



# Asymptotic equivalence for nonparametric regression with dependent errors: Gauss–Markov processes

Holger Dette<sup>1</sup> · Martin Kroll<sup>1</sup>

Received: 24 October 2021 / Revised: 24 February 2022 / Accepted: 22 March 2022 /

Published online: 9 May 2022

© The Institute of Statistical Mathematics, Tokyo 2022

## Abstract

For the class of Gauss–Markov processes we study the problem of asymptotic equivalence of the nonparametric regression model with errors given by the increments of the process and the continuous time model, where a whole path of a sum of a deterministic signal and the Gauss–Markov process can be observed. We derive sufficient conditions which imply asymptotic equivalence of the two models. We verify these conditions for the special cases of Sobolev ellipsoids and Hölder classes with smoothness index  $> 1/2$  under mild assumptions on the Gauss–Markov process. To give a counterexample, we show that asymptotic equivalence fails to hold for the special case of Brownian bridge. Our findings demonstrate that the well-known asymptotic equivalence of the Gaussian white noise model and the nonparametric regression model with i.i.d. standard normal errors (see Brown and Low (Ann Stat 24:2384–2398, 1996)) can be extended to a setup with general Gauss–Markov noises.

**Keywords** Asymptotic equivalence · Nonparametric regression · Dependent errors · Gauss–Markov process · Triangular kernel

## 1 Introduction

In the seminal paper (Brown and Low, 1996), the authors establish asymptotic equivalence of the nonparametric regression model with discrete observations

---

This research has been supported in part by the research grant DE 502/27-1 of the German Research Foundation (DFG)

---

✉ Holger Dette  
holger.dette@rub.de

<sup>1</sup> Fakultät für Mathematik, Lehrstuhl für Stochastik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Deutschland

$$Y_{i,n} = f\left(\frac{i}{n}\right) + \eta_i, \quad i = 1, \dots, n, \quad (1)$$

and the continuous time model defined by the stochastic differential equation

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in [0, 1], \quad (2)$$

where  $\eta_1, \dots, \eta_n$  are independent, standard Gaussian random variables,  $W$  denotes a standard Brownian motion (that is,  $dW$  is white noise) and  $f$  is the unknown nonparametric drift satisfying a smoothness assumption. Equation (2) is commonly referred to as the *Gaussian white noise model* and serves as an important benchmark model in nonparametric statistics. Often, due to the absence of discretization effects, statistical methods are easier to analyze in model (2) than in (1). Asymptotic equivalence between the two models then suggests that theoretical results obtained in the Gaussian white noise model hold true in the more realistic model (1) as well.

Since the contribution of Brown and Low numerous authors have worked on the problem of establishing asymptotic equivalence of various models from different perspectives. For example, (Grama and Nussbaum, 1998) investigate nonparametric generalized linear models, and (Brown and Zhang, 1998) prove nonequivalence when the smoothness of the function class is equal to  $1/2$ . (Brown et al., 2002) and (Reiß, 2008) study the random design in the one-dimensional and multivariate case, respectively (see also Carter, 2006 for some results in models with a multivariate fixed design). The general framework in Brown and Low (1996) is already formulated for heteroscedastic errors and (Carter, 2007) shows asymptotic equivalence for unknown variances and design density. We also refer to the work of (Reiß, 2011) and (Meister, 2011) who propose rate-optimal estimators of the volatility function and sharp minimax constants in the functional linear regression model as an application of asymptotic equivalence, respectively.

This list is by no means complete, but a common feature of most publications in this field consists in the fact that the random variables in the corresponding discrete nonparametric regression model are assumed to be independent. (Grama and Neumann, 2006) consider a nonparametric autoregression model but still use an i.i.d. assumption for the innovations. (Golubev et al., 2010) study a stationary process and show asymptotic equivalence in the context of spectral density estimation. However, when it comes to regression the assumption of i.i.d. errors is often made for the theoretical analysis. Motivated by the work of (Johnstone and Silverman, 1997) and (Johnstone, 1999), (Carter, 2009) considers asymptotic equivalence under the assumption that the noise process is given by a wavelet composition with stochastically independent coefficients. Another notable exception is the recent work of (Schmidt-Hieber, 2014), who considers the nonparametric regression model (1) with fractional Gaussian noise (fGN) and establishes the asymptotic equivalence to a model of the form (2), where the error process  $\frac{1}{\sqrt{n}}W_t$  is replaced by a fractional Brownian motion (fBM)  $n^{H-1}B_t^H$  with Hurst parameter  $H \in \left(\frac{1}{4}, 1\right)$  over periodic Sobolev ellipsoids containing sufficiently smooth functions. It is also shown that asymptotic equivalence fails to hold for certain combinations of Hurst parameter

and smoothness index leading to sharp results concerning the smoothness requirement in the case  $H \in \left[\frac{1}{2}, 1\right)$ .

The purpose of the present paper is to provide an essentially distinct way of investigating nonparametric regression models with dependent errors for asymptotic equivalence. Instead of replacing the Brownian motion by a fractional Brownian motion, we consider here arbitrary Gauss–Markov processes as error processes in model (2). To be precise let  $\Xi = (\Xi_t)_{t \in [0,1]}$  be such a Gauss–Markov process with initial state  $\Xi_0 = 0$ , and consider the continuous time model

$$Y_t = \int_0^t f(s)ds + \frac{1}{\sqrt{n}}\Xi_t, \quad t \in [0, 1]. \quad (3)$$

Assuming equidistant design points, the candidate regression model for asymptotic equivalence is given by

$$Y_{i,n} = f\left(\frac{i}{n}\right) + \sqrt{n}\xi_{i,n}, \quad i = 1, \dots, n, \quad (4)$$

where  $\xi_{i,n} = \Xi_{i/n} - \Xi_{(i-1)/n}$  denote the increments of the process  $\Xi$ . Note that in general the observation errors  $\xi_{i,n}$  are not uncorrelated. However, for the special case of Brownian motion, that is  $\Xi = W$ , Eqs. (1) and (4) are equivalent in distribution since Brownian motion has independent increments and satisfies the scaling property that  $W(\cdot)$  and  $c^{-1/2}W(c \cdot)$  have the same distribution for any  $c > 0$ . The main results of the present paper establish the asymptotic equivalence of the models (3) and (4) for a wide class of Gauss–Markov processes for functional parameters  $f$  belonging to a Sobolev or Hölder class of sufficiently smooth functions.

There are different ways to prove asymptotic equivalence between regression and white noise experiments in the literature. The original paper of Brown and Low (1996) considers the case where the regression  $f$  is an element of a function class, say  $\Theta$ , and uses the key assumption

$$\lim_{n \rightarrow \infty} \sup_{f \in \Theta} n \int_0^1 \left(f(t) - \tilde{f}_n(t)\right)^2 dt = 0, \quad (5)$$

where  $\tilde{f}_n(\cdot) = \sum_{j=1}^n f(j/n) \mathbf{1}_{((j-1)/n, j/n]}(\cdot)$  denotes a piecewise constant approximation of  $f$  and  $\mathbf{1}_A$  is the indicator function of the set  $A$ . These authors introduce an intermediate set of random variables that forms a sufficient statistic for the white noise model with the function  $f$  replaced with  $\tilde{f}_n$ . Then, the Hellinger distance between this sufficient statistic and the regression experiment with discrete sampling locations is shown to converge to zero. On the other hand, condition (5) guarantees that the Le Cam distance between the white noise models with parameters  $f$  and  $\tilde{f}_n$ , respectively, vanishes as  $n \rightarrow \infty$ . Note that this approach can be used to show asymptotic equivalence but cannot provide the optimal rate of convergence for the Le Cam distance between the two experiments. This issue has been solved by Rohde (2004), who considers a Gaussian sequence space model as an intermediate experiment between the two experiments of interest. (Schmidt-Hieber, 2014) generalizes the conditions

in Brown and Low (1996) and formalizes them in the framework of reproducing kernel Hilbert spaces (RKHSs), which is a suitable setup for the investigation in the case of fractional Brownian motion. Our analysis of the asymptotic equivalence for models with Gauss–Markov errors will also be based on the RKHS framework. As an essential ingredient we will use a characterization of Gauss–Markov processes introduced by Doob (1949) to derive an explicit representation of the RKHSs associated with the Gauss–Markov processes under consideration, which can be used to develop sufficient conditions for the asymptotic equivalence of the models (3) and (4).

The remaining part of the paper is organized as follows. In Sect. 2 we introduce Gauss–Markov processes and recap the characterization of such processes introduced by Doob (1949) that will be pivotal for our approach. Roughly speaking, these processes can be characterized by the property that the corresponding covariance kernel is *triangular* (see Sect. 2.1 for a precise definition). In Sect. 2.2 we study the reproducing kernel Hilbert spaces (RKHSs) associated with Gauss Markov processes and derive representations of these spaces via Hilbert space isomorphisms into a space of square-integrable functions. In Sect. 3 we recall the basic elements of Le Cam theory and provide sufficient conditions on the class of potential functions  $f$  and the Gauss–Markov process that imply asymptotic equivalence. The characterizations of the RKHSs are used in Sect. 4 where we establish asymptotic equivalence of the models (3) and (4) under mild assumptions on the Gauss–Markov process in model (3) for Sobolev ellipsoids and Hölder classes. In Sect. 4.3 we discuss a different approach to derive such results, which is not based on RKHS theory. It requires a (slightly) stronger assumption than those used in this paper and applies results from the seminal paper (Brown and Low, 1996) to suitable transformations of the experiments of interest.

Finally, in Sect. 4.4 we demonstrate that asymptotic equivalence cannot hold without any additional assumptions on the Gauss–Markov process. More precisely, for the special case of Brownian bridge we show that the Le Cam distance between the two experiments is bounded away from zero. The proofs of our results are deferred to the Appendix.

## 1.1 Notation

Vectors will be denoted with bold letters (i.e. we write  $\mathbf{x}_n = (x_{1,n}, \dots, x_{m(n),n})$  when both length and entries of a vector might vary with  $n$ ). We also use the shorthand  $\mathbf{x}_{1:k,n} = (x_{1,n}, \dots, x_{k,n})$ . We write  $a_n \lesssim b_n$  if  $a_n \leq Cb_n$  for a constant  $C$  that is independent of  $n$ . The shorthand  $a_n \asymp b_n$  is used when  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$  hold simultaneously.

## 2 The RKHS associated with a Gauss–Markov process

The purpose of this section is to lay the foundations for our main results in Sects. 3 and 4. In Sect. 2.1 we recapitulate a characterization of Gauss–Markov process going back to (Doob, 1949) which will be important for our further reasoning. As a consequence of this characterization we can give an explicit description of the RKHS associated with the covariance kernel of a Gauss–Markov process, which is of independent interest and presented in Sect. 2.2.

### 2.1 Gauss–Markov processes

By definition a Gauss–Markov process is a stochastic process that is both Gaussian and Markov. Such a process  $X = (X_t)_{t \in [0,1]}$  is, assuming that  $\mathbf{E}[X_t] = 0$  for all  $t \in [0, 1]$ , essentially characterized by the following factorization property of the covariance function:

$$K_X(s, t) := \text{Cov}(X_s, X_t) = \mathbf{E}[X_s X_t] = U(s)V(t) \quad \text{for } 0 \leq s \leq t \leq 1, \quad (6)$$

where  $U$  and  $V$  are (known) non-negative functions on the interval  $[0, 1]$ ; see (Doob, 1949; Mehr and McFadden, 1965) for details. Kernels with the factorization property (6) are sometimes referred to as *triangular kernels* in the literature. Examples of Gauss–Markov processes include standard Brownian motion ( $U(t) = t$ ,  $V \equiv 1$ ), the Ornstein–Uhlenbeck process ( $U(t) = \exp(Lt) - \exp(-Lt)$  and  $V(t) = \exp(-Lt)$  for some  $L > 0$ ) and the Brownian bridge ( $U(t) = t$ ,  $V(t) = 1 - t$ ).

For our results, we will further assume that the considered Gauss–Markov process starts in zero. Such a process will be denoted with  $\Xi = (\Xi_t)_{t \in [0,1]}$  from now on, and we impose the following assumption.

**Assumption 1** The process  $\Xi = (\Xi_t)_{t \in [0,1]}$  is a centered Gauss–Markov process with  $\Xi_0 = 0$ , and non-degenerate on the open interval  $(0, 1)$ .

