Stat Inference Stoch Process (2014) 17:99–120
DOI 10.1007/s11203-013-9084-z

A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise

Andreas Neuenkirch · Samy Tindel

Received: 27 January 2012 / Accepted: 3 April 2013 / Published online: 5 November 2013

Abstract We study a least square-type estimator for an unknown parameter in the drift coefficient of a stochastic differential equation with additive fractional noise of Hurst parameter $H > 1/2$. The estimator is based on discrete time observations of the stochastic differential equation, and using tools from ergodic theory and stochastic analysis we derive its strong consistency.

Keywords Fractional Brownian motion · Parameter estimation · Least square procedure · Ergodicity

Mathematics Subject Classification 62M09 · 62F12

1 Introduction and main results

In this article, we will consider the following $\mathbb{R}^d$-valued stochastic differential equation (SDE)

$$Y_t = y_0 + \int_0^t b(Y_s; \vartheta_0)\, ds + \sum_{j=1}^m \sigma_j B_t^j, \quad t \in [0, T]. \tag{1}$$

Here $y_0 \in \mathbb{R}^d$ is a given initial condition, $B = (B^1, \ldots, B^m)$ is an m-dimensional fractional Brownian motion (fBm) with Hurst parameter $H \in (0, 1)$, the unknown parameter ϑ_0 lies in a certain set Θ which will be specified later on, $\{b(\cdot; \vartheta),\ \vartheta \in \Theta\}$ is a known family of

S. Tindel is member of the BIGS (Biology, Genetics and Statistics) team at INRIA.

A. Neuenkirch (✉)
Institut für Mathematik, Universität Mannheim, A5, 6, 68131 Mannheim, Germany
e-mail: neuenkirch@kiwi.math.uni-mannheim.de

S. Tindel
Institut Élie Cartan, Université de Lorraine, B.P. 239, 54506 Vandœuvre-lès-Nancy Cedex, France
e-mail: samy.tindel@univ-lorraine.fr

drift coefficients with $b(\cdot;\vartheta):\mathbb{R}^d\to\mathbb{R}^d$, and $\sigma_1,\ldots,\sigma_m\in\mathbb{R}^d$ are assumed to be known diffusion coefficients.

Let us recall that B is a centered Gaussian process defined on a complete probability space $(\Omega,\mathcal{F},\mathbf{P})$. Its law is thus characterized by its covariance function, which is defined by

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

The variance of the increments of B is then given by

$$\mathbf{E}\,|B_t^i-B_s^i|^2=|t-s|^{2H},\quad s,t\in\mathbb{R},\quad i=1,\ldots,m,$$

and this implies that almost surely the fBm paths are γ-Hölder continuous for any $\gamma<H$. Furthermore, for $H=1/2$, fBm coincides with the usual Brownian motion, converting the family $\{B^H,\ H\in(0,1)\}$ into the most natural generalization of this classical process. Applications for SDEs driven by fractional Brownian motion can be found in various fields, which include electrical engineering, biophysics or financial modeling, see e.g. Bender et al. (2008), Denk et al. (2001), Kou (2008).

In the current article we assume that the Hurst coefficient satisfies $H>1/2$ and we focus on the estimation of the unknown parameter $\vartheta_0\in\Theta$. Note that the Hurst parameter and the diffusion coefficients can be estimated via the quadratic variation of Y, see e.g. Bégyn (2005), Coeurjolly (2001), Istas and Lang (1994) and also Remark 4.6.

Estimators for the unknown parameter in Eq. (1) based on continuous observation of Y have been studied e.g. in Belfadli et al. (2011), Hu and Nualart (2010), Kleptsyna and Le Breton (2002), Le Breton (1998), Papavasiliou and Ladroue (2011), Prakasa Rao (2010), Tudor and Viens (2007). Estimators based on discrete time data, which are important for practical applications, are then obtained via discretization. However, to the best of our knowledge no genuine estimators based on discrete time data have been analyzed yet.

We propose here a least square estimator for ϑ_0 based on discrete observations of the process Y at times $\{t_k;\ 0\le k\le n\}$. For simplicity, we shall take equally spaced observation times with $t_{k+1}-t_k=\kappa\,n^{-\alpha}:=\alpha_n$ with given $\alpha\in(0,1)$, $\kappa>0$. We call our method least square-type procedure, insofar as we consider a quadratic statistics of the form

$$Q_n(\vartheta)=\frac{1}{n\alpha_n^2}\sum_{k=0}^{n-1}\left(\left|\delta Y_{t_kt_{k+1}}-b(Y_{t_k};\vartheta)\alpha_n\right|^2-\|\sigma\|^2\alpha_n^{2H}\right),\tag{2}$$

where $\delta Y_{u_1u_2}:=Y_{u_2}-Y_{u_1}$ for any $0\le u_1\le u_2\le T$ and $\|\sigma\|^2=\sum_{j=1}^m|\sigma_j|^2$.

Let us now describe the assumptions under which we shall work, starting from a standard hypothesis on the parameter set Θ:

Hypothesis 1.1 *The set Θ is compactly embedded in $\mathbb{R}^q$ for a given $q\ge 1$.*

In order to describe the assumptions on our coefficients b, we will use the following notation for partial derivatives:

Notation 1.2 Let $f:\mathbb{R}^d\times\Theta\to\mathbb{R}$ be a $\mathcal{C}^{p_1,p_2}$ function for $p_1,p_2\ge 1$. Then for any tuple $(i_1,\ldots,i_p)\in\{1,\ldots,d\}^p$, we set $\partial_x^{i_1\ldots i_p}f$ for $\frac{\partial^p f}{\partial x_{i_1}\ldots\partial x_{i_p}}$. Analogously, we use the notation $\partial_\vartheta^{i_1\ldots i_p}f$ for $\frac{\partial^p f}{\partial\vartheta_{i_1}\ldots\partial\vartheta_{i_p}}$ for $(i_1,\ldots,i_p)\in\{1,\ldots,q\}^p$. Moreover, we will write $\partial_x f$ resp. $\partial_\vartheta f$ for the Jacobi-matrices $(\partial_{x_1}f,\ldots,\partial_{x_d}f)$ and $(\partial_{\vartheta_1}f,\ldots,\partial_{\vartheta_q}f)$.

With this notation in mind, our drift coefficients and their derivatives will satisfy a polynomial growth condition, plus an inward condition (also called one-sided dissipative Lipschitz condition) which is traditional for estimation procedures in the Brownian diffusion case, see e.g. Florens-Zmirou (1989), Kasonga (1988):

Hypothesis 1.3 *We have* $b \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ *and there exist constants* $c_1, c_2 > 0$ *and* $N \in \mathbb{N}$ *such that:*

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

$$\langle b(x;\vartheta) - b(y;\vartheta),\ x-y\rangle \le -c_1|x-y|^2.$$

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

$$|b(x;\vartheta)| \le c_2\left(1+|x|^N\right),\quad |\partial_x b(x;\vartheta)| \le c_2\left(1+|x|^N\right),\quad |\partial_\vartheta b(x;\vartheta)| \le c_2\left(1+|x|^N\right).$$

As a consequence of the above assumptions on the drift coefficient and the initial condition, for given $\vartheta_0 \in \Theta$ the solution of Eq. (1) converges for $t \to \infty$ almost surely to a stationary and ergodic stochastic process $(\overline{Y}_t, t \ge 0)$, see the next section.

Finally, we also assume that our drift coefficient is of gradient-type, i.e.:

Hypothesis 1.4 *There exists a function* $U \in \mathcal{C}^{2,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ *such that*

$$\partial_x U(x;\vartheta) = b(x;\vartheta), \quad x \in \mathbb{R}^d,\ \vartheta \in \Theta.$$

With those assumptions in mind, we obtain the following convergence result:

Theorem 1.5 *Assume that the Hypotheses 1.1, 1.3 and 1.4 are satisfied for Eq. (1) and that we moreover have* $H > 1/2$. *Let* $Q_n(\vartheta)$ *be defined by (2). Then we have*

$$\sup_{\vartheta\in\Theta}\left|(Q_n(\vartheta) - Q_n(\vartheta_0)) - \left(\mathbf{E}\,|b(\overline{Y}_0;\vartheta)|^2 - \mathbf{E}|b(\overline{Y}_0;\vartheta_0)|^2\right)\right| \to 0 \tag{3}$$

in the **P**-*almost sure sense.*

This convergence is in contrast to the case $H = 1/2$, i.e. to the case of SDEs with additive Brownian noise. There it holds

$$\sup_{\vartheta\in\Theta}\left|(Q_n(\vartheta) - Q_n(\vartheta_0)) - \mathbf{E}\left[|b(\overline{Y}_0;\vartheta) - b(\overline{Y}_0;\vartheta_0)|^2\right]\right| \to 0 \tag{4}$$

in the **P**-almost sure sense, and usually the consistent least squares estimator (which coincides also with a particular minimum contrast estimator)

$$\operatorname{argmin}_{\vartheta\in\Theta} \sum_{k=0}^{n-1}\left|\delta Y_{t_k t_{k+1}} - b(Y_{t_k};\vartheta)\alpha_n\right|^2,$$

is considered, see e.g. Florens-Zmirou (1989), Kasonga (1988), Kessler (2000).

