{aligned} \tag{5.27}$$

Owing to inequality (5.25) for a fixed $\vartheta \in \Theta$, we can deduce from (5.27) that

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta}\varphi(\vartheta)^q\right] \leq c_q\mathbf{E}\left[\sup_{\vartheta\in\Theta}|\varphi(\vartheta) - \varphi(\vartheta^{(\varepsilon)})|^q\right] + c'_q M_\varepsilon\left(T^{-q} + T^{-\frac{q}{2}(2-(2H\vee 1))}\right). \tag{5.28}$$

In the remainder of the step, we thus focus on the first right hand term in (5.28). Namely we will show the existence of an integrable random variable $\zeta_T > 0$ such that for all $\vartheta_1, \vartheta_2 \in \Theta$ we have

$$|\varphi(\vartheta_1) - \varphi(\vartheta_2)| \leq \zeta_T |\vartheta_1 - \vartheta_2| \quad a.s. \tag{5.29}$$

For this purpose, let us split the quantity $|\varphi(\vartheta_1) - \varphi(\vartheta_2)|$ in two terms

$$|\varphi(\vartheta_1) - \varphi(\vartheta_2)| \leq d\left(\nu_{\vartheta_1}, \nu_{\vartheta_2}\right) + d\left(\frac{1}{T}\int_0^T \delta_{Y_t^{\vartheta_1}} dt, \frac{1}{T}\int_0^T \delta_{Y_t^{\vartheta_2}} dt\right). \tag{5.30}$$

Then, one can show the following inequalities for any $d \in \mathcal{D}_1$

$$d\left(\nu_{\vartheta_1}, \nu_{\vartheta_2}\right) \leq C|\vartheta_1 - \vartheta_2| \sup_{\vartheta\in\Theta} \nu_\vartheta\left(|\cdot|^r\right) \tag{5.31}$$

$$d\left(\tfrac{1}{T}\int_0^T \delta_{Y_t^{\vartheta_1}} dt, \tfrac{1}{T}\int_0^T \delta_{Y_t^{\vartheta_2}} dt\right) \leq C|\vartheta_1 - \vartheta_2| \left(\tfrac{1}{T}\int_0^T \sup_{\vartheta\in\Theta} |Y_s^\vartheta|^r ds\right) \tag{5.32}$$

where r is given in assumption $(\mathbf{C_s})$ and where equation (5.31) is obtained by following the proof of Proposition 3.8 (see appendix B) for the stationary solutions $\bar{Y}^{\vartheta_1}$ and $\bar{Y}^{\vartheta_2}$. Plugging (5.31) and (5.32) into (5.30), we obtain that (5.29) holds true with

$$\zeta_T := C \max\left\{\sup_{\vartheta\in\Theta} \nu_\vartheta\left(|\cdot|^r\right), \ \frac{1}{T}\int_0^T \sup_{\vartheta\in\Theta} |Y_s^\vartheta|^r ds\right\}.$$

By Proposition A.1 (i), we get $\sup_{T>0} \mathbf{E}[\zeta_T^{\,q}] < +\infty$. Then plugging (5.29) into (5.28) we end up with

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} \varphi(\vartheta)^q\right] \leq c_q \sup_{T>0} \mathbf{E}[\zeta_T^{\,q}]\varepsilon^q + c_q' M_\varepsilon \left(T^{-q} + T^{-\frac{q}{2}(2-(2H\vee 1))}\right). \tag{5.33}$$

For all $\varepsilon > 0$, since Θ is a compact of $\mathbb{R}^d$, we can choose $M_\varepsilon \leq \frac{C_\Theta}{\varepsilon^d}$ and then if we choose $\varepsilon := T^{-\eta}$ for some $\eta > 0$, we finally get

$$\begin{aligned}\mathbf{E}\left[\sup_{\vartheta\in\Theta} \varphi(\vartheta)^q\right] &\leq c_q \sup_{T>0} \mathbf{E}[\zeta_T^{\,q}] T^{-q\eta} + c_q' C_\Theta \left(T^{-q+d\eta} + T^{-\frac{q}{2}(2-(2H\vee 1))+d\eta}\right) \\ &\leq c_q \sup_{T>0} \mathbf{E}[\zeta_T^{\,q}] T^{-q\eta} + c_q' C_\Theta T^{-\frac{q}{2}(2-(2H\vee 1))+d\eta}.\end{aligned}$$

It just remains to optimize in η to conclude that:

$$\mathbf{E}\left[\sup_{\vartheta\in\Theta} \varphi(\vartheta)^q\right] \leq C_q T^{-\tilde{\eta}} \tag{5.34}$$

with $\tilde{\eta} := \frac{q^2}{2(q+d)}(2 - (2H \vee 1))$. By putting together (5.18) with both (5.21) and (5.34), this concludes the proof of (5.17). □

Let us now conclude the section. Through inequality (5.2) and the control of the three right hand side terms, namely (5.8), (5.15) and (5.17), we are in position to conclude that Theorem 2.13 holds true.

6. Identifiability assumption

In this section we will provide some examples of equations of the form (1.2) for which the crucial assumptions (2.9) and (2.10) are satisfied. We first review briefly the diffusion case in Section 6.1, and then give a particular example in the fractional Brownian motion case in Section 6.2.

6.1. **Case of a diffusion process.** In this section we consider equation (1.2) in the case $H = \frac{1}{2}$, that is when the equation is driven by a d-dimensional Wiener process. Our considerations are summarized in the following proposition.

Proposition 6.1. *Consider equation (1.2) in the case $H = \frac{1}{2}$. We assume Hypothesis $(\mathbf{H_0})$ and $(\mathbf{C_w})$ to be met and call ν_ϑ the invariant measure corresponding to the coefficient b_ϑ. We pick $\vartheta_1, \vartheta_2 \in \Theta$ and set*

$$F = \mathrm{Span}\{\nabla f, \text{ for all } f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}) \text{ with compact support}\}$$

where F is considered as a subspace of $L^2(\nu_{\vartheta_1})$. We assume that $b_{\vartheta_1} - b_{\vartheta_2}$ is not an element of $F^\perp$ in $L^2(\nu_{\vartheta_1})$. Then

$$d(\nu_{\vartheta_1}, \nu_{\vartheta_2}) > 0.$$

If this condition is satisfied for all couples $(\vartheta_1, \vartheta_2) \in \Theta^2$, then $(\mathbf{I_w})$ holds true.

Proof. For $\vartheta \in \Theta$, let L_ϑ denote the linear operator defined on $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ by:

$$L_\vartheta f(x) = \langle \nabla f, b_\vartheta \rangle(x) + \frac{1}{2}(\sigma^* D^2 f(x) \sigma),$$

where $D^2 f$ denote the Hessian matrix of f. By a classical criterion, ν_ϑ is invariant for (1.2) when $H = \frac{1}{2}$ if and only if $\nu_\vartheta(L_\vartheta f) = 0$ for any $f \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ with compact support. As a consequence, $\nu_{\vartheta_1} = \nu_{\vartheta_2}$ if and only if $\nu_{\vartheta_1}((L_{\vartheta_1} - L_{\vartheta_2})f) = 0$ for any compactly supported $\mathcal{C}^2$-function $f : \mathbb{R}^d \to \mathbb{R}$. Now observe that

$$\nu_{\vartheta_1}((L_{\vartheta_1} - L_{\vartheta_2})f) = \int \langle \nabla f(x), b_{\vartheta_1}(x) - b_{\vartheta_2}(x) \rangle \nu_{\vartheta_1}(dx).$$

The result follows. □

In other words, this result says that in the diffusion setting, the identifiability assumption is true if $\mathrm{Proj}_F(b_\vartheta - b_{\vartheta_0})$ is not the null function for any $\vartheta \neq \vartheta_0$. Notice that this is always true in the one-dimensional case or if b_ϑ is a gradient. Unfortunately, the generalization of this simple characterization to SDEs driven by fBm is far from being straightforward.

6.2. **Fractional Brownian motion case.** In this section we wish to check $(\mathbf{I_s})$ for some specific examples of equation (2.7) and for the distance $d_{\mathrm{CF},p}$. Specifically, we shall consider a family $Y^{\vartheta,\lambda}$ of real valued processes defined by

$$dY_t^{\lambda,\vartheta} = \left[-\vartheta\, Y_t^{\lambda,\vartheta} + \lambda b_\vartheta(Y_t^{\lambda,\vartheta})\right] dt + \sigma\, dB_t \tag{6.1}$$

where B is a 1-dimensional fractional Brownian motion. In equation (6.1) the quantity λ is a small enough parameter, which is assumed to be known. The estimation procedure is still for ϑ only. The coefficient b_ϑ is bounded together with its derivatives with respect to y and ϑ. The process $Y^{\vartheta,\lambda}$ has to be seen as a small perturbation of a fractional Ornstein-Uhlenbeck process with parameter ϑ. We also assume that ϑ is a 1-dimensional parameter and:

$$\vartheta \in [m, M], \quad \text{with} \quad 0 < m < M < \infty. \tag{6.2}$$

Let us start our analysis by the case $X^\vartheta \equiv Y^{0,\vartheta}$, that is the fractional Ornstein-Uhlenbeck process itself, solution of the following equation:

$$dX_t^\vartheta = -\vartheta\, X_t^\vartheta\, dt + \sigma\, dB_t. \tag{6.3}$$

It is easily seen that X^ϑ is a centered Gaussian process whose variance is given (see e.g. [23, p.724]) by

$$\mathbf{E}[(X_t^\vartheta)^2] = 2\sigma^2 e^{-\vartheta t} \int_0^t s^{2H-1} \cosh(\vartheta(t-s)) ds. \tag{6.4}$$

In this case our assumption $(\mathbf{I_s})$ is easily satisfied, as shown in the following lemma.

Lemma 6.2. *Let $\vartheta \in [m, M]$ as in (6.2), and consider the fractional Ornstein-Uhlenbeck process X^ϑ defined by (6.3). We call μ_ϑ its invariant measure. Then for all $\vartheta_1, \vartheta_2 \in [m, M]$, we have*

$$d_{\mathrm{CF},p}(\mu_{\vartheta_1}, \mu_{\vartheta_2}) \geq c_{m,M,H}\, |\vartheta_1 - \vartheta_2|. \tag{6.5}$$

Proof. It is well-known (see e.g [6]) that for the fractional Ornstein-Uhlenbeck process we have $\mu_\vartheta = \mathcal{N}(0, \sigma_\vartheta^2)$, with

$$\sigma_\vartheta^2 = \frac{c_H}{\vartheta^{2H}}. \tag{6.6}$$

Taking expression (2.21) into account, this yields

$$d^2_{\mathrm{CF},p}(\mu_{\vartheta_1}, \mu_{\vartheta_2}) = \int_{\mathbb{R}} \left[\exp\left(-\frac{c_H}{2\vartheta_1^{2H}} \xi^2 \right) - \exp\left(-\frac{c_H}{2\vartheta_2^{2H}} \xi^2 \right) \right]^2 g_p(\xi) d\xi,$$

from which our claim (6.5) is easily proved. Notice that the fact that ϑ is bounded away from 0 is crucial here in order to ensure the continuity of $\vartheta \mapsto \sigma_\vartheta$ in (6.6) on the interval $[m, M]$. □

Let us also state an elementary bound on ordinary differential equations for further use.

