The Annals of Statistics
2014, Vol. 42, No. 6, 2557–2585
DOI: 10.1214/14-AOS1262

ASYMPTOTIC EQUIVALENCE FOR REGRESSION UNDER FRACTIONAL NOISE[1]

BY JOHANNES SCHMIDT-HIEBER

Leiden University

Consider estimation of the regression function based on a model with equidistant design and measurement errors generated from a fractional Gaussian noise process. In previous literature, this model has been heuristically linked to an experiment, where the anti-derivative of the regression function is continuously observed under additive perturbation by a fractional Brownian motion. Based on a reformulation of the problem using reproducing kernel Hilbert spaces, we derive abstract approximation conditions on function spaces under which asymptotic equivalence between these models can be established and show that the conditions are satisfied for certain Sobolev balls exceeding some minimal smoothness. Furthermore, we construct a sequence space representation and provide necessary conditions for asymptotic equivalence to hold.

1. Introduction. Suppose we have observations from the regression model

$$Y_{i,n} = f\left(\frac{i}{n}\right) + N_i^H, \qquad i = 1, \dots, n, \tag{1}$$

where $(N_i^H)_{i\in\mathbb{N}}$ denotes a fractional Gaussian noise (fGN) process with Hurst index $H \in (0,1)$, that is, a stationary Gaussian process with autocovariance function $\gamma(k) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H})$. This model can be viewed as a prototype of a nonparametric regression setting under dependent measurement errors. Corresponding to $H \le \frac{1}{2}$ and $H > \frac{1}{2}$, the noise process exhibits short- and long-range dependence, respectively. In the case $H = \frac{1}{2}$, fGN is just Gaussian white noise.

Although observing (1) is the "more realistic" model, one might be tempted to replace (1) by a continuous version which is more convenient to work with as it avoids discretization effects. Recall the definition of a fractional Brownian motion (fBM) with Hurst parameter $H \in (0,1)$ as a Gaussian process $(B_t^H)_{t\ge 0}$ with covariance function $(s,t) \mapsto \operatorname{Cov}(B_s^H, B_t^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$.

Received December 2013; revised June 2014.
[1]Supported by DFG Fellowship SCHM 2807/1-1 and ERC Grant 320637.
MSC2010 subject classifications. Primary 62G08; secondary 62G20.
Key words and phrases. Asymptotic equivalence, long memory, fractional Brownian motion, fractional Gaussian noise, fractional calculus, inverse problems, nonharmonic Fourier series, reproducing kernel Hilbert space (RKHS), stationarity.

In Wang [38], Johnstone and Silverman [20] and Johnstone [19] it has been argued that

$$Y_t = \int_0^t f(u)\,du + n^{H-1} B_t^H, \qquad t \in [0,1],\ B^H \text{ a fBM} \tag{2}$$

is a natural candidate for a continuous version of (1) for $H \geq 1/2$. By projecting $(Y_t)_{t\in[0,1]}$ onto a suitable basis, one is further interested in an equivalent sequence space representation $(\dots, Z_{-1}, Z_0, Z_1, \dots)$, where the weighted Fourier coefficients of f under additive white noise are observed, that is,

$$Z_k = \sigma_k^{-1}\theta_k(f) + n^{H-1}\varepsilon_k, \qquad k \in \mathbb{Z},\ \varepsilon_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{3}$$

Here, $\theta_k(f)$ denote the Fourier coefficients and $\sigma_k > 0$ are weights. Models of type (3) have been extensively studied in statistical inverse problems literature (cf. Cavalier [5]).

In this work, we investigate these approximations and its limitations under Le Cam distance (cf. Appendix E in the supplementary material [33] for a summary of the topic). The Le Cam distance allows to quantify the maximal error that one might encounter by changing the experiment. Indeed, it controls the largest possible difference of risks that could occur under bounded loss functions. Two experiments are said to be asymptotic equivalent, if the Le Cam distance converges to zero. Therefore, if we can establish asymptotic equivalence, then replacing (1) by (2) or (3) is harmless at least for asymptotic statements about the regression function f.

Our main finding is that for $H \in (\frac{1}{4}, 1)$ the experiments generated by model (1) and model (2) are asymptotic equivalent for Θ a periodic Sobolev space with smoothness index $\alpha > 1/2$, if $H \in [\frac{1}{2}, 1)$ and $\alpha > (1-H)/(H+1/2) + H - 1/2$, if $H \in (\frac{1}{4}, \frac{1}{2}]$. Moreover, we show that for any $H \in (0,1)$ asymptotic equivalence does not hold for $\alpha = 1/2$ and any $\alpha < 1 - H$, proving that the minimal smoothness requirement $\alpha > 1/2$ for $H \in [\frac{1}{2}, 1)$ is sharp in this sense. The asymptotic equivalence for $H \in (\frac{1}{4}, \frac{1}{2})$ is surprising and leads to better estimation rates than the heuristic continuous approximation presented in [20]. The case $H \in (0, \frac{1}{4}]$ remains open. Since the noise level in (2) and (3) decreases with H, discretization effects become more and more dominant. We conjecture that for small H asymptotic equivalence will not hold. For suitable $\sigma_k, \theta_k(f)$, equivalence between the experiments generated by model (2) and model (3) can be derived for all $H \in (0,1)$. We find that $\sigma_k \asymp |k|^{1/2-H}$. Generalization of the latter result is possible if the fBM is replaced by a Gaussian process with stationary increments.

One of the motivations for our work is to extend the asymptotic equivalence result for regression under independent Gaussian noise (in our framework the case $H = 1/2$). In Brown and Low [2], it was shown that the experiments generated by the standard regression model

$$Y_{i,n} = f\left(\frac{i}{n}\right) + \varepsilon_{i,n}, \qquad (\varepsilon_{i,n})_{i=1,\dots,n} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1),\ f \in \Theta$$

and the Gaussian white noise model

$$dY_t = f(t)\,dt + n^{-1/2}\,dW_t, \qquad t \in [0, 1],\ W \text{ a Brownian motion},\ f \in \Theta, \tag{4}$$

are asymptotically equivalent, provided that the parameter space $\Theta \subset L^2[0, 1]$ has the approximation property

$$n \sup_{f\in\Theta} \int_0^1 \big(f(u) - \overline{f}_n(u)\big)^2\,du \to 0, \tag{5}$$

with $\overline{f}_n := \sum_{j=1}^n f(j/n)\mathbb{I}_{((j-1)/n, j/n]}(\cdot)$. A natural choice for Θ would be the space of Hölder continuous functions with regularity index larger $1/2$ and Hölder norm bounded by a finite constant. One of the consequences of our results is that the smoothness constraints for general Hurst index, simplify to (5) if $H = 1/2$.

Organization of the work. In Section 2, we use RKHSs as an abstract tool in order to formulate sufficient approximation conditions for asymptotic equivalence of the experiments generated by the discrete and continuous regression models (1) and (2). Although these conditions appear naturally, they are very difficult to interpret. By using a spectral characterization of the underlying RKHS, we can reduce the problem to uniform approximation by step functions in homogeneous Sobolev spaces. This is described in Section 3. We further mention some orthogonality properties that reveal the structure of the underlying function spaces and explain the key ideas of the proofs. The main results together with some discussion are stated in Section 4. In this section, we construct a sequence space representation and prove equivalence with the continuous regression experiment. This allows to study the ill-posedness induced by the dependence of the noise. Thereafter, we study necessary conditions and outline a general scheme for deriving sequence space representation given a regression model with stationary noise. This scheme does not require knowledge of the Karhunen–Loeve expansion. Since the appearance of [2], many other asymptotic equivalence results have been established for related nonparametric problems and there are various strategies in order to bound the Le Cam distance. We provide a brief survey in Section 5 and relate the existing approaches to our techniques. Proofs are mainly deferred to the Appendix. Parts of the Appendix as well as a short summary of Le Cam distance and asymptotic equivalence can be found in the supplement [33].

Notation. If two experiments are equivalent, we write $=$ and $\simeq$ denotes asymptotic equivalence. The operator $\mathcal{F}$, defined on $L^1(\mathbb{R})$ or $L^2(\mathbb{R})$ (depending on the context) is the Fourier transform $(\mathcal{F}f)(\lambda) = \int e^{-i\lambda x} f(x)\,dx$. For the indicator function on $[0, t)$, we write $\mathbb{I}_t := \mathbb{I}_{[0,t)}(\cdot)$ (as a function on $\mathbb{R}$). The Gamma function is denoted by $\Gamma(\cdot)$. For a Polish space Ω, let $\mathcal{B}(\Omega)$ be the Borel sets. Further, $\mathcal{C}[T]$ denotes the space of continuous functions on T equipped with the uniform norm.

2. Abstract conditions for asymptotic equivalence. The goal of this section is to reduce asymptotic equivalence to approximation conditions (cf. Theorem 1). For that, tools from Gaussian processes and RKHS theory are required which are introduced in a first step.

The concept of a RKHS can be defined via the Moore–Aronszajn theorem. It states that for a given index set $T \subseteq \mathbb{R}$ and a symmetric, positive semi-definite function $K : T \times T \to \mathbb{R}$, there exists a unique Hilbert space $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ with:

(i) $K(\cdot, t) \in \mathbb{H}, \forall t \in T$,

(ii) $\langle f, K(\cdot, t)\rangle_{\mathbb{H}} = f(t), \forall f \in \mathbb{H}, \forall t \in T$.

The second condition is called the reproducing property. The Hilbert space $\mathbb{H}$ is called the reproducing kernel Hilbert space (with reproducing kernel K). RKHSs are a strict subset of Hilbert spaces, as, for instance, $L^2[0, 1]$ has no reproducing kernel.

A centered Gaussian process $(X_t)_{t\in T}$ can be associated with a RKHS $\mathbb{H}$ defined via the reproducing kernel $K : T \times T \to \mathbb{R}$, $K(s, t) := \mathbb{E}[X_s X_t]$. $\mathbb{H}$ can also be characterized by the completion of the function class

$$\left\{\phi : T \to \mathbb{C} \middle| \phi : t \mapsto \sum_{j=1}^{M} u_j K(s_j, t), (s_j, u_j) \in T \times \mathbb{C}, j = 1, \dots, M\right\} \tag{6}$$

with respect to the norm $\| \sum_{j=1}^{M} u_j K(s_j, \cdot)\|_{\mathbb{H}}^2 := \sum_{j,k\le M} u_j K(s_j, s_k)\overline{u_k}$. For a Gaussian process $(X_t)_{t\in T}$, there is a Girsanov formula with the associated RKHS $\mathbb{H}$ playing the role of the Cameron–Martin space.

LEMMA 1 (Example 2.2, Theorem 2.1 and Lemma 3.2 in van der Vaart and van Zanten [37]). *Let $(X_t)_{t\in T}$ be a Gaussian process with continuous sample paths on a compact metric space T and $\mathbb{H}$ the associated RKHS. Denote by P_f the probability measure of $t \mapsto f(t) + X_t$ on $(\mathcal{C}[T], \mathcal{B}(\mathcal{C}[T]))$. If $f \in \mathbb{H}$, then P_f and P_0 are equivalent measures and the Radon–Nikodym derivative is given by*

$$\frac{dP_f}{dP_0} = \exp\left(Uf - \frac{1}{2}\|f\|_{\mathbb{H}}^2\right), \qquad f \in \mathbb{H},$$

where U denotes the iso-Gaussian process, that is, the centered Gaussian process $(Uh)_{h\in\mathbb{H}}$ with $UK(t, \cdot) := X_t$ and covariance $\mathbb{E}[(Uh)(Ug)] = \langle h, g\rangle_{\mathbb{H}}$.

Given such a change of measure formula, it is straightforward to compute the Kullback–Leibler divergence $d_{\mathrm{KL}}(\cdot, \cdot)$ in terms of the RKHS norm.

LEMMA 2. *For $f, g \in \mathbb{H}$, and P_f, P_g as in Lemma 1,*

$$d_{\mathrm{KL}}(P_f, P_g) = \tfrac{1}{2}\|f - g\|_{\mathbb{H}}^2.$$

Throughout the following, the RKHS with reproducing kernel $(s,t) \mapsto K(s,t) = \mathbb{E}[B_s^H B_t^H]$ will be denoted by $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ (for convenience, the dependence of $\mathbb{H}$ and K on the Hurst index H is omitted). Before the main result of this section can be stated, we need to introduce the experiments generated by the models in Section 1.

Experiment $\mathcal{E}_{1,n}(\Theta)$: *Nonparametric regression under fractional noise*. Denote by $\mathcal{E}_{1,n}(\Theta) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_f^n : f \in \Theta))$ the experiment with P_f^n the distribution of $\mathbf{Y_n} := (Y_{1,n}, \dots, Y_{n,n})^t$, where

$$Y_{i,n} = f\left(\frac{i}{n}\right) + N_i^H, \qquad i = 1, \dots, n \text{ and } (N_i^H)_i \text{ is a fGN process.} \tag{7}$$

Experiment $\mathcal{E}_{2,n}(\Theta)$: Let $\mathcal{E}_{2,n}(\Theta) = (\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]), (Q_f^n : f \in \Theta))$ be the experiment with Q_f^n the distribution of

$$Y_t = \int_0^t f(u)\,du + n^{H-1} B_t^H, \qquad t \in [0,1], B^H \text{ a fBM.} \tag{8}$$

We write F_f for the anti-derivative of f on $[0,1]$, that is, $F_f(t) = \int_0^t f(u)\,du$ for all $t \in [0,1]$. The first result relates asymptotic equivalence to abstract approximation conditions.