Remark 1.6 The difference in the limits (3) and (4) is due to the higher smoothness and long-range dependence of fractional Brownian motion for $H > 1/2$. In order to give an intuition of this fact, let us focus on the case $m = d = 1$ and $\sigma = 1$. Then first note that

$$\sum_{k=0}^{n-1}\left|\delta Y_{t_k t_{k+1}} - b(Y_{t_k};\vartheta)\alpha_n\right|^2 - \sum_{k=0}^{n-1}\left|\delta Y_{t_k t_{k+1}} - b(Y_{t_k};\vartheta_0)\alpha_n\right|^2$$
$$= -2\sum_{k=0}^{n-1}\left[b(Y_{t_k};\vartheta) - b(Y_{t_k};\vartheta_0)\right]\alpha_n \delta Y_{t_k t_{k+1}} + \sum_{k=0}^{n-1}\left[b(Y_{t_k};\vartheta)^2 - b(Y_{t_k};\vartheta_0)^2\right]\alpha_n^2.$$

Furthermore, up to higher order terms, Eq. (1) yields

$$\delta Y_{t_k t_{k+1}} \approx b(Y_{t_k};\vartheta_0)\alpha_n + \delta B_{t_k t_{k+1}}.$$

Inserting the above relations into the definition (2) of Q_n we end up with

$$Q_n(\vartheta) - Q_n(\vartheta_0) \approx \frac{1}{n}\sum_{k=0}^{n-1} |b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta)|^2 + \frac{2}{n\alpha_n}\sum_{k=0}^{n-1}(b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta))\delta B_{t_k t_{k+1}}. \tag{5}$$

Let us now separate the Brownian case from the situation where $H > 1/2$:

(a) When B is a Brownian motion, the independence of its increments gives

$$\mathbf{E}\left[Q_n(\vartheta) - Q_n(\vartheta_0)\right] \approx \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}\left[|b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta)|^2\right],$$

while our ergodicity result, i.e. Proposition 2.3, yields

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}\left[|b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta)|^2\right] \longrightarrow \mathbf{E}\left[|b(\overline{Y}_0;\vartheta_0) - b(\overline{Y}_0;\vartheta)|^2\right].$$

Plugging those two relations into (5), this illustrates why (4) holds true in the Brownian motion case.

(b) In contrast, for $H > 1/2$ we have

$$\mathbf{E}\left[Q_n(\vartheta) - Q_n(\vartheta_0)\right] \approx \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}\left[|b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta)|^2\right] + \frac{2}{n\alpha_n}\sum_{k=0}^{n-1} \mathbf{E}\left[(b(Y_{t_k};\vartheta_0) - b(Y_{t_k};\vartheta))\delta B_{t_k t_{k+1}}\right]. \tag{6}$$

Now the first term in the right hand side of (6) still converges to $\mathbf{E}[|b(\overline{Y}_0;\vartheta_0) - b(\overline{Y}_0;\vartheta)|^2]$, but the dependence structure of our driving fBm B induces some non-negligible correction terms which are reflected in our formula (3). An important part of our computations below is devoted to quantify those correction terms.

Let us now explain how to obtain an estimation procedure from Theorem 1.5. In the classical least square setting, one should minimize $Q_n(\vartheta)$, which is equivalent to minimizing $Q_n(\vartheta) - Q_n(\vartheta_0)$. Convergence of the minimizer to the unknown parameter is then guaranteed by Proposition 4.1 below. However, in our case the parameter ϑ_0 can only be seen as a minimizer of $|Q_\infty(\vartheta)|$ for which $|Q_\infty(\vartheta_0)| = 0$, under an additional assumption on the sampling step size α_n. This leads to a different estimation procedure. Furthermore, in order to ensure that the set of conditions in Proposition 4.1 is satisfied in our case, some additional constraints on our parameters and coefficients will be given below. The first one is a natural identifiability assumption similar to Assumption A6 in Kessler (2000):

Hypothesis 1.7 *For any $\vartheta_0 \in \Theta$, we have*

$$\mathbf{E}\,|b(\overline{Y}_0;\vartheta_0)|^2 = \mathbf{E}|b(\overline{Y}_0;\vartheta)|^2 \quad \textit{iff} \quad \vartheta = \vartheta_0.$$

The condition on the sampling size α_n we have mentioned above, which is required to control the contribution of the quadratic variation of the fractional Brownian motions, is the following:

Hypothesis 1.8 *We have* $0 < \alpha < \min\left\{\frac{1}{4}\frac{1}{1-H}, 1\right\}$.

Notice that Hypothesis 1.8 is only a mild restriction. Indeed, since we will work under the assumption $H > 1/2$, the choices $\alpha = 1/2$ or $\alpha = H$ are always possible. Note also that for $H > 3/4$ the above condition simply reads as $\alpha \in (0, 1)$, so is in fact no restriction.

With these additional hypotheses, the main result of the current article is the consistency of the least squares-type estimator based on the statistics $|Q_n|$:

Theorem 1.9 *Assume that the Hypotheses 1.1, 1.3, 1.4, 1.7 and 1.8 are satisfied for Eq. (1) and let* $H > 1/2$. *Let* $Q_n(\vartheta)$ *be defined by (2), and let* $\widehat{\vartheta}_n = \mathrm{argmin}_{\vartheta\in\Theta} |Q_n(\vartheta)|$. *Then for any* $\vartheta_0 \in \Theta$, *we have* $\lim_{n\to\infty} \widehat{\vartheta}_n = \vartheta_0$ *in the* $\mathbf{P}$*-almost sure sense.*

Let us shortly compare Theorem 1.9 with the existing literature on estimation procedures for fBm driven equations:

(i) Most of the previous results, see e.g. Belfadli et al. (2011), Hu and Nualart (2010), Kleptsyna and Le Breton (2002), Prakasa Rao (2010), deal with the one-dimensional fractional Ornstein-Uhlenbeck process in a continuous observation setting. In particular, for this process simple continuous time least-square estimators are obtained in Belfadli et al. (2011), Hu and Nualart (2010), for which also convergence rates and asymptotic error distributions are derived. Compared to these results our estimation procedure covers a broad class of ergodic multi-dimensional equations and relies on discrete data only.

(ii) A general estimation procedure based on moment matching is established in Papavasiliou and Ladroue (2011). However, the main assumption in Papavasiliou and Ladroue (2011) is that many independent observations of sample paths over a short time interval are available, which is not the case in many practical situations where rather one sample path is discretely observed for a long time period. Let us also mention the article Chronopoulou and Tindel (2011), in which a general discrete data maximum likelihood type procedure has been designed for parameter estimation in both the drift and diffusion coefficients, however without proof of consistency.

(iii) Our current work probably compares best with the maximum likelihood estimator analyzed in Tudor and Viens (2007). The latter pioneering reference focused on one-dimensional SDEs of the form

$$dY_t = \vartheta_0 h(Y_t)\,dt + dB_t$$

with $h : \mathbb{R} \to \mathbb{R}$ satisfying suitable regularity assumptions. Strong consistency is obtained for the continuous time estimator and also for a discretized version of the estimator. However, the discretized estimator involves rather complicated operators related to the kernel functions arising in the Wiener-integral representation of fBm, which are avoided in our approach. Moreover, in contrast to Tudor and Viens (2007) the consistency proof for our estimator does not rely on Malliavin calculus methods.

So, in view of the existing results in the literature, Theorem 1.9 can be seen a step towards simple and implementable parameter estimation procedures for SDEs driven by fBm. Note that in the case of the fractional Ornstein-Uhlenbeck process a central limit theorem similar to

Hu and Nualart (2010) could be obtained using the now classical tools for random variables in a finite Gaussian chaos, see e.g. Nualart and Peccati (2005). However, in the general case of a non-linear drift such a theorem remains an open question and would require first a central limit theorem version of the ergodicity result given in Proposition 2.3.

Finally, let us comment on the assumptions we have imposed on the drift coefficient and on the Hurst parameter:

(a) The hypotheses of Theorem 1.5 are standard for the case $H = 1/2$, except Hypothesis 1.4 which restricts us to gradient-type drift coefficients. We require this condition to show an ergodic-type result for weighted sums of the increments of fBm, see Lemma 3.4. However, this Hypothesis 1.4 is also implicitly present in the additional condition of Theorem 1 in Kasonga (1988).
(b) It can easily be shown that whenever ϑ is a one-dimensional coefficient (namely for $q = 1$), Hypothesis 1.7 is satisfied if the drift coefficient is of the form $b(x; \vartheta) = \vartheta h(x)$ for some $h : \mathbb{R}^d \to \mathbb{R}^d$ and the stationary solution is non-degenerate, i.e. we have $\mathbf{E}|\overline{Y}_0|^2 \neq 0$. The latter conditions hold in particular in the case of the ergodic fractional Ornstein-Uhlenbeck process. It would be nice to obtain criteria for richer classes of examples, but this would rely on differentiability and non-degeneracy properties of the map $\vartheta \mapsto \mathbf{E}|b(\overline{Y}_0; \vartheta)|^2$ (see Hairer and Majda 2010 in the Markovian case). We wish to investigate this question in future works.
(c) Even if the noise enters additively in our equation, we still need the assumption $H > 1/2$ in order to prove Theorem 1.9. Indeed, this hypothesis ensures the convergence of some deterministic and stochastic Riemann sums in the computations below (see Remark 3.5 for further details). Whether an adaptation of the proposed zero squares estimator is also convergent in the case $H < 1/2$ remains an open problem.

