Asymptotic equivalence for nonparametric regression with dependent errors: Gauss-Markov processes*

Holger Dette and Martin Kroll

Ruhr-Universität Bochum

April 19, 2021

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. In particular we provide sufficient conditions such that asymptotic equivalence of the two models holds for functions from a given class, and we verify these for the special cases of Sobolev ellipsoids and Hölder classes with smoothness index $> 1/2$ under mild assumptions on the Gauss-Markov process at hand. To derive these results, we develop an explicit characterization of the reproducing kernel Hilbert space associated with the Gauss-Markov process, that hinges on a characterization of such processes by a property of the corresponding covariance kernel introduced by Doob [Doo49]. In order to demonstrate that the given assumptions on the Gauss-Markov process are in some sense sharp we also show that asymptotic equivalence fails to hold for the special case of Brownian bridge. Our results demonstrate that the well-known asymptotic equivalence of the Gaussian white noise model and the nonparametric regression model with independent standard normal distributed errors (see Brown and Low, [BL96]) can be extended to a broad class of models with dependent data.

Keywords: Asymptotic equivalence, nonparametric regression, dependent errors, Gauss-Markov process, triangular kernel

AMS Subject classification: 62B15, 62G08, 60G15, 62G20

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

1 Introduction

In a seminal paper Brown and Low [BL96] establish asymptotic equivalence of the nonparametric regression model with discrete observations

$$Y_{i,n} = f\left(\frac{i}{n}\right) + \eta_i, \qquad i = 1, \dots, n, \tag{1.1}$$

and the continuous time model defined by the stochastic differential equation

$$\mathrm{d}Y_t = f(t)\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W_t, \quad t \in [0,1], \tag{1.2}$$

where $\eta_1, \dots, \eta_n$ are independent, standard Gaussian random variables, W denotes a standard Brownian motion (that is, $\mathrm{d}W$ is white noise) and f is the unknown nonparametric drift satisfying a smoothness assumption. Equation (1.2) 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 (1.2) than in (1.1). 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 (1.1) 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 [GN98] investigate nonparametric generalized linear models, and Brown and Zhang [BZ98] prove nonequivalence when the smoothness of the function class is equal to 1/2. Brown et al. [Bro+02] and Reiß [Rei08] study the random design in the one-dimensional and multivariate case, respectively (see also Carter, [Car06] for some results in models with a multivariate fixed design). The general framework in [BL96] is already formulated for heteroscedastic errors and Carter [Car07] shows asymptotic equivalence for unknown variances and design density. We also refer to the work of Reiß [Rei11] and Meister [Mei11] 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 [GN06] consider a nonparametric autoregression model but still use an i.i.d. assumption for the innovations. Golubev, Nussbaum, and Zhou [GNZ10] 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 [JS97] and Johnstone [Joh99], Carter [Car09] 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 [SH14], who considers the nonparametric regression model (1.1) with fractional Gaussian noise (fGN) and establishes the asymptotic equivalence to a model of the form (1.2), 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 $H \in (\frac{1}{4}, 1)$ 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 $H \in [\frac{1}{2}, 1)$.

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 (1.2). 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

$$Y_t = \int_0^t f(s)\mathrm{d}s + \frac{1}{\sqrt{n}}\Xi_t, \quad t \in [0,1]. \tag{1.3}$$

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

$$Y_{i,n} = f\left(\frac{i}{n}\right) + \sqrt{n}\xi_{i,n}, \quad i = 1, \dots, n, \tag{1.4}$$

where $\xi_{i,n} \stackrel{\mathcal{L}}{=} \Xi_{i/n} - \Xi_{(i-1)/n}$ denote the increments of the process Ξ. Note that in general the observation errors $\xi_{i,n}$ are not uncorrelated. However, for the special case of Brownian motion, that is $\Xi \stackrel{\mathcal{L}}{=} W$, Equations (1.1) and (1.4) 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 (1.3) and (1.4) 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 [BL96] considers the case where the regression f is an element of a function class, say Θ, and uses the key assumption

$$\lim_{n\to\infty} \sup_{f\in\Theta} n \int_0^1 \Big(f(t) - \bar{f}_n(t)\Big)^2 \mathrm{d}t = 0, \tag{1.5}$$

where $\bar{f}_n(\cdot) = \sum_{j=1}^n f(j/n)\mathbf{1}_{((j-1)/n, j/n]}(\cdot)$ 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 $\bar{f}_n$. 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 (1.5) guarantees that the Le Cam distance between the white noise models with parameters f and $\bar{f}_n$, respectively, vanishes as $n \to \infty$. 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 [Roh04], who considers a Gaussian sequence space model as an intermediate experiment between the two experiments of interest. Schmidt-Hieber [SH14] in some sense generalizes the conditions in [BL96] 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 [Doo49] 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 (1.3) and (1.4).

The remaining part of the paper is organized as follows. In Section 2 we introduce Gauss-Markov processes and recap the characterization of such processes introduced by Doob [Doo49] 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 Section 2.1 for a precise definition). In Section 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 Section 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 Section 4 where we establish asymptotic equivalence of the models (1.3) and (1.4) under mild assumptions on the Gauss-Markov process in model (1.3) for Sobolev ellipsoids and Hölder classes. Finally, in Section 4.3 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.