THEOREM 1. *Let* $H \in (0,1)$. *Suppose that*:

(i) $(n^{1-2H} \vee 1) \sup_{f \in \Theta} \sum_{i=1}^n (n \int_{(i-1)/n}^{i/n} f(u)\,du - f(\frac{i}{n}))^2 \to 0$,
(ii) $n^{1-H} \sup_{f \in \Theta} \inf_{(\alpha_1,\dots,\alpha_n)^t \in \mathbb{R}^n} \|F_f - \sum_{j=1}^n \alpha_j K(\cdot, \frac{j}{n})\|_{\mathbb{H}} \to 0$.

Then,

$$\mathcal{E}_{1,n}(\Theta) \simeq \mathcal{E}_{2,n}(\Theta).$$

PROOF. The proof consists of three steps. Proposition A.1 in the Appendix states that, under condition (i), the values $f(\frac{i}{n})$ may be replaced by $\widetilde{f}_{i,n} := n \int_{(i-1)/n}^{i/n} f(u)\,du$ in model (7). Instead of $\mathcal{E}_{1,n}(\Theta)$, we can therefore consider the experiment $\mathcal{E}_{4,n}(\Theta) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_{4,f}^n : f \in \Theta))$ with $P_{4,f}^n$ the distribution of

$$\widetilde{Y}_{i,n} := \widetilde{f}_{i,n} + N_i^H, \qquad i = 1, \dots, n, f \in \Theta.$$

In order to link experiment $\mathcal{E}_{4,n}(\Theta)$ to the continuous model in $\mathcal{E}_{2,n}(\Theta)$, the crucial point is to construct a path on $[0,1]$ from the observations $\widetilde{Y}_{i,n}$, $i = 1, \dots, n$ with distribution "close" to (8). For this, let throughout the following $\mathbf{x}_n = (x_1, \dots, x_n)$, and $\mathbf{y}_n = (y_1, \dots, y_n)$ be vectors and consider the interpolation function

$$L(t|\mathbf{x}_n) := \mathbb{E}\big[B_t^H \big| B_{\ell/n}^H = x_\ell, \ell = 1, \dots, n\big], \qquad t \in [0,1],$$

with $(B_t^H)_{t\geq 0}$ a fBM. Let $\mathbf{B}_n^H$ denote the vector $(B_{\ell/n}^H)_{\ell=1,\dots,n}$. From the formula for conditional expectations of multivariate Gaussian random variables, we obtain the alternative representation

$$L(t|\mathbf{x}_n) = \left(K\left(t, \frac{1}{n}\right), K\left(t, \frac{2}{n}\right), \dots, K(t,1)\right) \operatorname{Cov}(\mathbf{B}_n^H)^{-1}\mathbf{x}_n^t \tag{9}$$

and it is easy to verify that

$$\begin{aligned} &\textit{linearity}: && L(\cdot|\mathbf{x}_n + \mathbf{y}_n) = L(\cdot|\mathbf{x}_n) + L(\cdot|\mathbf{y}_n),\\ &\textit{interpolation}: && L\left(\frac{j}{n}\middle|\mathbf{x}_n\right) = x_j \qquad \text{for } j \in \{0,1,\dots,n\} \text{ with } x_0 := 0. \end{aligned}$$

The key observations is that if B^H and $\breve{B}^H$ are two independent fBMs and $R_t^H = \breve{B}_t^H - L(t|(\breve{B}_{\ell/n}^H)_{\ell=1,\dots,n})$, then, by comparison of the covariance structure, the process

$$\left(L\left(t|(B_{\ell/n}^H)_{\ell=1,\dots,n}\right) + R_t^H\right)_{t\geq 0}$$

is a fBM as well. Define the vector of partial sums $\mathbf{S_n}\widetilde{\mathbf{Y}} := (S_k\widetilde{Y})_{k=1,\dots,n}$ with components $S_k\widetilde{Y} := \sum_{j=1}^k \widetilde{Y}_{j,n}$. Recall that $F_f(t) = \int_0^t f(u)\,du$, let

$$\mathbf{F}_{f,n} := \left(F_f\left(\frac{\ell}{n}\right)\right)_{\ell=1,\dots,n}, \tag{10}$$

and observe that $\mathbf{S_n}\widetilde{\mathbf{Y}} = n\mathbf{F}_{f,n} + \mathbf{B}_n^H$, in distribution. For $(R_t)_{t\geq 0}$ independent of $\mathbf{S_n}\widetilde{\mathbf{Y}}$, we find using the linearity property of L,

$$\widetilde{Y}_t := n^{H-1}\left(L\left(t|n^{-H}\mathbf{S_n}\widetilde{\mathbf{Y}}\right) + R_t^H\right) = L(t|\mathbf{F}_{f,n}) + n^{H-1}B_t^H, \qquad t \in [0,1],$$

for a fBM B^H. Consequently, we can construct paths $(\widetilde{Y}_t)_{t\in[0,1]}$ by interpolation of $\widetilde{Y}_{i,n}$, $i = 1,\dots,n$ and adding an uninformative process that match (8) up to the regression function. On the contrary, by the interpolation property of L, we can recover $\widetilde{Y}_{i,n}$, $i = 1,\dots,n$ from $(\widetilde{Y}_t)_{t\in[0,1]}$ and, therefore,

$$\mathcal{E}_{4,n}(\Theta) = \mathcal{E}_{5,n}(\Theta),$$

where $\mathcal{E}_{5,n}(\Theta) = (\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]), (Q_{5,f}^n : f \in \Theta))$ and $Q_{5,f}^n$ denotes the distribution of $(\widetilde{Y}_t)_{t\in[0,1]}$. In Proposition A.2, we prove that $\mathcal{E}_{5,n}(\Theta) \simeq \mathcal{E}_{2,n}(\Theta)$ under the approximation condition (ii). This shows that

$$\mathcal{E}_{1,n}(\Theta) \underset{\text{cond. (i)}}{\overset{\text{Prop. A.1}}{\simeq}} \mathcal{E}_{4,n}(\Theta) = \mathcal{E}_{5,n}(\Theta) \underset{\text{cond. (ii)}}{\overset{\text{Prop. A.2}}{\simeq}} \mathcal{E}_{2,n}(\Theta). \tag{11}$$ $\square$

Theorem 1 reduces proving asymptotic equivalence to verifying the imposed approximation conditions. Whereas (i) is of type (5) and well studied, the second condition requires that the anti-derivative of f can be approximated by linear combinations of the kernel functions in the RKHS $\mathbb{H}$. In particular, it implies that

$\{F_f : f \in \Theta\} \subset \mathbb{H}$. In Section 4.3 below, we give a heuristic, why the second condition appears naturally.

By Jensen's inequality, property (i) in Theorem 1 is satisfied, provided that $(n^{1-H} \vee n^{1/2}) \sup_{f\in\Theta} \|f - \overline{f}_n\|_{L^2[0,1]} \to 0$ with $\overline{f}_n$ being the step function $\sum_{j=1}^n f(\frac{j}{n})\mathbb{I}_{((j-1)/n,j/n]}(\cdot)$. In the case of Brownian motion, that is $H = 1/2$, we can simplify the conditions further. Recall that in this case $\|h\|_{\mathbb{H}} = \|h'\|_{L^2[0,1]}$ and $K(s,t) = s \wedge t$ (cf. van der Vaart and van Zanten [37], Section 10). Consequently, both approximation conditions hold if $n^{1/2} \sup_{f\in\Theta} \|f - \overline{f}_n\|_{L^2[0,1]} \to 0$. Thus, we reobtain the well-known Brown and Low condition (5).

The approximation conditions do not allow for straightforward construction of function spaces Θ on which asymptotic equivalence holds. In view of condition (ii) in Theorem 1, a natural class of functions to study in a first step would consists of all f such that $F_f = K(\cdot, x_0)$ with $x_0 \in [0,1]$ fixed, or equivalently $f : t \mapsto \partial_t K(t, x_0)$. In the case $H = \frac{1}{2}$, this is just the class of indicator functions $\{\mathbb{I}_s : s \in [0,1]\}$ and it is not difficult to see that $\mathcal{E}_{2,n}(\Theta)$ is strictly more informative than $\mathcal{E}_{1,n}(\Theta)$, implying $\mathcal{E}_{1,n}(\Theta) \not\simeq \mathcal{E}_{2,n}(\Theta)$.

Thus, we need to construct Θ containing smoother functions, which at the same time can be well approximated by linear combinations of kernel functions in the sense of condition (ii) of the preceding theorem. In order to find suitable function spaces, a refined analysis of the RKHS $\mathbb{H}$ is required. This will be the topic of the next section.

3. The RKHS associated to fBM. Using RKHS theory, we show in this section that condition (ii) of Theorem 1 can be rewritten as approximation by step functions in a homogeneous Sobolev space.

The RKHS of fBM can either be characterized in the time domain via fractional operators, or in the spectral domain using Fourier calculus. For our approach, we completely rely on the spectral representation as it avoids some technical issues. In principle, however, all results could equally well be described in the time domain. For more on that, cf. Pipiras and Taqqu [27]. Set $c_H := \sin(\pi H)\Gamma(2H+1)$. Recall that $K(s,t) = \mathbb{E}[B_s^H B_t^H]$, for $s, t \in [0,1]$. Then (cf. Yaglom [40] or Samorodnitsky and Taqqu [32], equation (7.2.9)),

$$K(s,t) = \int \mathcal{F}(\mathbb{I}_s)(\lambda)\overline{\mathcal{F}(\mathbb{I}_t)(\lambda)}\mu(d\lambda) \tag{12}$$

with

$$\mu(d\lambda) = \frac{c_H}{2\pi}|\lambda|^{1-2H}\,d\lambda.$$

Given this representation, it is straightforward to describe the corresponding RKHS as follows (cf. Grenander [17], page 97): let $\mathbb{M}$ denote the closed linear span of $\{\mathcal{F}(\mathbb{I}_t) : t \in [0,1]\}$ in the weighted L^2-space $L^2(\mu)$, then

$$\mathbb{H} = \{F : \exists F^* \in \mathbb{M}, \text{ such that } F(t) = \langle F^*, \mathcal{F}(\mathbb{I}_t)\rangle_{L^2(\mu)}, \forall t \in [0,1]\},$$

where $\langle g, h\rangle_{L^2(\mu)} := \int g(\lambda)\overline{h(\lambda)}\mu(d\lambda)$ denotes the $L^2(\mu)$ inner product. Further,

$$Q : \bigl(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}}\bigr) \to \bigl(\mathbb{M}, \langle\cdot,\cdot\rangle_{L^2(\mu)}\bigr), \qquad Q(F) = F^*$$

is an isometric isomorphism and

$$\langle g, h\rangle_{\mathbb{H}} = \bigl\langle Q(g), Q(h)\bigr\rangle_{L^2(\mu)}. \tag{13}$$

Let us show the use of this representation of $\mathbb{H}$ for the approximation condition (ii) of Theorem 1, that is,

$$n^{1-H} \sup_{f\in\Theta} \inf_{(\alpha_1,\dots,\alpha_n)^t\in\mathbb{R}^n} \left\| F_f - \sum_{j=1}^n \alpha_j K\Bigl(\cdot, \frac{j}{n}\Bigr)\right\|_{\mathbb{H}} \to 0. \tag{14}$$

By (12), we obtain $Q(K(\cdot,s)) = \mathcal{F}(\mathbb{I}_s)$ and, therefore,

$$\left\| F_f - \sum_{j=1}^n \alpha_j K\Bigl(\cdot, \frac{j}{n}\Bigr)\right\|_{\mathbb{H}} = \left\| Q(F_f) - \sum_{j=1}^n \alpha_j \mathcal{F}(\mathbb{I}_{j/n})\right\|_{L^2(\mu)}.$$

It is natural to consider now functions f for which there exists a g with $Q(F_f) = \mathcal{F}(g)$. Since $Q(F_f)$ lies in $\mathbb{M}$, the closure of the functions $\{\mathcal{F}(\mathbb{I}_s) : s \in [0,1]\}$, the support of g must be contained in $[0,1]$. If for any f such a g exists, (14) simplifies further to

$$n^{1-H} \sup_{f\in\Theta} \inf_{(\beta_1,\dots,\beta_n)^t\in\mathbb{R}^n} \left\| \mathcal{F}\Bigl(g - \sum_{j=1}^n \beta_j \mathbb{I}_{((j-1)/n, j/n]}\Bigr)\right\|_{L^2(\mu)} \to 0.$$

Instead of approximating functions F_f by linear combinations of kernel functions in $\mathbb{H}$, we have reduced the problem to approximation by step functions in a homogeneous Sobolev space. The difficulty relies in computing g given a function f. To see, how f and g are linked, observe that by the characterization of $\mathbb{H}$ above, $Q(F_f) = \mathcal{F}(g)$, and Parseval's theorem [assuming that $|\cdot|^{1-2H}\mathcal{F}(g) \in L^2(\mathbb{R})$ for the moment],

$$\begin{aligned} F_f(t) &= \bigl\langle \mathcal{F}(g), \mathcal{F}(\mathbb{I}_t)\bigr\rangle_{L^2(\mu)} \\ &= \frac{c_H}{2\pi}\int_{-\infty}^{\infty} |\lambda|^{1-2H}\mathcal{F}(g)(\lambda)\overline{\mathcal{F}(\mathbb{I}_t)(\lambda)}\, d\lambda \\ &= c_H \int_0^t \mathcal{F}^{-1}\bigl(|\cdot|^{1-2H}\mathcal{F}(g)\bigr)(u)\, du, \qquad t \in [0,1], \end{aligned}$$

implying

$$f = c_H \mathcal{F}^{-1}\bigl(|\cdot|^{1-2H}\mathcal{F}(g)\bigr)|_{[0,1]}. \tag{15}$$

Thus, f and g are connected via a Fourier multiplier restricted to the interval $[0,1]$. For $H = 1/2$, we obtain $f = g$ as a special case. For other values of H, the

Fourier multiplier acts like a fractional derivative/integration operator, in particular it is nonlocal.