Let us finish this introduction with the simplest example of an equation which satisfies the above assumptions, namely the one-dimensional fractional Ornstein-Uhlenbeck process:

Proposition 1.10 *Consider the solution Y to the linear equation*

$$dY_t = \vartheta_0 Y_t \, dt + dB_t, \quad Y_0 = y_0 \in \mathbb{R}, \tag{7}$$

with $\vartheta_0 < 0$. Then for $n \geq n_0(\omega)$ large enough, the least square-type estimator of ϑ takes the form

$$\widehat{\vartheta}_n = \frac{s_n^{(2)}}{s_n^{(3)}} - \sqrt{\left(\frac{s_n^{(2)}}{s_n^{(3)}}\right)^2 - \frac{s_n^{(1)}}{s_n^{(3)}}},$$

where

$$s_n^{(1)} = \sum_{k=0}^{n-1} \left(|\delta Y_{t_k t_{k+1}}|^2 - \alpha_n^{2H}\right), \quad s_n^{(2)} = \sum_{k=0}^{n-1} Y_{t_k} \delta Y_{t_k t_{k+1}} \alpha_n, \quad s_n^{(3)} = \sum_{k=0}^{n-1} |Y_{t_k}|^2 \alpha_n^2. \tag{8}$$

This estimator is consistent.

Clearly, this estimator is a simple function of the observations, analogously to Tudor and Viens (2007), Hu and Nualart (2010). Moreover, in the proof of the above result we will see that our estimator is in general only asymptotically unique.

The remainder of this article is structured as follows: In Sect. 2 we give some auxiliary results on stochastic calculus for fractional Brownian motion. Sections 3 and 4 are then devoted to the proof of our main theorems.

2 Auxiliary results

2.1 Ergodic properties of the SDE

To deduce the ergodic properties of SDE (1) we will use the theory of random dynamical systems, see Arnold (1997). We will work without loss of generality on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$, i.e. $\Omega = C_0(\mathbb{R}, \mathbb{R}^m)$ equipped with the compact open topology, $\mathcal{F}$ is the corresponding Borel-σ-algebra and $\mathbf{P}$ is the distribution of the fractional Brownian motion B, which is consequently given here by the canonical process $B_t(\omega) = \omega(t)$, $t \in \mathbb{R}$. Together with the shift operators $\theta_t : \Omega \to \Omega$ defined by

$$\theta_t\omega(\cdot) = \omega(\cdot + t) - \omega(t), \quad t \in \mathbb{R}, \quad \omega \in \Omega,$$

the canonical probability space is an ergodic metric dynamical system, see e.g. Garrido-Atienza and Schmalfuss (2011). In particular, the measure $\mathbf{P}$ is invariant to the shift operators θ_t, i.e. the shifted process $(B_s(\theta_t\cdot))_{s\in\mathbb{R}}$ is still an m-dimensional fractional Brownian motion and for any integrable random variable $F : \Omega \to \mathbb{R}$ we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T F(\theta_t(\omega))\, dt = \mathbf{E}[F],$$

for $\mathbf{P}$-almost all $\omega \in \Omega$.

These ergodic properties of fractional Brownian motion can be seen as a time-continuous extension of the ergodicity of fractional Gaussian noise: For

$$\delta B_{nn+1}(\omega) = B_{n+1}(\omega) - B_n(\omega) = B_1(\theta_n\omega), \quad \omega \in \Omega, \quad n = 0, 1, \ldots$$

we have

$$\mathbf{E}\big[\delta B_{kk+1}\delta B_{\ell\ell+1}\big] = \frac{1}{2}\left(|k-l+1|^{2H} + |k-l-1|^{2H} - 2|k-l|^{2H}\right),$$

so $(\delta B_{nn+1})_{n\in\mathbb{N}}$ is stationary. Moreover, since

$$\rho(k) = \frac{1}{2}\left(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}\right)$$

satisfies

$$\lim_{k\to\infty} \frac{\rho(k)}{H(2H-1)k^{2H-2}} = 1,$$

the ergodicity of the fractional Gaussian noise $(\delta B_{nn+1})_{n\in\mathbb{N}}$ is a consequence of a classical criterion for stationary Gaussian sequences, see e.g. chapter 5 in Shiryaev (1995). In this context Birkhoff's ergodic theorem reads as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(\delta B_{nn+1}(\omega)) = \mathbf{E}[f(B_1)]$$

for $\mathbf{P}$-almost all $\omega \in \Omega$ and any measurable $f : \mathbb{R} \to \mathbb{R}$ such that $f(B_1)$ is integrable.

Owing to the results in Section 4 of Garrido-Atienza et al. (2009) we have:

Theorem 2.1 *Let Hypothesis 1.3 hold. Then for any $\vartheta_0 \in \Theta$ we have the following:*

(i) Equation (1) admits a unique solution Y in $C^\lambda(\mathbb{R}_+; \mathbb{R}^d)$ for all $\lambda < H$.

(ii) There exists a random variable $\overline{Y} : \Omega \to \mathbb{R}^d$ *such that*

$$\lim_{t\to\infty} |Y_t(\omega) - \overline{Y}(\theta_t\omega)| = 0$$

for $\mathbf{P}$*-almost all* $\omega \in \Omega$*. Moreover, we have* $\mathbf{E}|\overline{Y}|^p < \infty$ *for all* $p \geq 1$.

Note that the law of $\overline{Y}$ must coincide with the attracting invariant measure for (1) given in Hairer (2005), see also Hairer and Ohashi (2007), Hairer and Pillai (2011).

To illustrate the above result consider the one-dimensional SDE

$$Y_t = y_0 + \int_0^t f(Y_s)ds + B_t, \quad t \geq 0,$$

where $f : \mathbb{R} \to \mathbb{R}$ is of polynomial growth and one-sided dissipative Lipschitz with constant $c_1 > 0$, i.e.

$$(x - y)(f(x) - f(y)) \leq -c_1|x - y|^2, \quad x, y \in \mathbb{R}.$$

Now let $Y^{(1)}$ and $Y^{(2)}$ be the solutions of the above SDE corresponding to the initial values $y_0^{(1)}$ and $y_0^{(2)}$. Their difference satisfies

$$Y_t^{(1)} - Y_t^{(2)} = y_0^{(1)} - y_0^{(2)} + \int_0^t (f(Y_s^{(1)}) - f(Y_s^{(2)}))\, ds, \quad t \geq 0,$$

and differentiation yields

$$\frac{d}{dt}(Y_t^{(1)} - Y_t^{(2)}) = f(Y_t^{(1)}) - f(Y_t^{(2)}), \quad t \geq 0.$$

The inward condition now gives

$$\frac{d}{dt}\left|Y_t^{(1)} - Y_t^{(2)}\right|^2 = 2\left\langle Y_t^{(1)} - Y_t^{(2)}, f(Y_t^{(1)}) - f(Y_t^{(2)})\right\rangle \leq -2c_1\left|Y_t^{(1)} - Y_t^{(2)}\right|^2$$

and so

$$\left|Y_t^{(1)} - Y_t^{(2)}\right| \leq \left|y_0^{(1)} - y_0^{(2)}\right| e^{-c_1 t}.$$

Thus solutions with different initial conditions converge exponentially pathwise to each other as $t \to \infty$.

The convergence of $(Y_t)_{t\geq 0}$ to a stationary solution $(\overline{Y}_t)_{t\geq 0}$, i.e. $\overline{Y}_t(\omega) := \overline{Y}(\theta_t\omega)$, $t \geq 0$, $\omega \in \Omega$, relies on the concept of pullback absorption. Once pullback absorption is established one obtains the existence of a pullback attractor, see e.g. Arnold (1997), Crauel et al. (1997). Due to the pathwise forward convergence this attractor is the desired stationary solution. For pullback absorption we have to analyse the behavior of

$$\lim_{t\to\infty} Y_t(\theta_{-t}\omega),$$

where Y is the solution of our SDE, and we have to find—roughly spoken—an appropriate random set $D(\omega)$ such that

$$Y_t(\theta_{-t}\omega) \subset D(\omega)$$

for t sufficiently large and arbitrary initial values of the SDE. For more details see Arnold (1997), Crauel et al. (1997).