Notation

Vectors will be denoted with bold letters (i.e., we write $\boldsymbol{x}_n = (x_{1,n}, \dots, x_{m(n),n})$ when both length and entries of a vector might vary with n). We also use the shorthand $\boldsymbol{x}_{1:k,n} = (x_{1,n}, \dots, x_{k,n})$. We write $a_n \lesssim b_n$ if $a_n \leq C b_n$ 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 Sections 3 and 4. In Section 2.1 we recapitulate a characterization of Gauss-Markov process going back to Doob 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 Section 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 essentially characterized by the following factorization property of the covariance function:

$$K_X(s,t) := \mathrm{Cov}(X_s, X_t) = \mathbf{E}[X_s X_t] = U(s)V(t) \quad \text{for } 0 \le s \le t \le 1, \tag{2.1}$$

where U and V are (known) non-negative functions on the interval $[0,1]$; see [Doo49, MM65] for details. Kernels with the factorization property (2.1) 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 ($U(t) = \exp(Lt)$ 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 2.1. The process $\Xi = (\Xi_t)_{t\in[0,1]}$ is a Gauss-Markov process with $\Xi_0 = 0$, and non-degenerate on the open interval $(0,1)$.

Note that Assumption 2.1 implies that there exist functions u and v in the representation

$$K_\Xi(s,t) = u(s)v(t) \tag{2.2}$$

of the covariance kernel satisfying $u(\cdot)v(\cdot) \ge 0$ 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 [MM65], p. 507). Moreover, under Assumption 2.1, the Gauss-Markov process Ξ can be written in distribution as

$$\Xi_t = v(t) \cdot W_{q(t)}. \tag{2.3}$$

Vice versa, this transformation of Brownian motion defines a Gaussian process with covariance kernel (2.2). Note that a simple calculation shows that the process Ξ has independent increments if and only if the function v is constant.

Let us shortly explain how processes Ξ satisfying Assumption 2.1 can be obtained. First, one can of course define such a process directly by means of the representation (2.3) provided that the function 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 (2.1), 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

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

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

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

where $Q(t) = U(t)/V(t)$. In particular, the process Ξ satisfies Assumption 2.1, 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 [MM65], 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 Ξ on the interval $[0,1]$ is a Gauss-Markov process satisfying Assumption 2.1, 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 by Slepian in [Sle61] where it was proved that the process fulfills a 'peculiar Markov-like property'. The restriction of the process Ξ 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 Section 3 and 4 hold in particular for this process.

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 of Berlinet and Thomas-Agnan [BTA04] and Paulsen and Raghupathi [PR16]. 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 $\ell_x \colon \mathcal{H} \to \mathbb{R}$ 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 $K \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

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

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

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

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 $K_\Xi \colon [0,1]\times[0,1] \to \mathbb{R}$ of a Gauss-Markov process Ξ satisfying Assumption 2.1 as introduced in Section 2. As a consequence

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

Let us from now on assume that $q(0) = 0$ for $q(x) = u(x)/v(x)$ which is under the validity of Assumption 2.1 certainly satisfied whenever $v(0) \neq 0$. This additional condition is valid for all the results in Section 4. As mentioned above, Ξ 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 $\psi \colon \mathcal{H}(\Xi) \to L^2([0,T])$ defined on the 'generators' $\{K_\Xi(\cdot,t),\, t \in [0,1]\}$ via

$$K_\Xi(\cdot, t) \overset{\psi}{\mapsto} v(t)\mathbf{1}_{[0,q(t)]}(\cdot) \tag{2.4}$$

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

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

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

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

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

$$\mathcal{H}(\Xi) = \left\{ F : [0,1] \to \mathbb{R} \;\middle|\; F(t) = v(t)\int_0^{q(t)} g(u)\mathrm{d}u \text{ for some } g \in L^2([0,T]) \right\}, \quad (2.5)$$

and this space is equipped with the norm $\|F\|_{\mathcal{H}(\Xi)} = \|g\|_{L^2([0,T])}$ for F as in (2.5). 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 $F(\cdot) = F_f(\cdot) = \int_0^{\cdot} f(s)\mathrm{d}s$, then f can be derived by differentiation:

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

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

$$g(q(t)) = \left(\frac{F_f(t)}{v(t)}\right)' \cdot \frac{1}{q'(t)}. \quad (2.6)$$

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

3 Abstract conditions for asymptotic equivalence

In this section, we derive sufficient conditions for an abstract function class Θ that guarantee asymptotic equivalence of nonparametric regression and a continuous time model with a . In Section 3.1 we gather the necessary notions from the asymptotic equivalence theory. In Section 3.2, we state the main result that provides the announced sufficient conditions for asymptotic equivalence of models (1.3) and (1.4).

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 [Rei11] 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 [Mei11] 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 [LC86, LCY00, Tor91] as well as in the recent survey paper [Mar16] by Mariucci, 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 [Nus96]. To be precise, let $\mathfrak{E}_i = (\mathcal{X}_i, \mathscr{X}_i, (\mathbf{P}_{i,\theta})_{\theta\in\Theta})$ for $i \in \{1, 2\}$ two such dominated experiments with Polish parameter spaces $(\mathcal{X}_i, \mathscr{X}_i)$, and sharing the same parameter space Θ. Then, the *deficiency* of $\mathfrak{E}_1$ with respect to $\mathfrak{E}_2$ is defined as the quantity

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

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

Building on the defintion 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 $\lim_{n\to\infty} \Delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n}) \to 0$.

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 Θ denote a given class of functions. The first experiment with discrete observations is given by