Lemma 6.3. *Let $f, g : \mathbb{R}_+ \to \mathbb{R}$ be two functions such that there exist some constants $\kappa, M > 0$ satisfying*

$$f_r \geq \kappa, \quad \text{and} \quad |g_r| \leq M, \quad \text{for all } r \in \mathbb{R}_+. \tag{6.7}$$

Let y be the solution of the following differential equation:

$$\dot{y}_t + f_t\, y_t = g_t. \tag{6.8}$$

Then y is uniformly bounded in t and verifies

$$|y_t| \leq \frac{M}{\kappa}.$$

Proof. Equation (6.8) admits an explicit solution under the form

$$y_t = \int_0^t \exp\left(-\int_s^t f_r\, dr \right) g_s\, ds.$$

Plugging the bounds (6.7) into the above expression, we easily get

$$|y_t| \leq M \int_0^t e^{-\kappa(t-s)}\, ds \leq \frac{M}{\kappa},$$

which is our claim. □

We now wish to extend Lemma 6.2 to the model given by equation (6.1). Namely we wish to prove the following proposition.

Proposition 6.4. *Let $Y^{\lambda,\vartheta}$ be the process defined by (6.1) and consider $p > 3/2$. We assume $\vartheta \in [m, M]$ and $\lambda \in (0, \lambda_0)$ with a small enough $\lambda_0 = \lambda_0(m, M, p)$. Also assume (without loss of generality) that b_ϑ, $\partial_y b_\vartheta$, $\partial_\vartheta b_\vartheta$, $\partial^2_{\vartheta y} b_\vartheta$ are all bounded by 1. Then the following lower bound holds true for any $\vartheta_1, \vartheta_2 \in [m, M]$:*

$$d_{\mathrm{CF},p}(\nu_{\vartheta_1}, \nu_{\vartheta_2}) \geq c_{m,M,H}\, |\vartheta_1 - \vartheta_2|. \tag{6.9}$$

Proof. Owing to the definition (2.21) of the distance $d_{\mathrm{CF},p}$, we have

$$d_{\mathrm{CF},p}(\nu_{\vartheta_1},\nu_{\vartheta_2}) = \left\{ \int_{\mathbb{R}} \left(\mathbf{E}[e^{i\xi \bar{Y}^{\lambda,\vartheta_1}}] - \mathbf{E}[e^{i\xi \bar{Y}^{\lambda,\vartheta_2}}] \right)^2 g_p(\xi)\, d\xi \right\}^{1/2}$$

We will decompose this quantity as follows:

$$d_{\mathrm{CF},p}(\nu_{\vartheta_1},\nu_{\vartheta_2}) \geq I_3^{1/2} - \left(I_2^{1/2} + I_{11}^{1/2} + I_{12}^{1/2} \right), \tag{6.10}$$

where the quantities I_{1j} and I_3 are defined by

$$\begin{aligned} I_{1j} &= \int_{\mathbb{R}} \left(\mathbf{E}\left[\exp\left(i\xi \bar{Y}^{\lambda,\vartheta_j}\right)\right] - \mathbf{E}\left[\exp\left(i\xi Y_t^{\lambda,\vartheta_j}\right)\right] \right)^2 g_p(\xi)\, d\xi \\ I_3 &= \int_{\mathbb{R}} \left(\mathbf{E}\left[\exp\left(i\xi X_t^{\vartheta_1}\right)\right] - \mathbf{E}\left[\exp\left(i\xi X_t^{\vartheta_2}\right)\right] \right)^2 g_p(\xi)\, d\xi, \end{aligned}$$

and where we recall that the Ornstein-Uhlenbeck process X^{ϑ} is given by (6.3). In equation (6.10), we also have

$$\begin{aligned} I_2 = \int_{\mathbb{R}} \Big(&\mathbf{E}\left[\exp\left(i\xi Y_t^{\lambda,\vartheta_1}\right)\right] - \mathbf{E}\left[\exp\left(i\xi X_t^{\vartheta_1}\right)\right] \\ &- \mathbf{E}\left[\exp\left(i\xi Y_t^{\lambda,\vartheta_2}\right)\right] + \mathbf{E}\left[\exp\left(i\xi X_t^{\vartheta_2}\right)\right] \Big)^2 g_p(\xi)\, d\xi. \end{aligned}$$

In the definitions above, t is an arbitrarily large time, to be determined later on. Our goal is now to lower bound I_3 and upper bound $I_{1,j}$ and I_2.

Lower bound for I_3. In order to lower bound I_3, we proceed as in Lemma 6.2. Indeed, Lemma 6.2 stems from a lower bound on

$$|\mathbf{E}[e^{i\xi X_\infty^{\vartheta_1}}] - \mathbf{E}[e^{i\xi X_\infty^{\vartheta_2}}]|,$$

while we are interested here in a lower bound on

$$|\mathbf{E}[e^{i\xi X_t^{\vartheta_1}}] - \mathbf{E}[e^{i\xi X_t^{\vartheta_2}}]|,$$

for a fixed t. However it is readily checked from (6.4) that there exists $t_0 > 0$ such that for all $t \geq t_0$ we have

$$I_3 \geq c_1 \, |\vartheta_1 - \vartheta_2|^2 \tag{6.11}$$

with a given constant $c_1 > 0$ depending on m, M.

Upper bound for I_{1j}. Recall that both b_ϑ and $\partial_y b_\vartheta$ are bounded by 1. We also assume that λ is small enough so that $\lambda \leq m(1-\varepsilon)$ with $\varepsilon > 0$. Then it is readily checked that $x \mapsto -\vartheta x + \lambda b_\vartheta(x)$ satisfies the condition $(\mathbf{C_s})$. Hence one can see as in (3.7) that

$$I_{1j} \leq C e^{-\frac{m\varepsilon}{2} t}.$$

If we wish to have $I_{1j} \leq c_3 \, |\vartheta_1 - \vartheta_2|^2$, with c_3 arbitrarily small, it is thus sufficient to pick $t \geq t_1$ with $t_1 = C \log\left(\frac{1}{|\vartheta_1 - \vartheta_2|^2}\right)$. In the sequel we choose this time t_1 such that

$$I_{11}^{1/2} + I_{12}^{1/2} \leqslant \frac{\sqrt{c_1}}{4} |\vartheta_1 - \vartheta_2|. \tag{6.12}$$

Upper bound for I_2. We start by recalling that $X^\vartheta = Y^{0,\vartheta}$. Next we set $R = [\vartheta_1, \vartheta_2] \times [0, \lambda]$ and for $r, \tau \in [0, 1]$ we define

$$a(r,\tau) = Y_t^{0,\vartheta_2} + r\left(Y_t^{0,\vartheta_1} - Y_t^{0,\vartheta_2}\right) + \tau\left(Y_t^{\lambda,\vartheta_2} - Y_t^{0,\vartheta_2}\right) + r\tau\Delta_R Y_t, \tag{6.13}$$

where the rectangular increment $\Delta_R Y_t$ is given by

$$\Delta_R Y_t = Y_t^{\lambda,\vartheta_1} - Y_t^{0,\vartheta_1} - Y_t^{\lambda,\vartheta_2} + Y_t^{0,\vartheta_2}. \tag{6.14}$$

Notice that I_2 can be expressed as

$$I_2 = \int_{\mathbb{R}} \left(\mathbf{E}\left[\Delta_R \psi_\xi(Y_t)\right]\right)^2 g_p(\xi)\, d\xi, \tag{6.15}$$

where ψ_ξ is the oscillating function $e^{i\xi x}$ and $\Delta_R \psi_\xi(Y_t)$ still denotes a rectangular increment as in (6.14). Moreover, resorting to the path a introduced in (6.13) we get

$$\Delta_R \psi_\xi(Y_t) = \int_{[0,1]^2} \partial^2_{r\tau}\left[\psi_\xi(a(r,\tau))\right] dr d\tau,$$

and computing the differential $\partial^2_{r\tau}\left[\psi_\xi(a(r,\tau))\right]$ explicitly we get

$$\Delta_R \psi_\xi(Y_t) = \int_{[0,1]^2} \left(\psi'_\xi(a(r,\tau))\partial^2_{r\tau} a(r,\tau) + \psi''_\xi(a(r,\tau))\partial_r a(r,\tau)\partial_\tau a(r,\tau)\right) dr d\tau. \tag{6.16}$$

Taking into account (6.16) and (6.15), plus the fact that ψ and its derivatives ψ', ψ'' are bounded by $|\xi|^2$ and $|\xi|^2 g_p$ is integrable if $p > 3/2$, we get that

$$\begin{aligned} I_2 &\le C_p \left\{\mathbf{E}\left[\left(|Y_t^{0,\vartheta_2} - Y_t^{0,\vartheta_1}| + |\Delta_R(Y_t)|\right)\left(|Y_t^{\lambda,\vartheta_2} - Y_t^{0,\vartheta_2}| + |\Delta_R(Y_t)|\right)\right] + \mathbf{E}\left[|\Delta_R(Y_t)|\right]\right\}^2 \\ &\le C_{p,m,M,\varepsilon}\mathbf{E}\left[\|\partial_\lambda Y^\vartheta\|^2_\infty + \|\partial_\vartheta Y^\vartheta\|^2_\infty + \|\partial^2_{\lambda\vartheta} Y^\vartheta\|^2_\infty\right]^2 \lambda^2 |\vartheta_1 - \vartheta_2|^2. \end{aligned} \tag{6.17}$$

In order to bound the right hand side of (6.17), we are now reduced to the estimation of $\partial_\lambda Y^{\lambda,\vartheta}$, $\partial_\vartheta Y^{\lambda,\vartheta}$ and $\partial^2_{\lambda\vartheta} Y^{\lambda,\vartheta}$. Let us thus show how to establish a bound for $\partial_\lambda Y^{\lambda,\vartheta}$. To this aim, differentiating formally equation (6.1), the process $\partial_\lambda Y^{\lambda,\vartheta}$ solves a system of the form

$$\frac{d[\partial_\lambda Y^{\lambda,\vartheta}]_t}{dt} = \left[-\vartheta + \lambda\partial_y b_\vartheta(Y_t^{\lambda,\vartheta})\right]\partial_\lambda Y_t^{\lambda,\vartheta} + b_\vartheta(Y_t^{\lambda,\vartheta}). \tag{6.18}$$

Whenever ϑ satisfies (6.2) and $\lambda \le m(1-\varepsilon)$ with $\varepsilon > 0$, equation (6.18) above fulfills the hypothesis of Lemma 6.3. Hence we get

$$|\partial_\lambda Y_t^{\vartheta,\lambda}| \le c_{m,M,\varepsilon},$$

uniformly in $t \ge 0$. We let the reader check that the same kind of inequality holds true for $\partial_\vartheta Y^{\lambda,\vartheta}$ and $\partial^2_{\lambda\vartheta} Y^{\lambda,\vartheta}$ as well, and thus we get

$$\|\partial_\lambda Y^\vartheta\|_\infty + \|\partial_\vartheta Y^\vartheta\|_\infty + \|\partial^2_{\lambda\vartheta} Y^\vartheta\|_\infty \le c_{m,M,\varepsilon}.$$

Plugging this inequality into (6.17), we end up with

$$|I_2| \le C_{p,m,M,\varepsilon}\lambda^2\, |\vartheta_1 - \vartheta_2|^2.$$

We now choose λ such that $\lambda \le \frac{1}{4}\left(\frac{c_1}{C_{p,m,M,\varepsilon}}\right)^{1/2}$, where c_1 is defined by (6.11) and $\varepsilon = 1/2$. This yields

$$I_2^{1/2} \le \frac{\sqrt{c_1}}{4}|\vartheta_1 - \vartheta_2|. \tag{6.19}$$

We now gather (6.19) and (6.12) into (6.10), which proves our claim with $\lambda_0 = \frac{1}{4}\left(\frac{c_1}{C_{p,m,M,\varepsilon}}\right)^{1/2}$. □

7. Numerical Discussions and Illustrations

In this section, we provide several numerical examples in order to illustrate our main results. To this end, we first investigate several numerical questions which are related to our theoretical results.