Note that Assumption 1 implies that there exist functions  $u$  and  $v$  in the representation

$$K_\Xi(s, t) = u(s)v(t) \quad (7)$$

of the covariance kernel satisfying  $u(\cdot)v(\cdot) \geq 0$  on the interval  $[0, 1]$ ,  $u(\cdot)v(\cdot) > 0$  on  $(0, 1)$  and that the function

$$q(t) = \frac{u(t)}{v(t)},$$

is continuous on the interval  $[0, 1)$ , non-negative and strictly increasing on  $[0, 1]$  (see Mehr and McFadden, 1965, p. 507). Moreover, under Assumption 1, the Gauss–Markov process  $\Xi$  can be written in distribution as

$$\Xi_t = v(t) \cdot W_{q(t)}. \quad (8)$$

Vice versa, this transformation of Brownian motion defines a Gaussian process with covariance kernel (7). Note that a simple calculation shows that the process  $\Xi$  has independent increments if and only if the function  $v$  is constant.

Let us shortly explain how processes  $\Xi$  satisfying Assumption 1 can be obtained. First, one can of course define such a process directly by means of the representation (8) provided that the functions  $u$  and  $v$  satisfy the properties stated above and  $u(0) = 0$  or  $v(0) = 0$  (this latter assumption guarantees that the process starts in 0). Second, starting with a general centered non-degenerate Gauss–Markov process  $X = (X_t)_{t \in [0,1]}$  with covariance kernel (6), and then conditioning on  $X_0 = 0$  does neither suspend Gaussianity nor the Markov property. More precisely, the process  $\Xi = (\Xi_t)_{t \geq 0}$  with  $\Xi_t \sim X_t | \{X_0 = 0\}$  is a centered Gaussian process with covariance kernel

$$K_{\Xi}(s, t) := \text{Cov}(\Xi_s, \Xi_t) = \mathbf{E}[\Xi_s \Xi_t] = u(s)v(t), \quad 0 \leq s \leq t \leq 1, \quad (9)$$

where the functions  $u$  and  $v$  are related to  $U$  and  $V$  through the identities

$$u(t) = U(t) - Q(0)V(t) \quad \text{and} \quad v(t) = V(t),$$

where  $Q(t) = U(t)/V(t)$ . In particular, the process  $\Xi$  satisfies Assumption 1, and  $q(t) = u(t)/v(t)$  can be represented as  $q(t) = Q(t) - Q(0)$ .

Surprisingly, Gauss–Markov processes starting at zero can also be obtained by conditioning a non-Markovian Gaussian process  $Y = (Y_t)_{t \geq 0}$  on the event  $\{Y_0 = 0\}$ . Then, the initial Gaussian process  $Y$  is called *conditionally Markov*. As one interesting example (further examples of conditionally Markov processes can again be found in Mehr and McFadden (1965), pp. 513 and 516) let us state the stationary Gaussian process  $(Y_t)_{t \in \mathbb{R}}$  defined on the whole real-line with zero mean and covariance kernel

$$K_Y(x, y) = \begin{cases} 1 - |x - y|, & \text{if } 0 \leq |x - y| \leq 1, \\ 0, & \text{if } |x - y| > 1, \end{cases}$$

for  $x, y \in \mathbb{R}$ . Then, the process  $(\Xi_t)_{t \geq 0}$  with  $\Xi_t \sim Y_t | \{Y_0 = 0\}$  is a centered Gaussian process with covariance kernel

$$K_{\Xi}(t_1, t_2) = \begin{cases} t_1(2 - t_2), & \text{if } 0 \leq t_1 \leq t_2 \leq 1, \\ 1 - (t_2 - t_1), & \text{if } 0 < t_2 - t_1 < 1, t_2 \geq 1, t_1 \geq 0, \\ 0, & \text{if } t_2 - t_1 \geq 0, t_1 \geq 0. \end{cases}$$

In particular, the restriction of  $\Xi$  on the interval  $[0, 1]$  is a Gauss–Markov process satisfying Assumption 1, and  $K_{\Xi}(s, t) = u(s)v(t)$  for  $0 \leq s \leq t \leq 1$  with  $u(x) = x$  and  $v(x) = 2 - x$ . This process has been considered in Slepian (1961) where it was proved that the process fulfills a ‘peculiar Markov-like property’. The restriction of the process  $\Xi$  on the unit interval fulfills all the technical assumptions made on the functions  $u$  and  $v$  below, and thus the results on asymptotic equivalence derived in Sects. 3 and 4 hold in particular for this process.

Let us finally mention that integrated Gauss–Markov processes of the form  $\Xi_t = \int_0^t X_s ds$  for a Gauss–Markov process  $X$  inherit the Gauss–Markov property from  $X$  (see Abundo, 2015, Prop. 2.2), and the theory derived in the sequel covers this special case. In particular, the increments in the discrete model (4) are then given by

$$\xi_{i,n} = \sqrt{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} X_s ds.$$

## 2.2 Reproducing kernel Hilbert spaces

We briefly recall some basic facts about reproducing kernel Hilbert spaces. For a detailed discussion the reader is referred to the monographs (Berlinet and Thomas-Agnan, 2004) and (Paulsen and Raghupathi, 2016). A subset  $\mathcal{H}$  of the set  $\mathcal{F}(\mathcal{X}, \mathbb{R})$  of real-valued functions on a domain  $\mathcal{X}$  is called a reproducing kernel Hilbert space (RKHS) on  $\mathcal{X}$  if

- $\mathcal{H}$  is a vector subspace of  $\mathcal{F}(\mathcal{X}, \mathbb{R})$ ,
- $\mathcal{H}$  has an inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  with respect to which  $\mathcal{H}$  is a Hilbert space,
- for any  $x \in \mathcal{X}$ , the linear evaluation functional  $\ell_x : \mathcal{H} \rightarrow \mathbb{R}$  with  $\ell_x(f) = f(x)$  is bounded.

The Riesz representation theorem implies that for any  $x \in \mathcal{X}$  there exists a unique function  $k_x \in \mathcal{H}$  such that for every  $f \in \mathcal{H}$  one has  $f(x) = \ell_x(f) = \langle f, k_x \rangle_{\mathcal{H}}$ . The function  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

$$K(x, y) = k_x(y)$$

is called the *reproducing kernel* for  $\mathcal{H}$ . Reproducing kernels are *positive definite* in the sense that the inequality

$$\sum_{i,j=1}^n a_i \bar{a}_j K(x_i, x_j) \geq 0$$

holds for all  $n \in \mathbb{N}$ ,  $x_i \in \mathcal{X}$  and  $a_i \in \mathbb{C}$ . Vice versa, for any positive definite function  $K$  on  $\mathcal{X}$  there exists a RKHS  $\mathcal{H}(K)$  on  $\mathcal{X}$  with reproducing kernel  $K$  and this RKHS is uniquely determined by the kernel. In particular, this holds true for covariance kernels. Finally, it is well-known that the linear span of the functions  $k_y = K(\cdot, y)$  is dense in  $\mathcal{H}$ .

In the following, we will consider the RKHS  $\mathcal{H}(\Xi)$  associated with the covariance kernel  $K_{\Xi} : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  of a Gauss–Markov process  $\Xi$  satisfying Assumption 1 as introduced in Sect. 2. As a consequence

$$K_{\Xi}(s, t) = u(s)v(t), \quad 0 \leq s \leq t \leq 1.$$

Let us from now on assume that  $q(0) = 0$  for  $q(x) = u(x)/v(x)$  which is under the validity of Assumption 1 certainly satisfied whenever  $v(0) \neq 0$ . This additional condition is valid for all the results in Sect. 4. As mentioned above,  $\Xi$  can be represented in distribution as

$$\Xi_t = v(t)W_{q(t)}$$

where  $W$  is standard Brownian motion. Denote  $T = q(1)$ , and let us consider the mapping  $\psi : \mathcal{H}(\Xi) \rightarrow L^2([0, T])$  defined on the 'generators'  $\{K_\Xi(\cdot, t), t \in [0, 1]\}$  via

$$K_\Xi(\cdot, t) \xrightarrow{\psi} v(t)\mathbf{1}_{[0, q(t)]}(\cdot) \quad (10)$$

(on general elements of  $\mathcal{H}(\Xi)$  is naturally defined by a limiting process). The mapping  $\psi$  is an isometry of Hilbert spaces since

$$\begin{aligned} \langle K_\Xi(\cdot, s), K_\Xi(\cdot, t) \rangle_{\mathcal{H}(\Xi)} &= K_\Xi(s, t) = v(s)v(t)q(s) \\ &= \int_0^T v(s)v(t)\mathbf{1}_{[0, q(s)]}(u)\mathbf{1}_{[0, q(t)]}(u)du \\ &= \langle \psi K_\Xi(\cdot, s), \psi K_\Xi(\cdot, t) \rangle_{L^2([0, T])} \end{aligned}$$

for  $0 \leq s \leq t \leq 1$  (this holds true due to the strict monotonicity and continuity of  $q$  combined with the fact that the indicator functions  $\mathbf{1}_{[0, t]}$ ,  $t \in [0, T]$ , are dense in  $L^2([0, T])$ ). We use the shorthand notation  $\psi_t = \psi(K_\Xi(\cdot, t))$ . For  $F \in \mathcal{H}(\Xi)$  we obtain with  $g = \psi F \in L^2([0, T])$  the representation

$$\begin{aligned} F(t) &= \langle F, K_\Xi(\cdot, t) \rangle_{\mathcal{H}(\Xi)} = \langle \psi F, \psi_t \rangle_{L^2([0, T])} = \langle g, \psi_t \rangle_{L^2([0, T])} \\ &= \int_0^T g(u)v(t)\mathbf{1}_{[0, q(t)]}(u)du = v(t) \int_0^{q(t)} g(u)du. \end{aligned}$$

This gives an explicit characterization of the RKHS  $\mathcal{H}(\Xi)$ , namely

$$\mathcal{H}(\Xi) = \left\{ F : [0, 1] \rightarrow \mathbb{R} \mid F(t) = v(t) \int_0^{q(t)} g(u)du \text{ for some } g \in L^2([0, T]) \right\}, \quad (11)$$

and this space is equipped with the norm  $\|F\|_{\mathcal{H}(\Xi)} = \|g\|_{L^2([0, T])}$  for  $F$  as in (11). Of course, in the special case of Brownian motion ( $u(t) = t, v \equiv 1$ ) these calculations yield the well-known fact that the RKHS corresponding to the kernel of standard BM contains exactly the primitives (starting at 0) of square-integrable functions. If  $F(\cdot) = F_f(\cdot) = \int_0^\cdot f(s)ds$ , then  $f$  can be derived by differentiation:

$$\begin{aligned} f(t) &= \frac{\partial}{\partial t} \left[ v(t) \int_0^{q(t)} g(u)du \right] \\ &= v'(t) \int_0^{q(t)} g(u)du + v(t)g(q(t))q'(t). \end{aligned}$$

Moreover, the function  $g$  can be obtained from  $F_f$  (and thus from  $f$ ) by



$$g(q(t)) = \left( \frac{F_f(t)}{v(t)} \right)' \cdot \frac{1}{q'(t)}. \quad (12)$$

This relation between  $f$  and  $g$  will be exploited when proving asymptotic equivalence results for concrete function classes in Sect. 4.

### 3 Abstract conditions for asymptotic equivalence

In this section, we derive sufficient conditions for an abstract function class  $\Theta$  that guarantee asymptotic equivalence of nonparametric regression with dependent errors and a continuous time model with an additive error process that is Gauss–Markov. In Sect. 3.1 we gather the necessary notions from asymptotic equivalence theory. In Sect. 3.2, we state the main result that provides the announced sufficient conditions for asymptotic equivalence of models (3) and (4).

#### 3.1 Asymptotic equivalence of experiments

Le Cam’s equivalence theory for statistical experiments has by now become a common tool in nonparametric statistics. As one of its major appeals one might consider the fact that complex statistical models can be shown equivalent to simple and well-studied benchmark models, at least asymptotically when the amount of information contained in the data increases. This is not only of interest in its own but has also turned out useful when proving optimality properties of estimation techniques. For instance, (Reiß, 2011) proposes rate-optimal estimators of the volatility function and simple efficient estimators of the integrated volatility as an application of Le Cam theory. Similarly, (Meister, 2011) derives sharp minimax constants in the functional linear regression model as an application of asymptotic equivalence of this model and an inverse problem in a white noise setup. The interested reader will find comprehensive introductions into Le Cam theory in the monographs (Le Cam, 1986; Le Cam and Yang, 2000; Torgersen, 1991) as well as in the recent survey paper (Mariucci, 2016), the latter focusing on nonparametric models that are also the topic of the present paper.