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

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

$$Y_{i,n} = f(t_{i,n}) + \sqrt{n}\xi_{i,n}, \quad i = 1, \dots, n, \tag{3.1}$$

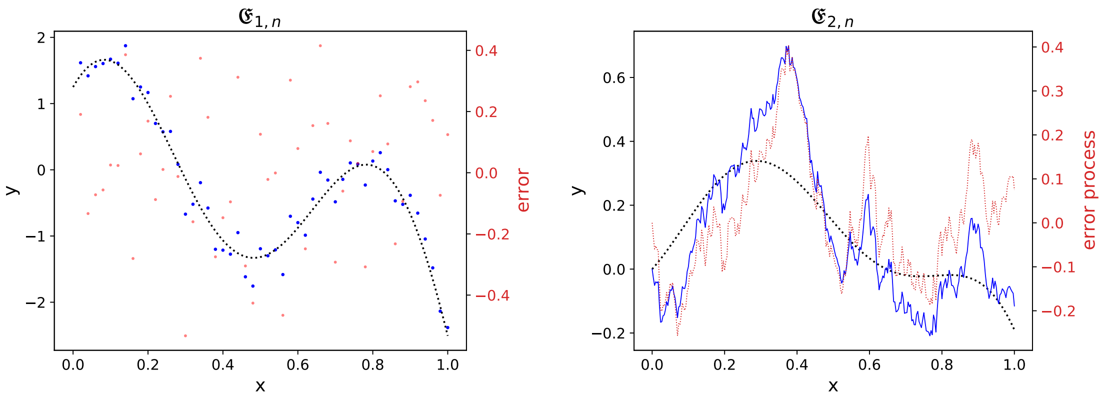

Figure 1: Exemplary realizations of the observations in the two experiments $\mathfrak{E}_{1,n}$ (left) and $\mathfrak{E}_{2,n}$ (right) for $n = 50$ where the Gauss-Markov process is given through the functions $u(x) = \exp(x)$ and $q(x) = \exp(2x) - 1$. In both figures the actual observations are plotted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

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 Section 2.

The second experiment with continuous observations is given through

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

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

$$Y_t = \int_0^t f(s)\mathrm{d}s + \frac{1}{\sqrt{n}}\Xi_t, \quad t \in [0,1]. \tag{3.2}$$

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

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

Theorem 3.1. *Let Ξ be a Gauss-Markov process on the interval $[0,1]$ with $\Xi_0 = 0$ such that Assumption 2.1 holds with $K_\Xi(s,t) = u(s)v(t)$ for $0 \le s \le t \le 1$, and $v(1) \neq 0$. Set $F_f(\cdot) = \int_0^\cdot f(s)\mathrm{d}s$ and assume that the conditions*

(i) $\frac{1}{n}\sup_{f\in\Theta}\sum_{i=1}^n \frac{(f(t_{i,n}) - n\int_{t_{i-1,n}}^{t_{i,n}} f(s)\mathrm{d}s)^2}{v^2(t_{i,n})(q(t_{i,n}) - q(t_{i-1,n}))} \to 0$ *and*

(ii) $\sqrt{n}\sup_{f\in\Theta}\inf_{\boldsymbol{\alpha}_n\in\mathbb{R}^n}\|F_f(\cdot) - \sum_{i=1}^n \alpha_i K_\Xi(\cdot, t_{i,n})\|_{\mathcal{H}(\Xi)} \to 0$

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 $(\mathfrak{E}_{1,n}(\Theta))_{n\in\mathbb{N}}$ and $(\mathfrak{E}_{2,n}(\Theta))_{n\in\mathbb{N}}$ are asymptotically equivalent, i.e.

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

4 Specific function classes and a counterexample

In this section, we first apply the abstract Theorem 3.1 in order to establish asymptotic equivalence over Sobolev ellipsoids (Section 4.1) and Hölder balls (Section 4.2) of sufficiently smooth functions. In Section 4.3 we give a counterexample which shows that asymptotic equivalence cannot hold (even for function classes containing very smooth functions) in general if only Assumption 2.1 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

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

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 $\theta_k = \langle f, \mathbf{e}_k\rangle_{L^2} = \int_0^1 f(s)\exp(2\pi \mathrm{i} k s)\mathrm{d}s$ 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

$$\Theta(\beta, L) = \left\{ f(\cdot) = \sum_{k=-\infty}^{\infty} \theta_k \mathbf{e}_k(\cdot) : \theta_k = \overline{\theta_{-k}} \text{ and } \sum_{k=-\infty}^{\infty} (1+|k|)^{2\beta}|\theta_k|^2 \leq L^2 \right\}. \quad (4.1)$$

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 Section 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 4.1. *Let Ξ be a Gauss-Markov process on the interval $[0,1]$ with $\Xi_0 = 0$ such that Assumption 2.1 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 $0 < \inf_{t\in[0,1]} q'(t) \leq \sup_{t\in[0,1]} q'(t) \leq C < \infty$,

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

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

The proof of Theorem 4.1 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 4.2. It is known that for the smoothness index $\beta = 1/2$ asymptotic equivalence between nonparametric regression model (1.1) and the white noise with drift experiment (1.2) (i.e., our experiment $\mathfrak{E}_{2,n}$ for the special case of Brownian motion) does not hold. See Remark 4.6 in [BL96] for further details. In Section 4.3 below we consider the example of a Brownian bridge as an error process and prove that asymptotic equivalence for arbitrary values of β cannot hold without additional assumptions on the Gauss-Markov process (like, for instance, assumptions on the functions v and q as made in Theorem 4.1).

Remark 4.3. In the proof of Theorem 4.1 it is shown that under our assumptions the Le Cam distance between the experiments $\mathfrak{E}_{1,n}(\Theta(\beta, L))$ and $\mathfrak{E}_{2,n}(\Theta(\beta, L))$ 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 γ 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, [Roh04] for further details. Since we have restricted ourselves to proving asymptotic equivalence, we have not further pursued this issue here.

Remark 4.4 (Sequence space representations). Note that for the special case of Brownian motion the relation between the functions $f = F'$ and g in in the reprsentation (2.5) 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 ψ defined in (2.4) (more precisely, its inverse ψ^{-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

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

where ξ_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 (2.6) does not 'transport' an orthonormal basis for the representation of f to such basis for g.