Using again the inward condition on the drift one obtains that

$$\left|Y_t - \overline{O}_t\right| \le \left|Y_0 - \overline{O}_0\right| e^{-c_1 t} + e^{-c_1 t} \int_0^t e^{c_1 s} \left(\left|f(\overline{O}_s)\right| + \left|\overline{O}_s\right|\right) ds,$$

where

$$\overline{O}_t = e^{-t} \int_{-\infty}^t e^s dB_s, \quad t \in \mathbb{R},$$

is a stationary fractional Ornstein-Uhlenbeck process, i.e. the stationary solution of

$$dO_t = -O_t dt + dB_t, \quad t \ge 0.$$

So setting

$$R(\omega) := 1 + \int_{-\infty}^0 e^{c_1 s} \left(\left|f(\overline{O}_s(\omega))\right| + \left|\overline{O}_s(\omega)\right|\right) ds$$

one has the desired pullback attraction

$$|X_t(\theta_{-t}\omega)| \le \left|\overline{O}_0(\omega)\right| + R(\omega)$$

for $t \ge t_0(\omega)$. For more details we refer to Garrido-Atienza et al. (2009).

Exploiting the integrability properties of the stationary fractional Ornstein-Uhlenbeck process we have:

Proposition 2.2 *Assume Hypothesis 1.3 holds true. Then for any $\vartheta_0 \in \Theta$ and $p \ge 1$ there exist constants $c_p, k_p > 0$ such that*

$$\mathbf{E}\,|Y_t|^p \le c_p, \qquad \mathbf{E}\,|Y_t - Y_s|^p \le k_p |t-s|^{pH}, \quad \textit{for all} \quad s, t \ge 0.$$

The integrability of $\overline{Y}$ now implies the ergodicity of Eq. (1):

Proposition 2.3 *Assume Hypothesis 1.3 holds true. Then for any $\vartheta_0 \in \Theta$ and any $f \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R})$ such that*

$$|f(x)| + |\partial_x f(x)| \le c\left(1 + |x|^N\right), \quad x \in \mathbb{R}^d,$$

for some $c > 0$, $N \in \mathbb{N}$, we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(Y_t)\, dt = \mathbf{E} f(\overline{Y}) \qquad \mathbf{P}\text{-}a.s. \tag{9}$$

Proof Since the shift operator is ergodic and f has polynomial growth, we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\overline{Y}(\theta_t))\, dt = \mathbf{E} f(\overline{Y}) \quad \mathbf{P}\text{-}a.s.$$

Moreover, Theorem 2.1 yields

$$\lim_{t\to\infty} |Y_t(\omega) - \overline{Y}(\theta_t\omega)| = 0,$$

and since f is polynomially Lipschitz, our assertion (9) easily follows. □

2.2 Generalized Riemann-Stieltjes integrals

We set

$$\|f\|_{\infty;[a,b]} = \sup_{t\in[a,b]} |f(t)|, \qquad |f|_{\lambda;[a,b]} = \sup_{s,t\in[a,b]} \frac{|f(t)-f(s)|}{|t-s|^{\lambda}}$$

where $f:\mathbb{R}\to\mathbb{R}^n$ and $\lambda\in(0,1)$.

Now, let $f\in C^{\lambda}([a,b];\mathbb{R})$ and $g\in C^{\mu}([a,b];\mathbb{R})$ with $\lambda+\mu>1$. Then it is well known that the Riemann-Stieltjes integral $\int_a^b f(x)\,dg(x)$ exists, see e.g. Young (1936). Also, the classical chain rule for the change of variables remains valid, see e.g. Zähle (2005): Let $f\in C^{\lambda}([a,b];\mathbb{R})$ with $\lambda>1/2$ and $F\in C^1(\mathbb{R};\mathbb{R})$. Then we have

$$F(f(y)) - F(f(a)) = \int_a^y F'(f(x))\,df(x), \quad y\in[a,b]. \tag{10}$$

Moreover, one has a density type formula: let $f,h\in C^{\lambda}([a,b];\mathbb{R})$ and $g\in C^{\mu}([a,b];\mathbb{R})$ with $\lambda+\mu>1$. Then for

$$\varphi:[a,b]\to\mathbb{R}, \quad \varphi(y) = \int_a^y f(x)\,dg(x), \quad y\in[a,b],$$

we have

$$\int_a^b h(x)\,d\varphi(x) = \int_a^b h(x)f(x)\,dg(x). \tag{11}$$

For later use, we also note the following estimate, which can be found e.g. in Young (1936).

Proposition 2.4 *Let f,g be as above. Then, there exists a constant $c_{\lambda,\mu}$ (independent of a,b) such that*

$$\left|\int_a^b (f(s)-f(a))dg(s)\right| \le c_{\lambda,\mu}|f|_{\lambda;[a,b]}|g|_{\mu;[a,b]}|b-a|^{\lambda+\mu}$$

holds for all $a,b\in[0,\infty)$.

2.3 Some limit theorems

We include here some general analytic and probabilistic tools which will be crucial for the proof of Theorem 1.5. Let us start by the following variant of the Garcia-Rodemich-Rumsey Lemma (Garcia et al. 1978):

Lemma 2.5 *Let $q>1$, $\alpha\in(1/q,1)$ and $f:[0,\infty)\to\mathbb{R}$ be a continuous function. Then there exists a constant $c_{\alpha,q}>0$, depending only on α,q, such that*

$$|f|^q_{\alpha-1/q;[s,t]} \le c_{\alpha,q}\int_s^t\int_s^t \frac{|f(u)-f(v)|^q}{|u-v|^{1+q\alpha}}\,du\,dv.$$

The following Lemma (see e.g. Kloeden and Neuenkirch 2007), which is a direct consequence of the Borel-Cantelli Lemma, allows us to turn convergence rates in the pth mean into pathwise convergence rates.

Lemma 2.6 *Let $\alpha > 0$, $p_0 \in \mathbb{N}$ and $c_p \in [0, \infty)$ for $p \geq p_0$. In addition, let Z_n, $n \in \mathbb{N}$, be a sequence of random variables such that*

$$(\mathbf{E}|Z_n|^p)^{1/p} \leq c_p \cdot n^{-\alpha}$$

for all $p \geq p_0$ and all $n \in \mathbb{N}$. Then for all $\varepsilon > 0$ there exists a random variable η_ε such that

$$|Z_n| \leq \eta_\varepsilon \cdot n^{-\alpha+\varepsilon} \qquad a.s.$$

for all $n \in \mathbb{N}$. Moreover, $\mathbf{E}|\eta_\varepsilon|^p < \infty$ for all $p \geq 1$.

Finally, we shall need the following well known result (see e.g. Tudor and Viens 2009) for the behavior of the quadratic variations of a one-dimensional fractional Brownian motion.

Proposition 2.7 *Let β be a fractional Brownian motion with Hurst parameter H, and set $\delta_{kk+1}\beta = \beta_{k+1} - \beta_k$. Then for $H < 3/4$ we have*

$$\lim_{n\to\infty} \mathbf{E}\left|\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left[|\delta_{kk+1}\beta|^2 - 1\right]\right|^2 = c_H, \tag{12}$$

while for $H = \frac{3}{4}$ it holds

$$\lim_{n\to\infty} \mathbf{E}\left|\frac{1}{\sqrt{n\log(n)}}\sum_{k=0}^{n-1}\left[|\delta_{kk+1}\beta|^2 - 1\right]\right|^2 = c_{3/4}. \tag{13}$$

Finally, if $H \in (\frac{3}{4}, 1)$ then we have

$$\lim_{n\to\infty} \mathbf{E}\left|\frac{1}{n^{2H-1}}\sum_{k=0}^{n-1}\left[|\delta_{kk+1}\beta|^2 - 1\right]\right|^2 = c_H. \tag{14}$$

In the above relations $c_H > 0$ denotes a constant depending only on H.

3 Proof of Theorem 1.5

This section is devoted to the proof of our main result Theorem 1.5. In the sequel, we denote constants, whose particular value is not important (and which do not depend on ϑ or n) by c, regardless of their value. Before we start with our computations, we will define a useful notation:

Notation 3.1 With the conventions of Sect. 1, we set $F_t = \sum_{j=1}^m \sigma_j B_t^j$ and

$$\delta F_{t_k t_{k+1}} = F_{t_{k+1}} - F_{t_k}. \tag{15}$$

Moreover, we set

$$\delta_{\vartheta_0\vartheta} b(x) = b(x; \vartheta) - b(x; \vartheta_0). \tag{16}$$

We now start by reducing the limiting behavior of $Q_n(\vartheta)$ to the study of two easier terms:

Lemma 3.2 *Let $Q_n(\vartheta)$ be the quantity defined by (2). Then we have*

$$Q_n(\vartheta) - Q_n(\vartheta_0) = Q_n^{(1)}(\vartheta) - 2Q_n^{(2)}(\vartheta) + R_n(\vartheta), \tag{17}$$

where

$$\lim_{n\to\infty} \sup_{\vartheta\in\Theta} |R_n(\vartheta)| = 0 \qquad \mathbf{P}\text{-}a.s.$$

and $Q_n^{(1)}(\vartheta)$, $Q_n^{(2)}(\vartheta)$ are given by

$$Q_n^{(1)}(\vartheta) = \frac{1}{n}\sum_{k=0}^{n-1} |\delta_{\vartheta_0\vartheta} b(Y_{t_k})|^2, \qquad Q_n^{(2)}(\vartheta) = \frac{1}{n\alpha_n}\sum_{k=0}^{n-1} \langle \delta_{\vartheta_0\vartheta} b(Y_{t_k}), \delta F_{t_k t_{k+1}}\rangle. \tag{18}$$

