The Annals of Statistics
2008, Vol. 36, No. 4, 1957–1982
DOI: 10.1214/07-AOS525
© Institute of Mathematical Statistics, 2008

ASYMPTOTIC EQUIVALENCE FOR NONPARAMETRIC REGRESSION WITH MULTIVARIATE AND RANDOM DESIGN

BY MARKUS REIß

University of Heidelberg

We show that nonparametric regression is asymptotically equivalent, in Le Cam's sense, to a sequence of Gaussian white noise experiments as the number of observations tends to infinity. We propose a general constructive framework, based on approximation spaces, which allows asymptotic equivalence to be achieved, even in the cases of multivariate and random design.

1. Introduction. Nonparametric regression is the model most often encountered in nonparametric statistics because of its widespread applications. However, for theoretical investigations, the Gaussian white noise (or sequence space) model is often preferred since it exhibits nice mathematical properties. The common wisdom that statistical decisions in the two models show the same asymptotic behavior was formalized and proven for the first time by Brown and Low (1996) in the one-dimensional case, using Le Cam's concept of equivalence of statistical experiments.

In this paper, we propose a unifying framework for establishing global asymptotic equivalence between Gaussian nonparametric regression and white noise experiments, based on constructive transitions with only minimal randomizations. This framework not only allows concise proofs of known results, but extends the asymptotic equivalence to the multivariate and random design situations. The multivariate result has often been alluded to, though it has never been proven; see, for example, Hoffmann and Lepski (2002). While Brown and Zhang (1998) remark that the regression and white noise experiments are not asymptotically equivalent for equidistant design on $[0,1]^d$ and Sobolev classes of regularity $s \le d/2$, the only positive result thus far, due to Carter (2006), ensures asymptotic equivalence for equidistant design in dimensions $d = 2$ and $d = 3$ when $s > d/2$. The difficulty in extending results to higher dimensions is that we have to go beyond piecewise constant or linear approximations. For the dynamic model of ergodic diffusions, Dalalyan and Reiß (2007) have established multidimensional asymptotic equivalence with a white noise model. For the case of univariate nonparametric regression, but with non-Gaussian errors, we refer to Grama and Nussbaum (1998).

Received June 20007; revised June 2007.
AMS 2000 subject classifications. 62G08, 62G20, 62B15.
Key words and phrases. Le Cam deficiency, equivalence of experiments, approximation space, interpolation, Gaussian white noise.

To obtain a first insight into the problem of asymptotic equivalence for nonparametric regression, let us consider the regression model

$$Y_i = f(x_i) + \sigma\varepsilon_i, \qquad i = 1, \dots, n,$$

with an unknown function $f : \mathscr{D} \subseteq \mathbb{R}^d \to \mathbb{R}$, $x_1, \dots, x_n \in \mathscr{D}$ and $\varepsilon_i \sim \mathscr{N}(0, 1)$ i.i.d. Then, using some orthonormal basis (φ_j) of $L^2(\mathscr{D})$ and writing $v_j = (\varphi_j(x_1), \dots, \varphi_j(x_n))^\top$, $Y = (Y_1, \dots, Y_n)^\top$, the observations can be transformed to

$$y_j := n^{-1} v_j^\top Y = n^{-1} \sum_{i=1}^n f(x_i)\varphi_j(x_i) + n^{-1} v_j^\top \varepsilon, \qquad j = 1, \dots, n.$$

The covariance matrix of (y_j) is given by $\sigma^2\Sigma_n$, with $\Sigma_n = (v_j^\top v_k/n)_{jk}$. On the other hand, the model of observing the function f in Gaussian white noise of level $\sigma/\sqrt{n}$ can be written as a sequence space model with respect to the basis (φ_j) as follows:

$$z_j = \int_{\mathscr{D}} f(x)\varphi_j(x)\,dx + \frac{\sigma}{\sqrt{n}}\tilde{\varepsilon}_j, \qquad j = 1, 2, \dots,$$

with $(\tilde{\varepsilon}_j) \sim \mathscr{N}(0, 1)$ i.i.d. In the so-called *isometric case*, where we can choose (φ_j) such that Σ_n is the identity matrix, we can realize the two experiments on the same probability space, setting $\varepsilon_j = \tilde{\varepsilon}_j$ for $j \ge 1$ and $y_j = \frac{\sigma}{\sqrt{n}}\varepsilon_j$ for $j > n$, and the total variation distance between the observation laws tends to zero for $n \to \infty$ if and only if

$$\lim_{n\to\infty} \frac{n}{\sigma^2}\left(\sum_{j=1}^n (\mathbb{E}[y_j] - \mathbb{E}[z_j])^2 + \sum_{j=n+1}^\infty (\mathbb{E}[z_j])^2\right) = 0.$$