One possibility to solve for g is to extend the regression function f to the real line and then to invert the Fourier multiplier (15). Recall, however, that g has to be supported on [0, 1] and because of the nonlocality of the operator, this strategy does not lead to "valid" functions g.

Another possibility is to interpret (15) as a source condition: we construct function spaces Θ on which asymptotic equivalence can be established by considering source spaces, $\mathcal{S}(\Theta)$ say, of sufficiently smooth functions g first and then defining Θ as all functions f, for which there exists a source element $g \in \mathcal{S}(\Theta)$ such that (15) holds. Source conditions are a central topic in the theory of deterministic inverse problems (cf. Engl et al. [12] for a general treatment and Tautenhahn and Gorenflo [35] for source conditions for inverse problems involving fractional derivatives). A similar construction is employed in fractional calculus, by defining the domain of a fractional derivative as the image of the corresponding fractional integration operator (cf., e.g., see Samko et al. [31], Section 6.1). Although, thinking about (15) as abstract smoothness condition itself makes things formally tractable, it has the obvious drawback, that it does not result in a good description of the function space Θ. We still cannot decide whether all functions of a given Hölder or Sobolev space are generated by a source condition or not.

Surprisingly, there are explicit solutions to (15), which satisfy some remarkable orthogonality relations, both in $L^2[0, 1]$ and $L^2(\mu)$. For that some notation is required. Denote by J_ν the Bessel function of the first kind with index $\nu > 0$. It is well known that the roots of J_ν are real, countable, nowhere dense, and also contain zero (cf. Watson [39]). Throughout the following, let $\cdots < \omega_{-1} < \omega_0 := 0 < \omega_1 < \cdots$ be the ordered (real) roots of the Bessel function J_{1-H} (for convenience, we omit the dependence on the Hurst index H). Define the functions

$$g_k : s \mapsto \mathbb{I}_{(0,1)}(s)\partial_s \int_0^s e^{i2\omega_k(s-u)}\big(u - u^2\big)^{1/2-H}\,du, \qquad k \in \mathbb{Z}. \tag{16}$$

As we will show below, $\mathcal{F}^{-1}(|\cdot|^{1-2H}\mathcal{F}(g_k))|_{[0,1]}$ equals (up to a constant factor)

$$f_k : t \mapsto e^{2i\omega_k t}.$$

This provides us with solutions of (15). It is now natural to expand functions f as nonharmonic Fourier series $f = \sum_{k=-\infty}^{\infty} \theta_k e^{2i\omega_k \cdot}$ and to study asymptotic equivalence with the parameter space Θ being a Sobolev ball

$$\begin{aligned} &\Theta_H(\alpha, C) \\ &\quad := \Bigg\{ f = \sum_{k=-\infty}^{\infty} \theta_k e^{2i\omega_k \cdot} : \theta_k = \overline{\theta_{-k}}, \forall k, \sum_{k=-\infty}^{\infty} \big(1 + |k|\big)^{2\alpha} |\theta_k|^2 \le C^2 \Bigg\}. \end{aligned} \tag{17}$$

The constraint $\theta_k = \overline{\theta_{-k}}$ implies that f is real-valued.

Orthogonality properties of $(f_k)_k$ *and* $(g_k)_k$. The advantage of this approach is that *any* $f \in L^2[0,1]$ can be expanded in a unique way with respect to $(e^{2i\omega_k \cdot})_k$. We even have the stronger result.

LEMMA 3. *Given* $H \in (0,1)$. *Then,* $(e^{2i\omega_k \cdot})_k$ *is a Riesz basis of* $L^2[0,1]$.

Recall that a Riesz basis is a "deformed" orthonormal basis. For these bases, Parseval's identity only holds up to constants in the sense of equivalence of norms. This norm equivalence is usually referred to as frame inequality or near-orthogonality. For more on the topic, cf. Young [41], Section 1.8. The proof of Lemma 3 is delayed until Appendix B. It relies on a standard result for nonharmonic Fourier series in combination with some bounds on the zeros ω_k. Using the previous lemma, the Sobolev balls $\Theta_H(\alpha, C)$ can be linked to classical Sobolev spaces for integer α; cf. Lemma 5.

Next, let us prove that f_k and g_k are (up to a constant) solutions of (15) and state the key orthogonality property of $(g_k)_k$. This part relies essentially on the explicit orthogonal decomposition of the underlying RKHS $\mathbb{H}$ due to Dzhaparidze and van Zanten [10] (cf. also Appendix B).

THEOREM 2 (Dzhaparidze and van Zanten [10], Theorem 7.2). *Recall that* $\cdots < \omega_{-1} < \omega_0 := 0 < \omega_1 < \cdots$ *are the ordered zeros of the Bessel function* J_{1-H}. *For* $k \in \mathbb{Z}$, *define*

$$\phi_k(2\lambda) = \sqrt{\frac{\pi}{c_H}} 2^{H-1}\bigl(1 + (\sqrt{2-2H} - 1)\delta_{k,0}\bigr) e^{i(\omega_k - \lambda)} \frac{\lambda^H J_{1-H}(\lambda)}{\lambda - \omega_k}, \tag{18}$$

where $\phi_k(2\omega_k) := \lim_{\lambda \to 2\omega_k} \phi_k(\lambda)$ *and* $\delta_{k,0}$ *is the Kronecker delta. Then,* $\{\phi_k(\cdot) : k \in \mathbb{Z}\}$ *is an orthonormal basis* (*ONB*) *of* $\mathbb{M}$ *and we have the sampling formula*

$$h = \sum_{k=-\infty}^{\infty} a_k h(2\omega_k)\phi_k \qquad \text{for all } h \in \mathbb{M}$$

with convergence in $L^2(\mu)$ *and*

$$\begin{aligned} a_k^{-1} &:= \phi_k(2\omega_k) \\ &= \sqrt{\frac{\pi}{c_H}} \times \begin{cases} \sqrt{1-H}\, 2^{2H-3/2}\Gamma(2-H)^{-1}, & \text{for } k = 0, \\ 2^{H-1}\omega_k^H J'_{1-H}(\omega_k), & \text{for } k \neq 0. \end{cases} \end{aligned} \tag{19}$$

Moreover, for any k, $a_k = a_{-k}$ *and there exists a constant* $\overline{c}_H$, *such that*

$$\overline{c}_H^{-1}\bigl(1+|k|\bigr)^{1/2-H} \le |a_k| \le \overline{c}_H\bigl(1+|k|\bigr)^{1/2-H}. \tag{20}$$

Bessel functions have a power series expansion $J_{1-H}(\lambda) = \sum_{k=0}^{\infty} \gamma_k \lambda^{2k+1-H}$, for suitable coefficients γ_k. This allows to show that $\omega_k = -\omega_{-k}$ for all integer k and to identify $\lambda^H J_{1-H}(\lambda)$ for $\lambda < 0$ with the real-valued function $\sum_{k=0}^{\infty} \gamma_k \lambda^{2k+1}$.

The previous theorem is stated in a slightly different form than in [10]; see also the proof in Appendix B. Let us shortly comment on the sampling formula. Equation (8.544) in Gradshteyn and Ryzhik [15] states that $\lambda^H J_{1-H}(\lambda) = 2^{H-1}\Gamma(2-H)^{-1}\lambda\prod_{k=1}^{\infty}(1-\frac{\lambda^2}{\omega_k^2})$. Due to $\omega_k = -\omega_{-k}$, the sampling formula in Theorem 2 can thus be rewritten as (infinite) Lagrange interpolation (cf. also Young [41], Chapter 4). For $H = 1/2$, we have $J_{1/2}(\lambda) = \sqrt{2/(\pi\lambda)}\sin(\lambda)$ and $\omega_k = k\pi$. In this case, the theorem coincides with a shifted and scaled version of Shannon's sampling formula.

In the following, we describe the implications of the previous theorem for our analysis. As an immediate consequence of the sampling formula, we find that $\langle h, \phi_k\rangle_{L^2(\mu)} = \overline{\langle \phi_k, h\rangle}_{L^2(\mu)} = a_k h(2\omega_k)$ and

$$\langle \phi_k, \mathcal{F}(\mathbb{I}_t)\rangle_{L^2(\mu)} = a_k\overline{\mathcal{F}(\mathbb{I}_t)(2\omega_k)}. \tag{21}$$

By Lemma D.2(ii) (supplementary material [33]),

$$\mathcal{F}(g_k) = c'_H e^{-i\omega_k}\phi_k \qquad \text{with } c'_H := \frac{\Gamma(3/2-H)\sqrt{c_H}}{1+(\sqrt{2-2H}-1)\delta_{0,k}} \tag{22}$$

and g_k as in (16). The dependence of k on c'_H is irrelevant and we can therefore treat it as a constant. Since $\int_0^t e^{2i\omega_k u}\,du = \overline{\mathcal{F}(\mathbb{I}_t)(2\omega_k)}$, we have the following chain of equivalences

$$\begin{aligned} f_k(t) = e^{2i\omega_k t} \quad &\Leftrightarrow \quad F_{f_k}(t) = \overline{\mathcal{F}(\mathbb{I}_t)(2\omega_k)} \\ &\Leftrightarrow \quad Q(F_{f_k}) = a_k^{-1}\phi_k \\ &\Leftrightarrow \quad Q(F_{f_k}) = \mathcal{F}\biggl(\frac{e^{i\omega_k}g_k}{a_k c'_H}\biggr). \end{aligned} \tag{23}$$

This finally shows not only that f_k and $e^{i\omega_k}g_k/(a_k c'_H)$ are solutions to (15) but has also two important further implications for our analysis.

LEMMA 4. *The function sequences $(e^{2i\omega_k\cdot})_k$ and $(a_k e^{i\omega_k}g_k/c'_H)_k$ are biorthogonal Riesz bases of $L^2[0,1]$.*

PROOF. By Lemma 3, $(e^{2i\omega_k\cdot})_k$ is a Riesz basis of $L^2[0,1]$. From (22) and (18), $\langle g_k, e^{2i\omega_\ell\cdot}\rangle_{L^2[0,1]} = \mathcal{F}(g_k)(2\omega_\ell) = c'_H e^{-i\omega_k}\phi_k(2\omega_k)\delta_{k,\ell}$, with $\delta_{k,\ell}$ the Kronecker delta. Consequently, $(e^{2i\omega_k\cdot})_k$ and $(a_k e^{i\omega_k}g_k/c'_H)_k$ are biorthogonal implying that $(a_k e^{i\omega_k}g_k/c'_H)_k$ is a Riesz basis of $L^2[0,1]$ as well (cf. Young [41], page 36). $\square$

Notice that if $f = \sum_k \theta_k e^{2i\omega_k\cdot}$, then in general, $\theta_k \neq \langle f, e^{2i\omega_k\cdot}\rangle_{L^2[0,1]}$, since the basis functions are not orthogonal. Thanks to the previous lemma, the coefficients θ_k can be computed from f via

$$\theta_k = \frac{a_k e^{-i\omega_k}}{c'_H}\langle f, g_k\rangle_{L^2[0,1]}. \tag{24}$$

Moreover, (21) implies the following explicit characterization of the RKHS $\mathbb{H}$.

THEOREM 3.
$$\mathbb{H} = \left\{ F : F(t) = \sum_{k=-\infty}^{\infty} \theta_k \overline{\mathcal{F}(\mathbb{I}_t)(2\omega_k)}, \sum_{k=-\infty}^{\infty} \bigl(1+|k|\bigr)^{1-2H} |\theta_k|^2 < \infty \right\}.$$

PROOF. Since $(\phi_k)_k$ is an ONB of $\mathbb{M}$, $F \in \mathbb{H}$ if and only if $Q(F) = \sum_{k=-\infty}^{\infty} c_k \phi_k$, with $\sum_{k=-\infty}^{\infty} |c_k|^2 < \infty$. By (23), this is equivalent to $F(t) = \sum_{k=-\infty}^{\infty} \theta_k \overline{\mathcal{F}(\mathbb{I}_t)(2\omega_k)}$ with $\sum_{k=-\infty}^{\infty} |a_k \theta_k|^2 < \infty$. The result follows from (20). □

Recall the definition of $\Theta_H(\alpha, C)$ in (17) and set

$$\Theta_H(\alpha) := \bigl\{ f : \exists C = C(f) < \infty \text{ with } f \in \Theta_H(\alpha, C) \bigr\}. \tag{25}$$

From the first equivalence in (23), we obtain

COROLLARY 1. $\Theta_H(\frac{1}{2} - H) \subset \{ f : \int_0^{\cdot} f(u)\, du \in \mathbb{H} \}$.

To conclude this section, notice that we have derived a system of functions $(f_k, g_k, \phi_k)_k$ with $(f_k)_k$ and $(g_k)_k$ solving (15) and nearly orthogonalizing Θ and its source space and ϕ_k being an ONB of the underlying RKHS $\mathbb{H}$. The simultaneous (near)-orthogonalization of the spaces is the crucial tool to verify the second approximation condition of Theorem 1 on Sobolev balls. A slightly simpler characterization of the RKHS $\mathbb{H}$ can be given (cf. Picard [26], Theorem 6.12), but it remains unclear whether it can lead to a comparable simultaneous diagonalization. For more, see the discussion in Section 5.

4. Asymptotic equivalence: Main results.

4.1. *Asymptotic equivalence between the experiments* $\mathcal{E}_{1,n}(\Theta)$ *and* $\mathcal{E}_{2,n}(\Theta)$. In this section, we state the theorems establishing asymptotic equivalence between the experiments generated by the discrete regression model with fractional measurement noise $Y_{i,n} = f(\frac{i}{n}) + N_i^H$, $i = 1, \dots, n$ and its continuous counterpart $Y_t = \int_0^t f(u)\, du + n^{H-1} B_t^H$, $t \in [0, 1]$.