Simulated data and Euler scheme: In order to test our results, we have chosen to simulate our observations. Nevertheless, the fractional SDE (1.2) cannot be simulated exactly, except in some particular cases. Therefore we have opted for a discretization procedure thanks to a simple first order Euler scheme with very small step $\underline{\gamma}$ (namely $\underline{\gamma} = 10^{-3}$) in order to get a sharp approximation of the true process.

Let us recall that in the additive setting of equation (1.2) the simple Euler scheme converges strongly to the true SDE, while this is not true in general in the multiplicative case (see *e.g.* [37]). The convergence of the scheme can be checked for instance through Proposition 3.7(i), applied with constant step $\underline{\gamma}$. Furthermore, taking the expectation in Proposition 3.7(i) leads to a marginal control of the L^2-distance between the Euler scheme and the true SDE (with same fBm) of order $\underline{\gamma}^H$ (independently of the horizon). This confirms that our approximation of the observations is reasonable when H is not too small (getting a control for the uniform distance is more involved).

Let us also recall that the increments of the fBm can be simulated through the Wood-Chan method (see [53]), which is based on the embedding of the covariance matrix of the fractional increments in a symmetric circulant matrix (whose eigenvalues can be computed using the Fast Fourier Transform). Therefore up to the approximation of the true SDE detailed above, we now assume that we are given a sequence $(Y_{k\underline{\gamma}})_{k\geq 0}$, where $(Y_t)_{t\geq 0}$ is a solution to (2.7) with a given θ_0. Then we select from this path a subsequence of observations $(Y_{t_k})_{k=1}^n$ where $t_k = k\gamma$, which means in particular that we assume γ to be of the form $k_0\underline{\gamma}$ with $k_0 \in \mathbb{N}^*$.

Computation of the distance between empirical measures: The theoretical construction of an estimator like (2.15) involves in practice the computation of the distance d between the empirical measures of the observed process and of the Euler scheme, for a distance $d \in \mathcal{D}_p$ as defined in (2.4). We briefly describe how to compute this kind of distance.

Whenever d is the p-Wasserstein distance, an explicit computation of the distance d in (2.15) is possible if the observation Y is 1-dimensional. To this aim, one can use the following representation (see [47]): if μ and ν are two one-dimensional probabilities with *c.d.f* F and G respectively, then for all $p > 0$ we have

$$\mathcal{W}_p^p(\mu,\nu) = \int_0^1 |F^-(t) - G^-(t)|^p dt,$$

where F^- and G^- denote the (left or right) pseudo-inverse of F and G. Moreover, when $\mu = \sum_{i=1}^{n_1} p_i \delta_{x_i}$ and $\nu = \sum_{j=1}^{n_2} q_j \delta_{y_j}$, the computation of the right hand side above can be made explicit through a reordering (and using the fact that F^{-1} and G^{-1} are stepwise constant). In particular, when $n_1 = n_2 = n$ and $p_i = q_j = 1/n$, the Wasserstein distance between μ and ν simply reads

$$\mathcal{W}_p^p(\mu,\nu) = \frac{1}{n}\sum_{i=1}^n |x_{(i)} - y_{(i)}|^p. \tag{7.1}$$

We will use this representation in our simulations.

In higher dimension, the computation of the Wasserstein distance generally requires approximation/optimization methods which are out of the scope of this paper. In this context it seems to be numerically simpler to work with an approximation of the distance $d_{\mathrm{CF},p}$ (defined in (2.21)),

which is also used for our analysis of the rate of convergence. Such an approximation can be obtained by a standard discretization of the integral which appears in the definition (2.21).

Minimization of the distance with respect to ϑ: Eventually the implementation of our estimation procedure relies on an optimization problem in order to compute the argmin in (2.15). More specifically, in case the estimator is built with a constant step Euler scheme, this consists in minimizing the function

$$\mathcal{F}_d : \vartheta \mapsto d\left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{Y_{t_k}}, \frac{1}{N}\sum_{k=0}^{N-1}\delta_{Z^{\vartheta,\gamma}_{k\gamma}}\right). \tag{7.2}$$

In this paper, we first use the naive approach which consists in evaluating the function on a (finite) grid and then computing the minimum on this finite set. This minimization algorithm is clearly restricted to a low dimensional setting. Secondly, we use a stochastic optimization algorithm to explore an example in dimension two (in space) for the distance $d_{CF,p}$ which is easier to manage. More precisions are given after the one-dimensional case.

Numerical illustrations: Let us now turn to some numerical tests, for which we consider two one-dimensional examples and one two-dimensional example. We begin with the classical

▷ Ornstein-Uhlenbeck (OU) process: We consider the process X^ϑ defined by (6.3), where ϑ is assumed to sit in a compact interval of $(0, +\infty)$ like in (6.2). Let us recall that this case is a toy example since the Gaussian linear structure of the OU-process allows to develop specific estimation methods (on this topic, see [32] or more recently [51] and [5]). The assumptions $(\mathbf{H_0})$ and $(\mathbf{C_s})$ are clearly satisfied, whereas $(\mathbf{I_s})$ follows from Lemma 6.2. Using the strategy described in the first part of this section, we get a discretely observed path of Y with the following parameters:

$$\vartheta_0 = 2,\ \underline{\gamma} = 10^{-3},\ \gamma = 10^{-2},\ n = 3.10^4,$$

and different values of H. In Figure 1, we depict the function $\mathcal{F}_d$ defined in (7.2) with $d = d_{CF,2}$ for $H = 0.3$ and $H = 0.7$ respectively. As in the next examples, the Euler scheme is computed with $N = n$ and $\gamma = 10^{-2}$ as specified above. We remark that the function attains its minimum very close to the true value of ϑ. We also observe that the function $\mathcal{F}_d$ is more flat when H is

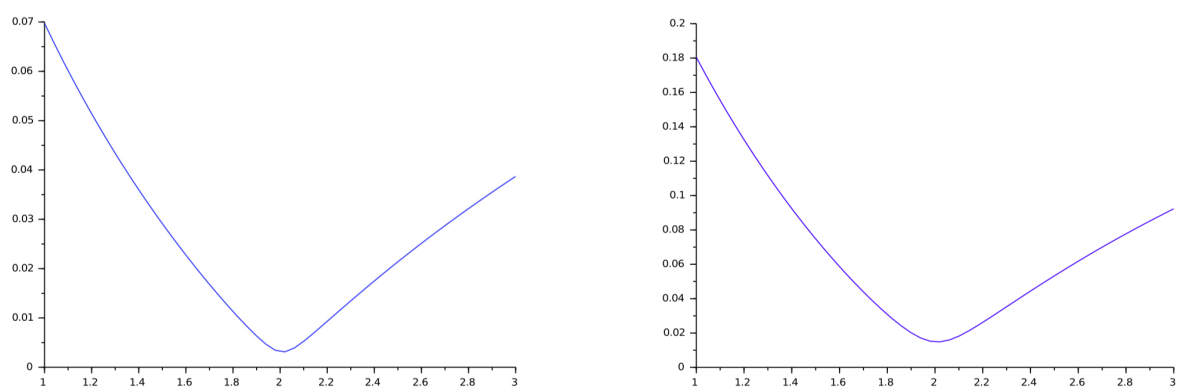

FIGURE 1. $\vartheta \mapsto \mathcal{F}_{d_{CF,2}}(\vartheta)$ for $H = 0.3$ (left) and $H = 0.7$ (right).

small, which is consistent with the fact that $H \mapsto \sigma^2_\vartheta$ in (6.6) is an increasing function.

Figure 2 below is devoted to a comparison between the different p-Wasserstein distances as p varies. Namely we fix $H = 0.3$ and we compute the function $\mathcal{F}_d$ defined by (7.2) with $d = \mathcal{W}_p$ for different values of p. Notice that in the 1-dimensional case we are considering we can resort

to formula (7.1), since we have chosen $N = n$. The true parameter is still $\vartheta = 2$. Our distances all perform correctly, although $p = 4$ seems to yields a slightly sharper contrast.

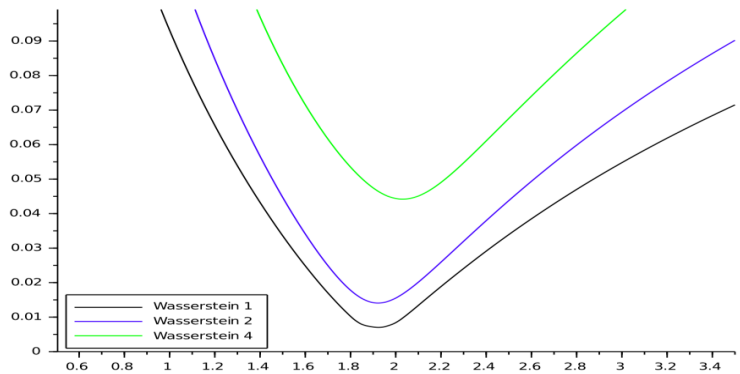

FIGURE 2. $\vartheta \mapsto \mathcal{F}_{\mathcal{W}_p}(\vartheta)$, $H = 0.3$.

▷ We now consider a second example with a non linear dependence in ϑ, namely an equation of the form:

$$dY_t^\vartheta = -Y_t^\vartheta(1 + \cos(\vartheta Y^\vartheta))dt + dB_t.$$

In this case, we only compute the Wasserstein distance for different values of p with the same choices of parameters. Once again, the minimum of the function $\mathcal{F}_{\mathcal{W}_p}$ is attained close to $\vartheta_0 = 2$. One also observes that the local behavior in the neighborhood of ϑ_0 is similar to the linear case.

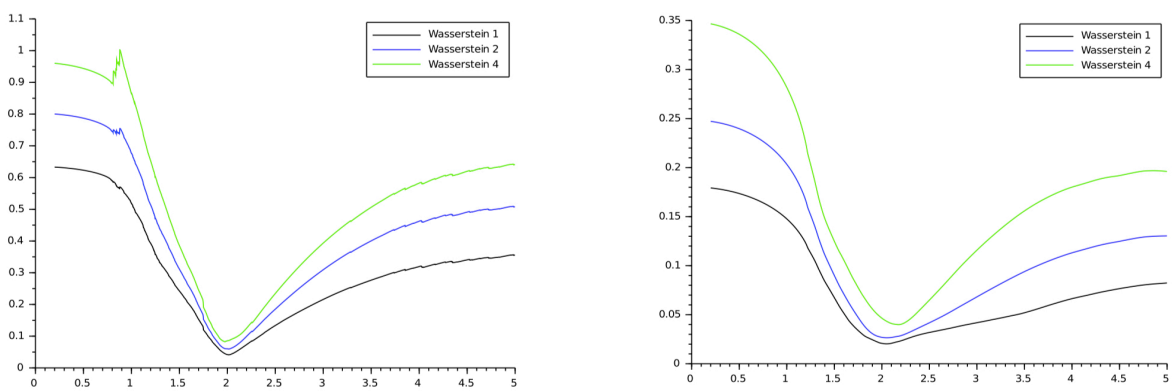

FIGURE 3. $\vartheta \mapsto \mathcal{F}_{\mathcal{W}_p}(\vartheta)$ for $H = 0.3$ (left) and $H = 0.7$ (right).