The second term is a classical approximation error and the first term can be regarded as an interpolation error due to the discretization of the integral. If the convergence can be shown uniformly over the class $\mathscr{F}$ of functions f under consideration, then we shall have established asymptotic equivalence in Le Cam's sense. In Section 2, this isometric case is presented in a slightly more general manner, using operator terminology. It is applied to the Haar and Fourier basis for equidistant observations, which is the framework for the results of Brown and Low (1996) and Rohde (2004) and which, more importantly, shows asymptotic equivalence in any dimension d for periodic Sobolev classes of regularity $s > d/2$.

If Σ_n is not the identity, we further transform to observing $\Sigma_n^{-1/2}Y$ and $\Sigma_n^{-1}Y$, respectively. The first transformation "whitens" the covariance structure such that only the observation means have to be matched asymptotically, whereas the second transformation better matches the mean at the cost of a heteroskedastic covariance structure. In Section 3, this *isomorphic* framework is presented. The spline approach of Carter (2006) emerges as an application of the second transformation.

The first transformation is applied to obtain a constructive asymptotic equivalence result on the basis of wavelet multiresolution analyses, which provides equivalence results also for nonperiodic function classes. Connections to asymptotic studies by Donoho and Johnstone (1999) and Johnstone and Silverman (2004) for wavelet estimators are discussed.

The case of a random design, uniform on a d-dimensional cube, is treated in Section 4. This setting is much more involved, but can also be cast in the isomorphic framework. The construction is based on a two-level procedure, generalizing an idea of Brown et al. (2002) and Brown et al. (2004). The general idea is to employ the Fourier basis and to match the means for low frequencies and the covariance structure for high frequencies. The high-frequency transformation, however, uses the Cholesky decomposition of the covariance matrix. Fine approximation and symmetry properties of the Fourier basis then yield that also in the case of random design asymptotic equivalence holds for Sobolev regularities $s > d/2$ and any dimension $d \geq 1$.

2. Isometric approximation.

2.1. *General theory.* We write $\mathscr{L}^2(\mathscr{D}) := \{f : \mathscr{D} \to \mathbb{K} \mid \|f\|_{L^2}^2 := \int |f|^2 < \infty\}$ with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ and $L^2(\mathscr{D})$ for the Hilbert space of equivalence classes with respect to $\|\bullet\|_{L^2}$. Although the observations are real-valued, we shall use complex-valued functions for simplicity when treating Fourier approximations.

DEFINITION 2.1. Let $\mathbb{E}_n^d$ be the regression experiment obtained from observing

$$Y_i = f(x_i) + \sigma\varepsilon_i, \qquad i = 1, \dots, n,$$

for $n \in \mathbb{N}$, $f : \mathscr{D} \to \mathbb{R}$ in some class $\mathscr{F}^d \subseteq \mathscr{L}^2(\mathscr{D})$, where $\mathscr{D} \subseteq \mathbb{R}^d$, for fixed design points $x_i \in \mathscr{D}$ and for independent random variables $\varepsilon_i \sim \mathscr{N}(0, 1)$.

Suppose we are given an n-dimensional space $S_n \subseteq \mathscr{L}^2(\mathscr{D})$ and a linear mapping $D_n : \mathscr{L}^2(\mathscr{D}) \to \mathbb{K}^n$ with the following isometric property on S_n:

$$\forall g_n \in S_n : \|g_n\|_{L^2} = \|g_n\|_n := n^{-1/2}|D_n g_n|_{\mathbb{K}^n}. \tag{2.1}$$

By $\langle \bullet, \bullet \rangle_n$, we denote the scalar product associated with $\|\bullet\|_n$. Usually, $D_n g = (g(x_i))_{1 \leq i \leq n}$ will be the point evaluation at the n design points, in which case $\|g\|_n^2 = \frac{1}{n}\sum_{i=1}^n |g(x_i)|^2$ is just the empirical norm. Let us further introduce the linear operator

$$\mathscr{I}_n : \mathscr{L}^2(\mathscr{D}) \to S_n, \qquad \mathscr{I}_n g := (D_n|_{S_n})^{-1}(g(x_1), \dots, g(x_n))^\top.$$

For $D_n g = (g(x_1), \dots, g(x_n))^\top$, we also have $D_n \mathscr{I}_n g = (g(x_1), \dots, g(x_n))^\top$ and therefore $\mathscr{I}_n = (D_n|_{S_n})^{-1} D_n$. Consequently, in this case, $\mathscr{I}_n$ is the $\|\bullet\|_n$-orthogonal projection onto S_n such that $\mathscr{I}_n g$ is the unique element of S_n interpolating g at the design points (x_i).