Proofs are provided in Appendix C (supplementary material [33]).

THEOREM 4. *Given* $H \in [1/2, 1)$. *Then, for any* $\alpha > 1/2$,
$$\mathcal{E}_{1,n}\bigl(\Theta_H(\alpha, C)\bigr) \simeq \mathcal{E}_{2,n}\bigl(\Theta_H(\alpha, C)\bigr).$$

THEOREM 5. *Given* $H \in (1/4, 1/2)$. *If* $\Theta_H^{(\mathrm{sym})}(\alpha, C) = \{ f \in \Theta_H(\alpha, C) : f = -f(1 - \cdot)\}$, *then, for any* $\alpha > (1 - H)/(H + 1/2) + H - 1/2$,
$$\mathcal{E}_{1,n}\bigl(\Theta_H^{(\mathrm{sym})}(\alpha, C)\bigr) \simeq \mathcal{E}_{2,n}\bigl(\Theta_H^{(\mathrm{sym})}(\alpha, C)\bigr).$$

In Section 4.3, we show that for any $H \in (0, 1)$, asymptotic equivalence fails to hold if $\alpha = 1/2$ or if $\alpha < 1 - H$. Therefore, for $H \geq 1/2$, the restriction $\alpha > 1/2$ is sharp in this sense. For $H < 1/2$, it is more difficult to prove asymptotic equivalence. If $H \in (1/4, 1/2]$, the minimal required smoothness in the previous result is slightly bigger than the lower bound $1 - H$ but always below $3/4$. In the case $H \downarrow 1/4$, the difference between the upper and lower smoothness assumption becomes arbitrarily small. For more on the case $H \leq 1/4$ and the restriction to $\Theta_H^{(\mathrm{sym})}(\alpha, C)$ for $H < 1/2$, see Section 4.3.

In the continuous fractional regression model (2), the noise term is $n^{H-1} \times$ fBM. Observe that the noise level n^{H-1} corresponds to an i.i.d. (regression) model with $N_n := n^{2-2H}$ observations. Thus, one can think about N_n as effective sample size of the problem. If $H < 1/2$, we find $N_n \gg n$ and if $H > 1/2$, $N_n \ll n$. The reason for that is the different correlation behavior in the discrete fractional regression model (1). If $H > 1/2$, any two observations in (1) are positively correlated, thus rewriting this as "independent" observations, we obtain $N_n \ll n$. On the contrary, if $H < 1/2$, observations are negatively correlated and errors cancel out, leading to smaller noise level and, therefore, $N_n \gg n$.

If short-range dependence is present, that is, $H < 1/2$, it has been argued in Johnson and Silverman [20], that for a specific choice of τ,

$$\widetilde{Y}_t = \int_0^t f(u)\, du + \tau n^{-1/2} B_t, \qquad t \in [0, 1], B \text{ a Brownian motion} \tag{26}$$

is a natural continuous approximation of the discrete model (1). The advantage is that this does not rely on the fGN and might hold for any stationary noise process with short-range dependence. For model (1), however, the asymptotically equivalent continuous model $Y_t = \int_0^t f(u)\, du + n^{H-1} B_t^H$, has the smaller noise level n^{H-1}, implying that (26) leads to a loss of information.

To conclude the discussion, let us relate the Sobolev ellipsoids $\Theta_H(\alpha, C)$ to classical Sobolev spaces. From that, we can establish asymptotic equivalence on a space that depends not on the choice of the basis. For any positive integer β, define

$$\begin{aligned} &\mathrm{Sob}_H(\beta, \widetilde{C}) \\ &\quad := \Big\{ f \in L^2[0, 1] : f^{(\beta-1)} \text{ is absolutely continuous and real-valued,} \\ &\qquad \|f\|_{L^2[0,1]} + \|f^{(\beta)}\|_{L^2[0,1]} \leq \widetilde{C}, \int_0^1 f^{(\ell)}(s)\big(s - s^2\big)^{1/2-H}\, ds = 0, \\ &\qquad \ell = 1, \ldots, \beta \Big\}. \end{aligned}$$

LEMMA 5. *Given $H \in (0, 1)$. Then, for any positive integer β and $\widetilde{C} < \infty$, there exists a finite constant C, such that*

$$\mathrm{Sob}_H(\beta, \widetilde{C}) \subset \Theta_H(\beta, C).$$

The proof is delayed until Appendix C (supplementary material [33]). For $H = \frac{1}{2}$, the constraints $\int_0^1 f^{(\ell)}(s)(s - s^2)^{1/2-H}\,ds = 0$, $\ell = 1, \dots, \beta$ simplify to the periodic boundary conditions $f^{(q)}(0) = f^{(q)}(1)$, $q = 0, \dots, \beta - 1$. In this case, Lemma 5 is well known; cf. Tsybakov [36], Lemma A.3. In the important case $\beta = 1$, the constraint in $\mathrm{Sob}_H(\beta, \widetilde{C})$ is satisfied whenever $f = f(1 - \cdot)$. If we restrict further to these functions, the definition of $\mathrm{Sob}_H(1, \widetilde{C})$ does not depend on the Hurst index H anymore. As a consequence of Theorem 4 and the embedding, we obtain the following.

COROLLARY 2. *Let $H \in [\frac{1}{2}, 1)$. Then, for any finite constant $\widetilde{C}$,*

$$\mathcal{E}_{1,n}\big(\mathrm{Sob}_H(1, \widetilde{C})\big) \simeq \mathcal{E}_{2,n}\big(\mathrm{Sob}_H(1, \widetilde{C})\big).$$

4.2. *Construction of equivalent sequence model.* Let $\Theta_H(\alpha)$ be as in (25) and write $f = \sum_{k=-\infty}^{\infty} \theta_k e^{2i\omega_k \cdot}$ for a generic element of $\Theta_H(\alpha)$. Define the experiment $\mathcal{E}_{3,n}(\Theta_H(\alpha)) = (\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}), (P_{3,f}^n : f \in \Theta_H(\alpha)))$. Here, $P_{3,f}^n$ is the joint distribution of $(Z_k)_{k \geq 0}$ and $(Z'_k)_{k \geq 1}$ with

$$Z_k = \sigma_k^{-1} \operatorname{Re}(\theta_k) + n^{H-1} \varepsilon_k \quad \text{and} \quad Z'_k = \sigma_k^{-1} \operatorname{Im}(\theta_k) + n^{H-1} \varepsilon'_k, \tag{27}$$

$(\varepsilon_k)_{k \geq 0}$ and $(\varepsilon'_k)_{k \geq 1}$ being two independent vectors of i.i.d. standard normal random variables. The scaling factors are $\sigma_k := a_k/\sqrt{2}$ for $k \geq 1$ and $\sigma_0 := a_0$, with $(a_k)_k$ as defined in (19).

THEOREM 6. $\mathcal{E}_{2,n}(\Theta_H(\frac{1}{2} - H)) = \mathcal{E}_{3,n}(\Theta_H(\frac{1}{2} - H))$.

The proof relies completely on RKHS theory and can be found in Appendix C (supplementary material [33]). To illustrate the result, let us give an informal derivation here. First, we may rewrite the continuous fractional regression model (2) in differential form $dY_t = f(t)\,dt + n^{H-1}\,dB_t^H$. Recall the definition of g_k in (16) and notice that $g_k = \overline{g_{-k}}$. Now, let $k \geq 1$ and consider the random variables $Z_k = \int \overline{(e^{i\omega_k} g_k(t) + e^{i\omega_{-k}} g_{-k}(t))}\,dY_t/(\sqrt{2}c'_H)$. Using (24),

$$Z_k := \frac{\sqrt{2}}{a_k} \operatorname{Re}(\theta_k) + \frac{n^{H-1}}{\sqrt{2}}(\eta_k + \eta_{-k}), \qquad k = 1, 2, \dots$$

with $\eta_k := \int \overline{e^{i\omega_k} g_k(t)}\,dB_t^H/c'_H$. From Pipiras and Taqqu [27], equation (3.4), $E[\int h_1(t)\,dB_t^H \cdot \int h_2(t)\,dB_t^H] = \langle \mathcal{F}(h_1), \mathcal{F}(h_2)\rangle_{L^2(\mu)}$ and together with (22) and the fact that $(\phi_k)_k$ is an ONB of $\mathbb{M}$,

$$\mathbb{E}[\eta_k \eta_\ell] = \langle \phi_\ell, \phi_k \rangle_{L^2(\mu)} = \delta_{k,\ell},$$

where $\delta_{k,\ell}$ denotes the Kronecker delta. Hence, $\varepsilon_k = (\eta_k + \eta_{-k})/\sqrt{2} \sim \mathcal{N}(0, 1)$, i.i.d. for $k = 1, 2, \dots$ and $Z_k = \sigma_k^{-1} \operatorname{Re}(\theta_k) + n^{H-1} \varepsilon_k$. Similarly, we can construct Z_0 and Z'_k, $k \geq 1$. This shows informally that the continuous model

$\mathcal{E}_{2,n}(\Theta_H(\frac{1}{2}-H))$ is not less informative than observing (27). The other direction follows from the completeness of $(g_k)_k$.

As an application of the previous theorem, let us study estimation of θ in the model

$$Y_{i,n} = \theta + N_i^H, \qquad i = 1, \dots, n,$$

that is, model (1) with $f = \theta$ constant. To estimate θ, one could consider the average $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n Y_{i,n} = \theta + n^{1-H}\xi$, with ξ a standard normal random variable. By Theorem 4 and Theorem 6, we find, however, that for $H \in [\frac{1}{2}, 1)$, the model is asymptotic equivalent to observing $\sigma_0 Z_0 = \theta + \sigma_0 n^{H-1}\varepsilon_0$. Recall that σ_0 depends on H. Clearly, $\sigma_0 \leq 1$ and $\sigma_0 = 1$ for $H = 1/2$. For $H > 1/2$, numerical evaluation shows that σ_0 is a bit smaller than 1 implying that the estimator $\widehat{\theta}$ can be slightly improved. Instead of the sample average, the construction of the asymptotic equivalence uses a weighted sum over the $Y_{i,n}$'s, where the weights are chosen proportional to $g_0(i/n)$ (excluding $i = n$) with $g_0(s) = (s - s^2)^{1/2-H}$ (cf. also the proof of Theorem 4). For $H > 1/2$, this gives far more weight to observations close to the boundaries. The choice of the weighting function is a consequence of the biorthogonality in Lemma 4.

To conclude the section, let us link the sequence model (27) to inverse problems. For that recall that by (20), $\sigma_k \propto |k|^{1/2-H}$. In the case of long-range dependence ($H > 1/2$), $\sigma_k \to 0$ and the problem is well-posed. The noise level, however, is n^{H-1} which is of larger order than the classical $n^{-1/2}$. The opposite happens if $H < 1/2$. In this case, the noise level is $o(n^{-1/2})$ but on the same time we face an inverse problem with degree of ill-posedness $1/2 - H$. In order to illustrate the effects of ill-posedness and noise level, let us study estimation of $f \in \Theta_H(\beta, C)$ with smoothness index $\beta > 0$ known. Then we might use the Fourier series estimator $\widehat{f} = \sum_{|k| \leq M_n} \widehat{\theta}_k e^{2i\omega_k \cdot}$, with M_n some cut-off frequency. Here, $\widehat{\theta}_k = \sigma_k(Z_k + iZ_k')$, for $k \geq 0$, with $Z_0' := 0$. For negative k set $\widehat{\theta}_k = \overline{\widehat{\theta}_{-k}}$. By Lemma 3, $(e^{2i\omega_k \cdot})_k$ is a Riesz basis for $L^2[0,1]$ and from the frame inequality

$$\mathbb{E}[\|\widehat{f} - f\|_{L^2[0,1]}^2] \lesssim \mathbb{E}\Bigg[\sum_{|k| \leq M_n} |\theta_k - \widehat{\theta}_k|^2\Bigg] + \sum_{|k| > M_n} |\theta_k|^2.$$

Choosing $M_n = O(n^{-(1-H)/(\beta+1-H)})$, the rate becomes $n^{-2\beta(1-H)/(\beta+1-H)}$ in accordance with Wang [38] and, for $\beta = 2$, $H \geq 1/2$, Hall and Hart [18]. Surprisingly, faster rates can be obtained if H is small. The ill-posedness is overcompensated by the gain in the noise level.

4.3. *Necessary conditions.* In this section, we provide necessary minimal smoothness assumptions for asymptotic equivalence.

The result below shows that asymptotic equivalence cannot hold for the (smaller) Sobolev space $\Theta_H^{(\mathrm{sym})}(\alpha, C)$ if $\alpha = 1/2$ or $\alpha < 1 - H$. This shows that $\alpha \geq \frac{1}{2} \vee (1 - H)$ is necessary in Theorems 4 and 5.

LEMMA 6. *For any $C > 0$, if $\alpha = 1/2$ or if $\alpha < 1 - H$, then*

$$\mathcal{E}_{1,n}\big(\Theta_H^{(\mathrm{sym})}(\alpha, C)\big) \not\simeq \mathcal{E}_{2,n}\big(\Theta_H^{(\mathrm{sym})}(\alpha, C)\big).$$

PROOF. We discuss the two cases (I) $\alpha = 1/2$ and (II) $\alpha < 1 - H$, separately.