▷ A two-dimensional example: we finally propose to focus on a family of two-dimensional fractional processes $(Z_t^\vartheta)_{t\geq 0}$ indexed by a real parameter $\vartheta \in \Theta = [m, M] \subset (0, +\infty)$, where m, M are two finite constants. Specifically, Z^ϑ is the solution to the stochastic system

$$dZ_t^\vartheta = b_\vartheta(Z_t^\vartheta)dt + dB_t^H, \tag{7.3}$$

where $(B_t^H)_{t\geq 0}$ is a two-dimensional standard fBm with Hurst parameter H. In equation (7.3), we also have that for all $\vartheta \in \Theta$ the function $b_\vartheta : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by:

$$\text{for all } z = (z_1, z_2) \in \mathbb{R}^2, \quad b_\vartheta(z) = -\vartheta z \psi\left((1 + |z|^2)^{1/2}\right) + \varepsilon z^\perp, \tag{7.4}$$

where $z^\perp = (-z_2, z_1)$ and $\psi(x) = \frac{e^x}{1+e^x}$ (namely ψ is the *sigmoid* function). Let us first remark that the coefficient b_ϑ above satisfies the main standing assumptions of the current paper and generalizes the framework of [36]. Indeed, the following holds true: (a) Due to the orthogonal component $z^\perp$ in equation (7.4), it is readily checked that $\partial_{z_2} b^1_\vartheta \neq \partial_{z_1} b^2_\vartheta$ (where we recall the notation of Section 2.1 for the partial derivatives of b). Therefore our family of drift functions b_ϑ cannot be of gradient type. (b) The functions b_ϑ are obviously smooth. In addition, we have $\langle x_1^\perp - x_2^\perp, x_1 - x_2\rangle = 0$. Thus in order to see that b_ϑ satisfies our Hypothesis $(\mathbf{C_s})$, it is enough to check that the eigenvalues of the Jacobian matrix of $z \mapsto z\varphi(|z|^2)$ are uniformly lower bounded by a positive constant, where we have set

$$\varphi(x) = \psi\left((1+x)^{1/2}\right), \quad \text{for} \quad x \geq 0,$$

and where ψ is the sigmoid function featuring in our definition (7.4). We leave to the patient reader the elementary task of computing those eigenvalues. Let us just mention that a lower bound is given by $\varphi(|z|^2)$, which is uniformly lower bounded by $\frac{1}{2}$ for all $z \in \mathbb{R}^2$. This implies that Assumption $(\mathbf{C_s})$ is fulfilled for the coefficient b_ϑ.

The multi-dimensional setting leads to new numerical difficulties. The first of those problems is due to the fact that the exact computation of the Wasserstein distance in $\mathbb{R}^2$ is still possible but costly (about n^3 computations for an empirical measure based on n points by the *Hungarian algorithm*), while approximations lead to numerical issues which are out of the scope of this paper (see *e.g.* [42]). We have thus decided to work with the distance $d_{\mathrm{CF},p}$ introduced in Section 2.5, which is simpler to compute. The second numerical problem we are facing comes from the underlying spatial dimension. Indeed, when $d = 2$ one can still optimize over the parameter ϑ by computing the contrast given by (2.15) over a discrete grid of parameters ϑ. However, it is clear that in larger dimensions such contrast estimators require to implement associated convex optimization algorithms.

With the above preliminary considerations in mind, we now briefly introduce a stochastic optimization algorithm for the approximation of $\hat{\vartheta}_{N,n,\gamma}$ related to $d_{\mathrm{CF},p}$. We will then implement this method on the coefficient b_ϑ given by (7.4). Here, the parameter ϑ is still one-dimensional but the method which is proposed can be adapted to the higher dimension. In order to describe our algorithm, let us first recast our expression (2.21) for the $d_{\mathrm{CF},p}$ distance. That is, observe that if $f_\xi(x) := e^{i\langle \xi, x\rangle}$ and if Ξ is a random variable with density g_p then for two probability measures μ_1 and μ_2 we have

$$d_{\mathrm{CF},p}(\mu_1, \mu_2) = \mathbf{E}\left[|\mu_1(f_\Xi) - \mu_2(f_\Xi)|^2\right].$$

Now recall from relation (2.15) that we are dealing with the following empirical measures:

$$\mu = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{Y_{t_k}} \quad \text{and} \quad \mu_\vartheta = \frac{1}{N}\sum_{k=0}^{N-1} \delta_{Z^{\vartheta,\gamma}_{k\gamma}}.$$

Then our aim in (2.15) is to minimize a functional F defined on $\Theta = [m, M]$ by

$$F(\vartheta) = d_{\mathrm{CF},p}(\mu, \mu_\vartheta) = \mathbf{E}\left[|\mu(f_\Xi) - \mu_\vartheta(f_\Xi)|^2\right].$$

Our gradient descent algorithm will thus be applied to the functional F, and we now provide some details about its computation. Indeed, the gradient ∇F is formally obtained, similarly to what we did in the proof of Proposition 6.4, as

$$\nabla F(\vartheta) = \mathbf{E}[\Lambda(\vartheta, \Xi)], \quad \text{where} \quad \Lambda(\vartheta, \xi) = \partial_\vartheta\left(|\mu(f_\xi) - \mu_\vartheta(f_\xi)|^2\right). \tag{7.5}$$

Furthermore, it is easily seen that $\Lambda(\vartheta,\xi)$ can be expressed as

$$\Lambda(\vartheta,\xi) = 2(\mu_\vartheta-\mu)(\cos(\langle\xi,\cdot\rangle))\rho_\vartheta(-\sin(\langle\xi,\cdot\rangle)) + 2(\mu_\vartheta-\mu)(\sin(\langle\xi,\cdot\rangle))\rho_\vartheta(\cos(\langle\xi,\cdot\rangle)), \tag{7.6}$$

where we resort to the following notation for a given function $g:\mathbb{R}\to\mathbb{R}$,

$$\rho_\vartheta(g(\langle\xi,.\rangle)) = \frac{1}{N}\sum_{k=0}^{N-1} g(\langle\xi, Z^{\vartheta,\gamma}_{k\gamma}\rangle)\,\langle\xi,\partial_\vartheta Z^{\vartheta,\gamma}_{k\gamma}\rangle. \tag{7.7}$$

Plugging (7.7) into (7.6) and then (7.5), one observes that the computation of $\nabla F(\vartheta)$ mostly relies on our ability to evaluate $(\partial_\vartheta Z^{\vartheta,\gamma}_{k\gamma})_{k\geq 0}$. Now it should be highlighted that this quantity can be computed *on the fly*, since $\partial_\vartheta Z^{\vartheta,\gamma}_0 = 0$ and for all $k\geq 0$ one can evaluate $\partial_\vartheta Z^{\vartheta,\gamma}_{(k+1)\gamma}$ recursively as

$$\partial_\vartheta Z^{\vartheta,\gamma}_{(k+1)\gamma} = \partial_\vartheta Z^{\vartheta,\gamma}_{k\gamma} + \gamma\left(\partial_\vartheta b_\vartheta(Z^{\vartheta,\gamma}_{k\gamma}) + \nabla_z b_\vartheta(Z^{\vartheta,\gamma}_{k\gamma})\partial_\vartheta Z^{\vartheta,\gamma}_{k\gamma}\right).$$

The convergence of the gradient descent algorithm to a minimum of F lies outside the scope of the current paper (see *e.g.* [16] for background on this topic). Let us just mention that the algorithm is recursively defined by

$$\theta_{\ell+1} = \theta_\ell - \eta_\ell\,\Lambda(\theta_\ell,\Xi_{\ell+1}), \tag{7.8}$$

where $(\eta_\ell)_\ell$ is a sequence of (positive) steps and $(\Xi_\ell)_{\ell\geq 1}$ is a sequence *i.i.d* random variables with common distribution g_p. Notice that in dimension 2 the random variables Ξ_ℓ can be simulated through a change of variable. More precisely, if R and Θ are two independent variables such that R has density $q_p(r) = \frac{2(p-1)r}{(1+r^2)^p}$ and Θ has uniform distribution on $[0,2\pi]$, then $(R\cos(\Theta), R\sin(\Theta))$ has density g_p. Furthermore, thanks to the inversion method we get that R can be simulated as $R = (U^{1/(1-p)}-1)^{1/2}$ where U has uniform distribution on $[0,1]$.

The illustration represented in Figure 4 is obtained with $\varepsilon = 1/10$ in equation (7.4) and $\eta_\ell = \gamma_0(1+\ell)^{-1/2}$ in relation (7.8). We have also considered two values of γ_0 and two different starting points y_0 for the process Z^ϑ in (7.3). Furthermore, in these simulations we used a *mini-batch* approach. This means that in equation (7.8) we replace the random simulated term $\Lambda(\theta_\ell,\Xi_{\ell+1})$ by an average over $m=30$ simulations. The mini-batch procedure is known to reduce the randomness of the algorithm. The results displayed in Figure 4 show that our approach yields a good convergence of the estimate to the true parameter ϑ. However, our simulations also reveal that the gradient gets quite flat near the true parameter $\vartheta_0 = 1$. Therefore the algorithm moves very slowly after a large number of iterations. This is the reason why in the illustration, the algorithm is stopped after $n = 100$ simulations. Also notice that we have provided several examples of initial values y_0 and step sizes γ_0, which exemplifies the sensibility of the current version of the algorithm with respect to the parameters. Those first attempts to combine our statistical procedure with gradient descent algorithms would certainly need to be enriched with deeper considerations, although this task is deferred to a subsequent publication for sake of conciseness.

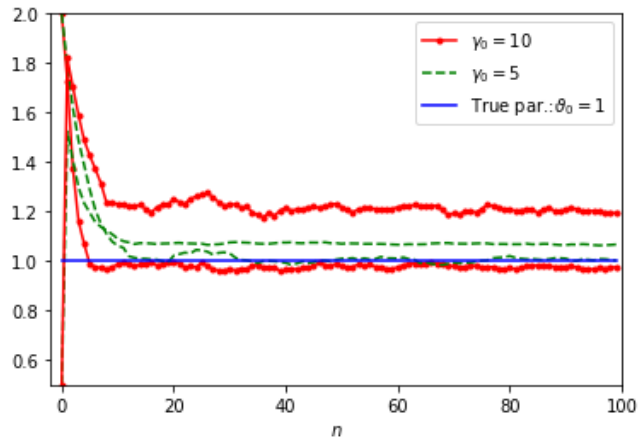

FIGURE 4. Example in dimension 2.

APPENDIX A. TIGHTNESS AND CONVERGENCE OF THE OCCUPATION MEASURES

In this Appendix, we first show some uniform estimates for Y^ϑ and $Z^{\vartheta,\gamma}$ and their respective occupation measures, as well as some convergence result for the occupation measure of Y^ϑ. Those results are all used in the proof of Proposition 3.3 and 3.4.

A.1. Moment estimates.