To state the first results, we refer to Le Cam and Yang (2000) for the notion of equivalence between experiments and of the Le Cam distance between two experiments $\mathbb{E}$ and $\mathbb{G}$, which, for the parameter class $\mathscr{F}$, will be denoted by $\Delta_{\mathscr{F}}(\mathbb{E}, \mathbb{G})$. The Gaussian law on a Hilbert space H with mean vector $\mu \in H$ and covariance operator $Q : H \to H$ will be denoted by $\mathscr{N}(\mu, Q)$.

The regression experiment $\mathbb{E}_n^d$ can be transformed to a functional Gaussian shift experiment by applying the isometry $(D_n|_{S_n})^{-1}$ to $Y = (Y_i) \in \mathbb{R}^n$ as follows:

$$Z := (D_n|_{S_n})^{-1} Y = \mathscr{I}_n f + \frac{\sigma}{\sqrt{n}} \zeta \in S_n, \tag{2.2}$$

where $\zeta := \sqrt{n}(D_n|_{S_n})^{-1}\varepsilon \sim \mathscr{N}(0, \mathrm{Id}_{S_n})$ is a Gaussian white noise in S_n because, for $g_n, h_n \in S_n$,

$$\begin{aligned} E[\langle \zeta, g_n \rangle_{L^2} \overline{\langle \zeta, h_n \rangle}_{L^2}] &= n^{-1} E[\langle \varepsilon, D_n g_n \rangle_{\mathbb{K}^n} \overline{\langle \varepsilon, D_n h_n \rangle}_{\mathbb{K}^n}] \\ &= \langle g_n, h_n \rangle_n = \langle g_n, h_n \rangle_{L^2}. \end{aligned}$$

By adding completely uninformative observations on the orthogonal complement of S_n in $L^2(\mathscr{D})$, the observation of Z in (2.2) is equivalent to observing

$$\langle \varphi, Z \rangle_{L^2} = \langle \varphi, \mathscr{I}_n f \rangle_{L^2} + \frac{\sigma}{\sqrt{n}} \langle \varphi, \bar{\zeta} \rangle_{L^2} \qquad \forall \varphi \in L^2(\mathscr{D}),$$

with $\langle \varphi, \bar{\zeta} \rangle_{L^2} \sim \mathscr{N}(0, \|\varphi\|_{L^2}^2)$. Here, we understand the scalar product with the white noise $\bar{\zeta}$ in a weak sense, for example, realized by a Brownian motion B (a Brownian sheet in dimension $d > 1$) via $\langle \varphi, \bar{\zeta} \rangle_{L^2} = \int_{\mathscr{D}} \varphi(x)\, dB(x)$. In differential notation, we have thus established the following equivalence.

PROPOSITION 2.2. *Let $\mathbb{F}_n^d$ be the Gaussian white noise experiment in $L^2(\mathscr{D})$ given by observing*

$$dY(x) = \mathscr{I}_n f(x)\, dx + \frac{\sigma}{\sqrt{n}} dB(x), \qquad x \in \mathscr{D},$$

where $f \in \mathscr{F}^d$ and dB is a Gaussian white noise in $L^2(\mathscr{D})$. Then, the regression experiment $\mathbb{E}_n^d$ is statistically equivalent to $\mathbb{F}_n^d$ for any functional class $\mathscr{F}^d$.

We are nearing the first main result.

DEFINITION 2.3. Let $\mathbb{G}_n^d$ be the Gaussian white noise experiment given by observing

$$dY(x) = f(x)\, dx + \frac{\sigma}{\sqrt{n}} dB(x), \qquad x \in \mathscr{D},$$

where $f \in \mathscr{F}^d$ and dB is a Gaussian white noise in $L^2(\mathscr{D})$.

THEOREM 2.4. *The Le Cam distance between $\mathbb{E}_n^d$ and $\mathbb{G}_n^d$ for the class $\mathscr{F}^d$ is bounded by*

$$\Delta_{\mathscr{F}^d}(\mathbb{E}_n^d, \mathbb{G}_n^d) \leq 1 - 2\Phi\left(-\frac{\sqrt{n}}{2\sigma} \sup_{f \in \mathscr{F}^d} \|f - \mathscr{I}_n f\|_{L^2}\right),$$

where Φ denotes the standard Gaussian cumulative distribution function.