An alternative sequence space representation (for general Gaussian processes) is obtained from (3.2) using the Karhunen-Loeve representation of the process Ξ. Note that the kernel K_Ξ gives rise to an integral operator defined through

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

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_Ξ is continuous, Mercer's theorem applies and (the operator) K_Ξ has a countable system of eigenfunctions φ_k with corresponding eigenvalues $\lambda_k \geq 0$. By Proposition 3.11 in Neveu [Nev68] an orthonormal basis of the RKHS $\mathcal{H}(\Xi)$ is given by the functions

$g_k = \sqrt{\lambda_k}\varphi_k$, $k = 1, 2, \ldots$ with $\lambda_k > 0$. In terms of this basis, (3.2) can be rewritten as

$$y_k = \theta_k + \frac{1}{\sqrt{n}}\eta_k, \quad k = 1, 2, \ldots \tag{4.2}$$

where $\theta_1, \theta_2, \ldots$ are the Fourier coefficients of the function F with respect to the basis $g_1, g_2, \ldots$ and $\eta_1, \eta_2, \ldots$ 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, [Bed87], p. 66)

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

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 (4.2) 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

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

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 4.1 we obtain asymptotic equivalence for functions from Hölder classes with smoothness index $\alpha > 1/2$.

Theorem 4.5. *Let Ξ be a Gauss-Markov process on the interval $[0,1]$ with $\Xi_0 = 0$ such that Assumption 2.1 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

$$\mathfrak{E}_{1,n}(\mathcal{F}(\alpha, L, M)) \approx \mathfrak{E}_{2,n}(\mathcal{F}(\alpha, L, M)). \tag{4.3}$$

Moreover, if the function $v \cdot q'$ is constant on the interval $[0,1]$, the asymptotic euqivalence (4.3) also holds in the case $M = \infty$.

The proof of Theorem 4.5 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 4.5 reproduces the result given in [BL96], at the top of page 2390.

4.3 Non-equivalence for Brownian bridge

We conclude this section by showing that the asymptotic equivalence result established in Theorem 4.1 cannot be transferred to arbitrary Gauss-Markov processes. To be more precise, Theorem 4.6 below shows that for the special case of Brownian bridge $(B_t)_{t\in[0,1]}$ 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 β and $L > 0$. Recall that for the Brownian Bridge we have $u(t) = t$, $v(t) = 1 - t$, and therefore $v(1) = 0$ and in addition Conditions a) and b) in Theorem 4.1 are not satisfied.

Theorem 4.6. *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* (3.1) *and* (3.2)*, respectively, where error process is given by a Brownian Bridge, that is* $(\Xi_t)_{t\in[0,1]} = (B_t)_{t\in[0,1]}$*. Then, for an arbitrary smoothness index* $\beta > 0$ *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* (4.1)*. 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.*

A Proof of Theorem 3.1

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

First step: We first show that under Condition (i) in Theorem 3.1 the model given by Equation (3.1) is asymptotically equivalent to observing

$$Y'_{i,n} = n\int_{t_{i-1,n}}^{t_{i,n}} f(s)\mathrm{d}s + \sqrt{n}\xi_{i,n}, \quad i = 1,\dots,n, \tag{A.1}$$

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 2.1, we can take advantage of the representation (2.3) 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 (3.1) 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

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

Similarly, (A.1) can be written as

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

For the sake of a transparent notation let $\mathbf{P}^{\mathbf{Y}_n} = \mathbf{P}^f_{1,n}$ 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 $\mathscr{F}_{i,n}$ the σ-algebra generated by $\left\{W_{q(t)}, t \le t_{i,n}\right\}$. We have

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

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

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

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

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

where

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

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

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

we have

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

This yields

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

and $\mathrm{KL}(\mathbf{P}^{\mathbf{Y}_n}, \mathbf{P}^{\mathbf{Y}'_n}) \to 0$ holds if and only if Condition (i) holds. Consequently, the experiments (3.1) and (A.1) are asymptotically equivalent in this case.

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

$$I(t|\boldsymbol{y}_n) = (K_\Xi(t, t_{1,n}), K_\Xi(t, t_{2,n}), \ldots, K_\Xi(t, t_{n,n}))\,\mathrm{Cov}(\boldsymbol{\Xi}_n)^{-1}\boldsymbol{y}_n^\top \tag{A.2}$$

for $\boldsymbol{y}_n = (y_{1,n}, \ldots, y_{n,n})$ and $\boldsymbol{\Xi}_n = (\Xi_{t_{1,n}}, \ldots, \Xi_{t_{n,n}})$ (the additional condition that $v(1) \neq 0$ guarantees the invertibility of $\boldsymbol{\Xi}_n$; see Lemma A.1 in [DPZ16] for an explicit formula for the entries of the inverse matrix). By definition the Kriging predictor is linear in the argument $\boldsymbol{y}_n$, and a simple argument shows the interpolation property

$$I(t_{j,n}|\boldsymbol{y}_n) = y_{j,n} \quad \text{for } j = 1, \ldots, n. \tag{A.3}$$

The second step now consists in proving (exact, that is, non-asymptotic) equivalence of the experiment defined by the discrete observations (A.1) and the experiment defined by the continuous path

$$\widetilde{Y}_t = I(t|\mathbf{F}_{f,n}) + n^{-1/2}\Xi_t, \tag{A.4}$$

where $\mathbf{F}_{f,n} = (F_{f,n}(t_{1,n}), \ldots, F_{f,n}(t_{n,n}))$. 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 the process $(\Xi_t)_{t\in[0,1]}$, we have

$$S'_{k,n} = \sum_{j=1}^k Y'_{j,n} = n\int_0^{t_{k,n}} f(s)\mathrm{d}s + \sqrt{n}\sum_{j=1}^k \xi_{j,n} = nF_f(t_{k,n}) + \sqrt{n}\Xi_{t_{k,n}} \tag{A.5}$$

where we used the interpolating property (A.3) and the the definition (A.4). Let $(\Xi'_t)_{t\in[0,1]}$ be an independent copy of $(\Xi_t)_{t\in[0,1]}$, and set $R_t = \Xi'_t - I(t|\boldsymbol{\Xi}'_n)$ with $\boldsymbol{\Xi}'_n = (\Xi'_{t_{1,n}}, \dots, \Xi'_{t_{n,n}})$. Then, the process

$$I(t|\boldsymbol{\Xi}_n) + R_t, \quad t \in [0,1],$$

follows the same law as $(\Xi_t)_{t\in[0,1]}$ and $(\Xi'_t)_{t\in[0,1]}$, which can 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 (A.4), we have