There exist various equivalent ways to introduce the concept of equivalence of statistical experiments. One approach in the general theory introduced by Le Cam is by means of the abstract concept of *transitions*. In the case of dominated statistical models with Polish sample spaces (which is exclusively considered in this article), however, this general notion essentially boils down to the class of *Markov kernels* as has been shown in Proposition 9.2 in Nussbaum (1996). To be precise, let  $\mathfrak{E}_i = (\mathcal{X}_i, \mathcal{X}_i, (\mathbf{P}_{i,\theta})_{\theta \in \Theta})$  for  $i \in \{1, 2\}$  be two such dominated experiments with Polish parameter spaces  $(\mathcal{X}_i, \mathcal{X}_i)$ , and sharing the same parameter space  $\Theta$ . Then, the *deficiency* of  $\mathfrak{E}_1$  with respect to  $\mathfrak{E}_2$  is defined as the quantity

$$\delta(\mathfrak{E}_1, \mathfrak{E}_2) = \inf_K \sup_{\theta \in \Theta} \|K\mathbf{P}_{1,\theta} - \mathbf{P}_{2,\theta}\|_{\text{TV}},$$

where  $\|\mathbf{P} - \mathbf{Q}\|_{\text{TV}}$  denotes the total variation distance between probability measures  $\mathbf{P}$  and  $\mathbf{Q}$ , and the infimum is taken over all Markov kernels  $K : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow [0, 1]$ .  $\delta(\mathfrak{E}_1, \mathfrak{E}_2) = 0$  has the interpretation that the experiment  $\mathfrak{E}_1$  is more informative than  $\mathfrak{E}_2$ .

Building on the definition of deficiency, the *Le Cam distance* between  $\mathfrak{E}_1$  and  $\mathfrak{E}_2$  is defined by symmetrization as

$$\Delta(\mathfrak{E}_1, \mathfrak{E}_2) = \max\{\delta(\mathfrak{E}_1, \mathfrak{E}_2), \delta(\mathfrak{E}_2, \mathfrak{E}_1)\},$$

which provides a pseudo-metric on the space of all statistical models with common parameter space. Two experiments  $\mathfrak{E}_1$  and  $\mathfrak{E}_2$  are called *equivalent* if  $\Delta(\mathfrak{E}_1, \mathfrak{E}_2) = 0$ . More generally, two sequences  $(\mathfrak{E}_{1,n})_{n \in \mathbb{N}}$  and  $(\mathfrak{E}_{2,n})_{n \in \mathbb{N}}$  are said to be *asymptotically equivalent* if  $\Delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n}) \rightarrow 0$ . In the latter case, following the notation used in (Golubev et al., 2010), we write  $\mathfrak{E}_{1,n} \approx \mathfrak{E}_{2,n}$ .

### 3.2 Sufficient conditions for asymptotic equivalence

We now rigorously define the two statistical experiments that will be examined for asymptotic equivalence in this paper. Let  $\Theta$  denote a given class of functions. The first experiment with discrete observations is given by

$$\mathfrak{E}_{1,n}(\Theta) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbf{P}_{1,n}^f)_{f \in \Theta}),$$

where  $\mathbf{P}_{1,n}^f$  denotes the distribution of the vector  $\mathbf{Y}_n = (Y_{1,n}, \dots, Y_{n,n})$  with components  $Y_{i,n}$  defined by

$$Y_{i,n} = f(t_{i,n}) + \sqrt{n}\xi_{i,n}, \quad i = 1, \dots, n, \quad (13)$$

where  $t_{i,n} = i/n$  are the sampling locations, and  $\xi_{i,n} = \Xi(t_{i,n}) - \Xi(t_{i-1,n})$  are the increments of a centered Gauss–Markov process  $\Xi = (\Xi_t)_{t \in [0,1]}$  with  $\Xi_0 = 0$  as introduced in Sect. 2.

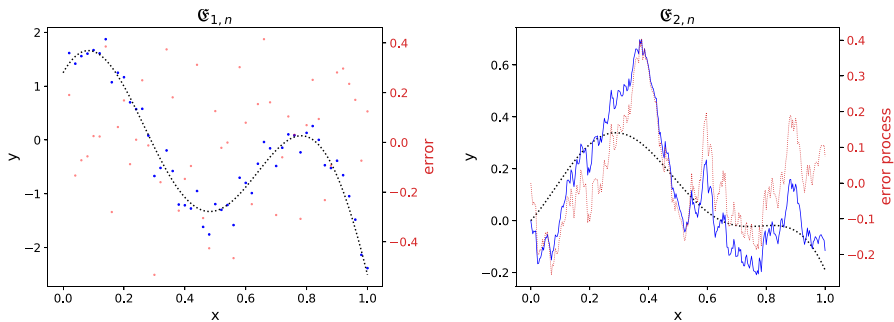
The second experiment with continuous observations is given through

$$\mathfrak{E}_{2,n}(\Theta) = (\mathcal{C}([0, 1], \mathbb{R}), \mathcal{B}(\mathcal{C}([0, 1], \mathbb{R})), (\mathbf{P}_{2,n}^f)_{f \in \Theta}),$$

where  $\mathbf{P}_{2,n}^f$  denotes the distribution of the continuous time process  $(Y_t)_{t \in [0,1]}$  defined by

$$Y_t = \int_0^t f(s)ds + \frac{1}{\sqrt{n}}\Xi_t, \quad t \in [0, 1]. \quad (14)$$

Typical realizations from the two experiments are visualized in Fig. 1 for the case of an Ornstein–Uhlenbeck process conditioned to start in zero.



**Fig. 1** Exemplary realizations of the observations in the two experiments  $\mathfrak{G}_{1,n}$  (left) and  $\mathfrak{G}_{2,n}$  (right) for  $n = 50$  where the Gauss–Markov process is given through the functions  $u(x) = \exp(x)$  and  $q(x) = \exp(2x) - 1$ . In both figures the actual observations are plotted in blue. The black dotted line represents the true function  $f$  in the left plot, and its anti-derivative  $F_f(\cdot) = \int_0^\cdot f(s)ds$  in the right plot. The error terms, that is, the increments of  $\Xi$  for  $\mathfrak{G}_{1,n}$  and the process  $\Xi$  itself for  $\mathfrak{G}_{2,n}$ , are plotted in red (note that there is a secondary y-axis in red for the error terms)

The following result provides sufficient conditions for asymptotic equivalence of the experiments  $\mathfrak{G}_{1,n}$  and  $\mathfrak{G}_{2,n}$  in terms of the class  $\Theta$  and the Gauss–Markov process  $\Xi$ . The proof is given in Appendix A.

**Theorem 2** *Let  $\Xi$  be a Gauss–Markov process on the interval  $[0, 1]$  with  $\Xi_0 = 0$  such that Assumption 1 holds with  $K_\Xi(s, t) = u(s)v(t)$  for  $0 \leq s \leq t \leq 1$ , and  $v(1) \neq 0$ . Set  $F_f(\cdot) = \int_0^\cdot f(s)ds$  and assume that the conditions*

- (i)  $\frac{1}{n} \sup_{f \in \Theta} \sum_{i=1}^n \frac{(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds)^2}{v^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n}))} \rightarrow 0$  and
- (ii)  $\sqrt{n} \sup_{f \in \Theta} \inf_{\alpha_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{i=1}^n \alpha_i K_\Xi(\cdot, t_{i,n})\|_{\mathcal{H}(\Xi)} \rightarrow 0$

are satisfied, where  $t_{i,n} = i/n$  and  $\|\cdot\|_{\mathcal{H}(\Xi)}$  is the norm in the RKHS  $\mathcal{H}(\Xi)$ . Then, the two sequences of experiments  $(\mathfrak{G}_{1,n}(\Theta))_{n \in \mathbb{N}}$  and  $(\mathfrak{G}_{2,n}(\Theta))_{n \in \mathbb{N}}$  are asymptotically equivalent, i.e.

$$\mathfrak{G}_{1,n}(\Theta) \approx \mathfrak{G}_{2,n}(\Theta).$$

**Remark 1** In this work, we have restricted ourselves to the one-dimensional setup, and, as pointed out by a reviewer, an interesting question is the extension to the multivariate case. In the standard white noise model, this was established in Reiß (2008). Another relevant reference in this direction is (Carter, 2006). In analogy with the last-mentioned paper a reasonable starting point to investigate whether and how our results can be extended to the multivariate case, is the consideration of processes whose covariance kernel is the product of  $d = 2$  Gauss–Markov kernels as considered in the present paper. This would be beyond the scope of the present paper and represents an interesting avenue for future research.

## 4 Specific function classes and a counterexample

In this section, we first apply the abstract Theorem 2 in order to establish asymptotic equivalence over Sobolev ellipsoids (Sect. 4.1) and Hölder balls (Sect. 4.2) of sufficiently smooth functions. In Sect. 4.3 we discuss an alternative proof strategy, which is based on some transformations of the different experiments and requires stronger assumptions as used in Sects. 4.1 and 4.2. Finally, in Section 4.4 we give a counterexample which shows that asymptotic equivalence cannot hold (even for function classes containing very smooth functions) in general if only Assumption 1 is required.

### 4.1 Asymptotic equivalence for Sobolev ellipsoids

Consider the (complex) trigonometric basis  $\{\mathbf{e}_k\}_{k \in \mathbb{Z}}$  of  $L^2([0, 1])$  given by

$$\mathbf{e}_k(\cdot) = \exp(-2\pi i k \cdot).$$

Any function  $f$  in  $L^2([0, 1])$  can be represented as a convergent series (in  $L^2$ -sense) of the form

$$f(\cdot) = \sum_{k=-\infty}^{\infty} \theta_k \mathbf{e}_k(\cdot)$$

where  $\theta_k = \langle f, \mathbf{e}_k \rangle_{L^2} = \int_0^1 f(s) \exp(2\pi i k s) ds$  are the uniquely defined Fourier coefficients of the function  $f$ . As potential function classes over which we would like to establish asymptotic equivalence we consider Sobolev ellipsoids

$$\Theta(\beta, L) = \left\{ f(\cdot) = \sum_{k=-\infty}^{\infty} \theta_k \mathbf{e}_k(\cdot) : \theta_k = \overline{\theta_{-k}} \text{ and } \sum_{k=-\infty}^{\infty} (1 + |k|)^{2\beta} |\theta_k|^2 \leq L^2 \right\}. \quad (15)$$

Note that the condition that  $\theta_k = \overline{\theta_{-k}}$  is used only to ensure that the functions in the set  $\Theta(\beta, L)$  are real-valued. The following result establishes asymptotic equivalence of the experiments  $\mathfrak{E}_{1,n}(\Theta(\beta, L))$  and  $\mathfrak{E}_{2,n}(\Theta(\beta, L))$  considered in Sect. 3 for  $\beta > 1/2$  under some assumptions on the functions  $v$  and  $q$  characterizing the Gauss–Markov process  $(\Xi_t)_{t \in [0,1]}$ .

**Theorem 3** *Let  $\Xi$  be a Gauss–Markov process on the interval  $[0, 1]$  with  $\Xi_0 = 0$  such that Assumption 1 holds with  $K_{\Xi}(s, t) = u(s)v(t)$  for  $0 \leq s \leq t \leq 1$ . If  $\beta > 1/2$  and the assumptions*

- (a)  $q \in C^1$  and  $0 < \inf_{t \in [0,1]} q'(t) \leq \sup_{t \in [0,1]} q'(t) < \infty$ ,
- (b)  $\inf_{t \in [0,1]} v(t) > 0$ , and
- (c)  $v'$  and  $q'$  are Hölder continuous with index  $\gamma > 1/2$

are satisfied, then

$$\mathfrak{G}_{1,n}(\Theta(\beta, L)) \approx \mathfrak{G}_{2,n}(\Theta(\beta, L)).$$

The proof of Theorem 3 is complicated and deferred to Appendix B.1. We proceed with some remarks that put the statement in the context of well-known results for the Brownian motion.

**Remark 2** It is known that for the smoothness index  $\beta = 1/2$  asymptotic equivalence between nonparametric regression model (1) and the white noise with drift experiment (2) (i.e. our experiment  $\mathfrak{G}_{2,n}$  for the special case of Brownian motion) does not hold. See Remark 4.6 in Brown and Low (1996) for further details. In Sect. 4.4 below we consider the example of a Brownian bridge as an error process and prove that asymptotic equivalence for arbitrary values of  $\beta$  cannot hold without additional assumptions on the Gauss–Markov process (like, for instance, assumptions on the functions  $v$  and  $q$  as made in Theorem 3).

**Remark 3** In the proof of Theorem 3 it is shown that under our assumptions the Le Cam distance between the experiments  $\mathfrak{G}_{1,n}(\Theta(\beta, L))$  and  $\mathfrak{G}_{2,n}(\Theta(\beta, L))$  converges to zero with a rate that cannot be faster than

$$\mathcal{O}(\max\{n^{-1/2}, n^{-\beta+1/2}\})$$

(indeed, the rate might even be slower due to the terms incorporating the Hölder index  $\gamma$  from Condition c) in the statement of the theorem). In the case where the Gauss–Markov is standard Brownian motion the rate  $n^{-\beta+1/2}$  can be established which is faster for  $\beta > 1$ ; see (Rohde, 2004) for further details. Since we have restricted ourselves to proving asymptotic equivalence, we have not further pursued this issue here.