REMARK 2.5. Note that $\|f - \mathscr{I}_n f\|_{L^2}^2 = \|f - P_n f\|_{L^2}^2 + \|P_n f - \mathscr{I}_n f\|_{L^2}^2$, where P_n is the L^2-orthogonal projection onto S_n. This means that the bound on the Le Cam distance is always larger than the same expression involving the classical bias estimate $\sup_{f \in \mathscr{F}^d} \|f - P_n f\|_{L^2}$. Because of $\Phi(0) = 1/2$, Proposition 2.4 yields the rate estimate

$$\Delta_{\mathscr{F}^d}(\mathbb{E}_n^d, \mathbb{G}_n^d) \lesssim \sigma^{-1} n^{1/2} \sup_{f \in \mathscr{F}^d} \|f - \mathscr{I}_n f\|_{L^2}.$$

Here and in the sequel, $A \lesssim B$ means $A \leq cB$ with a constant $c > 0$, independent of the other parameters involved, and $A \sim B$ is an abbreviation for $A \lesssim B$ and $B \lesssim A$.

PROOF OF THEOREM 2.4. Since $\mathbb{E}_n^d$ and $\mathbb{F}_n^d$ are equivalent, it suffices to establish the bound for $\Delta_{\mathscr{F}^d}(\mathbb{F}_n^d, \mathbb{G}_n^d)$. The two latter experiments are realized on the same sample space. Therefore, the Le Cam distance is bounded by the maximal total variation distance over the class $\mathscr{F}^d$ [Nussbaum (1996), Proposition 2.2]. For Gaussian white noise, the total variation distance is given by $1 - 2\Phi(-\frac{\sqrt{n}}{2\sigma}\|f - \mathscr{I}_n f\|_{L^2})$ [Carter (2006), Section 3.2] and the result follows. $\square$

2.2. *Piecewise constant approximation.* The original results of Brown and Low (1996) for equidistant design on $\mathscr{D} = (0, 1]$ fit into the proposed isometric framework. For design points $x_i = i/n$, $i = 1, \ldots, n$, we consider the n-dimensional space S_n of piecewise constant, left-continuous functions on $(0, 1]$ with possible jumps at i/n, $i = 1, \ldots, n-1$. Using $D_n g = (g(i/n))_{1 \leq i \leq n}$, we obtain, for $g_n \in S_n$,

$$\|g_n\|_n^2 = \frac{1}{n} \sum_{i=1}^n |g_n(i/n)|^2 = \sum_{i=1}^n \int_{(i-1)/n}^n |g_n(u)|^2 \, du = \|g_n\|_{L^2}^2,$$

such that D_n has the isometric property. To infer asymptotic equivalence by Proposition 2.4, we have to ensure that $\|f - \mathscr{I}_n f\|_{L^2} = o(n^{-1/2})$ uniformly over all f in some functional class $\mathscr{F}^d$. Considering the Hölder class of regularity $\alpha \in (0, 1]$,

$$\mathscr{F}_H(\alpha, R) := \left\{ f \in C^\alpha([0, 1]) \,\middle|\, \sup_{x \neq y} |f(x) - f(y)|/|x - y|^\alpha \leq R \right\},$$

we obtain, for $f \in \mathscr{F}_H(\alpha, R)$,

$$\|f - \mathscr{I}_n f\|_{L^2}^2 = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} |f(x) - f(i/n)|^2 \, dx \le R^2 (2\alpha+1)^{-1} n^{-2\alpha}.$$

Consequently, asymptotic equivalence between $\mathbb{E}_n^1$ and $\mathbb{G}_n^1$ holds for any Hölder class $\mathscr{F}_H(\alpha, R)$ with $\alpha > 1/2$ and $R > 0$ arbitrary. The approximation property of the Haar wavelet even yields asymptotic regularity for L^2-Sobolev classes of regularity $\alpha > 1/2$.

For nonuniform design $0 \le x_1 < \cdots < x_n \le 1$, consider the same setting as before, in particular, $D_n g = (g(i/n))_i \ne (g(x_i))_i$. We obtain, for $f \in \mathscr{F}_H(\alpha, R)$,

$$\begin{aligned}
\|f - \mathscr{I}_n f\|_{L^2}^2 &= \sum_{i=1}^n \int_{(i-1)/n}^{i/n} |f(x) - f(x_i)|^2 \, dx \\
&\le R^2 \sum_{i=1}^n \int_{(i-1)/n}^{i/n} |x - x_i|^{2\alpha} \, dx \\
&\le R^2 n^{-1} \sum_{i=1}^n (n^{-1} + |x_i - i/n|)^{2\alpha} \\
&\le 2R^2 n^{-2\alpha} + 2R^2 n^{-1} \sum_{i=1}^n |x_i - i/n|^{2\alpha}.
\end{aligned}$$

By Theorem 2.4, we have obtained the following result.