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

where we used the notation $\mathbf{S}'_n = (S'_{1.n}, \dots, S'_{n.n})$, equation (A.5) and the linearity of the Kriging estimator. Therefore, the process $(\widetilde{Y}_t)_{t\in[0,1]}$ 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 $(\widetilde{Y}_t)_{t\in[0,1]}$ since for $t = t_{k,n}$ the interpolation property (A.3) yields

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

and one obtains $Y'_{i,n}$ as $Y'_{i,n} = n\widetilde{Y}_{t_i} - n\widetilde{Y}_{t_{i-1}}$. Hence, the process $(\widetilde{Y}_t)_{t\in[0,1]}$ and the vector $\mathbf{Y}'_n$ contain the same information and the experiments (A.1) and (A.4) are equivalent.

Third step: It remains to show that the experiment $\widetilde{\mathfrak{E}}_{2,n}$ defined by the path in (A.4) is asymptotically equivalent to the experiment $\mathfrak{E}_{2,n}$ defined by

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

For this purpose we denote by $\mathbf{P}^{(Y_t)_{t\in[0,1]}}$ and $\mathbf{P}^{(\widetilde{Y}_t)_{t\in[0,1]}}$ the distributions of the processes $(Y_t)_{t\in[0,1]}$ and $(\widetilde{Y}_t)_{t\in[0,1]}$, respectively, where the dependence on the parameter f is again suppressed. First, note that the representation of the Kriging estimator shows that the function $t \to I(t|\mathbf{F}_{f,n})$ belongs to the RKHS $\mathcal{H}(\Xi)$ associated with the covariance kernel K_Ξ. Second, Condition (ii) in the statement of Theorem 3.1 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

$$\begin{aligned}
\Delta^2(\mathfrak{E}_{2,n}, \widetilde{\mathfrak{E}}_{2,n}) &\le \sup_{f\in\Theta} \mathrm{KL}(\mathbf{P}^{(Y_t)_{t\in[0,1]}}, \mathbf{P}^{(\widetilde{Y}_t)_{t\in[0,1]}}) \\
&= \frac{n}{2}\sup_{f\in\Theta} \|F_f(\cdot) - I(\cdot \mid \mathbf{F}_{f,n})\|^2_{\mathcal{H}(\Xi)}
\end{aligned}$$

$$= \frac{n}{2} \sup_{f\in\Theta} \inf_{\boldsymbol{\alpha}_n\in\mathbb{R}^n} \left\| F_f(\cdot) - \sum_{j=1}^n \alpha_j K_\Xi(\cdot, t_{j,n}) \right\|^2_{\mathcal{H}(\Xi)} \to 0$$

(the first equality follows from Lemma 2 in [SH14], the second from Theorem 13.1 in [Wen05]), which completes the proof of Theorem 3.1.

B Proofs of the results in Section 4

B.1 Proof of Theorem 4.1

The proof consists in checking the two conditions (i) and (ii) in Theorem 3.1.

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

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

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

$$\sup_{f\in\Theta(\beta,L)} \sum_{i=1}^n \left(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)\mathrm{d}s \right)^2 \to 0. \tag{B.1}$$

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

$$\begin{aligned} f_K(\cdot) &= \sum_{|k|\le K} \theta_k \mathbf{e}_k(\cdot) \qquad \text{and} \\ f_K^\top(\cdot) &= f(\cdot) - f_K(\cdot) = \sum_{|k|>K} \theta_k \mathbf{e}_k(\cdot), \end{aligned}$$

respectively. Consequently,

$$\sum_{i=1}^n \left(f(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f(s)\mathrm{d}s \right)^2 \le 3\sum_{i=1}^n A_{i,n}^2 + 3\sum_{i=1}^n B_{i,n}^2 + 3\sum_{i=1}^n C_{i,n}^2, \tag{B.2}$$

where

$$\begin{aligned} A_{i,n} &= f_K(t_{i,n}) - n \int_{t_{i-1,n}}^{t_{i,n}} f_K(s)\mathrm{d}s, \\ B_{i,n} &= f_K^\top(t_{i,n}), \quad \text{and} \\ C_{i,n} &= n \int_{t_{i-1,n}}^{t_{i,n}} f_K^\top(s)\mathrm{d}s. \end{aligned}$$

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

$$\sum_{i=1}^n A_{i,n}^2 = \mathcal{O}(\max\{n^{-1}, n^{1-2\beta}\}), \tag{B.3}$$