We start by bounding the moments of Y^ϑ and its occupation measure.

Proposition A.1. *Let Y^ϑ be the unique solution of (2.7). Assume $(\mathbf{H_0})$ and $(\mathbf{C_w})$. Then the following inequalities hold true for $p \geq 1$.*

(i) $\sup_{t\geqslant 0} \mathbf{E}[\sup_{\vartheta\in\Theta} |Y_t^\vartheta|^p] < +\infty$.

(ii) $\sup_{\vartheta\in\Theta} \sup_{n\geqslant 1} \frac{1}{n}\sum_{k=0}^{n-1} |Y_{k\kappa}^\vartheta|^p < +\infty$ *a.s.*

(iii) $\sup_{\vartheta\in\Theta} \sup_{t>0} \frac{1}{t}\int_0^t |Y_t^\vartheta|^p dt < +\infty$ *a.s.*

Proof. We treat our three items separately.

(i) The proof is done in [23]. It is based on a comparison with the moments of the fractional Ornstein-Uhlenbeck process, similarly to what is done in step (ii) below. Since the constants in $(\mathbf{C_w})$ are independent from ϑ, we get the uniformity with respect to ϑ.

(ii) Let $p > 0$. By Cauchy-Schwarz inequality it is enough to prove the result with $2p$ instead of p. Let us denote by $(X_t)_{t\geqslant 0}$ the fractional Ornstein Uhlenbeck process defined by (6.3) with $\vartheta = 1$ and starting from y_0, i.e.

$$\begin{cases} dX_t = -X_t\, dt + \sigma\, dB_t \\ X_0 = y_0. \end{cases} \tag{A.1}$$

In this proof, we set $\underline{t} := \inf\{k\kappa \mid k\kappa \leqslant t < (k+1)\kappa\}$ and $T := n\kappa$. With this notations, similarly to what is done in the proof of Proposition 4.2, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} |Y_{k\kappa}^\vartheta|^{2p} = \frac{1}{T}\int_0^T |Y_{\underline{t}}^\vartheta|^{2p} dt.$$

Let us also denote by $\tilde{X}$ the stationary Ornstein-Uhlenbeck process related to (A.1). Then the proof of item (ii) is based on the following inequality:

$$\frac{1}{T}\int_0^T |Y_{\underline{t}}^{\vartheta}|^{2p}dt \leqslant 3^{2p-1}\left(\frac{1}{T}\int_0^T |Y_{\underline{t}}^{\vartheta}-X_{\underline{t}}|^{2p}dt+\frac{1}{T}\int_0^T |X_{\underline{t}}-\tilde{X}_{\underline{t}}|^{2p}dt+\frac{1}{T}\int_0^T |\tilde{X}_{\underline{t}}|^{2p}dt\right) \quad (A.2)$$

Let us start by bounding the two last term in the right hand side of (A.2). First, since $(\tilde{X}_{k\kappa})_{k\geqslant 0}$ is a stationary Gaussian sequence and since $\mathbf{E}[\tilde{X}_\kappa \tilde{X}_{(n+1)\kappa}] \underset{n\to+\infty}{\longrightarrow} 0$ (according to [7, Theorem 2.3]), a criterion for Gaussian processes (see [35]) gives that $(\tilde{X}_{k\kappa})_{k\geqslant 0}$ is ergodic. Therefore, we get

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T |\tilde{X}_{\underline{t}}|^{2p}dt = \lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}|\tilde{X}_{k\kappa}|^{2p} = \mathbf{E}[|\tilde{X}_\kappa|^{2p}] < +\infty, \quad (A.3)$$

where the limit holds in the almost sure sense.
In order to treat the second term in the right hand side of (A.2), we resort to equation (A.1). From there it is readily checked that $\frac{d}{dt}|X_t-\tilde{X}_t|^2 = -2|X_t-\tilde{X}_t|^2$, and thus $|X_t-\tilde{X}_t|^2 = e^{-2t}|y_0-\tilde{X}_0|^2$. In particular, we have $\lim_{k\to+\infty}|X_{k\kappa}-\tilde{X}_{k\kappa}|^{2p} = \lim_{k\to+\infty} e^{-2pk\kappa}|y_0-\tilde{X}_0|^{2p} = 0$ and the corresponding Cesaro summation tends also to 0, i.e.

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T |X_{\underline{t}}-\tilde{X}_{\underline{t}}|^{2p}dt = \lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}|X_{k\kappa}-\tilde{X}_{k\kappa}|^{2p} = 0. \quad (A.4)$$

Now it remains to treat the first term in the right hand side of (A.2). To this end, we will invoke our hypothesis $(\mathbf{C_w})$ and a classical argument based on Gronwall's lemma. Namely, note that for all $\varepsilon > 0$,

$$\begin{aligned}\frac{d}{dt}|Y_t^{\vartheta}-X_t|^2 &= 2\langle Y_t^{\vartheta}-X_t, b_\vartheta(Y_t^{\vartheta})+X_t\rangle\\ &= 2\langle Y_t^{\vartheta}-X_t, b_\vartheta(Y_t^{\vartheta})-b_\vartheta(X_t)\rangle + 2\langle Y_t^{\vartheta}-X_t, b_\vartheta(X_t)+X_t\rangle\\ &\leqslant 2\beta - 2\alpha|Y_t^{\vartheta}-X_t|^2 + \frac{1}{\varepsilon}|Y_t^{\vartheta}-X_t|^2 + \varepsilon\tilde{C}(1+|X_t|^{2r}),\end{aligned}$$

where the last inequality stems from $(\mathbf{C_w})$, Young's inequality and the linear growth of b_ϑ entailed by (2.5). We then set $\varepsilon = 1/\alpha$, which gives

$$\frac{d}{dt}|Y_t^{\vartheta}-X_t|^2 \leqslant 2\beta - \alpha|Y_t^{\vartheta}-X_t|^2 + \frac{\tilde{C}}{\alpha}(1+|X_t|^{2r}).$$

Hence from Gronwall's lemma, we deduce that

$$|Y_t^{\vartheta}-X_t|^2 \leqslant \int_0^t e^{-\alpha(t-s)}\left(2\beta+\frac{\tilde{C}}{\alpha}(1+|X_s|^{2r})\right)ds. \quad (A.5)$$

Starting from (A.5), one can easily get a bound on $|Y_t^{\vartheta}-X_t|^{2p}$. Namely, due to the fact that $\int_0^t e^{-\alpha(t-s)}ds = \alpha^{-1}(1-e^{-\alpha t})$, a direct application of Jensen's inequality yields

$$\begin{aligned}|Y_t^{\vartheta}-X_t|^{2p} &\leqslant \left(\frac{1-e^{-\alpha t}}{\alpha}\right)^{p-1}\int_0^t e^{-\alpha(t-s)}\left(2\beta+\frac{\tilde{C}}{\alpha}(1+|X_s|^{2r})\right)^p ds\\ &\leqslant C_p\int_0^t e^{-\alpha(t-s)}(1+|X_s|^{2pr})ds.\end{aligned}$$

Then, by using Fubini theorem, it comes

$$\frac{1}{T}\int_0^T |Y^\vartheta_{\underline{t}} - X_{\underline{t}}|^{2p}dt \leqslant \frac{C_p}{T}\int_0^T\int_0^{\underline{t}} e^{-\alpha(\underline{t}-s)}(1+|X_s|^{2pr})ds\ dt$$
$$\leqslant \frac{C_p}{T}\int_0^T (1+|X_s|^{2pr})\int_s^T \mathbf{1}_{[0,\underline{t}]}(s)e^{-\alpha(\underline{t}-s)}dt\ ds. \tag{A.6}$$

In addition, for $t\in[0,T]$ we have $\mathbf{1}_{[0,\underline{t}]}(s)e^{-\alpha(\underline{t}-s)} \leqslant e^{\alpha\kappa}e^{-\alpha(t-s)}$. Hence integrating the right hand side of (A.6) we get

$$\frac{1}{T}\int_0^T |Y^\vartheta_{\underline{t}} - X_{\underline{t}}|^{2p}dt \leqslant \frac{C_pe^{\alpha\kappa}}{\alpha}\frac{1}{T}\int_0^T (1+|X_s|^{2pr})\left(1-e^{-\alpha(T-s)}\right)ds \leqslant \frac{C_pe^{\alpha\kappa}}{\alpha}\left(1+\frac{1}{T}\int_0^T |X_s|^{2pr}ds\right).$$

Finally, by the same type of arguments used in the first part of the proof (see (A.4)), we can show that $\lim_{T\to+\infty}\frac{1}{T}\int_0^T|X_s|^{2pr}ds = \mathbf{E}[|\tilde{X}_0|^{2pr}] < +\infty$. Thus we have obtained

$$\sup_{T>0}\sup_{\vartheta\in\Theta}\frac{1}{T}\int_0^T |Y^\vartheta_{\underline{t}} - X_{\underline{t}}|^{2p}dt < +\infty. \tag{A.7}$$

Plugging (A.3), (A.4) and (A.7) into (A.2), our claim (ii) is now proved.

(iii) We let the patient reader check the details for item (iii). It follows the same lines as item (ii), except for the fact that the discretization procedure is avoided. □

Our next result is an analog of Proposition A.1 for the Euler scheme $Z^{\vartheta,\gamma}$.

Proposition A.2. *Let $Z^{\vartheta,\gamma}$ be the Euler approximation scheme defined by (2.13). Assume $(\mathbf{H_0})$ and $(\mathbf{C_w})$. Then, there exists $\gamma_0>0$ such that for all $p>0$,*

- (i) $\sup_{\vartheta\in\Theta}\sup_{\gamma\in(0,\gamma_0]}\limsup_{N\to+\infty}\mathbf{E}[|Z^{\vartheta,\gamma}_{N\gamma}|^p] < +\infty.$
- (ii) $\sup_{\gamma\in(0,\gamma_0]}\sup_{N\geqslant 0}\gamma^{p(1-H)}\mathbf{E}[\sup_{\vartheta\in\Theta}|Z^{\vartheta,\gamma}_{N\gamma}|^p] < +\infty.$
- (iii) *Fix $\gamma\in(0,1]$. Then* $\sup_{\vartheta\in\Theta}\sup_{N\geqslant 1}\frac{1}{N}\sum_{k=0}^{N-1}|Z^{\vartheta,\gamma}_{k\gamma}|^p < +\infty$ *a.s.*

The strategy for the proof of Proposition A.2 is based on a comparison between $Z^{\vartheta,\gamma}$ and the Euler scheme Σ related to the Ornstein-Uhlenbeck process X given by (A.1). Namely define Σ recursively by $\Sigma_0 = y_0$ and:

$$\Sigma_{(k+1)\gamma} = (1-\gamma)\Sigma_{k\gamma} + \sigma((B_{(k+1)\gamma} - B_{k\gamma})), \quad \forall k\geqslant 0. \tag{A.8}$$

We first prove some bounds on Σ itself, which are labeled in the following lemma

Lemma A.3. *Let Σ be the Euler scheme defined by (A.8). There exists $\gamma_0>0$ such that for all $p>0$ we have*

- (i) $\sup_{\gamma\in(0,\gamma_0]}\limsup_{N\to+\infty}\mathbf{E}[|\Sigma_{N\gamma}|^p] < +\infty.$
- (ii) $\sup_{\gamma\in(0,\gamma_0]}\sup_{N\geqslant 0}\gamma^{p(1-H)}\mathbf{E}[|\Sigma_{N\gamma}|^p] < +\infty.$
- (iii) *Fix $\gamma\in(0,\gamma_0]$. Then* $\sup_{N\geqslant 1}\frac{1}{N}\sum_{k=0}^{N-1}|\Sigma_{k\gamma}|^p < +\infty$ *a.s.*

Proof. By Cauchy-Schwarz inequality, it is enough to show the three results for $2p$ instead of p. We treat again the three items separately.