**Remark 4** (Sequence space representations) Note that for the special case of Brownian motion the relation between the functions  $f = F'$  and  $g$  in the representation (11) of  $F \in \mathcal{H}(\Xi)$  reduces to the equality  $g = f$ . Hence, by representing  $f$  in terms of the coefficients of a series expansion with respect to an orthonormal basis (for instance, the complex trigonometric basis given through  $\{\mathbf{e}_k\}_{k \in \mathbb{Z}}$ ) the same representation is achieved for  $g$ . Via the isomorphism  $\psi$  defined in (10) (more precisely, its inverse  $\psi^{-1}$ ) this yields a representation of  $F_f$  in terms of an orthonormal basis in the RKHS leading to a Gaussian sequence space model of the form

$$y_k = \theta_k + \xi_k, \quad k \in \mathbb{Z},$$

where  $\xi_k$  i.i.d.  $\sim \mathcal{N}(0, 1)$  and the sequence  $\theta = (\theta_k)_{k \in \mathbb{Z}}$  satisfies the same smoothness conditions as in the definition of the Sobolev ellipsoid  $\Theta(\beta, L)$ . In the general case of an arbitrary Gauss–Markov process the relation (12) does not ‘transport’ an orthonormal basis for the representation of  $f$  to an orthonormal basis for  $g$ .

An alternative sequence space representation (for general Gaussian processes) is obtained from (14) using the Karhunen-Loeve representation of the process  $\Xi$ . Note that the kernel  $K_{\Xi}$  gives rise to an integral operator defined through

$$(K_{\Xi}f)(s) = \int_0^1 K_{\Xi}(s, t)f(t)dt$$

for  $f \in L^2([0, 1])$  (by slight abuse of notation, we use the same letter for both the kernel and the operator associated with it). Provided that (the kernel)  $K_{\Xi}$  is continuous, Mercer's theorem applies and (the operator)  $K_{\Xi}$  has a countable system of eigenfunctions  $\varphi_k$  with eigenvalues  $\lambda_k > 0$  (for square integrable symmetric kernels this can also be concluded directly from the spectral theorem for compact operators). By Proposition 3.11 in Neveu (1968) an orthonormal basis of the RKHS  $\mathcal{H}(\Xi)$  is given by the functions  $g_k = \sqrt{\lambda_k}\varphi_k$ ,  $k = 1, 2, \dots$ . In terms of this basis, (14) can be rewritten as

$$y_k = \theta_k + \frac{1}{\sqrt{n}}\eta_k, \quad k = 1, 2, \dots \quad (16)$$

where  $\theta_1, \theta_2, \dots$  are the Fourier coefficients of the function  $F$  with respect to the basis  $g_1, g_2, \dots$  and  $\eta_1, \eta_2, \dots$  are independent and standard normal distributed (note that the sequence model here is stated in terms of the coefficients of  $F_f$  instead of those of  $f$  itself). In the special case of Brownian motion, one has (see Beder, 1987, p. 66)

$$g_k(t) = \frac{\sqrt{2}}{(k - \frac{1}{2})\pi} \sin\left(\left(k - \frac{1}{2}\right)\pi t\right) = \sqrt{\lambda_k}\varphi_k, \quad \lambda_k = \frac{1}{\left(k - \frac{1}{2}\right)^2 \pi^2}.$$

In this case, the derivatives  $\psi_k = g'_k$  of the basis functions  $g_k$  in the RKHS form an orthonormal basis of  $L^2([0, 1])$ . Hence, the sequence space model (16) can be interpreted both as a perturbation of the coefficients of  $F_f$  or of  $f$  (note that  $f \in L^2([0, 1])$  if and only if  $F_f \in \mathcal{H}(\Xi)$ ).

## 4.2 Asymptotic equivalence for Hölder classes

For a smoothness index  $0 < \alpha \leq 1$  and constants  $0 < L < \infty$ ,  $0 < M \leq \infty$ , we introduce the Hölder class

$$\mathcal{F}(\alpha, L, M) = \left\{ f : [0, 1] \rightarrow \mathbb{R} : \sup_{0 \leq x < y \leq 1} \frac{|f(x) - f(y)|}{|x - y|^\alpha} \leq L \text{ and } \|f\|_\infty \leq M \right\},$$

where the case  $M = \infty$  means that the assumption  $\|f\|_\infty \leq M$  is omitted in the definition of  $\mathcal{F}(\alpha, L, M)$ . Under the same technical assumptions on the functions  $v$  and  $q$

as stated in Theorem 3 we obtain asymptotic equivalence for functions from Hölder classes with smoothness index  $\alpha > 1/2$ .

**Theorem 4** *Let  $\Xi$  be a Gauss–Markov process on the interval  $[0, 1]$  with  $\Xi_0 = 0$  such that Assumption 1 holds with  $K_\Xi(s, t) = u(s)v(t)$  for  $0 \leq s \leq t \leq 1$ . Suppose further that  $\alpha > 1/2$  and that the assumptions*

- (a)  $q \in C^1$  and  $0 < \inf_{t \in [0,1]} q'(t) \leq \sup_{t \in [0,1]} q'(t) \leq C < \infty$ ,
- (b)  $\inf_{t \in [0,1]} v(t) > 0$ , and
- (c)  $v'$  and  $q'$  are Hölder continuous with index  $\gamma > 1/2$

are satisfied. Then, if  $M < \infty$ , we have

$$\mathfrak{E}_{1,n}(\mathcal{F}(\alpha, L, M)) \approx \mathfrak{E}_{2,n}(\mathcal{F}(\alpha, L, M)). \quad (17)$$

Moreover, if the function  $v \cdot q'$  is constant on the interval  $[0, 1]$ , the asymptotic equivalence (17) also holds in the case  $M = \infty$ .

The proof of Theorem 4 is given in Appendix B.2. Note that in the Gaussian white noise model we have  $v \cdot q' \equiv 1$ , and in this case Theorem 4 reproduces the result given in Brown and Low (1996), at the top of page 2390.

### 4.3 Sketch of a different proof approach

The proofs of our results concerning Sobolev and Hölder classes in Sects. 4.1 and 4.2 both relied on the general conditions stated in Theorem 2. We will now sketch a different proof that is not based on the RKHS approach used in this paper. It requires a slightly stronger assumption on the function  $q$  and exploits results from the seminal paper (Brown and Low, 1996), which are applied to suitable transformations of the experiments of interest. Although checking all the details of this alternative approach might be a tedious affair, we will state the main steps here which yields a look on the asymptotic equivalence from a different angle.

The arguments used in the sequel make use of several transformations of the experiments. To be precise, starting with the discrete experiment  $\mathfrak{E}_{1,n}$ , we note that observing the  $Y_{i,n}$  in Eq. (13) is equivalent to observing

$$Y_{j,n}^* = \frac{\sum_{i=1}^j Y_{i,n}}{v(t_{j,n})} = \frac{\sum_{i=1}^j f(t_{i,n})}{v(t_{j,n})} + \sqrt{n} W_{q(t_{j,n})}, \quad j = 1, \dots, n.$$

By taking increments these observations are in turn equivalent to the experiment  $\mathfrak{E}_{1,n}^*$  with observations

$$Y_{j,n}^* - Y_{j-1,n}^* = \mu_{j,n} + \sigma_{j,n} \varepsilon_{j,n}, \quad j = 1, \dots, n, \quad (18)$$

where  $Y_{0,n}^* = 0$ , the  $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$  are i.i.d.  $\sim \mathcal{N}(0, 1)$ ,

$$\mu_{j,n} = \frac{\sum_{i=1}^j f(t_{i,n})}{v(t_{j,n})} - \frac{\sum_{i=1}^{j-1} f(t_{i,n})}{v(t_{j-1,n})},$$

and

$$\sigma_{j,n}^2 = n[q(t_{j,n}) - q(t_{j-1,n})].$$

We now transform the continuous experiment  $\mathfrak{E}_{2,n}$  to observations with a similar structure as in (18) but different mean and variance components. The observation of the path  $Y_t$  in (14) is, assuming that the function  $v$  does not vanish, equivalent to observing the path

$$Z_t^* = \frac{Y(t)}{v(t)} = \int_0^t \mu(s) ds + \frac{1}{\sqrt{n}} W_{q(t)}, \quad t \in [0, 1].$$

Here, the function  $\mu : [0, 1] \rightarrow \mathbb{R}$  is given through

$$\mu(s) = \frac{f(s)v(s) - v'(s) \int_0^s f(\tau) d\tau}{v^2(s)}.$$

Using time-change for martingales (see Karatzas and Shreve, 1991, Chapter 3, Theorem 4.6) the Gaussian process  $Z_t^*$  can be characterized by the SDE

$$dZ_t^* = \mu(t)dt + \frac{1}{\sqrt{n}} \sqrt{q'(t)} dW_t. \quad (19)$$

Now, under regularity assumptions on  $\mu$  and  $q$  Theorem 4.1 in Brown and Low (1996) can be applied to show asymptotic equivalence of the experiments where one either observes the path  $Z_t^*$  or the discrete data

$$Z_{j,n}^* = \mu(j/(n+1)) + \sqrt{q'(j/(n+1))} \varepsilon_{j,n}, \quad j = 1, \dots, n, \quad (20)$$

and  $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$  are i.i.d.  $\sim \mathcal{N}(0, 1)$ . Let us denote the latter experiment by  $\mathfrak{E}_{2,n}^*$ . Then, by the arguments sketched above the target equivalence  $\mathfrak{E}_{1,n} \approx \mathfrak{E}_{2,n}$  follows from the equivalence  $\mathfrak{E}_{1,n}^* \approx \mathfrak{E}_{2,n}^*$ . In order to prove this statement, however, it is sufficient to show that

$$n \cdot \sup_j (\mu(j/(n+1)) - \mu_{j,n})^2 + n \sup_j (q'(j/(n+1)) - \sigma_{j,n}^2) \rightarrow 0,$$

which can be established under mild regularity assumptions.

We emphasize that the assumptions for this alternative proof strategy and the ones imposed by us in Theorems 3 and 4, respectively, do not coincide entirely. In particular, Theorems 3 and 4 in the paper assume the existence of a Hölder continuous first derivative  $q'$ . On the other hand, Theorem 4.1 in Brown and Low (1996), which is used to establish the asymptotic equivalence of the models (19) and (20) in



this section, requires the existence of a second derivative of  $q$  (see Condition (4.1) in this reference). From this point of view, the RKHS approach uses (slightly) weaker assumptions.

One main motivation to study the theory of asymptotic equivalence is the transfer of decision rules from one (usually simpler) experiment to a second (usually more complicated) one. This is feasible whenever one explicitly knows Markov kernels that transfer observations between the two statistical experiments at hand. Corollary 4.1 from (Brown and Low, 1996) describes how to transfer decision rules between the standard regression model with i.i.d. noise and the standard Gaussian white noise model. Combining this result with the proof sketch given above allows the transfer of decision rules also between the models (3) and (4) studied in the present paper. In our specific case the regression model seems to be the ‘more realistic’ one, whereas the continuous model, for instance in the form (19) might allow for a precise analysis concerning the target parameter  $f$  (which is accessible through the function  $\mu$  in (19) since the functions  $v$  and  $q$  appearing in the definition of  $\mu$  assumed to be known). However, in absence of a specific application (including a specific choice of the functions  $v$  and  $q$ ) motivating the regression model (4), we do not pursue this issue here further.

#### 4.4 Non-equivalence for Brownian bridge

We conclude this section by showing that the asymptotic equivalence result established in Theorem 3 cannot be transferred to arbitrary Gauss–Markov processes. To be more precise, Theorem 5 below shows that for the special case of the Brownian bridge the deficiency  $\delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n})$  is bounded away from zero, and as a consequence the two experiments are not asymptotically equivalent over Sobolev spaces  $\Theta(\beta, L)$  for arbitrary smoothness  $\beta$  and  $L > 0$ . Recall that the covariance kernel of a Brownian Bridge  $(B_t)_{t \in [0,1]}$  satisfies the representation

$$\text{Cov}(B(s), B(t)) = u(s)v(t), \quad 0 \leq s \leq t \leq 1,$$

with  $u(s) = s$ ,  $v(t) = 1 - t$ . Therefore  $v(1) = 0$ , and in addition Conditions a) and b) in Theorem 3 are not satisfied.