(I) Define $f_{0,n} = cn^{-1/2}\sin(\omega_n(2\cdot -1))$ and $f_{1,n} = cn^{-1/2}\sin(\omega_{2n}(2\cdot -1))$. Because of $\sin(\omega_k(2\cdot -1)) = (e^{-i\omega_k}e^{2i\omega_k\cdot} - e^{i\omega_k}e^{-2i\omega_k\cdot})/(2i)$, the constant c can and will be chosen such that $f_{0,n}, f_{1,n} \in \Theta_H^{(\mathrm{sym})}(\frac{1}{2}, C)$ for all n. From the equivalent sequence space representation (27), and since by (20), $\sigma_k \asymp |k|^{1/2-H}$, we see that $f_{0,n}$ and $f_{1,n}$ are separable in experiment $\mathcal{E}_{2,n}(\Theta_H^{(\mathrm{sym})}(\frac{1}{2}, C)) = \mathcal{E}_{3,n}(\Theta_H^{(\mathrm{sym})}(\frac{1}{2}, C))$ with positive probability. Recall that P_f^n denotes the distribution of the observation vector in experiment $\mathcal{E}_{1,n}(\Theta_H^{(\mathrm{sym})}(\frac{1}{2}, C))$. It is enough to show that

$$d_{\mathrm{KL}}\big(P_{f_{0,n}}^n, P_{f_{1,n}}^n\big) \to 0, \tag{28}$$

since this implies that there exists no test in $\mathcal{E}_{1,n}(\Theta_H^{(\mathrm{sym})}(\frac{1}{2}, C))$ distinguishing $f_{0,n}$ and $f_{1,n}$ asymptotically with positive probability. Let $x_n = (n + \frac{1}{4}(1-2H))\pi$ and notice that $\sin(2x_n\frac{\ell}{n}) = \sin(2x_{2n}\frac{\ell}{n})$ for all integer ℓ. With Lemma A.2, Taylor approximation and Lemma D.1 in [33],

$$\begin{aligned}
&d_{\mathrm{KL}}\big(P_{f_0}^n, P_{f_{1,n}}^n\big) \\
&\quad\lesssim \big(n^{2-2H} \vee n\big) \max_{j=1,\dots,n}\left| f_{0,n}\left(\frac{j}{n}\right) - f_{1,n}\left(\frac{j}{n}\right)\right|^2 \\
&\quad= \big(n^{2-2H} \vee n\big)\frac{c^2}{n} \max_{j=1,\dots,n}\left|\sin\left(\omega_n\frac{2j-n}{n}\right) - \sin\left(x_n\frac{2j-n}{n}\right)\right. \\
&\qquad\qquad\qquad\qquad \left. + \sin\left(x_{2n}\frac{2j-n}{n}\right) - \sin\left(\omega_{2n}\frac{2j-n}{n}\right)\right|^2 \\
&\quad\lesssim n^{-2H-1} \vee n^{-2} \to 0.
\end{aligned}$$

Hence, (28) holds and this completes the proof for $\alpha = 1/2$.

(II) Let $\alpha < 1 - H$ and choose k_n as the smallest integer larger than $\log n$. Define $f_{0,n} = 0$ and $f_{1,n} = cn^{H-1}k_n^{1/2-H}[\sin(\omega_{k_n}(2\cdot -1)) - \sin(\omega_{k_n+n}(2\cdot -1))]$. If $c = c(C)$ is chosen sufficiently small, then $f_{0,n}, f_{1,n} \in \Theta_H^{(\mathrm{sym})}(\alpha, C)$. If $f = f_{1,n}$, the k_nth coefficient is $\theta_{k_n} = cn^{H-1}k_n^{1/2-H}e^{-i\omega_{k_n}}/(2i)$ and in experiment $\mathcal{E}_{3,n}(\Theta_H^{(\mathrm{sym})}(\alpha, C))$, we observe $Z_{k_n} = \sigma_{k_n}^{-1}\operatorname{Re}(\theta_{k_n}) + n^{H-1}\varepsilon_{k_n}$ and $Z'_{k_n} = \sigma_{k_n}^{-1}\operatorname{Im}(\theta_{k_n}) + n^{H-1}\varepsilon'_{k_n}$. Due to $|\sigma_{k_n}| \asymp k_n^{1/2-H}$ the functions $f_{0,n}$ and $f_{1,n}$ can be distinguished in experiment $\mathcal{E}_{2,n}(\Theta_H^{(\mathrm{sym})}(\alpha, C)) = \mathcal{E}_{3,n}(\Theta_H^{(\mathrm{sym})}(\alpha, C))$ with positive probability. In contrast, by the same argument as for case (I), we find

$d_{\mathrm{KL}}(P^n_{f_{0,n}}, P^n_{f_{1,n}}) \to 0$ in $\mathcal{E}_{1,n}(\Theta_H^{(\mathrm{sym})}(\alpha, C))$. This shows that asymptotic equivalence does not hold. $\square$

The previous lemma shows essentially that the approximation condition (i) in Theorem 1, which controls the discretization effects of the regression function, is necessary. Next, we give a heuristic argument explaining why asymptotic equivalence requires also an approximation condition in the RKHS, that is, why condition (ii) in Theorem 1 is necessary. Since under condition (i), $\mathcal{E}_{1,n}(\Theta) \simeq \mathcal{E}_{5,n}(\Theta)$ [cf. (11)], it is sufficient to study asymptotic equivalence between $\mathcal{E}_{5,n}(\Theta)$ and $\mathcal{E}_{2,n}(\Theta)$. In $\mathcal{E}_{5,n}(\Theta)$, we observe $\widetilde{Y}_t = L(t|\mathbf{F}_{f,n}) + n^{H-1}B_t^H$, $t \in [0, 1]$ with $L(\cdot|\mathbf{F}_{f,n})$ as in (9). With the change of measure formula in Lemma 1, it is not hard to see that for the likelihood ratio test $\phi_n = \mathbb{I}_{\{dQ_f^n/dQ_0^n \le 1\}}$, $Q_f^n \phi_n + Q_0^n(1 - \phi_n) \le 2\exp(-\frac{1}{8}n^{2-2H}\|F_f\|_{\mathbb{H}}^2)$. Thus, in $\mathcal{E}_{2,n}(\Theta)$, we can distinguish with positive probability between f and 0 if $n^{1-H}\|F_f\|_{\mathbb{H}}$ is larger than some constant. With the same argument, we can distinguish with positive probability between f and 0 in experiment $\mathcal{E}_{5,n}(\Theta)$ provided that $n^{1-H}\|L(\cdot|\mathbf{F}_{f,n})\|_{\mathbb{H}}$ is larger than some constant. For asymptotic equivalence, we should have therefore that $n^{1-H}\|F_f - L(\cdot|\mathbf{F}_{f,n})\|_{\mathbb{H}}$ is small uniformly over Θ and this is just a reformulation of condition (ii) (cf. also the proof of Proposition A.2).

Necessary conditions for $H < 1/2$. Let us derive a heuristic indicating that for $H < 1/2$, asymptotic equivalence cannot hold on the unrestricted Sobolev ball $\Theta_H(\alpha, C)$. This motivates the use of $\Theta_H^{(\mathrm{sym})}(\alpha, C)$ in Theorem 5. Moreover, we give an argument why asymptotic equivalence fails for $H \le 1/4$. First, recall that from (11), the discrete regression experiment $\mathcal{E}_{1,n}(\Theta)$ is asymptotically equivalent to $\mathcal{E}_{5,n}(\Theta)$ under approximation condition (i) of Theorem 1. Therefore, $\mathcal{E}_{1,n}(\Theta) \simeq \mathcal{E}_{5,n}(\Theta)$, whenever Θ is a Hölder ball with index larger $1 - H$, for example. To study $\mathcal{E}_{1,n}(\Theta) \not\simeq \mathcal{E}_{2,n}(\Theta)$, it is thus sufficient to show $\mathcal{E}_{5,n}(\Theta) \not\simeq \mathcal{E}_{2,n}(\Theta)$. In $\mathcal{E}_{5,n}(\Theta)$, we observe $\widetilde{Y}_t = L(t|\mathbf{F}_{f,n}) + n^{H-1}B_t^H$, $t \in [0, 1]$. Using (9), $L(\cdot|\mathbf{F}_{f,n})$ is a linear combination of the functions $K(\cdot, \frac{j}{n}) = \mathrm{Cov}(B_\cdot^H, B_{j/n}^H) = \frac{1}{2}(|\cdot|^{2H} + |\frac{j}{n}|^{2H} - |\cdot - \frac{j}{n}|^{2H})$. Thus, we can write $L(\cdot|\mathbf{F}_{f,n}) = \sum_{j=1}^n \gamma_{j,n} K(\cdot, \frac{j}{n})$ for suitable weights $(\gamma_{j,n})_j$. In the continuous regression experiment $\mathcal{E}_{2,n}(\Theta)$, $Y_t = \int_0^t f(u)\,du + n^{H-1}B_t^H$, $t \in [0, 1]$ is observed. Informally, we can consider differentials

$$\text{in } \mathcal{E}_{5,n}(\Theta): \qquad d\widetilde{Y}_t = \partial_t L(t|\mathbf{F}_{f,n})\,dt + n^{H-1}\,dB_t^H, \qquad t \in [0, 1],$$

$$\text{in } \mathcal{E}_{2,n}(\Theta): \qquad dY_t = f(t)\,dt + n^{H-1}\,dB_t^H, \qquad t \in [0, 1].$$

The experiments $\mathcal{E}_{5,n}(\Theta)$ and $\mathcal{E}_{2,n}(\Theta)$ will be close if $\partial_t L(t|\mathbf{F}_{f,n})$ well approximates $f(t)$. Notice, however, that for $H < 1/2$, the function $t \mapsto \partial_t L(t|\mathbf{F}_{f,n}) = \sum_{j=1}^n \gamma_{j,n}\partial_t K(t, \frac{j}{n})$ has singularities at $t = \frac{j}{n}$ for $j = 0, 1, \ldots, n$ and for $H \le 1/4$, it is not in L^2 anymore. More precisely, $t \mapsto \partial_t K(t, \frac{j}{n})$ has a singularities at $t = 0$ and $t = \frac{j}{n}$. Since $\partial_t L(t|\mathbf{F}_{f,n})$ and $f(t)$ must be of the same order, we have $\sum_{j=1}^n \gamma_{j,n} = O(1)$. Typically, $|\gamma_{j,n}| \lesssim 1/n$ for $j = 1, \dots, n$ and this downweights the singular behavior of $\partial_t L(t|\mathbf{F}_{f,n})$ at j/n for $j = 1, \dots, n$ but not at $t = 0$. Since all summands contribute to the singularity at $t = 0$, we find

$$\partial_t L(t|\mathbf{F}_{f,n}) \sim H \sum_{j=1}^n \gamma_{j,n} t^{2H-1} \qquad \text{for } t \downarrow 0.$$

If $a = a_n \downarrow 0$, then $\widetilde{Y}_{a/n} = \frac{1}{2} a^{2H} \sum_{j=1}^n \gamma_{j,n} n^{-2H}(1 + o(1)) + \eta/(na^H)$ and $Y_{a/n} = o(n^{-1/2}) + \eta/(na^H)$ with $\eta \sim \mathcal{N}(0, 1)$. For $H < 1/2$, $n^{-2H} \gg n^{-1}$ and the paths of $(\widetilde{Y}_t)_t$ and $(Y_t)_t$ can be distinguished if $a \downarrow 0$ not too fast. Asymptotic equivalence will thus not hold unless we additionally assume that $\sum_{j=1}^n \gamma_{j,n} = o(n^{2H-1})$. Via (9), this can be expressed as a constraint on f indicating why restriction of the Sobolev ball $\Theta_H(\alpha, C)$ in Theorem 5 is necessary for $H < 1/2$.

Second, let us give a heuristic argument which shows that asymptotic equivalence fails to hold for $H \leq 1/4$. We compare the decay of Fourier coefficients for $(\widetilde{Y}_t)_t$ and $(Y_t)_t$. The absolute values of the Fourier coefficients $q(a, k) := \int_0^1 e^{2\pi kt} \operatorname{sign}(t - a)|t - a|^{2H-1} \, dt = e^{2\pi i ka} \int_{-a}^{1-a} e^{2\pi ikt} \operatorname{sign}(t)|t|^{2H-1} \, dt$ decay like $|k|^{-2H}$ if $a \in [0, 1]$. In particular, the Fourier coefficients depend in a nontrivial way on a. Write $p(a, k) = q(a, k)|k|^{2H}$. Then the kth Fourier coefficient of $t \mapsto \partial_t L(t|\mathbf{F}_{f,n})$ is $|k|^{-2H} \sum_{j=1}^n \gamma_{j,n} p(\frac{j}{n}, k)$, already assuming that $\sum_{j=1}^n \gamma_{j,n} = 0$. The decay is unaffected by the smoothness of f. To compute the Fourier coefficients of the fBM, we find using [27], equation (3.4) that $\int_0^1 e^{2\pi ikt} \, dB_t^H \sim \mathcal{N}(0, \|e^{2\pi ik\cdot}\|_H^2)$. Since $\|e^{2\pi ik\cdot}\|_H^2 = c_H/2\pi \int |\mathcal{F}(e^{2\pi ik\cdot})(\lambda)|^2 |\lambda|^{1-2H} \, d\lambda \asymp |k|^{1-2H}$ (for the last approximation consider a neighborhood of $\lambda = 2\pi k$), roughly,

$$\text{in } \mathcal{E}_{5,n}(\Theta): \qquad \int_0^1 e^{2\pi ikt} \, d\widetilde{Y}_t \approx k^{-2H} + n^{H-1} k^{1/2-H} \xi_k, \qquad k = 1, 2, \dots,$$

$$\text{in } \mathcal{E}_{2,n}(\Theta): \qquad \int_0^1 e^{2\pi ikt} \, dY_t \approx \int_0^1 e^{2\pi ikt} f(t) \, dt + n^{H-1} k^{1/2-H} \xi_k, \qquad k = 1, 2, \dots,$$

where ξ_k are centered, normally distributed random variables with variance bounded in k. If $k \asymp n$, then in $\mathcal{E}_{5,n}(\Theta)$, the Fourier coefficients are in first order $n^{-2H} + n^{-1/2} \xi_k$, whereas if f is smooth, we observe $o(n^{-1/2}) + n^{-1/2} \xi_k$ in $\mathcal{E}_{2,n}(\Theta)$. If $H \leq 1/4$, we can therefore distinguish $\mathcal{E}_{5,n}(\Theta)$ and $\mathcal{E}_{2,n}(\Theta)$. The only possibility to avoid this is to add further constraints to Θ that ensure that $\sum_{j=1}^n \gamma_{j,n} q(\frac{j}{n}, k)$ is small. Since the argument above applies to any $k \asymp n$ and $q(\frac{j}{n}, k)$ depends on k, we exclude more and more subspaces. This indicates $\mathcal{E}_{5,n}(\Theta) \not\simeq \mathcal{E}_{2,n}(\Theta)$ and, therefore, also $\mathcal{E}_{1,n}(\Theta) \not\simeq \mathcal{E}_{2,n}(\Theta)$.