(i) Let $\gamma_0 \in (0,1)$. By Lemma 2 in [9], we know that

$$\sup_{\gamma\in(0,\gamma_0]} \limsup_{N\to+\infty} \mathbf{E}[|\Sigma_{N\gamma}|^2] < +\infty. \tag{A.9}$$

Since Σ is a Gaussian process, the extension from a second order moment to a moment of order $2p$ is trivial. We thus omit details for sake of conciseness.

(ii) Let $\gamma_0 \in (0,1)$ and $\gamma \in (0,\gamma_0)$. First note that by induction, we have for all $k \geq 1$:

$$\Sigma_{k\gamma} = (1-\gamma)^k y_0 + \sigma \sum_{j=0}^{k-1} (1-\gamma)^j \Delta_{k-j}, \tag{A.10}$$

where $\Delta_{k-j} := (B_{(k-j)\gamma} - B_{(k-1-j)\gamma})$. In equation (A.10), we apply the triangular inequality for the norm in $L^2(\Omega)$ and we invoke the fact that $\mathbf{E}[(\Delta_k^i)^2]^{1/2} = \gamma^H$ for all $i \in \{1,\dots,d\}$. This yields

$$\begin{aligned}
\mathbf{E}[|\Sigma_{k\gamma}|^2]^{1/2} &\leqslant (1-\gamma)^k |y_0| + |\sigma| \sum_{j=0}^{k-1} (1-\gamma)^j \mathbf{E}[|\Delta_{k-j}|^2]^{1/2} \\
&\leqslant (1-\gamma)^k |y_0| + |\sigma| \sum_{j=0}^{k-1} (1-\gamma)^j \sum_{i=1}^{d} \mathbf{E}[(\Delta_{k-j}^i)^2]^{1/2} \\
&\leqslant (1-\gamma)^k |y_0| + d|\sigma|\gamma^H \sum_{j=0}^{k-1} (1-\gamma)^j \\
&\leqslant |y_0| + d|\sigma| c \gamma^{H-1}.
\end{aligned}$$

Therefore, we easily get that

$$\gamma^{2-2H}\mathbf{E}[|\Sigma_{k\gamma}|^2] \leqslant 2\gamma^{2-2H}\left(|y_0|^2 + d^2|\sigma|^2\gamma^{2H-2}\right) \leqslant 2\gamma_0^{2-2H}|y_0|^2 + 2d^2|\sigma|^2$$

which gives the following bound

$$\sup_{\gamma\in(0,\gamma_0]} \sup_{N\geqslant 0} \gamma^{2-2H}\mathbf{E}[|\Sigma_{N\gamma}|^2] < +\infty.$$

Exactly as in item (i), we now resort to the Gaussian nature of Σ in order to conclude that

$$\sup_{\gamma\in(0,\gamma_0]} \sup_{N\geqslant 0} \gamma^{2p(1-H)}\mathbf{E}[|\Sigma_{N\gamma}|^{2p}] < +\infty, \qquad \text{for all } p \geq 1. \tag{A.11}$$

(iii) We start from identity (A.10). Applying Jensen's inequality we deduce that

$$\begin{aligned}
|\Sigma_{k\gamma}|^{2p} &\leqslant 2^{2p-1}(1-\gamma)^{2kp}|y_0|^{2p} + \left(\frac{2}{\gamma}\right)^{2p-1} |\sigma|^{2p} \sum_{j=0}^{k-1} (1-\gamma)^j |\Delta_{k-j}|^{2p} \\
&\leqslant 2^{2p-1}|y_0|^{2p} + \left(\frac{2}{\gamma}\right)^{2p-1} |\sigma|^{2p} \sum_{j=0}^{k-1} (1-\gamma)^j |\Delta_{k-j}|^{2p}.
\end{aligned} \tag{A.12}$$

We have thus obtained that

$$\frac{1}{N}\sum_{k=0}^{N-1} |\Sigma_{k\gamma}|^{2p} \leqslant (2^{2p-1}+1)|y_0|^{2p} + \left(\frac{2}{\gamma}\right)^{2p-1} |\sigma|^{2p} \frac{1}{N}\sum_{k=1}^{N-1}\sum_{j=0}^{k-1} (1-\gamma)^j |\Delta_{k-j}|^{2p}.$$

Therefore our claim (iii) is reduced to show that

$$\sup_{N\geqslant 1}\ \frac{1}{N}\sum_{k=1}^{N-1}\sum_{j=0}^{k-1}(1-\gamma)^j|\Delta_{k-j}|^{2p}<+\infty. \tag{A.13}$$

In order to prove (A.13) we make a change of variable $\ell=k-1-j$ and apply Fubini's theorem. This gives

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N-1}\sum_{j=0}^{k-1}(1-\gamma)^j|\Delta_{k-j}|^{2p}&=\frac{1}{N}\sum_{\ell=0}^{N-2}|\Delta_{\ell+1}|^{2p}\sum_{k=\ell+1}^{N-1}(1-\gamma)^{k-1-\ell}\\&\leqslant\frac{1}{\gamma}\frac{1}{N}\sum_{\ell=0}^{N-2}|\Delta_{\ell+1}|^{2p}\leqslant\frac{d^{p'-1}}{\gamma}\sum_{i=1}^{d}\frac{1}{N}\sum_{\ell=0}^{N-2}(\Delta^i_{\ell+1})^{2p}.\end{aligned}$$

Since the sequence $(\Delta^i_n)_{n\geqslant 1}$ is ergodic for every $i\in\{1,\dots,d\}$, we have

$$\frac{1}{N}\sum_{\ell=0}^{N-2}(\Delta^i_{\ell+1})^{2p}\underset{N\to+\infty}{\longrightarrow}\mathbf{E}[(\Delta^i_1)^{2p}]=c_p\gamma^{2pH}$$

and our claim (A.13) follows. This finishes the proof. □

With those preliminary considerations on Σ in hand, we can now prove our Proposition A.2.

Proof of Proposition A.2. As previously, we will prove the result for $2p$ instead of p. Moreover, for sake of simplicity, we write Z^ϑ instead of $Z^{\vartheta,\gamma}$. According to the dynamics (A.8) for Σ and (2.13) for Z^ϑ, we have for all $k\geqslant 1$,

$$\begin{aligned}|Z^\vartheta_{k\gamma}-\Sigma_{k\gamma}|^2&=|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2\\&\quad+2\gamma\langle Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma},\Sigma_{(k-1)\gamma}+b_\vartheta(Z^\vartheta_{(k-1)\gamma})\rangle+\gamma^2|\Sigma_{(k-1)\gamma}+b_\vartheta(Z^\vartheta_{(k-1)\gamma})|^2.\end{aligned} \tag{A.14}$$

In order to treat the second term in (A.14), we recast it as

$$\begin{aligned}&\langle Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma},\Sigma_{(k-1)\gamma}+b_\vartheta(Z^\vartheta_{(k-1)\gamma})\rangle\\&\quad=\langle Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma},b_\vartheta(Z^\vartheta_{(k-1)\gamma})-b_\vartheta(\Sigma_{(k-1)\gamma})\rangle+\langle Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma},\Sigma_{(k-1)\gamma}+b_\vartheta(\Sigma_{(k-1)\gamma})\rangle\end{aligned}$$

Then, we invoke our condition $(\mathbf{C_w})$ to bound the first term in the right hand side above and Young's inequality for the second term. Similar manipulations can be performed for the third term in (A.14). We let the patient reader check that for some arbitrary parameters ε,ε' we get

$$\begin{aligned}&|Z^\vartheta_{k\gamma}-\Sigma_{k\gamma}|^2\\&\quad\leqslant 2\beta\gamma+(1-2\gamma\alpha)|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2+\frac{\gamma}{\varepsilon}|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2\\&\quad+\gamma\varepsilon|\Sigma_{(k-1)\gamma}+b_\vartheta(\Sigma_{(k-1)\gamma})|^2+\frac{\gamma^2}{2\varepsilon'}|\Sigma_{(k-1)\gamma}+b_\vartheta(\Sigma^\vartheta_{(k-1)\gamma})|^2+\frac{\gamma^2L^2\varepsilon'}{2}|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2.\end{aligned}$$

We now choose $\varepsilon=\frac{1}{\alpha}$ and $\varepsilon'=\frac{\alpha}{\gamma L^2}$, which yields

$$|Z^\vartheta_{k\gamma}-\Sigma_{k\gamma}|^2\leqslant 2\beta\gamma+\left(1-\frac{\gamma\alpha}{2}\right)|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2+\left(\frac{\gamma}{\alpha}+\frac{\gamma^3L^2}{2\alpha}\right)|\Sigma_{(k-1)\gamma}+b_\vartheta(\Sigma_{(k-1)\gamma})|^2.$$

Observe that under $(\mathbf{C_w})$, b_ϑ is sublinear. Hence there exists $C>0$ depending only on α,β,L such that

$$|Z^\vartheta_{k\gamma}-\Sigma_{k\gamma}|^2\leqslant(1-\gamma\tilde\alpha)\,|Z^\vartheta_{(k-1)\gamma}-\Sigma_{(k-1)\gamma}|^2+C\gamma\left(1+\gamma^2\right)\left(1+|\Sigma_{(k-1)\gamma}|^2\right) \tag{A.15}$$

where we have set $\tilde{\alpha} := \alpha/2$ in order to ease our next computations. We now choose $0 < \gamma_0 < 1/\tilde{\alpha}$. By a direct induction, it comes

$$|Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^2 \leqslant C\gamma \left(1+\gamma^2\right) \sum_{j=0}^{k-1} (1-\gamma\tilde{\alpha})^{k-1-j} \left(1 + |\Sigma_{j\gamma}|^2\right).$$

Finally, applying Jensen's inequality similarly to what is done in (A.12), we end up with

$$\begin{aligned} |Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^{2p} &\leqslant C^p \frac{\gamma}{\tilde{\alpha}^{p'-1}} \left(1+\gamma^2\right)^p \sum_{j=0}^{k-1} (1-\gamma\tilde{\alpha})^{k-1-j} \left(1 + |\Sigma_{j\gamma}|^2\right)^p \\ &\leqslant C^p \left(\frac{2}{\tilde{\alpha}}\right)^{p-1} \gamma \left(1+\gamma^2\right)^p \sum_{j=0}^{k-1} (1-\gamma\tilde{\alpha})^{k-1-j} \left(1 + |\Sigma_{j\gamma}|^{2p}\right). \end{aligned} \tag{A.16}$$

With those preliminary considerations in hand, we now prove the three items in Proposition A.2 separately.