Proof Analogously to the one-dimensional case highlighted in the introduction, we can write

$$\begin{aligned} n\alpha_n^2\,[Q_n(\vartheta) - Q_n(\vartheta_0)] &= \sum_{k=0}^{n-1} \left|\delta Y_{t_k t_{k+1}} - b(Y_{t_k};\vartheta)\alpha_n\right|^2 - \sum_{k=0}^{n-1} \left|\delta Y_{t_k t_{k+1}} - b(Y_{t_k};\vartheta_0)\alpha_n\right|^2 \\ &= -2\sum_{k=0}^{n-1} \langle b(Y_{t_k};\vartheta) - b(Y_{t_k};\vartheta_0), \delta Y_{t_k t_{k+1}}\rangle\alpha_n + \sum_{k=0}^{n-1}\Big(|b(Y_{t_k};\vartheta)|^2 - |b(Y_{t_k};\vartheta_0)|^2\Big)\alpha_n^2. \end{aligned} \tag{19}$$

Recalling our notation (15) and setting $r_k = \int_{t_k}^{t_{k+1}} \big(b(Y_u;\vartheta_0) - b(Y_{t_k};\vartheta_0)\big)\, du$, Eq. (1) easily yields

$$\delta Y_{t_k t_{k+1}} = \delta F_{t_k t_{k+1}} + b(Y_{t_k};\vartheta_0)\alpha_n + r_k.$$

Hence, using notation (16) for $\delta_{\vartheta_0\vartheta} b(x)$, it is readily checked from (19) that relation (17) holds true, with

$$R_n(\vartheta) = -\frac{2}{n\alpha_n}\sum_{k=0}^{n-1} \langle \delta_{\vartheta_0\vartheta} b(Y_{t_k}), r_k\rangle.$$

It now remains to prove that R_n is a negligible term. To this aim, note that our assumptions on the drift coefficient imply that

$$\sup_{\vartheta\in\Theta} |b(x;\vartheta) - b(y;\vartheta)| \le c\big(1 + |x|^N + |y|^N\big)\cdot|x-y|$$

for all $x, y \in \mathbb{R}^d$ and

$$|b(x;\vartheta_1) - b(x;\vartheta_2)| \le c\big(1 + |x|^N\big)\cdot|\vartheta_1 - \vartheta_2|$$

for all $x \in \mathbb{R}^d$ and $\vartheta_1, \vartheta_2 \in \Theta$. So, straightforward estimations using Proposition 2.2 give

$$\mathbf{E}|r_k|^p \le c\cdot\alpha_n^{p(1+H)}.$$

Hence for all $p \ge 1$ it holds

$$\mathbf{E}\left|\sup_{\vartheta\in\Theta} \frac{1}{n\alpha_n^2}\sum_{k=0}^{n-1} \langle \delta_{\vartheta_0\vartheta} b(Y_{t_k}), r_k\rangle\alpha_n\right|^p \le c\cdot\alpha_n^{pH},$$

so Lemma 2.6 implies

$$\lim_{n\to\infty}\sup_{\vartheta\in\Theta}\frac{2}{n\alpha_n^2}\left|\sum_{k=0}^{n-1}\langle\delta_{\vartheta_0\vartheta}b(Y_{t_k}),r_k\rangle\alpha_n\right|=0 \qquad \mathbf{P}\text{-}a.s. \tag{20}$$

which finishes our proof. □

According to Lemma 3.2, our limit theorem can be reduced to determine the behavior of the terms $Q_n^{(1)}(\vartheta)$ and $Q_n^{(2)}(\vartheta)$. This task will be carried out in the following two Lemmata. We first show a discrete version of Proposition 2.3:

Lemma 3.3 *Let $f\in\mathcal{C}^{1,1}(\mathbb{R}^d\times\Theta;\mathbb{R}^d)$ be a function such that*

$$|f(x;\vartheta)|\le c\left(1+|x|^N\right),\quad |\partial_x f(x;\vartheta)|\le c\left(1+|x|^N\right),\quad |\partial_\vartheta f(x;\vartheta)|\le c\left(1+|x|^N\right)$$

for some $c>0$, $N\in\mathbb{N}$, independent of $\vartheta\in\Theta$. Then we have

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{n}\sum_{k=0}^{n-1}|f(Y_{t_k};\vartheta)|^2-\mathbf{E}|f(\overline{Y};\vartheta)|^2\right|\to 0 \;\; \mathbf{P}\text{-}a.s.$$

In particular, we have

$$\sup_{\vartheta\in\Theta}\left|Q_n^{(1)}(\vartheta)-\mathbf{E}|\delta_{\vartheta_0\vartheta}b(\overline{Y})|^2\right|\to 0 \;\; \mathbf{P}\text{-}a.s.,$$

where $Q_n^{(1)}(\vartheta)$ is defined by (18).

Proof Let $T_n=n\alpha_n$ and set

$$V_n(\vartheta)=\frac{1}{T_n}\int_0^{T_n}|f(Y_s;\vartheta)|^2\,ds.$$

The ergodicity of Y yields that there exists a set $A_1\in\mathcal{F}$ with full measure such that

$$\lim_{n\to\infty}V_n(\vartheta)(\omega)=\mathbf{E}|f(\overline{Y};\vartheta)|^2$$

for all $\vartheta\in\Theta\cap\mathbb{Q}^q$ and all $\omega\in A_1$. The assumptions on f give

$$|V_n(\vartheta_1)-V_n(\vartheta_2)|\le c\cdot\left(1+\frac{1}{T_n}\int_0^{T_n}|Y_s|^{2N}\,ds\right)\cdot|\vartheta_1-\vartheta_2|, \tag{21}$$

so V_n is Lipschitz continuous in ϑ and thus

$$\sup_{\vartheta\in\Theta}\left|V_n(\vartheta)-\mathbf{E}|f(\overline{Y};\vartheta)|^2\right|=\sup_{\vartheta\in\Theta\cap\mathbb{Q}^q}\left|V_n(\vartheta)-\mathbf{E}|f(\overline{Y};\vartheta)|^2\right|.$$

However, from (21) and the ergodicity of Y, it also follows that there exists a set $A_2\in\mathcal{F}$ with $\mathbf{P}(A_2)=1$ in which the family of random functions $V_n:\Theta\to\mathbb{R}$, $n\in\mathbb{N}$, is equicontinuous, and hence the Arzela–Ascoli Theorem yields the desired uniform convergence, i.e.

$$\lim_{n\to\infty}\sup_{\vartheta\in\Theta}\left|V_n(\vartheta)-\mathbf{E}|f(\overline{Y};\vartheta)|^2\right|=0 \qquad \mathbf{P}\text{-}a.s. \tag{22}$$

Setting

$$G_n(t;\vartheta) = |f(Y_t;\vartheta)|^2 - |f(Y_{t_k};\vartheta)|^2, \quad t \in [t_k, t_{k+1}), \quad k = 0, 1, \ldots,$$

it remains to show that

$$\frac{1}{T_n}\int_0^{T_n} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|\, dt \to 0 \quad \mathbf{P}\text{-}a.s.$$

To this aim, the assumptions on f imply that

$$\sup_{\vartheta\in\Theta} |G_n(t;\vartheta)| \leq c \cdot (1 + Y_t^{2N} + Y_{t_k}^{2N}) \cdot |Y_t - Y_{t_k}|.$$

Using Proposition 2.2 and Hölder's inequality we obtain

$$\sup_{t\geq 0} \mathbf{E} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|^p \leq c \cdot \alpha_n^{pH} \tag{23}$$

for all $p \geq 1$. Now, Jensen's inequality gives

$$\mathbf{E}\left|\frac{1}{T_n}\int_0^{T_n} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|\, dt\right|^p \leq \frac{1}{T_n}\int_0^{T_n} \mathbf{E} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|^p\, dt,$$

and so (23) yields

$$\mathbf{E}\left|\frac{1}{T_n}\int_0^{T_n} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|\, dt\right|^p \leq c \cdot \alpha_n^{pH}$$

for all $p \geq 1$. Lemma 2.6 implies

$$\frac{1}{T_n}\int_0^{T_n} \sup_{\vartheta\in\Theta} |G_n(t;\vartheta)|\, dt \to 0 \qquad \mathbf{P}\text{-}a.s.$$

for $n \to \infty$. □

We now state a similar ergodic result for weighted sums of the increments of the process F defined at Notation 3.1, which yields the convergence of our term $Q_n^{(2)}(\vartheta)$.