$$\sum_{i=1}^{n} B_{i,n}^2 = \mathcal{O}(n^{1-2\beta}), \tag{B.4}$$

$$\sum_{i=1}^{n} C_{i,n}^2 = \mathcal{O}(n^{1-2\beta}), \tag{B.5}$$

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

Proof of (B.3)*:* The quantity

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

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

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

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

$$\begin{aligned}
F_j &= \frac{1}{n}\sum_{k=1}^{n} \Big(f_K(t_{k,n}) - n\int_{t_{k-1}}^{t_k} f_K(s)\mathrm{d}s \Big) e^{-2\pi \mathrm{i} kj/n} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left(\sum_{1\le|l|\le K} \theta_l e^{-2\pi \mathrm{i} lk/n} - n\int_{t_{k-1}}^{t_k} \sum_{1\le|l|\le K} \theta_l e^{-2\pi \mathrm{i} ls}\mathrm{d}s \right) e^{-2\pi \mathrm{i} kj/n} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left(\sum_{1\le|l|\le K} \theta_l e^{-2\pi \mathrm{i} lk/n} + n\sum_{1\le|l|\le K} \frac{\theta_l}{2\pi \mathrm{i} l}\left[e^{-2\pi \mathrm{i} lk/n} - e^{-2\pi \mathrm{i} l(k-1)/n}\right] \right) e^{-2\pi \mathrm{i} kj/n} \\
&= \frac{1}{n}\sum_{k=1}^{n} \left(\sum_{1\le|l|\le K} \theta_l e^{-2\pi \mathrm{i} lk/n} + n\sum_{1\le|l|\le K} \frac{\theta_l}{2\pi \mathrm{i} l}\left[1 - e^{2\pi \mathrm{i} l/n}\right] e^{-2\pi \mathrm{i} lk/n} \right) e^{-2\pi \mathrm{i} kj/n} \\
&= \frac{1}{n}\sum_{k=1}^{n} \sum_{1\le|l|\le K} \theta_l \left[1 + \frac{n}{2\pi \mathrm{i} l}(1 - e^{2\pi \mathrm{i} l/n}) \right] e^{-2\pi \mathrm{i} kl/n} e^{-2\pi \mathrm{i} kj/n}.
\end{aligned}$$

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

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

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

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

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

$$\sum_{k=1}^n e^{-2\pi \mathrm{i} km/n} = \begin{cases} n, & \text{if } m \in n\mathbb{Z}, \\ 0, & \text{if } m \notin n\mathbb{Z}. \end{cases} \tag{B.6}$$

Thus, we obtain (uniformly with respect to Θ)

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

An analogous argument for the term $\sum_{j=1}^n |F_j^-|^2$ proves the estimate (B.3).

Proof of (B.4)*:* We have

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

and it is again sufficient to consider the sum running over $k > K$. Using (B.6) again, we get

$$\sum_{j=1}^n \left|\sum_{k>K} \theta_k \exp(-2\pi \mathrm{i} kj/n)\right|^2 = \sum_{k,l>K} \theta_k \overline{\theta_l} \sum_{j=1}^n \exp(-2\pi \mathrm{i}(k-l)j/n)$$

$$= n \sum_{\substack{k,l>K \\ k-l\in n\mathbb{Z}}} \theta_k \overline{\theta_l}.$$

Taking $K = n$ here as well yields

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

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 (B.4).

Proof of (B.5)*:* Using Jensen's inequality and Parseval's identity we obtain

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

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

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

$$D_n := \min_{\boldsymbol{\alpha}_n \in \mathbb{R}^n} \| F_f(\cdot) - \sum_{j=1}^{n} \alpha_j K_\Xi(\cdot, t_{j,n}) \|_{\mathcal{H}(\Xi)}^2 = o(n^{-1}) \tag{B.7}$$

uniformly over all $f \in \Theta(\beta, L)$. Via the isomorphism ψ introduced in Equation (2.4) we have for $g = \psi F_f$

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

Assuming that q' is bounded from above we obtain

$$D_n \leq C(q) \min_{\boldsymbol{\alpha}_n \in \mathbb{R}^n} \int_0^1 (g(q(w)) - \sum_{j=1}^n \alpha_j \mathbf{1}_{(t_{j-1,n}, t_{j,n}]}(w))^2 \mathrm{d}w.$$

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

$$D_n = \min_{\boldsymbol{\alpha}_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{j=1}^n \alpha_j K_\Xi(\cdot, t_{j,n})\|^2_{\mathcal{H}(\Xi)} \lesssim I_1 + I_2 \tag{B.8}$$

where

$$I_1 = \int_0^1 \left(g_1(q(w)) - \sum_{j=1}^n \alpha_j^{(1)} \mathbf{1}_{[t_{j-1,n}, t_{j,n})}(w)\right)^2 \mathrm{d}w, \tag{B.9}$$

$$I_2 = \int_0^1 \left(g_2(q(w)) - \sum_{j=1}^n \alpha_j^{(2)} \mathbf{1}_{[t_{j-1,n}, t_{j,n})}(w)\right)^2 \mathrm{d}w. \tag{B.10}$$

We investigate the two terms I_1 and I_2 separately.

Bound for I_1: We use the estimate

$$I_1 \lesssim I_{11} + I_{12}, \tag{B.11}$$

where

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

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

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

$$\begin{aligned} I_{11} &= \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left(\frac{f_n(w)}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})}\right)^2 \mathrm{d}w \\ &\lesssim \sum_{j=1}^n \int_{t_{j-1,n}}^{t_{j,n}} \left(\frac{f_n(w)}{v(w)q'(w)} - \frac{f_n(t_{j,n})}{v(w)q'(w)}\right)^2 \mathrm{d}w \end{aligned} \tag{B.12}$$

$$+\sum_{j=1}^{n}\int_{t_{j-1,n}}^{t_{j,n}}\left(\frac{f_n(t_{j,n})}{v(w)q'(w)}-\frac{f_n(t_{j,n})}{v(t_{j,n})q'(t_{j,n})}\right)^2\mathrm{d}w. \tag{B.13}$$