(i) We start from relation (A.15). Since $\gamma \in (0, \gamma_0]$, $C(1+\gamma^2) \leq \tilde{C}$ for a constant $\tilde{C} > 0$

$$|Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^2 \leqslant (1-\gamma\tilde{\alpha}) \, |Z^{\vartheta}_{(k-1)\gamma} - \Sigma_{(k-1)\gamma}|^2 + \gamma\tilde{\alpha}\frac{\tilde{C}}{\tilde{\alpha}} \left(1 + |\Sigma_{(k-1)\gamma}|^2\right).$$

We now invoke the convexity of $x \mapsto |x|^p$ in order to get

$$\begin{aligned} |Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^{2p} &\leqslant (1-\gamma\tilde{\alpha}) \, |Z^{\vartheta}_{(k-1)\gamma} - \Sigma_{(k-1)\gamma}|^{2p} + \gamma\tilde{\alpha}\frac{\tilde{C}^p}{\tilde{\alpha}^p} \left(1 + |\Sigma_{(k-1)\gamma}|^2\right)^p \\ &\leqslant (1-\gamma\tilde{\alpha}) \, |Z^{\vartheta}_{(k-1)\gamma} - \Sigma_{(k-1)\gamma}|^{2p} + \tilde{C}_p\gamma \left(1 + |\Sigma_{(k-1)\gamma}|^{2p}\right). \end{aligned}$$

Then we take expectations and upper limits in k, which gives

$$\limsup_{k\to+\infty} \mathbf{E}[|Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^{2p}] \leqslant \frac{\tilde{C}_p}{\tilde{\alpha}} \left(1 + \limsup_{k\to+\infty} \mathbf{E}[|\Sigma_{k\gamma}|^{2p}]\right).$$

With this inequality in hand and Proposition A.3 (i), inequality (i) is proved.

(ii) First, we take successively the supremum over Θ and the expectation in (A.16). We then multiply by $\gamma^{2p(1-H)}$ and perform the same kind of manipulations as for (i). With the help of Proposition A.3 (ii), we get that

$$\sup_{\gamma\in(0,\gamma_0]} \sup_{k\geqslant 0} \gamma^{2p(1-H)} \mathbf{E}[\sup_{\vartheta\in\Theta} |Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^{2p}] < +\infty. \tag{A.17}$$

Having achieved the upper bound (A.17), our claim (ii) now stems from another direct application of Proposition A.3 (ii).

(iii) Here, we just need to sum (A.16) for k from 0 to $N-1$ and divide by N. Then, we use Fubini theorem on the right hand side combined with Proposition A.3 (iii) to conclude that $\sup_{\vartheta\in\Theta} \sup_{N\geqslant 1} \frac{1}{N} \sum_{k=0}^{N-1} |Z^{\vartheta}_{k\gamma} - \Sigma_{k\gamma}|^{2p} < +\infty$. Finally, by using again Proposition A.3 (iii), the result follows.

□

A.2. **Proof of Proposition 3.3.** We will focus on the proof of (3.8) only, the discrete counterpart of Proposition 3.3 being obtained by similar argument. We also note that (3.8) cannot be obtained as a consequence of Birkhoff's theorem. Indeed, Birkhoff's theorem would yield a $\bar{\nu}_\vartheta$-a.s. convergence in (3.8), where $\bar{\nu}_\vartheta$ is the invariant measure alluded to in Remark 2.3. In

order to avoid any reference to the support of the invariant measure $\bar{\nu}_\vartheta$, our relation (3.8) gives a weak convergence result which does not include any condition on the initial value of Y_0^ϑ.

In order to prove (3.8), we first observe that according to (3.9), the family of measures $\{\frac{1}{t}\int_0^t \delta_{Y_s^\vartheta} ds; t \geq 0\}$ is $a.s.$ tight. Therefore it only remains to prove that any limit is ν_ϑ. We now focus on the convergence of $\frac{1}{t}\int_0^t \delta_{Y_s^\vartheta} ds$. To this aim, we consider an additional family of probability measures on $\mathcal{C}([0,\infty),\mathbb{R}^d)$ in the following way:

$$\pi_t^\vartheta = \frac{1}{t}\int_0^t \delta_{Y_{s+.}^\vartheta}\, ds$$

where $Y_{s+.}^\vartheta = (Y_{s+u}^\vartheta)_{u\geq 0}$. We will first prove that $\{\pi_t^\vartheta; t \geq 0\}$ is an $a.s.$ tight family in the set $\mathcal{M}_1(\mathcal{C}([0,\infty),\mathbb{R}^d))$, where we recall that the notation $\mathcal{M}_1$ is introduced in Section 2.1. The tightness of $\{\pi_t^\vartheta; t \geq 0\}$ can be handled in the following way: we have seen that $\{\frac{1}{t}\int_0^t \delta_{Y_s^\vartheta} ds; t \geq 0\}$ is $a.s.$ tight in $\mathcal{M}_1(\mathbb{R}^d)$. Therefore a classical criterion (see *e.g.* [4, Theorem 8.3]) ensures that $\{\pi_t^\vartheta; t \geq 0\}$ is $a.s.$ tight if for every positive T, η and ε, there exists $\delta > 0$ such that for all $t_0 \in [0,T]$,

$$\limsup_{t\to+\infty} \frac{1}{t}\int_0^t \frac{1}{\delta}\mathbf{1}_{\{\sup_{u\in[t_0,t_0+\delta]}|Y_{s+u}^\vartheta - Y_{s+t_0}^\vartheta|\geq\varepsilon\}} ds \leq \eta \quad a.s. \tag{A.18}$$

Moreover, inequality (A.18) holds true as long as there exist some positive r and ρ such that

$$\limsup_{t\to+\infty} \frac{1}{t}\int_0^t \sup_{u\in[t_0,t_0+\delta]} |Y_{s+u}^\vartheta - Y_{s+t_0}^\vartheta|^r ds \leq C_{r,T}\delta^{1+\rho} \quad a.s. \tag{A.19}$$

Let us now prove (A.19). On the interval $[t_0, t_0+\delta]$, we have

$$|Y_{s+u}^\vartheta - Y_{s+t_0}^\vartheta| \leq \int_{t_0}^{t_0+\delta} |b_\vartheta(Y_{s+u}^\vartheta)| du + \sup_{u\in[t_0,t_0+\delta]} |B_{s+u} - B_{s+t_0}|. \tag{A.20}$$

Using Jensen's inequality and Fubini's Theorem, we get that for any $r > 0$,

$$\frac{1}{t}\int_0^t \left(\int_{t_0}^{t_0+\delta} |b_\vartheta(Y_{s+u}^\vartheta)| du\right)^r ds \leq \delta^{r-1}\left(\frac{1}{t}\int_0^{t+t_0+\delta} |b(Y_s^\vartheta)|^r ds\right).$$

Furthermore, since b_ϑ is sublinear according to $(\mathbf{C_w})$, it follows from (3.10) that

$$\limsup_{t\to+\infty} \frac{1}{t}\int_0^t \left(\int_{t_0}^{t_0+\delta} |b_\vartheta(Y_{s+u}^\vartheta)| du\right)^r ds \leq C\delta^{r-1} \tag{A.21}$$

for a constant $C > 0$. On the other hand, by the ergodicity of the increments of the fBm and its self-similarity,

$$\frac{1}{t}\int_0^t \sup_{u\in[t_0,t_0+\delta]} |B_{s+u} - B_{s+t_0}|^r ds \xrightarrow{t\to+\infty} \mathbf{E}[\sup_{u\in[0,\delta]} |B_u|^r] = C_r\delta^{Hr}, \tag{A.22}$$

where the limit in (A.22) holds in the $a.s.$ sense. Choosing $r > \sup\{2, 1/H\}$ in (A.21) and (A.22), then plugging (A.21) and (A.22) into (A.20), we get that (A.19) is satisfied. We have thus proved the $a.s.$ tightness of $\{\pi_t^\vartheta; t \geq 0\}$.

The second step is then to show that any limiting distribution of $\left(\frac{1}{t}\int_0^t \delta_{Y_{s+.}^\vartheta}\, ds\right)$ is necessarily the law of a stationary solution $\bar{Y}^\vartheta$ to SDE (2.7). This step in turn implies the result by uniqueness of the stationary solutions.

Let $\left(\frac{1}{t_n}\int_0^{t_n}\delta_{Y^{\vartheta}_{s+.}}ds\right)_n$ be a (pathwise) convergent sequence with limiting distribution μ where $\{t_n; n\geq 1\}$ is an increasing sequence converging to $+\infty$. We first prove that μ is the law of a stationary process. Namely, for any bounded functional $F:\mathcal{C}([0,\infty),\mathbb{R}^d)\to\mathbb{R}$, we have

$$\mu(F\circ\theta_T)-\mu(F)=\lim_{n\to+\infty}\frac{1}{t_n}\int_0^{t_n}[F(Y^{\vartheta}_{s+T+.})-F(Y^{\vartheta}_{s+.})]ds.$$

However, a simple change of variables reveals that $a.s.$ we have

$$\lim_{n\to+\infty}\frac{1}{t_n}\int_0^{t_n}[F(Y^{\vartheta}_{s+T+.})-F(Y^{\vartheta}_{s+.})]ds=\lim_{n\to+\infty}\frac{1}{t_n}\left(\int_{t_n}^{t_n+T}F(Y^{\vartheta}_{s+.})ds-\int_0^T F(Y^{\vartheta}_{s+.})ds\right)=0.$$

We thus easily get that μ is stationary. Now, let us prove that μ is the law of a solution to (2.7). Without loss of generality, we can say that a process $(x_t)_{t\geq 0}$ is a solution to (2.7) if $x_.-x_0-\int_0^. b_\vartheta(x_u)du$ is a fBm. In other words, we have to prove that $\mu\circ G^{-1}$ is the law of a fBm where $G(x)=x_.-x_0-\int_0^. b_\vartheta(x_u)du$. Since G is continuous for the $u.s.c$ topology, it is readily checked that

$$\mu\circ G^{-1}=\lim_{n\to+\infty}\frac{1}{t_n}\int_0^{t_n}\delta_{G(Y^{\vartheta}_{s+.})}ds.$$

In addition, by construction $G(Y^{\vartheta}_{s+.})=B_{s+.}-B_s$. Hence the fact that $\mu\circ G^{-1}$ is the law of a fBM follows again from the ergodicity of the increments of the fBM.
Summarizing our considerations, we have proved that μ is a stationary measure related to the system (2.7). Therefore we have $\mu=\mathcal{L}((\bar{Y}^{\vartheta}_t)_{t\geq 0})$, which concludes the proof.