4.4. *Generalization.* Let us shortly remark on possible extensions of our method. First notice that Theorem 1 relies on the specific self-similarity properties of fractional Brownian motion and a straightforward generalization is only partially possible (cf. Remark A.1 below). Passing from the continuous model to the sequence space representation, however, can be stated in a much more general framework.

Generalizing $\mathcal{E}_{2,n}(\Theta)$, denote by $\mathcal{G}_{2,n}(\Theta) = (\mathcal{C}[0,1], \mathcal{B}(\mathcal{C}[0,1]), (Q_{2,f}^n : f \in \Theta))$ the experiment with $Q_{2,f}^n$ the distribution of

$$Y_t = \int_0^t f(u)\,du + n^{-\beta} X_t, \qquad t \in [0,1]. \tag{29}$$

Here, f is the regression function, $\beta > 0$, and $X := (X_t)_{t\in[0,1]}$ is a continuous, centered Gaussian process with stationary increments. In particular, this contains model (2) if X is a fBM and $\beta = 1 - H$. The aim of this section is to construct an equivalent sequence space representation for (29).

With the Karhunen–Loeve expansion of X, this can be done in a straightforward way. The drawback of this approach is that closed form formulas for the basis functions are known only for some specific choices of X. Therefore, we propose a different construction leading again to nonharmonic Fourier series. This approach is based on the one-to-one correspondence between mass distributions of vibrating strings and certain measures which was developed in Kreĭn [21], de Branges [6], Dym and McKean [9] and Dzhaparidze et al. [11].

Let us sketch the construction. Recall that X is a continuous, centered Gaussian process with stationary increments and let $K_X(s,t) = \operatorname{Cov}(X_s, X_t)$. We have the representation $K_X(s,t) = \int_{-\infty}^{\infty} \mathcal{F}(\mathbb{I}_s)(\lambda)\overline{\mathcal{F}(\mathbb{I}_t)(\lambda)}\,d\mu_X(\lambda)$, where μ_X is a symmetric Borel measure on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty}(1+\lambda)^{-2}\,d\mu_X(\lambda) < \infty$ (cf. Doob [8], Section XI.11). If $\mathbb{M}_X = \overline{\operatorname{span}}\{\mathcal{F}(I_t) : t \in [0,1]\} \subset L^2(\mu_X)$, then the RKHS $\mathbb{H}_X$ associated to the Gaussian process X is given by

$$\mathbb{H}_X = \{F : \exists F^* \in \mathbb{M}_X, \text{ such that } F(t) = \langle F^*, \mathcal{F}(\mathbb{I}_t)\rangle_{L^2(\mu_X)}, \forall t \in [0,1]\}$$

and we have the isometric isomorphism

$$\begin{aligned} Q_X : (\mathbb{H}_X, \langle\cdot,\cdot\rangle_{\mathbb{H}_X}) &\to (\mathbb{M}_{\mu_X}, \langle\cdot,\cdot\rangle_{L^2(\mu_X)}), \\ F &\mapsto F^*. \end{aligned}$$

In order to extend Theorem 2, the crucial observation is that there is a one-to-one correspondence between measures μ_X with $\int_{-\infty}^{\infty}(1+\lambda)^{-2}\,d\mu_X(\lambda) < \infty$ and mass distribution functions, say m, of a vibrating string. Computation of m is quite involved and thus omitted here. For a detailed explanation, see [11]. If m is continuously differentiable and strictly positive, we have the following theorem.

THEOREM 7. *Denote by* $(\lambda,\omega) \mapsto S(\lambda,\omega)$ *the reproducing kernel of* $\mathbb{M}_X$. *There exist real numbers* $\cdots < \nu_{-1} < \nu_0 = 0 < \nu_1 < \cdots$ *such that* $(S(\nu_k,\cdot)/$

$\sqrt{S(\nu_k,\nu_k)})_k$ *is an ONB of* $\mathbb{M}_X$ *and*

$$h=\sum_{k=-\infty}^{\infty} h(\nu_k)\frac{S(\nu_k,\cdot)}{S(\nu_k,\nu_k)} \qquad \textit{for all } h\in\mathbb{M}_X$$

with convergence of the sum in $L^2(\mu_X)$. *Furthermore, for any* $k\in\mathbb{Z}$, $\nu_k=-\nu_{-k}$, $S(\nu_k,\nu_k)=S(\nu_{-k},\nu_{-k})$ *and* $|\nu_k|=2|k|\pi(1+o(1))$, *for* $|k|\to\infty$.

PROOF. This follows largely from Dzhaparidze et al. [11], Theorem 3.5 and Zareba [42], Lemma 2.8.7. It remains to show that $\nu_k=-\nu_{-k}$ and $S(\nu_k,\nu_k)=S(\nu_{-k},\nu_{-k})$. Notice that by the reproducing property $S(\nu_k,\nu_k)=\|S(\nu_k,\lambda)\|^2_{L^2(\mu_X)}\geq 0$. To see that $\nu_k=-\nu_{-k}$, observe that from [11], equation (2.2), $A(x,\lambda)=A(x,-\lambda)$. Therefore, $B(x,\lambda)=-\frac{1}{\lambda}A^+(x,\lambda)$ (cf. [11], Section 2.3) satisfies $B(x,\lambda)=-B(x,-\lambda)$. The numbers $\cdots<\nu_{-1}<\nu_0=0<\nu_1<\cdots$ are the zeros of $B(x,\cdot)$ for a specific value of x (cf. [11], Theorem 2.10, equation (3.2) and Theorem 3.5). Hence, $\nu_{-k}=-\nu_k$. Using [11], equation (2.10),

$$\begin{aligned} K_T(\nu_k,\lambda)&=\frac{A(x(T),\nu_k)B(x(T),\lambda)}{\pi(\lambda-\nu_k)}\\ &=\frac{A(x(T),\nu_{-k})B(x(T),-\lambda)}{\pi(-\lambda-\nu_{-k})}\\ &=K_T(\nu_{-k},\lambda). \end{aligned}$$

Together with [11], equation (3.1) this shows $S(\nu_k,\nu_k)=S(\nu_{-k},\nu_{-k})$. The proof is complete. $\square$

Write $\psi_{k,X}=S(\nu_k,\cdot)/\sqrt{S(\nu_k,\nu_k)}$ and $\rho_k=S(\nu_k,\nu_k)^{-1/2}$ and notice that by the preceding theorem, $\rho_k=\rho_{-k}$. The sampling formula reads then $h=\sum_{k=-\infty}^{\infty}\rho_k h(\nu_k)\psi_{k,X}$ and $\langle h,\psi_{k,X}\rangle_{L^2(\mu_X)}=\rho_k h(\nu_k)$. Generalizing (23), we find that if $F(t)=\sum_{k=-\infty}^{\infty}\theta_k\mathcal{F}(\mathbb{I}_t)(\nu_k)$, then $F^*=Q_X(F)=\sum_{k=-\infty}^{\infty}\theta_k\rho_k^{-1}\psi_{k,X}$. Analogously to Theorem 3 and Corollary 1, we have the characterization

$$\mathbb{H}_X=\left\{F:F(t)=\sum_{k=-\infty}^{\infty}\theta_k\overline{\mathcal{F}(\mathbb{I}_t)(\nu_k)},\ \sum_{k=-\infty}^{\infty}\rho_k^2|\theta_k|^2<\infty\right\}$$

and that $\Theta_X:=\{f:f=\sum_{k=-\infty}^{\infty}\theta_k e^{i\nu_k\cdot},\theta_k=\overline{\theta}_{-k}\sum_{k=-\infty}^{\infty}\rho_k^2|\theta_k|^2<\infty\}$ is a subset of $\{f:\int_0^{\cdot} f(u)\,du\in\mathbb{H}_X\}$. Due to $\nu_k=-\nu_{-k}$, a function is real-valued within this class iff $\theta_k=\overline{\theta_{-k}}$ for all k.

Generalizing $\mathcal{E}_{2,n}$, define the experiment $\mathcal{G}_{3,n}(\Theta_X)=(\mathbb{R}^{\mathbb{Z}},\mathcal{B}(\mathbb{R}^{\mathbb{Z}}),(Q^n_{3,f}:f\in\Theta_X))$. Here, $Q^n_{3,f}$ is the joint distribution of $(Z_k)_{k\geq 0}$ and $(Z'_k)_{k\geq 1}$ with

$$Z_k=\sigma_k^{-1}\operatorname{Re}(\theta_k)+n^{H-1}\varepsilon_k \quad\text{and}\quad Z'_k=\sigma_k^{-1}\operatorname{Im}(\theta_k)+n^{H-1}\varepsilon'_k,$$

$(\varepsilon_k)_{k\geq 0}$ and $(\varepsilon'_k)_{k\geq 1}$ are two independent vectors of Gaussian noise. The scaling factors are $\sigma_k:=\rho_k/\sqrt{2}$ for $k\geq 1$ and $\sigma_0:=\rho_0$.

THEOREM 8. $\mathcal{G}_{2,n}(\Theta_X) = \mathcal{G}_{3,n}(\Theta_X)$.

The proof is the same as for Theorem 6.

The advantage of this approach is that by following the program outlined in [11] or [42], closed form expressions for σ_k can be derived even if the Karhunen–Loeve decomposition is unknown. The difference is that f is not expanded in an ONB but again as a nonharmonic Fourier series with respect to $(e^{i\nu_k \cdot})_k$. By Theorem 7, $|\nu_k| = 2|k|\pi(1+o(1))$. Therefore, the functions $(e^{i\nu_k \cdot})_k$ are "close" to the harmonic basis $(e^{2\pi i k \cdot})_k$.

5. Discussion. In this section, we give a short summary of related work on regression under dependent noise and asymptotic equivalence.

Optimal rates of convergence for regression under long-range dependent noise were first considered by Hall and Hart [18] using kernel estimators.

Inspired by the asymptotic equivalence result of Brown and Low [2], Wang [38] makes the link between discrete regression under dependent noise and experiment $\mathcal{E}_{2,n}(\Theta)$, in which the path of the integral of f is observed plus a scaled fBM. Passing from the discrete to the continuous model is done by adding uninformative Brownian bridges. It is argued that this will lead to good approximations of the continuous path. From an asymptotic equivalence perspective this interpolation scheme leads, however, to dependencies in the errors which are difficult to control. To prove Theorem 1, we used instead the interpolation function

$$t \mapsto L(t|\mathbf{x}_n) = \mathbb{E}\big[B_t^H \,|\, B_{\ell/n}^H = x_\ell, \ell = 1, \dots, n\big], \tag{30}$$

which has the advantage that the interpolated discrete observations have the exact error distribution of the continuous model resulting in the equivalence $\mathcal{E}_{4,n}(\Theta) = \mathcal{E}_{5,n}(\Theta)$ in the proof of Theorem 1. The use of the interpolation function (30) for asymptotic equivalence appears implicitly already in the proof of Theorem 2.2 in Reiß [29].

Approximation of discrete regression under dependent errors by $\int_0^{\cdot} f(u)\,du + n^{H-1}B_{\cdot}^H$ and sequence model representations were further studied in Johnstone and Silverman [20] and more detailed in Johnstone [19].

Donoho [7] investigates the wavelet-vaguelette decomposition for inverse problems. Since this is very close to the simultaneous orthogonalization presented in Section 3, the connection is discussed in more detail here. Let $f_k = e^{2i\omega_k \cdot}$ and $\breve{g}_k = a_k e^{i\omega_k} g_k / c'_H$. By Lemma 4,

$$(f_k)_k \quad \text{and} \quad (\breve{g}_k)_k \qquad \text{are biorthogonal bases of } L^2[0,1].$$

Next, define the operator $Sh = \sqrt{c_H/2\pi}|\cdot|^{1/2-H}\mathcal{F}(h)$ and notice that (15) can be rewritten as $f = S^* S g$ with S^* the adjoint operator. By Theorem 2, the functions $\lambda \mapsto \psi_k(\lambda) = \sqrt{c_H/2\pi}|\lambda|^{1/2-H}\phi_k(\lambda)$ are orthonormal with respect to $L^2(\mathbb{R})$. Using (22) and $f_k = S^* S(a_k^{-2}\breve{g}_k)$, we have the quasi-singular value decomposition

$$S\breve{g}_k = a_k \psi_k \quad \text{and} \quad S^*\psi_k = a_k f_k \qquad \text{for all } k \in \mathbb{Z}.$$

This should be compared to the wavelet-vaguelette decomposition in [7], Section 1.5 which proposes, within a general framework, to start with a wavelet decomposition $(\psi_{j,k})_{j,k}$, replacing the orthonormal functions $(\psi_k)_k$ above. This allows to consider more general spaces than Sobolev balls but cannot be applied here as we need to work in the RKHS $\mathbb{H}$. The restriction to $\mathbb{H}$ implies that the underlying functions ϕ_k (or $\phi_{j,k}$ in a multi-resolution context) have to be functions in $\mathbb{M}$, that is, the closed linear span of $\{\mathcal{F}(\mathbb{I}_t) : t \in [0, 1]\}$ in $L^2(\mu)$. Because fBM on $[0, 1]$ is considered, the index t has to be in $[0, 1]$ and $\mathbb{M}$ is very difficult to characterize. In particular, it is strictly smaller than $L^2(\mu)$. This shows that finding an ONB of $\mathbb{M}$ (cf. Theorem 2) is highly nontrivial and it remains unclear in which sense $\mathbb{M}$ could admit a multi-resolution decomposition.