First, we further decompose (B.12) as

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

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

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

From this identity we obtain (exploiting (B.6) again)

$$\begin{aligned}&\sum_{j=1}^{n}\int_{t_{j-1,n}}^{t_{j,n}}|f_n^+(w)-f_n^+(t_{j,n})|^2\mathrm{d}w\\&=\sum_{j=1}^{n}\sum_{k,l=1}\theta_k\overline{\theta}_l\exp(2\pi\mathrm{i}(l-k)j/n)\int_{\frac{j-1}{n}}^{\frac{j}{n}}[\exp(2\pi\mathrm{i}k(j/n-w))-1]\\&\qquad\cdot[\exp(-2\pi\mathrm{i}l(j/n-w))-1]\mathrm{d}w\\&=\sum_{j=1}^{n}\sum_{k,l=1}\theta_k\overline{\theta}_l\exp(2\pi\mathrm{i}(l-k)j/n)\int_{0}^{\frac{1}{n}}[\exp(2\pi\mathrm{i}k(1/n-w))-1]\\&\qquad\cdot[\exp(-2\pi\mathrm{i}l(1/n-w))-1]\mathrm{d}w\\&=n\sum_{k=1}^{n}|\theta_k|^2\int_0^{\frac{1}{n}}[\exp(2\pi\mathrm{i}k(1/n-w))-1][\exp(-2\pi\mathrm{i}k(1/n-w))-1]\mathrm{d}w\\&\le Cn\sum_{k=1}^{n}|\theta_k|^2k^2n^{-3}\\&\le Cn^{-2}\sum_{k=1}^{n}|\theta_k|^2k^{2\beta}k^{-2\beta+2}\\&\le C(L)n^{-2}\max\{1,n^{-2\beta+2}\}\lesssim\max\{n^{-2},n^{-2\beta}\}=o(n^{-1})\end{aligned}$$

To derive an estimate of (B.13) we note that for any $n\in\mathbb{N}$,

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

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

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

Because the product of a γ_1-Hölder function and a γ_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,

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

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 (B.11) can bounded as follows:

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

Observing the estimate (B.1) 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 (B.11) we have

$$\begin{aligned}
I_2 &= \int_0^1 \left(\frac{F_f(w)v'(w)}{v^2(w)q'(w)} - \sum_{j=1}^{n} \alpha_j^{(2)} \mathbf{1}_{[t_{j-1,n},t_{j,n})}(w) \right)^2 \mathrm{d}w \\
&\lesssim \sum_{j=1}^{n} \int_{t_{j-1,n}}^{t_{j,n}} \left(\frac{F_f(w)v'(w)}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} \right)^2 \mathrm{d}w \qquad \text{(B.14)} \\
&\quad + \sum_{j=1}^{n} \int_{t_{j-1,n}}^{t_{j,n}} \left(\frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(w)q'(w)} - \frac{F_f(t_{j,n})v'(t_{j,n})}{v^2(t_{j,n})q'(t_{j,n})} \right)^2 \mathrm{d}w. \qquad \text{(B.15)}
\end{aligned}$$

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 (B.14) can be bounded as

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

which is of order $o(n^{-1})$. For the term (B.15) we obtain using our assumptions that

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

Using the same arguments as for the bound of (B.13), 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 (B.7) follows from (B.8).

B.2 Proof of Theorem 4.5

As in the proof of Theorem 4.1, we have to verify the the two conditions (i) and (ii) from Theorem 3.1.

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

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

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

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

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

$$\min_{\boldsymbol{\alpha}_n \in \mathbb{R}^n} \|F_f(\cdot) - \sum_{j=1}^{n} \alpha_j K_\Xi(\cdot, t_{j,n})\|^2_{\mathcal{H}(\Xi)} \lesssim I_1 + I_2 \tag{B.16}$$

where we define I_1 and I_2 as in Appendix B.1 (see the equations (B.9) and (B.10)) 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,

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

The first integral can be bounded as

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

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 (B.16) can be bounded exactly as the corresponding term in the Sobolev ellipsoid case. This finishes the proof of the theorem.

B.3 Proof of Theorem 4.6

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

$$\ell \colon \Theta \times \mathcal{A} \to [0,\infty), \qquad (\theta, a) \mapsto \ell(\theta, a)$$

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 $\rho \colon \mathcal{X} \times \mathscr{A} \to [0,1]$, and the associated *risk* is

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

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

$$\delta(\mathfrak{E}_1, \mathfrak{E}_2) = \inf_{\rho_1} \sup_{\rho_2} \sup_{\theta} \sup_{\ell} |R_\theta(\mathfrak{E}_1, \rho_1, \ell) - R_\theta(\mathfrak{E}_2, \rho_2, \ell)| \tag{B.17}$$

(see [Mar16], Theorem 2.7, and the references cited there), where the supremum is taken over all loss functions ℓ with $0 \le \ell(\theta, a) \le 1$ for all $\theta \in \Theta$ and $a \in \mathcal{A}$, all admissible parameters $\theta \in \Theta$, and decision rules ρ_2 in the second experiment. The infimum is taken over all decision rules ρ_1 in the first experiment.

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

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

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

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

and we consider the (non-randomized) decision rule ρ_2 defined by

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

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

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

for all $f \in \Theta(\beta, L)$ and $n \in \mathbb{N}$. From (B.17) we thus obtain

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

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 $j = 1, \ldots, n$ 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

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

where ρ_1 is any (potentially randomized) decision rule. Because

$$\int_0^1 f_0(x)\mathrm{d}x = 0 \neq \sqrt{2/3} L n^{-\beta} = \int_0^1 f_n(x)\mathrm{d}x$$

at least one of the terms $\mathbf{P}[\rho_1(\mathbf{Y}_n) \neq \int_0^1 f_0(x)\mathrm{d}x]$ and $\mathbf{P}[\rho_1(\mathbf{Y}_n) \neq \int_0^1 f_n(x)\mathrm{d}x]$ must be $\ge 1/2$ for any $\mathbf{Y}_n$ (otherwise there is a contradiction). Thus, setting $A_\diamond = \{\mathbf{Y}_n \in$