**Theorem 5** *Consider the two sequences of experiments  $(\mathfrak{E}_{1,n})_{n \in \mathbb{N}}$  and  $(\mathfrak{E}_{2,n})_{n \in \mathbb{N}}$  defined in (13) and (14), respectively, where the error process  $(\Xi_t)_{t \in [0,1]}$  is a Brownian Bridge. Then, for an arbitrary smoothness index  $\beta > 0$ , any  $L > 0$ , and all  $n \in \mathbb{N}$ ,*

$$\delta(\mathfrak{E}_{1,n}(\Theta(\beta, L)), \mathfrak{E}_{2,n}(\Theta(\beta, L))) \geq 1/4,$$

*where  $\Theta(\beta, L)$  is the Sobolev ellipsoid introduced in (15). In particular, the experiments  $(\mathfrak{E}_{1,n})_{n \in \mathbb{N}}$  and  $(\mathfrak{E}_{2,n})_{n \in \mathbb{N}}$  are not asymptotically equivalent.*

## Proof of Theorem 2

The proof consists in the consideration of two intermediate experiments, given through Eqs. (21) and (22) below, that lie between  $\mathfrak{E}_{1,n}$  and  $\mathfrak{E}_{2,n}$ .

*First step:* We first show that under Condition (i) in Theorem 2 the model given by Equation (13) is asymptotically equivalent to observing

$$Y'_{i,n} = n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds + \sqrt{n}\xi_{i,n}, \quad i = 1, \dots, n, \quad (21)$$

where  $\xi_{i,n} = \Xi_{t_{i,n}} - \Xi_{t_{i-1,n}}$  are the increments of the process  $(\Xi_t)_{t \in [0,1]}$ . Under Assumption 1, we can take advantage of the representation (8) and write

$$\xi_{i,n} = v(t_{i,n})W_{q(t_{i,n})} - v(t_{i-1,n})W_{q(t_{i-1,n})},$$

and the experiment (13) can be written as

$$Y_{i,n} = f(t_{i,n}) + \sqrt{n} \left[ v(t_{i,n})W_{q(t_{i,n})} - v(t_{i-1,n})W_{q(t_{i-1,n})} \right].$$

Adding and subtracting  $v(t_{i,n})W_{q(t_{i-1,n})}$ , we get

$$\begin{aligned} Y_{i,n} &= f(t_{i,n}) + \sqrt{n}(v(t_{i,n}) - v(t_{i-1,n}))W_{q(t_{i-1,n})} \\ &\quad + \sqrt{n}v(t_{i,n})(W_{q(t_{i,n})} - W_{q(t_{i-1,n})}). \end{aligned}$$

Similarly, (21) can be written as

$$\begin{aligned} Y'_{i,n} &= n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds + \sqrt{n}(v(t_{i,n}) - v(t_{i-1,n}))W_{q(t_{i-1,n})} \\ &\quad + \sqrt{n}v(t_{i,n})(W_{q(t_{i,n})} - W_{q(t_{i-1,n})}). \end{aligned}$$

For the sake of a transparent notation let  $\mathbf{P}^{\mathbf{Y}_n} = \mathbf{P}_{1,n}^f$  denote the distribution of the vector  $\mathbf{Y}_n = (Y_{1,n}, \dots, Y_{n,n})$ , where we do not reflect the dependence on  $f$  in the notation. Similarly, let  $\mathbf{P}^{\mathbf{Y}'_n}$  denote the distribution of the vector  $\mathbf{Y}'_n = (Y'_{1,n}, \dots, Y'_{n,n})$ . Note that the squared total variation distance can be bounded by the Kullback-Leibler divergence, and therefore we will derive a bound for the Kullback-Leibler distance between the distributions  $\mathbf{P}^{\mathbf{Y}_n}$  and  $\mathbf{P}^{\mathbf{Y}'_n}$  by suitable conditioning. Denote with  $\mathcal{F}_{i,n}$  the  $\sigma$ -algebra generated by  $\{W_{q(t)}, t \leq t_{i,n}\}$ . We have

$$\begin{aligned} \text{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n}) &= \mathbf{E}[\text{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n} | \mathcal{F}_{n-1,n})] \\ &= \mathbf{E}[\text{KL}(\mathbf{P}^{\mathbf{Y}_{1:n-1,n}}, \mathbf{P}^{\mathbf{Y}'_{1:n-1,n}} | \mathcal{F}_{n-1,n}) + \text{KL}(\mathbf{P}^{\mathbf{Y}_{n,n}}, \mathbf{P}^{\mathbf{Y}'_{n,n}} | \mathcal{F}_{n-1,n})] \\ &= \mathbf{E}[\text{KL}(\mathbf{P}^{\mathbf{Y}_{1:n-1,n}}, \mathbf{P}^{\mathbf{Y}'_{1:n-1,n}} | \mathcal{F}_{n-1,n})] + \mathbf{E}[\text{KL}(\mathbf{P}^{\mathbf{Y}_{n,n}}, \mathbf{P}^{\mathbf{Y}'_{n,n}} | \mathcal{F}_{n-1,n})] \\ &= \text{KL}(\mathbf{P}^{\mathbf{Y}_{1:n-1,n}}, \mathbf{P}^{\mathbf{Y}'_{1:n-1,n}}) + \mathbf{E}[\text{KL}(\mathbf{P}^{\mathbf{Y}_{n,n}}, \mathbf{P}^{\mathbf{Y}'_{n,n}} | \mathcal{F}_{n-1,n})], \end{aligned}$$

where we use the notation  $\mathbf{Y}_{1:n-1,n} = (Y_{1,n}, \dots, Y_{n-1,n})$  and  $\mathbf{Y}'_{1:n-1,n} = (Y'_{1,n}, \dots, Y'_{n-1,n})$ . Repeating the same argument, one obtains

$$\text{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n}) = \sum_{i=1}^n \mathbf{E}[\text{KL}(\mathbf{P}^{Y_{i,n}}, \mathbf{P}^{Y'_{i,n}} | \mathcal{F}_{i-1,n})].$$

In order to study the terms  $\mathbf{E}[\text{KL}(\mathbf{P}^{Y_{i,n}}, \mathbf{P}^{Y'_{i,n}} | \mathcal{F}_{i-1,n})]$ , note that

$$Y_{i,n} | \mathcal{F}_{i-1,n} \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \text{and} \quad Y'_{i,n} | \mathcal{F}_{i-1,n} \sim \mathcal{N}(\mu'_i, \sigma_i^2),$$

where

$$\begin{aligned} \mu_i &= f(t_{i,n}) + \sqrt{n}(v(t_{i,n}) - v(t_{i-1,n}))W_{q(t_{i-1,n})}, \\ \mu'_i &= n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds + \sqrt{n}(v(t_{i,n}) - v(t_{i-1,n}))W_{q(t_{i-1,n})}, \quad \text{and} \\ \sigma_i^2 &= nv^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n})). \end{aligned}$$

Here and in the following,  $\mathcal{N}(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Using the fact that the Kullback–Leibler divergence between two normal distributions with common variance is given by

$$\text{KL}(\mathcal{N}(\mu_1, \sigma^2), \mathcal{N}(\mu_2, \sigma^2)) = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2},$$

we have

$$\text{KL}(\mathbf{P}^{Y_{i,n}}, \mathbf{P}^{Y'_{i,n}} | \mathcal{F}_{i-1,n}) = \frac{(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds)^2}{2nv^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n}))}.$$

This yields

$$\text{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n}) = \frac{1}{2n} \sum_{i=1}^n \frac{(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds)^2}{v^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n}))},$$

and  $\text{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n}) \rightarrow 0$  holds if and only if Condition (i) holds. Consequently, the experiments (13) and (21) are asymptotically equivalent in this case.

*Second step:* Let  $I(t|\mathbf{y}_n)$  denote the Kriging interpolator which is defined as

$$I(t|\mathbf{y}_n) = (K_{\Xi}(t, t_{1,n}), K_{\Xi}(t, t_{2,n}), \dots, K_{\Xi}(t, t_{n,n})) \text{Cov}(\Xi_n)^{-1} \mathbf{y}_n^{\top} \quad (22)$$

for  $\mathbf{y}_n = (y_{1,n}, \dots, y_{n,n})$  and  $\Xi_n = (\Xi_{t_{1,n}}, \dots, \Xi_{t_{n,n}})$  (the additional condition that  $v(1) \neq 0$  guarantees the invertibility of  $\Xi_n$ ; see Lemma A.1 in Dette et al. (2016) for an explicit formula for the entries of the inverse matrix). By definition the Kriging predictor is linear in the argument  $\mathbf{y}_n$ , and a simple argument shows the interpolation property

$$I(t_{j,n}|\mathbf{y}_n) = y_{j,n} \quad \text{for } j = 1, \dots, n. \quad (23)$$

The second step now consists in proving (exact, that is, non-asymptotic) equivalence of the experiment defined by the discrete observations (21) and the experiment defined by the continuous path

$$\tilde{Y}_t = I(t|\mathbf{F}_{f,n}) + n^{-1/2}\Xi_t, \quad (24)$$

where  $\mathbf{F}_{f,n} = (F_{f,n}(t_{1,n}), \dots, F_{f,n}(t_{n,n}))$ . Defining the partial sums  $S'_{k,n} = \sum_{j=1}^k Y'_{j,n}$  and recalling the notation  $\xi_{k,n} = \Xi_{t_{k,n}} - \Xi_{t_{k-1,n}}$  for the increments of the process  $(\Xi_t)_{t \in [0,1]}$ , we have

$$S'_{k,n} = \sum_{j=1}^k Y'_{j,n} = n \int_0^{t_{k,n}} f(s)ds + \sqrt{n} \sum_{j=1}^k \xi_{j,n} = nF_f(t_{k,n}) + \sqrt{n}\Xi_{t_{k,n}} \quad (25)$$

where we used the interpolating property (23) and the definition (24). Let  $(\Xi'_t)_{t \in [0,1]}$  be an independent copy of  $(\Xi_t)_{t \in [0,1]}$ , and set  $R_t = \Xi'_t - I(t|\Xi'_n)$  with  $\Xi'_n = (\Xi'_{t_{1,n}}, \dots, \Xi'_{t_{n,n}})$ . Then, the process

$$I(t|\Xi_n) + R_t, \quad t \in [0, 1],$$

follows the same law as  $(\Xi_t)_{t \in [0,1]}$  and  $(\Xi'_t)_{t \in [0,1]}$ , which can be checked by a comparison of the covariance structure (indeed, this kind of construction is valid for any centered Gaussian process). Then, observing the definition (24), we have

$$\begin{aligned} \tilde{Y}_t &= I(t|\mathbf{F}_{f,n}) + n^{-1/2}\Xi_t \\ &\stackrel{\mathcal{L}}{=} I(t|\mathbf{F}_{f,n}) + n^{-1/2}I(t|\Xi_n) + n^{-1/2}R_t \\ &= n^{-1}(I(t|\mathbf{S}'_n) + \sqrt{n}R_t), \end{aligned}$$

where we used the notation  $\mathbf{S}'_n = (S'_{1,n}, \dots, S'_{n,n})$ , eq. (25) and the linearity of the Kriging estimator. Therefore, the process  $(\tilde{Y}_t)_{t \in [0,1]}$  can be constructed from the vector  $\mathbf{Y}'_n$ . On the other hand, the observations  $Y'_{1,n}, \dots, Y'_{n,n}$  can be recovered from the trajectory  $(\tilde{Y}_t)_{t \in [0,1]}$  since for  $t = t_{k,n}$  the interpolation property (23) yields

$$\begin{aligned} n\tilde{Y}_{t_{k,n}} &= nF_{f,n}(t_{k,n}) + \sqrt{n}\Xi_{t_{k,n}} \\ &= n \int_0^{t_{k,n}} f(s)ds + \sqrt{n}(\Xi_{t_{k,n}} - \Xi_0) \\ &= n \int_0^{t_{k,n}} f(s)ds + \sqrt{n} \sum_{j=1}^k \xi_{j,n}, \end{aligned}$$

and one obtains  $Y'_{i,n}$  as  $Y'_{i,n} = n\tilde{Y}_{t_i} - n\tilde{Y}_{t_{i-1}}$ . Hence, the process  $(\tilde{Y}_t)_{t \in [0,1]}$  and the vector  $\mathbf{Y}'_n$  contain the same information and the experiments (21) and (24) are equivalent.