Appendix B. Proof of Proposition 3.8

For sake of conciseness, we will focus on the proof of Proposition 3.8 (i). The proof of item (ii) relies on the same kind of tools, plus the discrete computations invoked in the proof of Proposition A.2. In order to prove item (i), let us consider $t\geq 1$ and a parameter $\rho>0$ to be chosen later on. An easy elaboration of (3.14) shows that

$$e^{\rho t}|Y^{\vartheta_1}_t-Y^{\vartheta_2}_t|^2=\int_0^t e^{\rho s}\left(\rho|Y^{\vartheta_1}_s-Y^{\vartheta_2}_s|^2+2\langle b_{\vartheta_1}(Y^{\vartheta_1}_s)-b_{\vartheta_2}(Y^{\vartheta_2}_s),Y^{\vartheta_1}_s-Y^{\vartheta_2}_s\rangle\right)ds. \tag{B.1}$$

In addition, one can write

$$b_{\vartheta_1}(Y^{\vartheta_1}_s)-b_{\vartheta_2}(Y^{\vartheta_2}_s)=b_{\vartheta_1}(Y^{\vartheta_1}_s)-b_{\vartheta_1}(Y^{\vartheta_2}_s)+b_{\vartheta_1}(Y^{\vartheta_2}_s)-b_{\vartheta_2}(Y^{\vartheta_2}_s).$$

We now combine the assumption $(\mathbf{C_s})$ (including the contraction property, the fact $x\mapsto b_\vartheta(x)$ is uniformly Lipschitz in ϑ and the fact that $\partial_\vartheta b_\vartheta(x)$ has polynomial growth) and Young's inequality $|ab|\leq\frac{1}{2\varepsilon}|a|^2+\frac{\varepsilon}{2}|b|^2$ for an arbitrary $\varepsilon>0$. This yields the existence of a constant $L>0$ such that

$$\langle b_{\vartheta_1}(Y^{\vartheta_1}_s)-b_{\vartheta_2}(Y^{\vartheta_2}_s),Y^{\vartheta_1}_s-Y^{\vartheta_2}_s\rangle\leq\left(-\alpha+\frac{L^2\varepsilon}{2}\right)|Y^{\vartheta_1}_s-Y^{\vartheta_2}_s|^2+\frac{C|\vartheta_1-\vartheta_2|^2(1+|Y_s|^r)^2}{2\varepsilon}.$$

Plugging this inequality into (B.1) and setting $\varepsilon=L^2/\alpha$ and $\rho=\alpha/2$, we have thus obtained

$$e^{\rho t}|Y^{\vartheta_1}_t-Y^{\vartheta_2}_t|^2\leq C|\vartheta_1-\vartheta_2|^2\int_0^t e^{\rho s}(1+|Y^{\vartheta_2}_s|^{2r}ds. \tag{B.2}$$

Thus, using Fubini's theorem, one deduces that

$$\begin{aligned}\frac{1}{t}\int_0^t|Y^{\vartheta_1}_s-Y^{\vartheta_2}_s|^2ds&\leq\frac{C|\vartheta_1-\vartheta_2|^2}{t}\int_0^t(1+|Y^{\vartheta_2}_s|^{2r}\int_u^t e^{\rho(s-u)}dsdu\\&\leq C_\rho|\vartheta_1-\vartheta_2|^2\left(1+\frac{1}{t}\int_0^t(1+|Y^{\vartheta_2}_s|^{2r}du\right).\end{aligned}$$

Hence a direct application of inequality (3.9) yields the existence of a random variable $C = C(\omega)$ such that for all $(\vartheta_1, \vartheta_2) \in \Theta^2$ we have

$$\sup_{t\geq 1} \frac{1}{t} \int_0^t |Y_s^{\vartheta_1} - Y_s^{\vartheta_2}|^2 ds \leq C|\vartheta_1 - \vartheta_2|^2.$$

This concludes the proof of Proposition 3.8 (i) for a distance $d \in \mathcal{D}_2$. To extend the result to any $p \geq 2$, one can apply Jensen's inequality to (B.2) and follow the same lines as for $d \in \mathcal{D}_2$.

Acknowledgements

The authors would like to thank the anonymous referee for his careful reading and his suggestions, which substantially improved the quality of the paper.

References

[1] L. Arnold (1997): *Random Dynamical Systems.* Berlin, Springer.
[2] B. Arras, G.Mijoule, G.Poly and Y.Swan (2016): A new approach to the Stein-Tikhomirov method: with applications to the second Wiener chaos and Dickman convergence, *arXiv preprint arXiv:1605.06819.*
[3] R. Belfadli, K. Es-Sebaiy, Y. Ouknine (2011): Parameter Estimation for Fractional Ornstein-Uhlenbeck Processes: Non-ergodic Case. In: *Frontiers in Science and Engineering.* Edited by Hassan II Academy of Science and Technology. **1**, 1–16.
[4] P. Billingsley (1968): *Convergence of Probability Measures*, Wiley.
[5] A. Brouste, M. Kleptsyna, A. Popier (2012): Design for estimation of the drift parameter in fractional diffusion systems. *Stat. Inference Stoch. Process.* **15**, no. 2, 133–149.
[6] B. Buchmann, C. Klüppelberg (2005): Maxima of stochastic processes driven by fractional Brownian motion. *Adv. in Appl. Probab.* **37** (2005), no. 3, 743-764.
[7] P. Cheridito, H. Kawaguchi and M.Maejima (2003): Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability* **8**, no. 3.
[8] J.F. Coeurjolly (2001): Estimating the Parameters of a Fractional Brownian Motion by Discrete Variations of its Sample Paths. *Stat. Infer. Stoch. Process.* **4**, no. 2, 199–227.
[9] S. Cohen, F. Panloup (2011): Approximation of stationary solutions of Gaussian driven stochastic differential equations. *Stochastic Processes and their Applications* **12**, no. 12, 2776–2801.
[10] S. Cohen, F. Panloup, S. Tindel (2014): Approximation of stationary solutions to SDEs driven by multiplicative fractional noise. *Stochastic Processes and their Applications*, 124(3), 1197-1225.
[11] H. Crauel, A. Debussche, F. Flandoli (1997): Random attractors. *J. Dyn. Diff. Eqns.* **9**, 307–341.
[12] A. Chronopoulou, S. Tindel (2013): On inference for fractional differential equations. *Stat. Infer. Stoch. Process.* **16**, no. 1, 29–61.
[13] F. Comte, V. Genon-Catalot, Y. Rozenholc (2007): Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli.* **13**, no. 2, 514–543.
[14] F. Comte, N. Marie (2018): Nonparametric estimation in fractional SDE. *Arxiv* preprint.
[15] A. Dalalyan (2005): Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33**, no. 6, 2507–2528.
[16] M. Duflo (1997). *Random iterative models.* Springer-Verlag, 385 p.
[17] R.J. Elliot (2001): A continuous time Kronecker's lemma and martingale convergence. *Stochastic analysis and applications.*
[18] D. Florens-Zmirou (1989): Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20**, no. 4, 547–557.
[19] R. Frydman (1980): A proof of the consistency of maximum likelihood estimators of non-linear regression models with autocorellated errors. *Econometrica* **48**, 853–860.
[20] A. M. Garcia, E. Rodemich, H. Rumsey Jr. (1978): A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Math. J.* **20**, 565–578.
[21] M. Garrido-Atienza, P. Kloeden, A. Neuenkirch (2009): Discretization of stationary solutions of stochastic systems driven by fractional Brownian motion. *Appl. Math. Optim.* **60**, no. 2, 151–172.
[22] M. Garrido-Atienza, B. Schmalfuss (2011): Ergodicity of the infinite-dimensional fractional Brownian motion. *J. Dynam. Differential Equations.* **23**, no. 3, 671–681.

[23] M. Hairer (2005): Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.* **33**, no. 2, 703–758.
[24] M. Hairer, A. Majda (2010): A simple framework to justify linear response theory. *Nonlinearity* **23**, no. 4, 909–922.
[25] M. Hairer, A. Ohashi (2007): Ergodicity theory of SDEs with extrinsic memory. *Ann. Probab.* **35**, no. 5, 1950–1977.
[26] M. Hairer, S. Pillai (2011): Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré, Probab. Stat.* **47**, no. 2, 601–628.
[27] M. Hoffmann (1999): Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79**, no. 1, 135–163.
[28] Y. Hu, D. Nualart (2010): Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Stat. Prob. Lett.* **80**, 1030–1038.
[29] J. Istas, G. Lang (1997): Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.* **33**, no. 4, 407–436.
[30] M. Kessler (1997): Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, no. 2, 211–229.
[31] M. Kessler (2000): Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process. *Scand. J. Statist.* **27**, no. 1, 65–82.
[32] M. Kleptsyna, A. Le Breton (2002): Statistical analysis of the fractional Ornstein-Uhlenbeck type process. *Stat. Inference Stoch. Process.* **5**, no. 3, 229–248.
[33] Y. A. Kutoyants (2004): Statistical inference for ergodic diffusion processes. *Springer Series in Statistics. Springer-Verlag London*, 481 p.
[34] A. Le Breton (1998): Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion. *Stat. Probab. Lett.* **38**, no. 3, 263–274.
[35] Maruyama, Gisiro (1949):The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyusyu Univ.* **4**, 45–106.
[36] A. Neuenkirch, S. Tindel (2011): A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise. *Stat. Inference Stoch. Process.* **17**, no. 1, 99–120.
[37] I. Nourdin (2005). Approximation schemes associated with a differential equation governed by a Hölder function; the case of fractional Brownian motion. *C. R. Math. Acad. Sci. Paris*, **340**, no. 8, 611–614.
[38] A. Papavasiliou, G.A. Pavliotis, A. Stuart (2009): Maximum likelihood drift estimation for multiscale diffusions. *Stochastic Process. Appl.* **119**, no. 10, 3173–3210.
[39] Prakasa Rao, B. L. S. (2010): Statistical inference for fractional diffusion processes. Wiley Series in Probability and Statistics, Chichester, John Wiley & Sons.
[40] A.Rackauskas, C.Suquet (1999): Central limit theorem in Hölder spaces. *Probability and mathematical statistics-Wroclaw University.*
[41] B.Saussereau (2012): Transportation inequalities for stochastic differential equations driven by a fractional Brownian motion, *Bernoulli* **18**, no. 1, 1–23.
[42] M. A. Schmitz, M. Heitz, N. Bonneel, F. Ngolè, D. Coeurjolly, M. Cuturi, G. Peyré, J.L. Starck (2018). Wasserstein dictionary learning: optimal transport-based unsupervised nonlinear dictionary learning. *SIAM J. Imaging Sci.* **11**, no. 1, 643–678.
[43] A.N. Shiryaev (1995): *Probability.* 2nd ed. Graduate Texts in Mathematics. New York, Springer.
[44] C. Strauch (2015). Sharp adaptive drift estimation for ergodic diffusions: the multivariate case. *Stochastic Process. Appl.* **125**, no. 7, 2562–2602.
[45] C. Tudor, F. Viens (2007): Statistical aspects of the fractional stochastic calculus. *Ann. Statist.* **35**, no. 3, 1183–1212.
[46] C. Tudor, F. Viens (2009): Variations and estimators for self-similarity parameters via Malliavin calculus. *Ann. Probab.* **37**, no. 6, 2093–2134.
[47] S.S. Vallander (1973): Calculations of the Vasserstein distance between probability distributions on the line (Russian). *Teor. Verojatnost. i Primenen.* **18**, 824–827.
[48] M. Varvenne (2019): Concentration inequalities for Stochastic Differential Equations with additive fractional noise. *Electron. J. Probab.* **24**, no. 124, 1–22.
[49] M. Varvenne (2019): Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory. *Bernoulli* **25**, no. 4B, 3234–3275.
[50] C. Villani (2008): Optimal transport: old and new. *Springer Science & Business Media*
[51] W. Xiao, W. Zhang, W. Xu (2011): Parameter estimation for fractional Ornstein-Uhlenbeck processes at discrete observation. *Appl. Math. Model.* **35**, no. 9, 4196–4207.
[52] J.Xiong (2008): An introduction to stochastic filtering theory. *Oxford University Press on Demand.*

[53] A.T. Wood, G. Chan (1994). Simulation of stationary Gaussian processes in $[0,1]^d$. *J. Comput. Graph. Statist.* **3**, no.4, 409–432.

Fabien Panloup: Larema, Université d'Angers, 2 Bd Lavoisier, 49045 Angers Cedex 01, France

Samy Tindel: Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, United States

Maylis Varvenne : Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 9, France