Lemma 3.4 *Let $f \in \mathcal{C}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ be a function such that*

$$|f(x;\vartheta)| \leq c\left(1 + |x|^N\right), \quad |\partial_x f(x;\vartheta)| \leq c\left(1 + |x|^N\right), \quad |\partial_\vartheta f(x;\vartheta)| \leq c\left(1 + |x|^N\right)$$

for some $c > 0$, $N \in \mathbb{N}$, independent of $\vartheta \in \Theta$. Assume moreover that there exists a function $U \in \mathcal{C}^{2,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ such that

$$\partial_x U(x;\vartheta) = f(x;\vartheta), \quad x \in \mathbb{R}^d, \ \vartheta \in \Theta,$$

i.e. f is of gradient type. Then, for $H > 1/2$, we have

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{n\alpha_n}\sum_{k=0}^{n-1}\langle f(Y_{t_k};\vartheta), \delta F_{t_k t_{k+1}}\rangle + \mathbf{E}\langle b(\overline{Y};\vartheta_0), f(\overline{Y};\vartheta)\rangle\right| \to 0 \qquad \mathbf{P}\text{-}a.s.$$

In particular,

$$\sup_{\vartheta\in\Theta}\left|Q_n^{(2)}(\vartheta)+\mathbf{E}\langle b(\overline{Y};\vartheta_0),\delta_{\vartheta_0\vartheta}b(\overline{Y})\rangle\right|\to 0\quad \mathbf{P}\text{-}a.s.$$

Proof Let $T_n = n\alpha_n$. First note that the change of variable and density formulae for Riemann-Stieltjes integrals, see (10) and (11) in Sect. 2.2, give that

$$\frac{1}{T_n}\left(U(Y_{T_n};\vartheta)-U(y_0;\vartheta)\right)=\frac{1}{T_n}\int_0^{T_n}\langle f(Y_u;\vartheta),b(Y_u;\vartheta_0)\rangle\,du+\frac{1}{T_n}\int_0^{T_n}\langle f(Y_u;\vartheta),dF_u\rangle.$$

Now the properties of f, Proposition 2.2 and Lemma 2.6 imply that

$$\sup_{\vartheta\in\Theta}\frac{1}{T_n}\left|U(Y_{T_n};\vartheta)-U(y_0;\vartheta)\right|\to 0\qquad \mathbf{P}\text{-}a.s.$$

Moreover, we have

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{T_n}\int_0^{T_n}\langle f(Y_u;\vartheta),b(Y_u;\vartheta_0)\rangle\,du-\mathbf{E}\langle f(\overline{Y};\vartheta),b(\overline{Y};\vartheta_0)\rangle\right|\to 0\qquad \mathbf{P}\text{-}a.s.,$$

which can be derived completely analogously to (22). It follows

$$\sup_{\vartheta\in\Theta}\left|\frac{1}{T_n}\int_0^{T_n}\langle f(Y_u;\vartheta),dF_u\rangle+\mathbf{E}\langle f(\overline{Y};\vartheta),b(\overline{Y};\vartheta_0)\rangle\right|\to 0\quad \mathbf{P}\text{-}a.s.$$

So, it remains to show that

$$\sup_{\vartheta\in\Theta}\frac{1}{T_n}\left|\int_0^{T_n}\langle G_n(t;\vartheta),dF_t\rangle\right|\to 0\quad \mathbf{P}\text{-}a.s. \tag{24}$$

where

$$G_n(t;\vartheta)=f(Y_t;\vartheta)-f(Y_{t_k};\vartheta),\quad t\in[t_k,t_{k+1}),\quad k=0,1,\ldots.$$

Applying Proposition 2.4 and using the polynomial Lipschitz continuity of f yields, for all $\lambda\in(1/2,H)$,

$$\left|\int_0^{T_n}\langle G_n(t;\vartheta),dF_t\rangle\right|\le c\cdot\alpha_n^{2\lambda}\cdot\sum_{j=1}^m\sum_{k=0}^{n-1}\sup_{t\in[t_k,t_{k+1}]}(1+|Y_t|^N)|Y|_{\lambda;[t_k;t_{k+1}]}|B^j|_{\lambda;[t_k;t_{k+1}]}.$$

From the Garcia-Rodemich-Rumsey inequality, see Lemma 2.5, and Proposition 2.2 we have that

$$\left(\mathbf{E}|Y|^p_{\lambda;[t_k;t_{k+1}]}\right)^{1/p}\le c\cdot\alpha_n^{H-\lambda}$$

and also

$$\left(\mathbf{E}|B^j|^p_{\lambda;[t_k;t_{k+1}]}\right)^{1/p}\le c\cdot\alpha_n^{H-\lambda}.$$

Since moreover

$$\sup_{t\in[t_k,t_{k+1}]} (1+|Y_t|^N) \leq c\cdot \left(Y_{t_k}^N + \alpha_n^{\lambda N}\cdot |Y|_{\lambda;[t_k,t_{k+1}]}^N\right)$$

and $\sup_{t\geq 0} \mathbf{E}|Y_t|^p < \infty$ for all $p \geq 1$, it follows that

$$\left(\mathbf{E}\sup_{\vartheta\in\Theta}\left|\frac{1}{T_n}\int_0^{T_n}\langle G_n(t;\vartheta), dF_t\rangle\right|^p\right)^{1/p} \leq c\cdot\alpha_n^{2H-1}. \tag{25}$$

Now Lemma 2.6 implies (24), since $H > 1/2$. □

We can now turn to the main aim of this section, namely:

Proof of Theorem 1.5 As asserted by Lemma 3.2, we have

$$\lim_{n\to\infty}(Q_n(\vartheta) - Q_n(\vartheta_0)) = \lim_{n\to\infty}(Q_n^{(1)}(\vartheta) - 2Q_n^{(2)}(\vartheta)),$$

in the $\mathbf{P}$-almost sure sense, and uniformly in ϑ. Furthermore, combining Lemmata 3.3 and 3.4, we obtain

$$\begin{aligned}\lim_{n\to\infty}(Q_n^{(1)}(\vartheta) - 2Q_n^{(2)}(\vartheta)) &= \mathbf{E}\left[|\delta_{\vartheta_0\vartheta} b(\overline{Y})|^2\right] + 2\,\mathbf{E}\left[\langle b(\overline{Y};\,\vartheta_0), \delta_{\vartheta_0\vartheta} b(\overline{Y})\rangle\right]\\ &= \mathbf{E}\left[|b(\overline{Y};\,\vartheta)|^2\right] - \mathbf{E}\left[|b(\overline{Y};\,\vartheta_0)|^2\right],\end{aligned}$$

uniformly in ϑ, which is our claim in Theorem 1.5. □

Remark 3.5 As mentioned in the introduction, the condition $H > 1/2$ is used in our proofs. Specifically, it is used to derive (25).

4 Proof of Theorem 1.9 and Proposition 1.10

We now turn to the proof of our estimation results, namely Theorem 1.9 and its application to Ornstein-Uhlenbeck processes given at Proposition 1.10. These results will be based on the following general proposition borrowed from Frydman (1980), Kasonga (1988):

Proposition 4.1 *Assume that the family of random variables $L_n(\vartheta)$, $n\in\mathbb{N}$, $\vartheta\in\Theta$, satisfies:*

(1) With probability one, $L_n(\vartheta)\to L(\vartheta)$ uniformly in $\vartheta\in\Theta$ as $n\to\infty$.
(2) The limit L is non-random and $L(\vartheta_0)\leq L(\vartheta)$ for all $\vartheta\in\Theta$.
(3) It holds $L(\vartheta) = L(\vartheta_0)$ if and only if $\vartheta = \vartheta_0$.

Then, we have

$$\mathbf{P}\text{-}a.s.\text{-}\lim_{n\to\infty}\widehat{\vartheta}_n = \vartheta_0 \quad \textit{where} \quad L_n(\widehat{\vartheta}_n) = \min_{\vartheta\in\Theta} L_n(\vartheta).$$

In order to apply Proposition 4.1, we now show that $Q_n(\vartheta_0)\to 0$ for $n\to\infty$.