*Third step:* It remains to show that the experiment  $\tilde{\mathfrak{E}}_{2,n}$  defined by the path in (24) is asymptotically equivalent to the experiment  $\mathfrak{E}_{2,n}$  defined by

$$Y_t = \int_0^t f(s)ds + n^{-1/2} \Xi_t.$$

For this purpose we denote by  $\mathbf{P}^{(Y_t)}_{t \in [0,1]}$  and  $\mathbf{P}^{(\tilde{Y}_t)}_{t \in [0,1]}$  the distributions of the processes  $(Y_t)_{t \in [0,1]}$  and  $(\tilde{Y}_t)_{t \in [0,1]}$ , respectively, where the dependence on the parameter  $f$  is again suppressed. First, note that the representation of the Kriging estimator shows that the function  $t \rightarrow I(t|\mathbf{F}_{f,n})$  belongs to the RKHS  $\mathcal{H}(\Xi)$  associated with the covariance kernel  $K_\Xi$ . Second, Condition (ii) in the statement of Theorem 2 yields that the same holds true for the function  $F_{f,n}(\cdot)$ . Using the fact that the squared total variation distance can be bounded by the Kullback-Leibler divergence, we obtain

$$\begin{aligned} \Delta^2(\mathfrak{G}_{2,n}, \tilde{\mathfrak{G}}_{2,n}) &\leq \sup_{f \in \Theta} \text{KL}(\mathbf{P}^{(Y_t)}_{t \in [0,1]}, \mathbf{P}^{(\tilde{Y}_t)}_{t \in [0,1]}) \\ &= \frac{n}{2} \sup_{f \in \Theta} \|F_f(\cdot) - I(\cdot | \mathbf{F}_{f,n})\|_{\mathcal{H}(\Xi)}^2 \\ &= \frac{n}{2} \sup_{f \in \Theta} \inf_{\alpha_n \in \mathbb{R}^n} \left\| F_f(\cdot) - \sum_{j=1}^n \alpha_j K_\Xi(\cdot, t_{j,n}) \right\|_{\mathcal{H}(\Xi)}^2 \rightarrow 0 \end{aligned}$$

(the first equality follows from Lemma 2 in Schmidt-Hieber (2014), the second from Theorem 13.1 in Wendland (2005)), which completes the proof of Theorem 2.  $\square$

## Proofs of the results in Section 4

### Proof of Theorem 3

The proof consists in checking the two conditions (i) and (ii) in Theorem 2.

*Verification of condition (i):* We have to show that the expression

$$\frac{1}{n} \sup_{f \in \Theta(\beta, L)} \sum_{i=1}^n \frac{(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds)^2}{v^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n}))}$$

converges to zero as  $n \rightarrow \infty$ . By the assumptions regarding the functions  $v$  and  $q$  and an application of the mean value theorem this is equivalent to the condition

$$\sup_{f \in \Theta(\beta, L)} \sum_{i=1}^n \left( f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)ds \right)^2 \rightarrow 0. \quad (26)$$

Let  $f \in \Theta(\beta, L)$  with Fourier expansion  $f(\cdot) = \sum_{k \in \mathbb{Z}} \theta_k \mathbf{e}_k(\cdot)$ . For any  $K \in \mathbb{N}$  (the appropriate value of  $K = K(n)$  for our purposes will be specified below) we define the functions

$$f_K(\cdot) = \sum_{|k| \leq K} \theta_k \mathbf{e}_k(\cdot) \text{ and}$$

$$f_K^\top(\cdot) = f(\cdot) - f_K(\cdot) = \sum_{|k| > K} \theta_k \mathbf{e}_k(\cdot),$$

respectively. Consequently,

$$\sum_{i=1}^n \left( f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s) ds \right)^2 \leq 3 \sum_{i=1}^n A_{i,n}^2 + 3 \sum_{i=1}^n B_{i,n}^2 + 3 \sum_{i=1}^n C_{i,n}^2, \quad (27)$$

where

$$A_{i,n} = f_K(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f_K(s) ds,$$

$$B_{i,n} = f_K^\top(t_{i,n}), \quad \text{and}$$

$$C_{i,n} = n \int_{t_{i-1,n}}^{t_{i,n}} f_K^\top(s) ds.$$

We now show for  $K = n$  that the estimates

$$\sum_{i=1}^n A_{i,n}^2 = \mathcal{O}(\max\{n^{-1}, n^{1-2\beta}\}), \quad (28)$$

$$\sum_{i=1}^n B_{i,n}^2 = \mathcal{O}(n^{1-2\beta}), \quad (29)$$

$$\sum_{i=1}^n C_{i,n}^2 = \mathcal{O}(n^{1-2\beta}), \quad (30)$$

hold uniformly with respect to  $f \in \Theta(\beta, L)$ . Then, assertion (26) follows from (27) and the assumption  $\beta > 1/2$ .

**Proof of (28):** The quantity

$$\frac{1}{n} \sum_{j=1}^n |A_{j,n}|^2$$

can be interpreted as the (average) energy of the discrete signal  $A^{(n)} = (A_{1,n}, \dots, A_{n,n})$ . Define

$$F_j = \frac{1}{n} \sum_{k=1}^n A_{k,n} e^{-2\pi i k j / n}$$

as the discrete Fourier transform of the signal  $A^{(n)}$ , then Parseval's identity for the discrete Fourier transform yields

$$\frac{1}{n} \sum_{j=1}^n |A_{j,n}|^2 = \sum_{j=1}^n |F_j|^2,$$

and we have to derive an estimate for  $n \sum_{j=1}^n |F_j|^2$ . For this purpose, we recall the notation of  $A_{j,n}$  and note that

$$\begin{aligned} F_j &= \frac{1}{n} \sum_{k=1}^n \left( f_K(t_{k,n}) - n \int_{t_{k-1}}^{t_k} f_K(s) ds \right) e^{-2\pi i k j / n} \\ &= \frac{1}{n} \sum_{k=1}^n \left( \sum_{1 \leq |l| \leq K} \theta_l e^{-2\pi i l k / n} - n \int_{t_{k-1}}^{t_k} \sum_{1 \leq |l| \leq K} \theta_l e^{-2\pi i l s} ds \right) e^{-2\pi i k j / n} \\ &= \frac{1}{n} \sum_{k=1}^n \left( \sum_{1 \leq |l| \leq K} \theta_l e^{-2\pi i l k / n} + n \sum_{1 \leq |l| \leq K} \frac{\theta_l}{2\pi i l} [e^{-2\pi i l k / n} - e^{-2\pi i l (k-1) / n}] \right) e^{-2\pi i k j / n} \\ &= \frac{1}{n} \sum_{k=1}^n \left( \sum_{1 \leq |l| \leq K} \theta_l e^{-2\pi i l k / n} + n \sum_{1 \leq |l| \leq K} \frac{\theta_l}{2\pi i l} [1 - e^{2\pi i l / n}] e^{-2\pi i l k / n} \right) e^{-2\pi i k j / n} \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{1 \leq |l| \leq K} \theta_l \left[ 1 + \frac{n}{2\pi i l} (1 - e^{2\pi i l / n}) \right] e^{-2\pi i k l / n} e^{-2\pi i k j / n}. \end{aligned}$$

From now on, we take  $K = n$  and write

$$\begin{aligned} F_j^+ &= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \theta_l \left[ 1 + \frac{n}{2\pi i l} (1 - e^{2\pi i l / n}) \right] e^{-2\pi i k l / n} e^{-2\pi i k j / n}, \quad \text{and} \\ F_j^- &= \frac{1}{n} \sum_{k=1}^n \sum_{l=-n}^{-1} \theta_l \left[ 1 + \frac{n}{2\pi i l} (1 - e^{2\pi i l / n}) \right] e^{-2\pi i k l / n} e^{-2\pi i k j / n}. \end{aligned}$$

Since  $\sum_{j=1}^n |F_j|^2 \leq 2 \sum_{j=1}^n |F_j^+|^2 + 2 \sum_{j=1}^n |F_j^-|^2$ , it is sufficient to consider  $\sum_{j=1}^n |F_j^+|^2$  (the term involving  $F_j^-$  is treated analogously). We have

$$\begin{aligned} F_j^+ &= \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n \theta_l \left[ 1 + \frac{n}{2\pi i l} (1 - e^{2\pi i l / n}) \right] e^{-2\pi i k l / n} e^{-2\pi i k j / n} \\ &= \frac{1}{n} \sum_{l=1}^n \theta_l \left[ 1 + \frac{n}{2\pi i l} (1 - e^{2\pi i l / n}) \right] \sum_{k=1}^n e^{-2\pi i k l / n} e^{-2\pi i k j / n} \\ &= \theta_{l(j)} \left[ 1 + \frac{n}{2\pi i l(j)} (1 - e^{2\pi i l(j) / n}) \right] \end{aligned}$$

where  $l(j) = n - j$  for  $j = 1, \dots, n - 1$  and  $l(n) = n$ . Here, we used the well-known fact that for any integer  $m \in \mathbb{Z}$

$$\sum_{k=1}^n e^{-2\pi i k m/n} = \begin{cases} n, & \text{if } m \in n\mathbb{Z}, \\ 0, & \text{if } m \notin n\mathbb{Z}. \end{cases} \quad (31)$$

Thus, we obtain (uniformly with respect to  $\Theta$ )

$$\begin{aligned} \sum_{j=1}^n |F_j^+|^2 &= \sum_{j=1}^n |\theta_j|^2 \left| 1 + \frac{n}{2\pi i j} (1 - e^{2\pi i j/n}) \right|^2 \\ &= \sum_{j=1}^n |\theta_j|^2 \left| \frac{n}{2\pi i j} (e^{2\pi i j/n} - 1 - 2\pi i j/n) \right|^2 \\ &\asymp n^2 \sum_{j=1}^n \frac{|\theta_j|^2}{|j|^2} |e^{2\pi i j/n} - 1 - 2\pi i j/n|^2 \\ &\lesssim n^2 \sum_{j=1}^n \frac{|\theta_j|^2}{|j|^2} \cdot |j/n|^4 = n^{-2} \sum_{j=1}^n |\theta_j|^2 |j|^2 \\ &= n^{-2} \sum_{j=1}^n |\theta_j|^2 |j|^{2\beta} |j|^{2-2\beta} \leq n^{-2} L^2 \max\{1, n^{2-2\beta}\} \\ &\lesssim \max\{n^{-2}, n^{-2\beta}\}. \end{aligned}$$

An analogous argument for the term  $\sum_{j=1}^n |F_j^-|^2$  proves the estimate (28).  $\square$

**Proof of (29):** We have

$$\begin{aligned} \sum_{j=1}^n |f_K^\top(t_{j,n})|^2 &= \sum_{j=1}^n \left| \sum_{|k|>K} \theta_k \exp(-2\pi i k j/n) \right|^2 \\ &\leq 2 \sum_{j=1}^n \left| \sum_{k>K} \theta_k \exp(-2\pi i k j/n) \right|^2 + 2 \sum_{j=1}^n \left| \sum_{k<-K} \theta_k \exp(-2\pi i k j/n) \right|^2, \end{aligned}$$

and it is again sufficient to consider the sum running over  $k > K$ . Using (31) again, we get

$$\begin{aligned} \sum_{j=1}^n \left| \sum_{k>K} \theta_k \exp(-2\pi i k j/n) \right|^2 &= \sum_{k,l>K} \theta_k \bar{\theta}_l \sum_{j=1}^n \exp(-2\pi i (k-l)j/n) \\ &= n \sum_{\substack{k,l>K \\ k-l \in n\mathbb{Z}}} \theta_k \bar{\theta}_l. \end{aligned}$$

Taking  $K = n$  here as well yields

$$\sum_{j=1}^n \left| \sum_{k>K} \theta_k \exp(-2\pi i k j/n) \right|^2 = \sum_{m=n+1}^{2n} \left| \sum_{r=0}^{\infty} \theta_{m+r} \right|^2.$$



Now,

$$\left| \sum_{r=0}^{\infty} \theta_{m+rn} \right|^2 \leq \left( \sum_{r=0}^{\infty} |\theta_{m+rn}|^2 (m+rn)^{2\beta} \right) \left( \sum_{r=0}^{\infty} (m+rn)^{-2\beta} \right),$$

and we obtain

$$\sum_{m=n+1}^{2n} \left| \sum_{r=0}^{\infty} \theta_{m+rn} \right|^2 \leq nL^2 \sum_{r=0}^{\infty} (n+rn)^{-2\beta} \lesssim n^{1-2\beta}$$

uniformly over  $f \in \Theta(\beta, L)$ , which establishes (29).  $\square$

**Proof of (30):** Using Jensen's inequality and Parseval's identity we obtain

$$\sum_{i=1}^n \left( n \int_{t_{i-1,n}}^{t_{i,n}} f_K^{\top}(s) ds \right)^2 \leq n \int_0^1 (f_K^{\top}(s))^2 ds = n \sum_{|k|>K} |\theta_k|^2 \leq nL^2 K^{-2\beta}$$

uniformly over  $f \in \Theta(\beta, L)$ , which is of order  $n^{1-2\beta}$  if we choose  $K = n$ .  $\square$