Besides that, Brown and Low [2] and Nussbaum [25] established nonparametric asymptotic equivalence as own research field. Since then, there has been considerable progress in this area. Asymptotic equivalence for regression models was further generalized to random design in Brown et al. [1], non-Gaussian errors in Grama and Nussbaum [16] and higher-dimensional settings in Carter [3] and Reiß [28]. Rohde [30] considers periodic Sobolev classes, improving on condition (5) in this case. Carter [4] establishes asymptotic equivalence for regression under dependent errors. The result, however, is derived under the strong assumption that the noise process is completely decorrelated by a wavelet decomposition. Multiscale representations that nearly whiten fBM are known (cf. Meyer et al. [24], Section 7), but it is unclear whether fBM admits an exact wavelet decomposition. One possibility to extend the result to regression under fractional noise is to give up on orthogonality and to deal with nearly orthogonal wavelet decompositions instead. This, however, causes various new issues that are very delicate and technical. One might view the methods developed in Golubev et al. [14] and Reiß [29] as first steps toward such a theory, as both deal with similar problems, however in very specific settings.

APPENDIX A: PROOFS FOR SECTION 2

PROOF OF LEMMA 2. Write $\nu_f = dP_f/dP_0$ and $\nu_g = dP_g/dP_0$. Moreover, denote by $E_0[\cdot]$ expectation with respect to P_0. We have $E_0[\nu_f] = 1$ and by Lemma 1,

$$d_{\mathrm{KL}}(P_f, P_g) = E_0\left[\log\left(\frac{\nu_f}{\nu_g}\right)\nu_f\right] = E_0\big[(Uf - Ug)\nu_f\big] - \frac{1}{2}\|f\|_{\mathbb{H}}^2 + \frac{1}{2}\|g\|_{\mathbb{H}}^2.$$

Note that $g = g_1 + g_2$ with $g_1 := \langle g, f\rangle_{\mathbb{H}}\|f\|_{\mathbb{H}}^{-2} f$ and $g_2 := g - g_1$. Clearly, $\operatorname{Cov}(Uf, Ug_2) = 0$ and since Uf, Ug_2 are Gaussian, ν_f and Ug_2 are independent. For a centered normal random variable ξ with variance σ^2,

$$\mathbb{E}\big[\xi \exp(\xi)\big] = \partial_t \mathbb{E}\big[\exp(t\xi)\big]\big|_{t=1} = \partial_t \exp\left(\frac{\sigma^2}{2}t^2\right)\bigg|_{t=1} = \sigma^2 \exp\left(\frac{\sigma^2}{2}\right)$$

and hence,

$$E_0\big[(Uf - Ug)\nu_f\big] = \left(1 - \frac{\langle g, f\rangle_{\mathbb{H}}}{\|f\|_{\mathbb{H}}^2}\right) E_0\big[(Uf)\nu_f\big] = \|f\|_{\mathbb{H}}^2 - \langle g, f\rangle_{\mathbb{H}}.$$

Plugging this into the formula for $d_{\mathrm{KL}}(P_f, P_g)$, the result follows. $\square$

For a similar result, cf. Gloter and Hoffmann [13], Lemma 8.

A.1. Completion of Theorem 1. The remaining parts for the proof of Theorem 1 follows from Propositions A.1 and A.2 below.

If $\mathbf{X} = (X_1, \ldots, X_n)$ is a stationary process with spectral density f, it is well known that the eigenvalues of the Toeplitz matrix $\mathrm{Cov}(\mathbf{X})$ lie between the minimum and maximum of f on $[0, \pi]$. These bounds become trivial if $f(\lambda)$ converges to 0 or ∞ for $\lambda \downarrow 0$. The first lemma gives a sharper lower bound for the smallest eigenvalue of a Toeplitz matrix which is of independent interest.

LEMMA A.1. *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a stationary process with spectral density f and denote by $\lambda_i(\cdot)$ the ith eigenvalue. Then*

$$\lambda_n\big(\mathrm{Cov}(\mathbf{X})\big) \geq \left(1 - \frac{1}{\pi}\right) \inf_{\lambda \in [1/n, \pi]} f(\lambda).$$

PROOF. For any vector $v = (v_1, \ldots, v_n)$,

$$v^t \,\mathrm{Cov}(\mathbf{X}) v = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k=1}^n v_k e^{ik\lambda} \right|^2 f(\lambda)\, d\lambda = \frac{1}{\pi} \int_0^{\pi} \left| \sum_{k=1}^n v_k e^{ik\lambda} \right|^2 f(\lambda)\, d\lambda.$$

In particular, $\|v\|^2 = v^t v = \frac{1}{\pi} \int_0^{\pi} |\sum_{k=1}^n v_k e^{ik\lambda}|^2 \, d\lambda$. The estimate,

$$v^t \,\mathrm{Cov}(\mathbf{X}) v \geq \left(\|v\|^2 - \frac{1}{\pi} \int_0^{1/n} \left| \sum_{k=1}^n v_k e^{ik\lambda} \right|^2 d\lambda \right) \inf_{\lambda \in [1/n, \pi]} f(\lambda)$$

together with $|\sum_{k=1}^n v_k e^{ik\lambda}|^2 \leq (\sum_{k=1}^n |v_k|)^2 \leq n\|v\|^2$ yields the result. $\square$

Along the line of the proof, one can also show that $\sup_{\lambda \in [1/n, \pi)} f(\lambda) + \frac{n}{\pi} \int_0^{1/n} f(\lambda)\, d\lambda$ is an upper bound of the eigenvalues.

LEMMA A.2. *For a vector $v \in \mathbb{R}^n$, let P_v denote the distribution of $(Y_{1,n}, \ldots, Y_{n,n})$ with $Y_{i,n} = v_i + N_i^H$, $i = 1, \ldots, n$ and $(N_i^H)_i$ fGN. Then there exists a constant $c = c(H)$, such that*

$$d_{\mathrm{KL}}(P_v, P_w) \leq c\big(n^{1-2H} \vee 1\big)(v - w)^t (v - w).$$

PROOF. Denote the spectral density of fractional Gaussian noise with Hurst index H by f_H. fGN is stationary and from the explicit formula of f_H (cf. Sinaĭ [34]), we find that $f_H(\lambda) \sim c_H \lambda^{1-2H}$ for $\lambda \downarrow 0$, and that f_H is bounded away from zero elsewhere. Using Lemma A.1,

$$\lambda_n\big(\operatorname{Cov}(\mathbf{Y}_n)\big) \geq \left(1 - \frac{1}{\pi}\right) \inf_{\lambda \in [1/n, \pi]} f_H(\lambda) \gtrsim n^{2H-1} \wedge 1.$$

From the general formula for the Kullback–Leibler distance between two multivariate normal random variables (or by applying Lemma 2), we obtain

$$d_{\mathrm{KL}}(P_1, P_2) = \tfrac{1}{2}(\mu_1 - \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2)$$

whenever P_1 and P_2 denote the probability distributions corresponding to $\mathcal{N}(\mu_1, \Sigma)$ and $\mathcal{N}(\mu_2, \Sigma)$, respectively. This proves the claim. $\square$

As a direct consequence of (57) in [33], Lemma A.2, and condition (i) of Theorem 1, we obtain the following.

PROPOSITION A.1. *Given $H \in (0, 1)$ suppose that the parameter space Θ satisfies condition* (i) *of Theorem* 1. *Then*,

$$\mathcal{E}_{1,n}(\Theta) \simeq \mathcal{E}_{4,n}(\Theta).$$

REMARK A.1. The previous proposition can be easily extended to more general stationary noise processes and does not require RKHS theory as only condition (i) of Theorem 1 is involved.

PROPOSITION A.2. *Given $H \in (0, 1)$ suppose that the parameter space Θ satisfies condition* (ii) *of Theorem* 1. *Then*

$$\mathcal{E}_{5,n}(\Theta) \simeq \mathcal{E}_{2,n}(\Theta).$$

PROOF. Recall that $\mathbb{H}$ denotes the RKHS associated with $(B_t^H)_{t \in [0,1]}$. From the Moore–Aronszajn theorem, we can conclude $L(\cdot|\mathbf{F}_{f,n}) \in \mathbb{H}$ since by (9) it is a linear combination of functions $K(\cdot, j/n)$. Condition (ii) of Theorem 1 ensures $F_f \in \mathbb{H}$. Define $\mathbb{L}_n \subset \mathbb{H}$ as the space of functions

$$\sum_{j=1}^n \alpha_j K\left(\cdot, \frac{j}{n}\right) \qquad \text{with } (\alpha_1, \ldots, \alpha_n)^t \in \mathbb{R}^n.$$

From the reproducing property in the Moore–Aronszajn theorem and the interpolation property of $L(\cdot|\mathbf{F}_{f,n})$, it follows that $L(\cdot|\mathbf{F}_{f,n}) \in \mathbb{L}_n$ is the projection of F on $\mathbb{L}_n$, that is,

$$\langle F, h\rangle_{\mathbb{H}} = \big\langle L(\cdot|\mathbf{F}_{f,n}), h\big\rangle_{\mathbb{H}} \qquad \text{for all } h \in \mathbb{L}_n.$$

In particular, $\langle F, L(\cdot|\mathbf{F}_{f,n})\rangle_{\mathbb{H}} = \|L(\cdot|\mathbf{F}_{f,n})\|_{\mathbb{H}}^2$, and thus

$$\|F - L(\cdot|\mathbf{F}_{f,n})\|_{\mathbb{H}}^2 \le \|F - h\|_{\mathbb{H}}^2 \qquad \text{for all } h \in \mathbb{L}_n.$$

Together with Lemma 2, (57) and condition (ii) in Theorem 1,

$$\begin{aligned}
\Delta\big(\mathcal{E}_{5,n}(\Theta), \mathcal{E}_{2,n}(\Theta)\big)^2 &\le \sup_{f\in\Theta} d_{\mathrm{KL}}\big(Q_{5,f}^n, Q_f^n\big) \\
&= \frac{1}{2} n^{2-2H} \sup_{f\in\Theta} \|F - L(\cdot|\mathbf{F}_{f,n})\|_{\mathbb{H}}^2 \\
&= \frac{1}{2} n^{2-2H} \sup_{f\in\Theta} \inf_{(\alpha_1,\dots,\alpha_n)^t\in\mathbb{R}^n} \bigg\| F_f - \sum_{j=1}^n \alpha_j K\Big(\cdot, \frac{j}{n}\Big)\bigg\|_{\mathbb{H}}^2 \\
&\to 0.
\end{aligned}$$

This proves the assertion. □

APPENDIX B: PROOFS FOR SECTION 3

PROOF OF LEMMA 3. From Kadec's $\frac{1}{4}$-theorem (cf. Young [41], Theorem 14), we conclude that $(e^{2i\omega_k\cdot})_k$ is a Riesz basis if $|\omega_k/\pi - k| < 1/4$ for all $k \in \mathbb{Z}$. Using Lemma D.1(ii) and (iii) (supplementary material [33]), we find that $|\omega_k/\pi - k| \le \frac{1}{8} \vee \frac{|1-2H|}{4} < \frac{1}{4}$ and this proves the claim. □

REMARK B.1. The constant $\frac{1}{4}$ in Kadec's $\frac{1}{4}$-theorem is known to be sharp (cf. [41], Section 3.3). Since $\omega_k = (k + \frac{1}{4}(1-2H))\pi + O(1/k)$ by Lemma D.1(i), the LHS comes arbitrarily close to this upper bound at the boundaries $H \downarrow 0$ and $H \uparrow 1$.

PROOF OF THEOREM 2. Recall that $c_H = \sin(\pi H)\Gamma(2H+1)$. In a first step, we prove the identity

$$\frac{c_H}{2\pi} = \frac{2^{4H-3} H\Gamma(H+1/2)\Gamma(3-2H)}{(1-H)\Gamma^2(1-H)\Gamma(3/2-H)}. \tag{31}$$

Application of the replication formula $\Gamma(1-z)\Gamma(z) = \pi/\sin(\pi z)$ for $z = H$ and the duplication formula $\Gamma(2z) = 2^{2z-1}\pi^{-1/2}\Gamma(z+1/2)\Gamma(z)$ for $z = H$ and $z = 1 - H$ gives

$$\begin{aligned}
\frac{c_H}{2\pi} &= \frac{\sin(\pi H)\Gamma(2H+1)}{2\pi} = \frac{H\sin(\pi H)\Gamma(2H)}{\pi} \\
&= \frac{H\Gamma(2H)}{\Gamma(1-H)\Gamma(H)} = \frac{2^{2H-1} H\Gamma(H+1/2)}{\sqrt{\pi}\Gamma(1-H)} \\
&= \frac{2^{2H-2} H\Gamma(H+1/2)\Gamma(3-2H)}{\sqrt{\pi}(1-H)\Gamma(1-H)\Gamma(2-2H)} = \frac{2^{4H-3} H\Gamma(H+1/2)\Gamma(3-2H)}{(1-H)\Gamma^2(1-H)\Gamma(3/2-H)}.
\end{aligned}$$

This proves (31).