$\mathbb{R}^n : \mathbf{P}[\rho_1(\mathbf{Y}_n) \neq \int_0^1 f_\diamond(x)\mathrm{d}x] \geq 1/2\}$ for $\diamond \in \{0, n\}$ one has $A_0 \cup A_n = \mathbb{R}^n$. As a consequence, either $\mathbf{P}^{f_0}_{1,n}(A_0) = \mathbf{P}^{f_n}_{1,n}(A_0) \geq 1/2$ or $\mathbf{P}^{f_0}_{1,n}(A_n) = \mathbf{P}^{f_n}_{1,n}(A_n) \geq 1/2$ holds. Without loss of generality, we assume that $\mathbf{P}^{f_0}_{1,n}(A_0) = \mathbf{P}^{f_n}_{1,n}(A_0) \geq 1/2$ (the other case follows by exactly the same argument). In this case, using the definition of the set A_0,

$$
\begin{aligned}
\delta(\mathfrak{E}_{1,n}, \mathfrak{E}_{2,n}) &\geq \inf_{\rho_1} \sup_{f \in \{f_0, f_n\}} R_f(\mathfrak{E}_{1,n}, \rho_1, \ell) \\
&\geq \inf_{\rho_1} \int_{\mathbb{R}^n} \mathbf{P}[\rho_1(\mathbf{Y}_n) \neq \int_0^1 f_0(x)\mathrm{d}x] \mathbf{P}^{f_0}_{1,n}(\mathrm{d}\mathbf{Y}_n) \\
&\geq \inf_{\rho_1} \int_{A_0} \mathbf{P}[\rho_1(\mathbf{Y}_n) \neq \int_0^1 f_0(x)\mathrm{d}x] \mathbf{P}^{f_0}_{1,n}(\mathrm{d}\mathbf{Y}_n) \\
&\geq \frac{1}{2} \mathbf{P}^{f_0}_{1,n}(A_0) \\
&\geq \frac{1}{4},
\end{aligned}
$$

which proves the assertion.

References

[Bed87] Beder, J. H. A sieve estimator for the mean of a Gaussian process. *Ann. Statist.* 15 (1987), 59–78.

[BL96] Brown, L. D. and Low, M. G. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* 24 (1996), 2384–2398.

[Bro+02] Brown, L. D. et al. Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* 30 (2002). Dedicated to the memory of Lucien Le Cam, 688–707.

[BTA04] Berlinet, A. and Thomas-Agnan, C. *Reproducing kernel Hilbert spaces in probability and statistics.* With a preface by Persi Diaconis. Kluwer Academic Publishers, Boston, MA, 2004, pp. xxii+355.

[BZ98] Brown, L. D. and Zhang, C.-H. Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Ann. Statist.* 26 (1998), 279–287.

[Car06] Carter, A. V. A continuous Gaussian approximation to a nonparametric regression in two dimensions. *Bernoulli* 12 (2006), 143–156.

[Car07] Carter, A. V. Asymptotic approximation of nonparametric regression experiments with unknown variances. *Ann. Statist.* 35 (2007), 1644–1673.

[Car09] Carter, A. V. Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise. *J. Probab. Stat.* (2009), Art. ID 275308, 19.

[Doo49] Doob, J. L. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* 20 (1949), 393–403.

[DPZ16] Dette, H., Pepelyshev, A., and Zhigljavsky, A. Optimal designs in regression with correlated errors. *Ann. Statist.* 44 (2016), 113–152.

[GN06] Grama, I. G. and Neumann, M. H. Asymptotic equivalence of nonparametric autoregression and nonparametric regression. *Ann. Statist.* 34 (2006), 1701–1732.

[GN98] Grama, I. and Nussbaum, M. Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields* 111 (1998), 167–214.

[GNZ10] Golubev, G. K., Nussbaum, M., and Zhou, H. H. Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* 38 (2010), 181–214.

[Joh99] Johnstone, I. M. Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sinica* 9 (1999), 51–83.

[JS97] Johnstone, I. M. and Silverman, B. W. Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* 59 (1997), 319–351.

[LC86] Le Cam, L. *Asymptotic methods in statistical decision theory.* Springer Series in Statistics. Springer-Verlag, New York, 1986, pp. xxvi+742.

[LCY00] Le Cam, L. and Yang, G. L. *Asymptotics in statistics.* Second. Springer Series in Statistics. Some basic concepts. Springer-Verlag, New York, 2000, pp. xiv+285.

[Mar16] Mariucci, E. Le Cam theory on the comparison of statistical models. *Grad. J. Math.* 1 (2016), 81–91.

[Mei11] Meister, A. Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* 39 (2011), 1471–1495.

[MM65] Mehr, C. B. and McFadden, J. A. Certain properties of Gaussian processes and their first-passage times. *J. Roy. Statist. Soc. Ser. B* 27 (1965), 505–522.

[Nev68] Neveu, J. *Processus aléatoires gaussiens.* Séminaire de Mathématiques Supérieures, No. 34 (Été, 1968). Les Presses de l'Université de Montréal, Montreal, Que., 1968, p. 224.

[Nus96] Nussbaum, M. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* 24 (1996), 2399–2430.

[PR16] Paulsen, V. I. and Raghupathi, M. *An introduction to the theory of reproducing kernel Hilbert spaces.* Vol. 152. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2016, pp. x+182.

[Rei08] Reiß, M. Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* 36 (2008), 1957–1982.

[Rei11] Reiß, M. Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* 39 (2011), 772–802.

[Roh04] Rohde, A. On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statist. Decisions* 22 (2004), 235–243.

[SH14] Schmidt-Hieber, J. Asymptotic equivalence for regression under fractional noise. *Ann. Statist.* 42 (2014), 2557–2585.

[Sle61] Slepian, D. First passage time for a particular Gaussian process. *Ann. Math. Statist.* 32 (1961), 610–612.

[Tor91] Torgersen, E. *Comparison of statistical experiments.* Vol. 36. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1991, pp. xx+675.

[Wen05] Wendland, H. *Scattered data approximation.* Vol. 17. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2005, pp. x+336.