*Verification of condition (ii):* We have to show that

$$D_n := \min_{\alpha_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{j=1}^n \alpha_j K_{\Xi}(\cdot, t_{j,n})\|_{\mathcal{H}(\Xi)}^2 = o(n^{-1}) \quad (32)$$

uniformly over all  $f \in \Theta(\beta, L)$ . Via the isomorphism  $\psi$  introduced in Eq. (10) we have for  $g = \psi F_f$

$$\begin{aligned} D_n &= \min_{\alpha_n \in \mathbb{R}^n} \|g(\cdot) - \sum_{j=1}^n \alpha_j v(t_{j,n}) \mathbf{1}_{[0, q(t_{j,n})]}(\cdot)\|_{L^2([0, T])}^2 \\ &= \min_{\alpha_n \in \mathbb{R}^n} \|g(\cdot) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(q(t_{j-1,n}), q(t_{j,n})]}(\cdot)\|_{L^2([0, T])}^2 \\ &= \min_{\alpha_n \in \mathbb{R}^n} \int_0^T (g(u) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(q(t_{j-1,n}), q(t_{j,n})]}(u))^2 du \\ &= \min_{\alpha_n \in \mathbb{R}^n} \int_0^1 (g(q(w)) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(q(t_{j-1,n}), q(t_{j,n})]}(q(w)))^2 q'(w) dw \\ &= \min_{\alpha_n \in \mathbb{R}^n} \int_0^1 (g(q(w)) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(t_{j-1,n}, t_{j,n})]}(w))^2 q'(w) dw. \end{aligned}$$

Assuming that  $q'$  is bounded from above we obtain

$$D_n \leq C(q) \min_{\alpha_n \in \mathbb{R}^n} \int_0^1 \left( g(q(w)) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(t_{j-1,n}, t_{j,n}]}(w) \right)^2 dw.$$

Note that

$$g(q(w)) = g_1(q(w)) - g_2(q(w))$$

with

$$g_1(q(w)) = \frac{f(w)}{v(w)q'(w)}, \quad \text{and} \quad g_2(q(w)) = \frac{F_f(w)v'(w)}{v^2(w)q'(w)}.$$

With  $f_n(\cdot) = \sum_{|k| \leq n} \theta_k \mathbf{e}_k(\cdot)$  (and  $f_n^\top(\cdot) = f(\cdot) - f_n(\cdot)$ ) we define

$$\alpha_j^{(1)} = \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \quad \text{and} \quad \alpha_j^{(2)} = \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(t_{j,n})q'(t_{j,n})}.$$

Using these notations we get

$$D_n = \min_{\alpha_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{j=1}^n \alpha_j K_\Xi(\cdot, t_{j,n})\|_{\mathcal{H}(\Xi)}^2 \lesssim I_1 + I_2 \quad (33)$$

where

$$I_1 = \int_0^1 \left( g_1(q(w)) - \sum_{j=1}^n \alpha_j^{(1)} \mathbf{1}_{[t_{j-1,n}, t_{j,n}]}(w) \right)^2 dw, \quad (34)$$

$$I_2 = \int_0^1 \left( g_2(q(w)) - \sum_{j=1}^n \alpha_j^{(2)} \mathbf{1}_{[t_{j-1,n}, t_{j,n}]}(w) \right)^2 dw. \quad (35)$$

We investigate the two terms  $I_1$  and  $I_2$  separately.

*Bound for  $I_1$ :* We use the estimate

$$I_1 \lesssim I_{11} + I_{12}, \quad (36)$$

where

$$I_{11} = \int_0^1 \left( \frac{f_n(w)}{v(w)q'(w)} - \sum_{j=1}^n \alpha_j^{(1)} \mathbf{1}_{[t_{j-1,n}, t_{j,n}]}(w) \right)^2 dw,$$

$$I_{12} = \int_0^1 \left( \frac{f_n^\top(w)}{v(w)q'(w)} \right)^2 dw.$$

For the first integral  $I_{11}$  on the right-hand side of (36), we have

$$I_{11} = \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(w)}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw \quad (37)$$

$$\lesssim \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(w)}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(w)q'(w)} \right)^2 dw$$

$$+ \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(t_{j,n})}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw. \quad (38)$$

First, we further decompose (37) as

$$\begin{aligned} & \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(w)}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(w)q'(w)} \right)^2 dw \\ & \lesssim C(v, q) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} (f_n(w) - f_n(t_{j,n}))^2 dw \\ & \lesssim \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} |f_n^+(w) - f_n^+(t_{j,n})|^2 dw + \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} |f_n^-(w) - f_n^-(t_{j,n})|^2 dw, \end{aligned}$$

where

$$f_n^+ = \sum_{k=1}^n \theta_k \mathbf{e}_k(\cdot) \quad \text{and} \quad f_n^- = \sum_{k=-n}^{-1} \theta_k \mathbf{e}_k(\cdot).$$

In the sequel, we consider only the term involving  $f_n^+$  since the sum involving  $f_n^-$  can be bounded using the same argument. We have the identity

$$\begin{aligned} |f_n^+(w) - f_n^+(t_{j,n})|^2 &= \left| \sum_{k=1}^n \theta_k (\mathbf{e}_k(w) - \mathbf{e}_k(t_{j,n})) \right|^2 \\ &= \sum_{k,l=1}^n \theta_k \bar{\theta}_l \exp(-2\pi i k j / n) [\exp(2\pi i k (j/n - w)) - 1] \\ &\quad \cdot \exp(2\pi i l j / n) [\exp(-2\pi i l (j/n - w)) - 1]. \end{aligned}$$

From this identity we obtain (exploiting (31) again)

$$\begin{aligned}
& \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} |f_n^+(w) - f_n^+(t_{j,n})|^2 dw \\
&= \sum_{j=1}^n \sum_{k,l=1}^n \theta_k \bar{\theta}_l \exp(2\pi i(l-k)j/n) \int_{\frac{j-1}{n}}^{\frac{j}{n}} [\exp(2\pi i k(j/n - w)) - 1] \\
&\quad \cdot [\exp(-2\pi i l(j/n - w)) - 1] dw \\
&= \sum_{j=1}^n \sum_{k,l=1}^n \theta_k \bar{\theta}_l \exp(2\pi i(l-k)j/n) \int_0^{\frac{1}{n}} [\exp(2\pi i k(1/n - w)) - 1] \\
&\quad \cdot [\exp(-2\pi i l(1/n - w)) - 1] dw \\
&= n \sum_{k=1}^n |\theta_k|^2 \int_0^{\frac{1}{n}} [\exp(2\pi i k(1/n - w)) - 1] [\exp(-2\pi i k(1/n - w)) - 1] dw \\
&\leq Cn \sum_{k=1}^n |\theta_k|^2 k^2 n^{-3} \\
&\leq Cn^{-2} \sum_{k=1}^n |\theta_k|^2 k^{2\beta} k^{-2\beta+2} \\
&\leq C(L)n^{-2} \max\{1, n^{-2\beta+2}\} \lesssim \max\{n^{-2}, n^{-2\beta}\} = o(n^{-1}).
\end{aligned}$$

To derive an estimate of (38) we note that for any  $n \in \mathbb{N}$ ,

$$\|f_n\|_\infty^2 = \sup_{x \in [0,1]} \left| \sum_{|k| \leq n} \theta_k \mathbf{e}_k(x) \right|^2 \leq \left( \sum_{|k| \leq n} |\theta_k|^2 k^{2\beta} \right) \left( \sum_{|k| \leq n} k^{-2\beta} \right) \leq C(L, \beta)$$

(the same estimate holding true for  $f$  instead of  $f_n$  which formally corresponds to  $n = \infty$ ). Hence,

$$\begin{aligned}
& \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(t_{j,n})}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw \\
&= \sum_{j=1}^n |f_n(t_{j,n})|^2 \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{1}{v(w)q'(w)} - \frac{1}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw \\
&\leq C(L, \beta) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left| \frac{v(t_{j,n})q'(t_{j,n}) - v(w)q'(w)}{v(w)q'(w)v(t_{j,n})q'(t_{j,n})} \right|^2 dw \\
&\leq C(L, \beta, v, q) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} |v(t_{j,n})q'(t_{j,n}) - v(w)q'(w)|^2 dw.
\end{aligned}$$

Because the product of a  $\gamma_1$ -Hölder function and a  $\gamma_2$ -Hölder function is (at least) Hölder with index  $\min\{\gamma_1, \gamma_2\}$  we obtain from our assumptions that the function  $vq'$  is Hölder continuous with index  $\gamma > 1/2$ . Thus,

$$\begin{aligned} \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f_n(t_{j,n})}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw &\leq C(L, \beta, v, q) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} n^{-2\gamma} dw \\ &= C(L, \beta, v, q) n^{-2\gamma}, \end{aligned}$$

and this is  $o(n^{-1})$  since  $\gamma > 1/2$ . Combining these arguments we obtain

$$I_{11} = o(n^{-1}).$$

Finally, the second integral  $I_{12}$  on the right-hand side of (36) can be bounded as follows:

$$\begin{aligned} I_{12} &= \int_0^1 \left( \frac{f_n^\top(w)}{v(w)q'(w)} \right)^2 dw \leq C(v, q) \int_0^1 |f_n^\top(w)|^2 dw \\ &\leq C(v, q) \sum_{|k|>n} |\theta_k|^2 \leq C(v, q) L^2 n^{-2\beta} = o(n^{-1}). \end{aligned}$$

Observing the estimate (26) we finally obtain  $I_1 = o(n^{-1})$ .

*Bound for  $I_2$ :* In analogy to the decomposition of the term  $I_{11}$  on the right-hand side of (36) we have

$$I_2 = \int_0^1 \left( \frac{F_f(w)v'(w)}{v^2(w)q'(w)} - \sum_{j=1}^n \alpha_j^{(2)} \mathbf{1}_{[t_{j-1,n}, t_{j,n})}(w) \right)^2 dw \quad (39)$$

$$\begin{aligned} &\lesssim \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{F_f(w)v'(w)}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} \right)^2 dw \\ &+ \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(t_{j,n})q'(t_{j,n})} \right)^2 dw. \end{aligned} \quad (40)$$

Note that  $F_f$  is Lipschitz since it is continuously differentiable (recall that  $f$  itself is continuous since  $\beta > 1/2$ ) and  $v'$  is Hölder with index  $\gamma > 1/2$  due to our assumptions. Thus, the term (39) can be bounded as

$$\sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{F_f(w)v'(w)}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} \right)^2 dw \leq C(v, q) n^{-2\gamma}$$

which is of order  $o(n^{-1})$ . For the term (40) we obtain using our assumptions that

$$\begin{aligned} &\sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(t_{j,n})q'(t_{j,n})} \right)^2 dw \\ &\leq C(\beta, L, v, q) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} (v^2(t_{j,n})q'(t_{j,n}) - v^2(w)q'(w))^2 dw. \end{aligned}$$

Using the same arguments as for the bound of (38), this term can be shown to be of order  $o(n^{-1})$ . Since both terms  $I_1$  and  $I_2$  are of order  $o(n^{-1})$  the assertion (32) follows from (33).  $\square$

#### Proof of Theorem 4

As in the proof of Theorem 3, we have to verify the two conditions (i) and (ii) from Theorem 2.

*Verification of condition (i):* As in the Sobolev case it is sufficient to show that

$$\sup_{f \in \mathcal{F}(\alpha, L, M)} \sum_{i=1}^n \left( f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s) ds \right)^2 \rightarrow 0.$$

By the mean value theorem  $n \int_{t_{i-1,n}}^{t_{i,n}} f(s) ds = f(\zeta_{i,n})$  for some  $t_{i-1,n} \leq \zeta_{i,n} \leq t_{i,n}$ . Thus, since  $f \in \mathcal{F}(\alpha, L, M)$ ,

$$\begin{aligned} \sum_{i=1}^n \left( f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s) ds \right)^2 &= \sum_{i=1}^n (f(t_{i,n}) - f(\zeta_{i,n}))^2 \\ &\leq \sum_{i=1}^n L^2 |t_{i,n} - \zeta_{i,n}|^2 \\ &\leq L^2 n^{-2\alpha+1}, \end{aligned}$$

and the last term converges to zero uniformly over  $f \in \mathcal{F}(\alpha, L, M)$  whenever  $\alpha > 1/2$ .