Next, let us show that $(\phi_k)_k$ is $L^2(\mu)$-normalized, that is $\|\phi_k\|_{L^2(\mu)} = 1$. This is immediately clear for $k = 0$ since (cf. Luke [23], Section 13.2)

$$\int_0^\infty \left|J_{1-H}(\lambda)\right|^2 \lambda^{-1}\, d\lambda = 1/(2-2H).$$

To compute the normalization constant for $k \neq 0$, the last equality in the proof of [10], Theorem 7.2 gives $\|S_1(2\omega_k, \cdot)\|^2_{L^2(\mu)} = \sigma^{-2}(\omega_k)$, where for $k \neq 0$, using identity (31), $\sigma^{-2}(\omega_k) = \pi c_H^{-1} 2^{2H-2} |\omega_k|^{2H} J_{-H}^2(\omega_k)$ and $S_1(2\omega_k, 2\lambda) = p(H, \omega_k) e^{i(\lambda-\omega_k)} \lambda^H J_{1-H}(\lambda)/(\lambda - \omega_k)$ with $p(H, \omega_k) := \pi c_H^{-1} 2^{2H-2} \omega_k^H \times J_{-H}(\omega_k)$. By definition of ϕ_k we can write $\phi_k = (\pi/c_H)^{1/2} 2^{H-1} \overline{S_1(2\omega_k, \lambda)}/p(H, \omega_k)$ and

$$\begin{aligned}
\|\phi_k\|^2_{L^2(\mu)} &= \frac{\pi}{c_H} p\big(H, \omega_k\big)^{-2} 2^{2H-2} \|S_1(2\omega_k, \cdot)\|^2_{L^2(\mu)} \\
&= \frac{\pi}{c_H} p(H, \omega_k)^{-2} 2^{2H-2} \sigma^{-2}(\omega_k) \\
&= 1.
\end{aligned}$$

Since $\lambda \mapsto \lambda^H J_{1-H}(\lambda)$ is an odd function, we obtain $\overline{S_1(2\omega_{-k}, -\lambda)} = \overline{S_1(2\omega_k, \lambda)}$ implying $\phi_k = \psi_{-k}(-\cdot)$ with ψ_k as in Theorem 7.2 of [10]. Notice that the space $\mathcal{L}_T$ is defined as the closure of the functions $\overline{\mathcal{F}(\mathbb{I}_t)}$, $t \in [0, 1]$, whereas $\mathbb{M}$ is the closure of the functions $\mathcal{F}(\mathbb{I}_t)$, $t \in [0, 1]$, Therefore, a function h is in $\mathbb{M}$ if and only if $h(-\cdot)$ is in $\mathcal{L}_T$. This shows that $\{\phi_k : k \in \mathbb{Z}\}$ is a basis of $\mathbb{M}$ and that the sampling formula $h = \sum_k a_k h(2\omega_k)\phi_k$ is equivalent to the corresponding result in Theorem 7.2 of [10].

We obtain the expression for a_0, using Lebedev [22], Formula (5.16.1), $\lim_{\lambda \to 0} (\lambda/2)^{-\alpha} J_\alpha(\lambda) = \Gamma(\alpha+1)^{-1}$, for all $\alpha \geq 0$.

Furthermore, $a_k = a_{-k}$ follows from $\omega_k = -\omega_{-k}$ and the fact that a_k^{-1} is just a constant times the derivative of $\lambda \mapsto \lambda^H J_{1-H}(\lambda)$ evaluated at ω_k. Since $\lambda \mapsto \lambda^H J_{1-H}(\lambda)$ is an odd and smooth function, the derivative must be an even function (cf. the remarks after Theorem 2) and this gives $a_k = a_{-k}$.

To prove (20), let us first derive some inequalities. The symbol $\lesssim$ means up to a constant depending on H only.

From the asymptotic expansion of Bessel functions (cf. Gradshteyn and Ryzhik [15], formulas 8.451.1 and 7), $\sum_{r=0}^2 |J_{r-H}(\lambda)| \lesssim |\lambda|^{-1/2}$, for all $|\lambda| \geq \omega_1/2$. Together with Lemma D.2(ii) (supplementary material [33]) applied for $k = 0$ and the inequality $|\mathcal{F}(g_0)| \leq \|g_0\|_{L^1(\mathbb{R})} < \infty$, we find $|J_{1-H}(\lambda)| \lesssim \lambda^{1-H} \wedge \lambda^{-1/2}$. Using Taylor expansion and the recursion formula $2\frac{d}{d\lambda} J_{1-H}(\lambda) = J_{-H}(\lambda) - J_{2-H}(\lambda)$, for any $k \geq 1$ and any $\lambda \in [\omega_k/2, 2\omega_k]$,

$$\left|\frac{J_{1-H}(\lambda)}{\lambda - \omega_k}\right| \leq \frac{1}{2} \sup_{\xi \in [\omega_k/2, 2\omega_k]} \left|J_{-H}(\xi) + J_{2-H}(\xi)\right| \lesssim |\omega_k|^{-1/2}. \tag{32}$$

In a second step of the proof, we show that for $k \neq 0$, $G_k(-\infty, \infty) \leq \text{const.} \times |k|^{1/2-H}$, where

$$G_k(a,b) := \int_a^b \left| \lambda^{-H} \frac{J_{1-H}(\lambda)}{\lambda - \omega_k} \right| d\lambda, \qquad k = 1, 2, \ldots.$$

Notice that it is enough to prove $G_k(0, \infty) \lesssim |k|^{1/2-H}$ for $k = 1, 2, \ldots$. Decompose $[0, \infty) = [0, \omega_k/2] \cup [\omega_k/2, 2\omega_k] \cup [2\omega_k, \infty)$. To bound $G_k(0, \omega_k/2)$, use that $|\lambda - \omega_k| \geq \omega_k/2$ and that $|J_{1-H}(\lambda)| \lesssim \lambda^{1-H} \wedge \lambda^{-1/2}$; to bound $G_k(\omega_k/2, 2\omega_k)$, use (32); to bound $G_k(2\omega_k, \infty)$, use that $|\lambda - \omega_k| \geq \lambda/2$ and $|J_{1-H}(\lambda)| \lesssim \lambda^{-1/2}$. Together with Lemma D.1, this shows that $G_k(0, \infty) \lesssim |k|^{1/2-H}$.

Next, we show that for $k \neq 0$, $\gamma_{k,H} = e^{i\omega_k}/(\sqrt{2-2H} - 1)$,

$$a_k = \gamma_{k,H} a_0 + \frac{\sqrt{c_H}\omega_k}{2^H \sqrt{\pi}} \int e^{i(\omega_k - \lambda)} \frac{\lambda^{H-1} J_{1-H}(\lambda)}{\lambda - \omega_k} |\lambda|^{1-2H} \, d\lambda. \tag{33}$$

Notice that $t^{-1}\mathcal{F}(\mathbb{I}_t)(\lambda) \to 1$ for $t \to 0$ and λ fixed. Since $\phi_k(\lambda) = \gamma_{k,H}\phi_0(\lambda) + 2\omega_k \lambda^{-1}\phi_k(\lambda)$, we have by (21),

$$\begin{aligned} a_k &= \lim_{t \to 0} \left\langle \phi_k, \frac{1}{t}\mathcal{F}(\mathbb{I}_t) \right\rangle_{L^2(\mu)} \\ &= \gamma_{k,H} a_0 + \lim_{t \to 0} \frac{c_H 2^{1-2H}}{\pi} \int_{-\infty}^{\infty} \frac{\omega_k}{\lambda} \phi_k(2\lambda) \frac{1}{t} \overline{\mathcal{F}(\mathbb{I}_t)(2\lambda)} |\lambda|^{1-2H} \, d\lambda. \end{aligned}$$

Because of $|\frac{1}{t}\overline{\mathcal{F}(\mathbb{I}_t)}| \leq 1$, the integrand can be bounded by

$$\text{const.} \times \left| \lambda^{-2H} \phi_k(\lambda) \right| \lesssim \left| \frac{\lambda^{-H} J_{1-H}(\lambda)}{\lambda - \omega_k} \right|.$$

The $L^1(\mathbb{R})$-norm of this function is smaller than a constant multiple of $G_k(-\infty, \infty)$ and we may apply dominated convergence, that is, $\lim_{t \to 0}$ and the integral can be interchanged. The definition of ϕ_k gives then (33).

To prove (20), notice that the lower bound follows from (19), (32) and Lemma D.1 (supplementary material [33]). For the upper bound, we can restrict ourselves to $k = 1, \ldots$ since $a_k = a_{-k}$. The statement follows from (33) and $G_k(-\infty, \infty) \lesssim |k|^{1/2-H}$. □

Acknowledgement. The author would like to thank Harrison Zhou for bringing the problem to his attention and helpful discussion. The article was revised based on valuable comments by Yazhen Wang, an Associate Editor and three anonymous referees.

SUPPLEMENTARY MATERIAL

Asymptotic equivalence for regression under fractional noise (DOI: 10.1214/14-AOS1262SUPP; .pdf). The supplement contains proofs for Section 4, some technical results and a brief summary of the Le Cam distance.

REFERENCES

[1] BROWN, L. D., CAI, T. T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688–707. MR1922538

[2] BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958

[3] CARTER, A. V. (2006). A continuous Gaussian approximation to a nonparametric regression in two dimensions. *Bernoulli* **12** 143–156. MR2202326

[4] CARTER, A. V. (2009). Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. *J. Probab. Stat.* Art. ID 275308, 19. MR2602894

[5] CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems* **24** 034004, 19. MR2421941

[6] DE BRANGES, L. (1968). *Hilbert Spaces of Entire Functions*. Prentice Hall, Englewood Cliffs, NJ. MR0229011

[7] DONOHO, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126. MR1325535

[8] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York. MR0058896

[9] DYM, H. and MCKEAN, H. (1976). *Gaussian Processes, Function Theory, and the Inverse Spectral Measure*. Academic Press, San Diego, CA.

[10] DZHAPARIDZE, K. and VAN ZANTEN, H. (2005). Krein's spectral theory and the Paley–Wiener expansion for fractional Brownian motion. *Ann. Probab.* **33** 620–644. MR2123205

[11] DZHAPARIDZE, K., VAN ZANTEN, H. and ZAREBA, P. (2005). Representations of fractional Brownian motion using vibrating strings. *Stochastic Process. Appl.* **115** 1928–1953. MR2178502

[12] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems*. Kluwer Academic, Dordrecht. MR1408680

[13] GLOTER, A. and HOFFMANN, M. (2004). Stochastic volatility and fractional Brownian motion. *Stochastic Process. Appl.* **113** 143–172. MR2078541

[14] GOLUBEV, G. K., NUSSBAUM, M. and ZHOU, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* **38** 181–214. MR2589320

[15] GRADSHTEYN, I. S. and RYZHIK, I. M. (1994). *Table of Integrals, Series, and Products*, Russian ed. Academic Press, Boston, MA. MR1243179

[16] GRAMA, I. and NUSSBAUM, M. (2002). Asymptotic equivalence for nonparametric regression. *Math. Methods Statist.* **11** 1–36. MR1900972

[17] GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York. MR0599175

[18] HALL, P. and HART, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351. MR1084984

[19] JOHNSTONE, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statist. Sinica* **9** 51–83. MR1678881

[20] JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* **59** 319–351. MR1440585

[21] KREĬN, M. G. (1953). On some cases of effective determination of the density of an inhomogeneous cord from its spectral function. *Doklady Akad. Nauk SSSR* (*N.S.*) **93** 617–620. MR0061247

[22] LEBEDEV, N. N. (1972). *Special Functions and Their Applications*. Dover, New York. MR0350075

[23] LUKE, Y. L. (1962). *Integrals of Bessel Functions*. McGraw-Hill, New York. MR0141801

[24] MEYER, Y., SELLAN, F. and TAQQU, M. S. (1999). Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion. *J. Fourier Anal. Appl.* **5** 465–494. MR1755100

[25] NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

[26] PICARD, J. (2011). Representation formulae for the fractional Brownian motion. In *Séminaire de Probabilités XLIII. Lecture Notes in Math.* **2006** 3–70. Springer, Berlin. MR2790367

[27] PIPIRAS, V. and TAQQU, M. S. (2000). Integration questions related to fractional Brownian motion. *Probab. Theory Related Fields* **118** 251–291. MR1790083

[28] REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982. MR2435461

[29] REISS, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39** 772–802. MR2816338

[30] ROHDE, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statist. Decisions* **22** 235–243. MR2125610

[31] SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional Integrals and Derivatives*. Gordon and Breach, New York. MR1347689

[32] SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York. MR1280932

[33] SCHMIDT-HIEBER, J. (2014). Supplement to "Asymptotic equivalence for regression under fractional noise." DOI:10.1214/14-AOS1262SUPP.

[34] SINAĬ, JA. G. (1976). Self-similar probability distributions. *Theory Probab. Appl.* **21** 64–80.

[35] TAUTENHAHN, U. and GORENFLO, R. (1999). On optimal regularization methods for fractional differentiation. *Z. Anal. Anwend.* **18** 449–467. MR1701364

[36] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

[37] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. Collect.* **3** 200–222. IMS, Beachwood, OH. MR2459226

[38] WANG, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24** 466–484. MR1394972

[39] WATSON, G. N. (1995). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, Cambridge. Reprint of the second (1944) edition. MR1349110

[40] YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Vol. I*. Springer, New York. MR0893393

[41] YOUNG, R. M. (1980). *An Introduction to Nonharmonic Fourier Series*. Academic Press, New York. MR0591684

[42] ZAREBA, P. (2007). Representations of Gaussian processes with stationary increments. Ph.D. thesis, Vrije Univ. Amsterdam.

MATHEMATICAL INSTITUTE
LEIDEN UNIVERSITY
NIELS BOHRWEG 1
2333 CA LEIDEN
THE NETHERLANDS
E-MAIL: schmidthieberaj@math.leidenuniv.nl