Lemma 4.2 *Let Q_n be the quantity defined by (2). Then under the assumptions of Theorem 1.9 we have $\lim_{n\to\infty} Q_n(\vartheta_0) = 0$ $\mathbf{P}$-almost surely.*

Proof Recall that

$$Q_n(\vartheta_0) = \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} \left(\left| \delta Y_{t_k t_{k+1}} - b(Y_{t_k}; \vartheta_0)\alpha_n \right|^2 - \|\sigma\|^2 \alpha_n^{2H} \right).$$

Using our Notation 3.1 for $\delta F_{t_k t_{k+1}}$ and recalling that

$$r_k = \int_{t_k}^{t_{k+1}} (b(Y_u; \vartheta_0) - b(Y_{t_k}; \vartheta_0))\, du,$$

we have

$$Q_n(\vartheta_0) = \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} \left(|\delta F_{t_k t_{k+1}}|^2 - \|\sigma\|^2 \alpha_n^{2H} \right) + \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} |r_k|^2 + \frac{2}{n\alpha_n^2} \sum_{k=0}^{n-1} \langle \delta F_{t_k t_{k+1}}, r_k \rangle. \tag{26}$$

Since

$$\mathbf{E}|r_k|^p \leq c \cdot \alpha_n^{p(1+H)}$$

for all $p \geq 1$, it holds

$$\mathbf{E}\left| \sum_{k=0}^{n-1} |r_k|^2 \right|^p \leq c \cdot n^p \alpha_n^{2p(1+H)},$$

and Lemma 2.6 implies

$$\lim_{n\to\infty} \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} |r_k|^2 = 0 \quad \mathbf{P}\text{-}a.s. \tag{27}$$

Using Proposition 2.2, Lemma 2.6 and the fact that $H > 1/2$, it follows similarly

$$\lim_{n\to\infty} \frac{2}{n\alpha_n^2} \left| \sum_{k=0}^{n-1} \langle \delta F_{t_k t_{k+1}}, r_k \rangle \right| = 0 \quad \mathbf{P}\text{-}a.s. \tag{28}$$

Plugging (27) and (28) into (26), we thus get $\lim_{n\to\infty} Q_n(\vartheta_0) = \lim_{n\to\infty} Q_n^{(3)}$, where $Q_n^{(3)}$ is defined by

$$Q_n^{(3)} = \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} \left(|\delta F_{t_k t_{k+1}}|^2 - \|\sigma\|^2 \alpha_n^{2H} \right).$$

We will show in Lemma 4.3 that $\lim_{n\to\infty} Q_n^{(3)} = 0$, which finishes our proof. □

Lemma 4.3 *Let* $\alpha < \min\left\{\frac{1}{4(1-H)}, 1\right\}$. *We have*

$$\lim_{n\to\infty} Q_n^{(3)} = \lim_{n\to\infty} \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} \left(|\delta F_{t_k t_{k+1}}|^2 - \|\sigma\|^2 \alpha_n^{2H} \right) = 0 \qquad \mathbf{P}\text{-}a.s.$$

with $\|\sigma\|^2 = \sum_{j=1}^m |\sigma_j|^2$.

Proof We can decompose $|\delta F_{t_k t_{k+1}}|^2 - \|\sigma\|^2 \alpha_n^{2H}$ as $I_k^{(1)} + I_k^{(2)}$, where

$$I_k^{(1)} = \sum_{j=1}^{m} |\sigma_j|^2 \left(|\delta B^j_{t_k t_{k+1}}|^2 - \alpha_n^{2H} \right), \quad I_k^{(2)} = \sum_{i,j=1,\, i \neq j}^{m} \langle \sigma_i, \sigma_j \rangle \delta B^i_{t_k t_{k+1}} \delta B^j_{t_k t_{k+1}}.$$

We will now treat $\sum_{k=0}^{n-1} I_k^{(1)}$ and $\sum_{k=0}^{n-1} I_k^{(2)}$ separately.

In order to bound $\sum_{k=0}^{n-1} I_k^{(1)}$, notice that owing to the scaling property of fBm we have

$$\mathbf{E} \left| \sum_{k=0}^{n-1} I_k^{(1)} \right|^p = \alpha_n^{2Hp}\, \mathbf{E} \left| \sum_{k=0}^{n-1} \sum_{j=1}^{m} |\sigma_j|^2 \left[|\delta B^j_{kk+1}|^2 - 1 \right] \right|^p$$
$$\leq \alpha_n^{2Hp}\, \|\sigma\|^{2p}\, \mathbf{E} \left| \sum_{k=0}^{n-1} \left[|\delta B^1_{kk+1}|^2 - 1 \right] \right|^p .$$

Since all moments of random variables in a finite Gaussian chaos are equivalent, it follows from (12)–(14) that

$$\left(\mathbf{E} \left| \sum_{k=0}^{n-1} \left[|\delta B^j_{kk+1}|^2 - 1 \right] \right|^p \right)^{1/p} \leq c \cdot \left(|\log(n)| n^{1/2} + n^{2H-1} \right)$$

and consequently

$$\left(\mathbf{E} \left| \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} I_k^{(1)} \right|^p \right)^{1/p} \leq c \cdot \alpha_n^{2H-2} \cdot \left(|\log(n)| n^{-1/2} + n^{2H-2} \right).$$

Since $\alpha_n = \kappa \cdot n^{-\alpha}$ we have

$$\left(\mathbf{E} \left| \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} I_k^{(1)} \right|^p \right)^{1/p} \leq c \cdot \left(|\log(n)| n^{-\alpha(2H-2)-1/2} + n^{(1-\alpha)(2H-2)} \right)$$

and Lemma 2.6 plus the condition $\alpha < \min \left\{ \frac{1}{4(1-H)}, 1 \right\}$ implies that

$$\lim_{n\to\infty} \frac{1}{n\alpha_n^2} \left| \sum_{k=0}^{n-1} I_k^{(1)} \right| = 0 \qquad \mathbf{P}\text{-}a.s.$$

So it remains to consider the off-diagonal terms, i.e. $I_k^{(2)}$. Here we can exploit the following trick: Let β and $\tilde{\beta}$ be two independent fractional Brownian motions with the same Hurst index. From (12)–(14) we have again that

$$V_n = \sum_{k=0}^{n-1} \left(|\delta_{t_k t_{k+1}} \beta|^2 - |\delta_{t_k t_{k+1}} \tilde{\beta}|^2 \right)$$

satisfies

$$\left(\mathbf{E} |V_n|^p \right)^{1/p} \leq c \cdot \alpha_n^{2H} \cdot \left(|\log(n)| n^{1/2} + n^{2H-1} \right).$$

However, setting $B^i = (\beta + \widetilde{\beta})/\sqrt{2}$ and $B^j = (\beta - \widetilde{\beta})/\sqrt{2}$, then B^i and B^j are two independent fractional Brownian motions and

$$V_n \stackrel{\mathcal{L}}{=} 2 \sum_{k=0}^{n-1} \delta_{t_k t_{k+1}} B^i \delta_{t_k t_{k+1}} B^j .$$

Now we can easily conclude that

$$\lim_{n\to\infty} \frac{1}{n\alpha_n^2} \left| \sum_{k=0}^{n-1} I_k^{(2)} \right| = 0 \qquad \mathbf{P}\text{-}a.s.$$

Gathering our bounds on $\sum_{k=0}^{n-1} I_k^{(1)}$ and $\sum_{k=0}^{n-1} I_k^{(2)}$, the proof of our lemma is now completed. □

Remark 4.4 Note that for $\alpha \geq \frac{1}{4(1-H)}$ the expression $Q_n^{(3)}$ gives a non-zero contribution or diverges. This follows again from (12)–(14).

We can now turn to the main aim of this section:

Proof of Theorem 1.5 Recall that Lemma 4.2 asserts that under our standing assumptions we have $\lim_{n\to\infty} Q_n(\vartheta_0) = 0$ almost surely. Using Theorem 1.5 we conclude that

$$\lim_{n\to\infty} \sup_{\vartheta\in\Theta} \left| Q_n(\vartheta) - \left(\mathbf{E}|b(\overline{Y};\vartheta)|^2 - \mathbf{E}|b(\overline{Y};\vartheta_0)|^2\right) \right| = 0$$

Now our theorem follows by a direct application of Proposition 4.1 to $L_n(\vartheta) = |Q_n(\vartheta)|$ and $L(\vartheta) = \left|\mathbf{E}[|b(\overline{Y};\vartheta)|^2] - \mathbf{E}[|b(\overline{Y};\vartheta_0)|^2]\right|$. □

Remark 4.5 The following corrected quadratic variation of our process Y will be needed in the analysis of the fractional Ornstein-Uhlenbeck process:

$$V_n = \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} \left(|\delta Y_{t_k t_{k+1}}|^2 - \|\sigma\|^2 \alpha_n^{2H} \right). \tag{29}$$

Using the techniques, which we have introduced so far, we obtain

$$\begin{aligned} V_n &= Q_n^{(3)} + \frac{1}{n} \sum_{k=0}^{n-1} |b(Y_{t_k};\vartheta_0)|^2 + \frac{2}{n\alpha_n} \sum_{k=0}^{n-1} \langle \delta F_{t_k t_{k+1}}, b(Y_{t_k};\vartheta_0)\rangle \\ &\quad + \frac{2}{n\alpha_n} \sum_{k=0}^{n-1} \langle b(Y_{t_k};\vartheta_0), r_k\rangle + \frac{2}{n\alpha_n^2} \sum_{k=0}^{n-1} \langle \delta F_{t_k t_{k+1}}, r_k\rangle + \frac{1}{n\alpha_n^2} \sum_{k=0}^{n-1} |r_k|^2. \end{aligned}$$

So the Lemmata 3.3, 3.4 and 4.3 and (27), (28) and an analogous estimate to (20) give

$$\lim_{n\to\infty} V_n = -\mathbf{E}|b(\overline{Y}_0;\vartheta_0)|^2.$$

Remark 4.6 Estimations of the coefficients H and σ are available for some special cases of Eq. (1), some nice examples are e.g. provided in Berzin and León (2008). If we use a plug-in-estimator $\widehat{\|\sigma\|_n^2}$ for $\|\sigma\|^2$, i.e. if we replace $\|\sigma\|^2$ by this estimate in our estimator, convergence to ϑ_0 is still guaranteed, if

$$\lim_{n\to\infty} \left(\widehat{\|\sigma\|_n^2} - \|\sigma\|^2 \right) \alpha_n^{2H-2} = 0$$

almost surely. Similarly, a plug-in-estimator for $\widehat{H}_n$ for H has to satisfy

$$\lim_{n\to\infty}\left(\alpha_n^{2\widehat{H}_n-2}-\alpha_n^{2H-2}\right)=0$$

almost surely. Whenever these conditions are satisfied, our procedure leads to a consistent estimator of the triple (H, σ, ϑ).