*Verification of condition (ii):* The proof is based on nearly the same reduction as in the Sobolev case. Again, we consider the bound

$$\min_{\alpha_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{j=1}^n \alpha_j K_{\Xi}(\cdot, t_{j,n})\|_{\mathcal{H}(\Xi)}^2 \lesssim I_1 + I_2 \quad (41)$$

where we define  $I_1$  and  $I_2$  as in Appendix B.1 (see the equations (34) and (35)) with the only exception that we now put

$$\alpha_j^{(1)} = \frac{f(t_{j,n})}{v(t_{j,n})q'(t_{j,n})}$$

in the definition of  $I_1$ . Then,

$$\begin{aligned}
I_1 &\lesssim \int_0^1 \left( \frac{f(w)}{v(w)q'(w)} - \sum_{j=1}^n \alpha_j^{(1)} \mathbf{1}_{[t_{j-1,n}, t_{j,n}]}(w) \right)^2 dw \\
&= \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f(w)}{v(w)q'(w)} - \frac{f(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw \\
&\lesssim \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f(w)}{v(w)q'(w)} - \frac{f(t_{j,n})}{v(w)q'(w)} \right)^2 dw \\
&\quad + \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f(t_{j,n})}{v(w)q'(w)} - \frac{f(t_{j,n})}{v(t_{j,n})q'(t_{j,n})} \right)^2 dw.
\end{aligned}$$

The first integral can be bounded as

$$\begin{aligned}
\sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left( \frac{f(w)}{v(w)q'(w)} - \frac{f(t_{j,n})}{v(w)q'(w)} \right)^2 dw &\leq C(v, q) \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} (f(w) - f(t_{j,n}))^2 dw \\
&\leq C(v, q) L^2 n^{-2\alpha},
\end{aligned}$$

which converges to zero as  $n$  increases. The second integral can be bounded as in the Sobolev case using the assumption that  $f$  is bounded (as one can easily see, the assumption of uniform boundedness can be dropped if  $v \cdot q'$  is constant; this is for instance satisfied in the case of Brownian motion). Hence,  $I_1$  converges to 0.

The term  $I_2$  in (41) can be bounded exactly as the corresponding term in the Sobolev ellipsoid case. This finishes the proof of the theorem.  $\square$

### Proof of Theorem 5

To prepare the proof, we recall an alternative (but equivalent) characterization of asymptotic equivalence in the framework of statistical decision theory. Let  $\mathfrak{E} = (\mathcal{X}, \mathcal{X}, (\mathbf{P}_\theta)_{\theta \in \Theta})$  be a statistical experiment. In decision theory one considers a decision space  $(\mathcal{A}, \mathcal{A})$  where the set  $\mathcal{A}$  contains the potential decisions (or actions) that are at the observers disposal and  $\mathcal{A}$  is a  $\sigma$ -field on  $\mathcal{A}$ . In addition, there is a *loss function*

$$\ell : \Theta \times \mathcal{A} \rightarrow [0, \infty), \quad (\theta, a) \mapsto \ell(\theta, a)$$

with the interpretation that a loss  $\ell(\theta, a)$  occurs if the statistician chooses the action  $a \in \mathcal{A}$  and  $\theta \in \Theta$  is the true state of nature. A (randomized) decision rule is a Markov kernel  $\rho : \mathcal{X} \times \mathcal{A} \rightarrow [0, 1]$ , and the associated *risk* is

$$R_\theta(\mathfrak{E}, \rho, \ell) = \int_{\mathcal{X}} \left( \int_{\mathcal{A}} \ell(\theta, a) \rho(x, da) \right) \mathbf{P}_\theta(dx).$$

Then, the deficiency between two experiments  $\mathfrak{E}_1$  and  $\mathfrak{E}_2$  is exactly the quantity

$$\delta(\mathfrak{E}_1, \mathfrak{E}_2) = \inf_{\rho_1} \sup_{\rho_2} \sup_{\theta} \sup_{\ell} |R_{\theta}(\mathfrak{E}_1, \rho_1, \ell) - R_{\theta}(\mathfrak{E}_2, \rho_2, \ell)| \quad (42)$$

(see Mariucci, 2016, Theorem 2.7, and the references cited there), where the supremum is taken over all loss functions  $\ell$  with  $0 \leq \ell(\theta, a) \leq 1$  for all  $\theta \in \Theta$  and  $a \in \mathcal{A}$ , all admissible parameters  $\theta \in \Theta$ , and decision rules  $\rho_2$  in the second experiment. The infimum is taken over all decision rules  $\rho_1$  in the first experiment.

After these preliminaries, let us go on to the proof of the theorem. We consider the decision space  $(\mathcal{A}, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and the loss function

$$\ell : \Theta(\beta, L) \times \mathcal{A} \rightarrow \{0, 1\}, \quad (f, a) \mapsto \ell(f, a) = \begin{cases} 1, & \text{if } \int_0^1 f(x) dx \neq a, \\ 0, & \text{if } \int_0^1 f(x) dx = a. \end{cases}$$

In the experiment  $\mathfrak{E}_{2,n}$  we observe the whole path  $Y = \{Y_t, t \in [0, 1]\}$  satisfying

$$Y_t = \int_0^t f(s) ds + \frac{1}{\sqrt{n}} B_t, \quad t \in [0, 1],$$

where  $(B_t)_{t \in [0, 1]}$  is a Brownian Bridge, and we consider the (non-randomized) decision rule  $\rho_2$  defined by

$$\rho_2(h) = h(1) - h(0), \quad h \in \mathcal{C}([0, 1], \mathbb{R}).$$

This directly yields  $\rho_2(Y) = \int_0^1 f(s) ds$  since  $\Xi_0 = \Xi_1 = 0$  for the Brownian bridge. Hence,

$$R_f(\mathfrak{E}_{2,n}, \rho_2, \ell) = \int \ell(f, \rho_2(Y)) \mathbf{P}_{2,n}^f(dY) = 0$$

for all  $f \in \Theta(\beta, L)$  and  $n \in \mathbb{N}$ . From (42) we thus obtain

$$\delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n}) \geq \inf_{\rho_1} \sup_{f \in \Theta(\beta, L)} |R_f(\mathfrak{E}_{1,n}, \rho_1, \ell)|.$$

Recall the notation  $\mathbf{e}_n(x) = \exp(-2\pi i n x)$  and introduce the functions  $f_0 \equiv 0$  and

$$f_n(x) = \sqrt{\frac{2}{3}} \frac{L}{n^\beta} \left[ 1 - \frac{1}{2} \mathbf{e}_n(x) - \frac{1}{2} \mathbf{e}_{-n}(x) \right] = \sqrt{\frac{2}{3}} \frac{L}{n^\beta} \left[ 1 - \frac{1}{2} \cos(2\pi n x) \right]$$

for  $n \in \mathbb{N}$ . It is easily seen that  $f_n$  belongs to  $\Theta(\beta, L)$ . Note that by construction  $f_n(j/n) = 0$  for  $j = 1, \dots, n$  and thus in the experiment  $\mathfrak{E}_{1,n}$  the identity  $\mathbf{P}_{1,n}^{f_0} = \mathbf{P}_{1,n}^{f_n}$  holds. For the considered loss function we have

$$R_f(\mathfrak{E}_{1,n}, \rho_1, \ell) = \int_{\mathbb{R}^n} \rho_1(\mathbf{Y}_n) \left( \mathbb{R} \setminus \left\{ \int_0^1 f(x) dx \right\} \right) \mathbf{P}_{1,n}^f(d\mathbf{Y}_n),$$

where  $\rho_1$  is any (potentially randomized) decision rule. Because



$$\int_0^1 f_0(x)dx = 0 \neq \sqrt{2/3}Ln^{-\beta} = \int_0^1 f_n(x)dx$$

at least one of the quantities  $\rho_1(\mathbf{Y}_n)(\mathbb{R} \setminus \{\int_0^1 f_0(x)dx\})$  and  $\rho_1(\mathbf{Y}_n)(\mathbb{R} \setminus \{\int_0^1 f_n(x)dx\})$  must be  $\geq 1/2$  for any  $\mathbf{Y}_n$  (otherwise there is a contradiction). Thus, setting  $A_\diamond = \{\mathbf{Y}_n \in \mathbb{R}^n : \rho_1(\mathbf{Y}_n)(\mathbb{R} \setminus \{\int_0^1 f_\diamond(x)dx\}) \geq 1/2\}$  for  $\diamond \in \{0, n\}$  one has  $A_0 \cup A_n = \mathbb{R}^n$ . As a consequence, either  $\mathbf{P}_{1,n}^{f_0}(A_0) = \mathbf{P}_{1,n}^{f_n}(A_0) \geq 1/2$  or  $\mathbf{P}_{1,n}^{f_0}(A_n) = \mathbf{P}_{1,n}^{f_n}(A_n) \geq 1/2$  holds. Without loss of generality, we assume that  $\mathbf{P}_{1,n}^{f_0}(A_0) = \mathbf{P}_{1,n}^{f_n}(A_0) \geq 1/2$  (the other case follows by exactly the same argument). In this case, using the definition of the set  $A_0$ ,

$$\begin{aligned} \delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n}) &\geq \inf_{\rho_1} \sup_{f \in \{f_0, f_n\}} R_f(\mathfrak{E}_{1,n}, \rho_1, \mathcal{C}) \\ &\geq \inf_{\rho_1} \int_{\mathbb{R}^n} \rho_1(\mathbf{Y}_n) \left( \mathbb{R} \setminus \left\{ \int_0^1 f_0(x)dx \right\} \right) \mathbf{P}_{1,n}^{f_0}(d\mathbf{Y}_n) \\ &\geq \inf_{\rho_1} \int_{A_0} \rho_1(\mathbf{Y}_n) \left( \mathbb{R} \setminus \left\{ \int_0^1 f_0(x)dx \right\} \right) \mathbf{P}_{1,n}^{f_0}(d\mathbf{Y}_n) \\ &\geq \frac{1}{2} \mathbf{P}_{1,n}^{f_0}(A_0) \\ &\geq \frac{1}{4}, \end{aligned}$$

which proves the assertion.  $\square$

## References

- Abundo, M. (2014). On the representation of an integrated Gauss–Markov process. *Scientiae Mathematicae Japonicae*, 77, 357–361.
- Beder, J. H. (1987). A sieve estimator for the mean of a Gaussian process. *The Annals of Statistics*, 15, 59–78.
- Berlinet, A., Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Boston, MA: Kluwer Academic Publishers.
- Brown, L. D., Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, 24, 2384–2398.
- Brown, L. D., Zhang, C. H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is  $1/2$ . *The Annals of Statistics*, 26, 279–287.
- Brown, L. D., Cai, T. T., Low, M. G., Zhang, C. H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics*, 30, 688–707.
- Carter, A. V. (2006). A continuous Gaussian approximation to a nonparametric regression in two dimensions. *Bernoulli*, 12, 143–156.
- Carter, A. V. (2007). Asymptotic approximation of nonparametric regression experiments with unknown variances. *The Annals of Statistics*, 35, 1644–1673.
- Carter, A.V. (2009) Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. *Journal of Probability and Statistics*, Article ID 275308.
- Dette, H., Pepelyshev, A., Zhigljavsky, A. (2016). Optimal designs in regression with correlated errors. *The Annals of Statistics*, 44, 113–152.
- Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 20, 393–403.

- Golubev, G. K., Nussbaum, M., Zhou, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *The Annals of Statistics*, 38, 181–214.
- Grama, I., Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probability Theory and Related Fields*, 111, 167–214.
- Grama, I. G., Neumann, M. H. (2006). Asymptotic equivalence of nonparametric autoregression and nonparametric regression. *The Annals of Statistics*, 34, 1701–1732.
- Johnstone, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statistica Sinica*, 9, 51–83.
- Johnstone, I. M., Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 319–351.
- Karatzas, I., Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. New York: Springer.
- Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. New York: Springer.
- Le Cam, L., Yang, G. L. (2000). *Asymptotics in statistics*. New York: Springer.
- Mariucci, E. (2016). Le Cam theory on the comparison of statistical models. *The Graduate Journal of Mathematics*, 1, 81–91.
- Mehr, C. B., McFadden, J. A. (1965). Certain properties of Gaussian processes and their first-passage times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 27, 505–522.
- Meister, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *The Annals of Statistics*, 39, 1471–1495.
- Neveu, J. (1968). *Processus aléatoires gaussiens*. Montréal: Les Presses de l'Université de Montréal, Séminaire de Mathématiques Supérieures, No. 34 (Été, 1968).
- Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *The Annals of Statistics*, 24, 2399–2430.
- Paulsen, V. I., Raghupathi, M. (2016). *An introduction to the theory of reproducing kernel Hilbert spaces*. Cambridge: Cambridge University Press.
- Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *The Annals of Statistics*, 36, 1957–1982.
- Reiß, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics*, 39, 772–802.
- Rohde, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statistics & Decisions*, 22, 235–243.
- Schmidt-Hieber, J. (2014). Asymptotic equivalence for regression under fractional noise. *The Annals of Statistics*, 42, 2557–2585.
- Slepian, D. (1961). First passage time for a particular Gaussian process. *The Annals of Mathematical Statistics*, 32, 610–612.
- Torgersen, E. (1991). *Comparison of statistical experiments*. Cambridge: Cambridge University Press.
- Wendland, H. (2005). *Scattered data approximation*. Cambridge: Cambridge University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.