We close this section by analyzing the application of our method to a fractional Ornstein-Uhlenbeck process:

Proof of Proposition 1.10 Let Y be the solution to Eq. (7) with $\vartheta_0 < 0$. It is well known that an explicit expression for Y is given by

$$Y_t = y_0 \exp(\vartheta_0 t) + \exp(\vartheta_0 t)\int_0^t \exp(-\vartheta_0 s)\, dB_s.$$

For $t \to \infty$, this process converges to the stationary fractional Ornstein-Uhlenbeck process

$$\exp(\vartheta_0 t)\int_{-\infty}^t \exp(-\vartheta_0 s)\, dB_s, \quad t \geq 0,$$

see e.g. Garrido-Atienza et al. (2009). Furthermore, straightforward computations yield that our expression $Q_n(\vartheta)$ defined by (2) can be written as

$$\sum_{k=0}^{n-1}\left(|\delta Y_{t_k t_{k+1}} - \vartheta Y_{t_k}\alpha_n|^2 - \alpha_n^{2H}\right) = \left(\vartheta\sqrt{s_n^{(3)}} - \frac{s_n^{(2)}}{\sqrt{s_n^{(3)}}}\right)^2 + s_n^{(1)} - \frac{|s_n^{(2)}|^2}{s_n^{(3)}}$$

where $s_n^{(1)}, s_n^{(2)}, s_n^{(3)}$ are defined by relation (8). Now we have to distinguish two cases.

Case I: If $s_n^{(1)} s_n^{(3)} \geq |s_n^{(2)}|^2$, then the minimum of $|Q_n(\vartheta)|$ is obtained for $\widehat{\vartheta}_n = s_n^{(2)}/s_n^{(3)}$ as in the classical case $H = 1/2$.

Case II: For $s_n^{(1)} s_n^{(3)} \leq |s_n^{(2)}|^2$ the minimum of $|Q_n(\vartheta)|$ is obtained for

$$\widehat{\vartheta}_n = \frac{s_n^{(2)}}{s_n^{(3)}} \pm \sqrt{\left(\frac{s_n^{(2)}}{s_n^{(3)}}\right)^2 - \frac{s_n^{(1)}}{s_n^{(3)}}}.$$

Note that for the Ornstein-Uhlenbeck process the quantity $s_n^{(1)}/n\alpha_n^2$ coincides with V_n defined at Remark 4.5, so that

$$\lim_{n\to\infty}\frac{1}{n\alpha_n^2}s_n^{(1)} = -\vartheta_0^2 \mathbf{E}|\overline{Y}_0|^2 < 0 \qquad \mathbf{P}\text{-a.s.}$$

holds true. Moreover, Lemma 3.3 also yields

$$\lim_{n\to\infty}\frac{1}{n\alpha_n^2}s_n^{(3)} = \mathbf{E}|\overline{Y}_0|^2 \qquad \mathbf{P}\text{-a.s.}$$

and thus

$$\lim_{n\to\infty}\frac{s_n^{(1)}}{s_n^{(3)}} = -\vartheta_0^2 < 0 \qquad \mathbf{P}\text{-a.s.}$$

So for n large enough, case II will always be the correct option. Moreover, since $\vartheta_0 < 0$, this limit behavior leads to the asymptotically unique estimator

$$\widehat{\vartheta}_n = \frac{s_n^{(2)}}{s_n^{(3)}} - \sqrt{\left(\frac{s_n^{(2)}}{s_n^{(3)}}\right)^2 - \frac{s_n^{(1)}}{s_n^{(3)}}}.$$

□

References

Arnold L (1997) Random Dynamical Systems. Springer, Berlin

Bégyn A (2005) Quadratic variations along irregular subdivisions for Gaussian processes. Electronic J. Probab. 10:691–717

Belfadli R, Es-Sebaiy K, Ouknine Y (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

Bender C, Sottinen T, Valkeila E (2008) Pricing by hedging and no-arbitrage beyond semimartingales. Finance Stoch. 12:441–468

Berzin C, León JR (2008) Estimation in models driven by fractional Brownian motion. Ann. Inst. Henri Poincaré Probab. Stat. 44(2):191–213

Chronopoulou A, Tindel S (2011) On inference for fractional differential equations. arXiv:1104.3966 Preprint. To appear in Stat. Infer. Stoch. Process.

Coeurjolly JF (2001) Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. Stat. Infer. Stoch. Process. 4(2):199–227

Crauel H, Debussche A, Flandoli F (1997) Random attractors. J. Dyn. Diff. Eqns. 9:307–341

Denk G, Meintrup D, Schäffler S (2001) Transient noise simulation: Modeling and simulation of $1/f$-noise. In: Antreich K (ed) et al. Modeling, simulation and optimization of integrated circuits. Birkhäuser, Basel, pp 251–267

Florens-Zmirou D (1989) Approximate discrete-time schemes for statistics of diffusion processes. Statistics 20(4):547–557

Frydman R (1980) A proof of the consistency of maximum likelihood estimators of non-linear regression models with autocorellated errors. Econometrica 48:853–860

Garrido-Atienza M, Kloeden P, Neuenkirch A (2009) Discretization of stationary solutions of stochastic systems driven by fractional Brownian motion. Appl. Math. Optim. 60(2):151–172

Garcia AM, Rodemich E, Rumsey H Jr (1978) A real variable lemma and the continuity of paths of some Gaussian processes. Indiana Math. J. 20:565–578

Garrido-Atienza M, Schmalfuss B (2011) Ergodicity of the infinite-dimensional fractional Brownian motion. J. Dyn. Diff. Equat. 23(3):671–681

Hairer M (2005) Ergodicity of stochastic differential equations driven by fractional Brownian motion. Ann. Probab. 33(2):703–758

Hairer M, Majda A (2010) A simple framework to justify linear response theory. Nonlinearity 23(4):909–922

Hairer M, Ohashi A (2007) Ergodicity theory of SDEs with extrinsic memory. Ann. Probab. 35(5):1950–1977

Hairer M, Pillai S (2011) Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. Ann. Inst. Henri Poincaré Probab. Stat. 47(2):601–628

Hu Y, Nualart D (2010) Parameter estimation for fractional Ornstein-Uhlenbeck processes. Stat. Prob. Lett. 80:1030–1038

Istas J, Lang G (1994) Quadratic variations and estimation of the local Hölder index of a Gaussian process. Ann. Inst. Poincaré Probab. Stat. 33:407–436

Kasonga RA (1988) The consistency of a non-linear least squares estimator from diffusion processes. Stoch. Proc. Appl. 30:263–275

Kessler M (2000) Simple and explicit estimating functions for a discretely observed diffusion process. Scand. J. Statist. 27(1):65–82

Kleptsyna M, Le Breton A (2002) Statistical analysis of the fractional Ornstein-Uhlenbeck type process. Stat. Inference Stoch. Process. 5(3):229–248

Kou S (2008) Stochastic modeling in nanoscale physics: subdiffusion within proteins. Ann. Appl. Stat. 2(2): 501–535

Le Breton A (1998) Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion. Stat. Probab. Lett. 38(3):263–274
Kloeden P, Neuenkirch A (2007) The pathwise convergence of approximation schemes for stochastic differential equations. LMS J. Comp. Math. 10:235–253
Nualart D, Peccati G (2005) Central limit theorems for sequences of multiple stochastic integrals. Ann. Probab. 33(1):177–193
Papavasiliou A, Ladroue C (2011) Parameter estimation for rough differential equations. Ann. Statist. 39(4): 2047–2073
B.L.S. Prakasa Rao (2010) Statistical inference for fractional diffusion processes. Wiley Series in Probability and Statistics. Wiley, Chichester
Shiryaev AN (1995) Probability, 2nd edn. Graduate Texts in Mathematics. Springer, New York
Tudor C, Viens F (2007) Statistical aspects of the fractional stochastic calculus. Ann. Statist. 35(3):1183–1212
Tudor C, Viens F (2009) Variations and estimators for self-similarity parameters via Malliavin calculus. Ann. Probab. 37(6):2093–2134
Young LC (1936) An inequality of Hölder type connected with Stieltjes integration. Acta Math. 67:251–282
Zähle M (2005) Stochastic differential equations with fractal noise. Math. Nachr. 278(9):1097–1106