THEOREM 2.6. *On the Hölder class $\mathscr{F}_H(\alpha, R)$, the Le Cam distance between nonparametric regression with design $0 < x_1^{(n)} < \cdots < x_n^{(n)} \le 1$ and the white noise experiment satisfies*

$$\Delta_{\mathscr{F}_H(\alpha,R)}(\mathbb{E}_n^1, \mathbb{G}_n^1) \lesssim \sigma^{-1} R \left(n^{1-2\alpha} + \sum_{i=1}^n |x_i^{(n)} - i/n|^{2\alpha} \right)^{1/2}.$$

Consequently, asymptotic equivalence holds whenever $\alpha \in (1/2, 1]$ *and the design satisfies* $\lim_{n\to\infty} \sum_{i=1}^n |x_i^{(n)} - i/n|^{2\alpha} = 0$, *for example, if* $\max_i |x_i^{(n)} - i/n| = o(n^{-1/(2\alpha)})$.

REMARK 2.7. This approach does not permit the establishment of global equivalence for the random design case in Section 4 because the standard deviations of the order statistics $X_{(j)}$ only decrease with rate $n^{-1/2}$. Treating the random design as if equidistant nevertheless yields, for estimation purposes, nearly optimal asymptotic L^2-risk when $\alpha > 1/2$ [Cai and Brown (1999)].

2.3. *Fourier series approximation.* In the case of $\mathscr{D}=[0,1]^d$, $d\geq 1$, and of an equidistant design $(k/m)_{k\in\{1,\dots,m\}^d}$ with $m=n^{1/d}\in\mathbb{N}$ and odd, the Fourier system ($\iota:=\sqrt{-1}$)

$$\varphi_\ell(x):=\exp(2\pi\iota\langle x,\ell\rangle),\qquad \ell=(\ell_1,\dots,\ell_d),\ |\ell|_\infty:=\max_i|\ell_i|\leq\frac{m-1}{2},$$

is not only L^2-orthonormal, but also orthonormal with respect to $\langle\bullet,\bullet\rangle_n$ for $D_n g:=(g(k/m))_k$:

$$\begin{aligned}\langle\varphi_\ell,\varphi_{\ell'}\rangle_n&=\frac{1}{n}\sum_{k\in\{1,\dots,m\}^d}\varphi_\ell(k/m)\overline{\varphi_{\ell'}(k/m)}\\&=m^{-d}\sum_{k_1,\dots,k_d=1}^{m}\prod_{i=1}^{d}\exp\bigl(2\pi\iota k_i(\ell_i-\ell_i')/m\bigr)\\&=\prod_{i=1}^{d}\Biggl(\frac{1}{m}\sum_{\kappa=1}^{m}\exp\bigl(2\pi\iota\kappa(\ell_i-\ell_i')/m\bigr)\Biggr)\\&=\begin{cases}1, & \text{if } m|(\ell_i-\ell_i') \text{ for all } i,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}\tag{2.3}$$

Consequently, the space of trigonometric polynomials $S_n:=\operatorname{span}(\varphi_\ell,\ |\ell|_\infty\leq\frac{m-1}{2})$ satisfies the isometric property (2.1).

The periodic Sobolev class of regularity s and radius R on $[0,1]^d$ is given by

$$\mathscr{F}^d_{\mathrm{S,per}}(s,R):=\Biggl\{f\in L^2([0,1]^d)\Biggm|\sum_{\ell\in\mathbb{Z}^d}|\ell|_\infty^{2s}\bigl|\langle f,\varphi_\ell\rangle\bigr|_{L^2}^2\leq R^2\Biggr\}.$$

Due to the strong cancellation property (2.3) of the scalar product $\langle\bullet,\bullet\rangle_n$, we explicitly derive $(\mathscr{I}_n f)(x)=\sum_{|\ell|_\infty\leq(m-1)/2}(\sum_{k\in\mathbb{Z}^d}\langle f,\varphi_{\ell+km}\rangle_{L^2})\varphi_\ell(x)$. In view of Remark 2.5, we first bound the classical bias:

$$\begin{aligned}\sup_{f\in\mathscr{F}^d_{\mathrm{S,per}}(s,R)}\|f-P_nf\|_{L^2}^2&=\sup_{f\in\mathscr{F}^d_{\mathrm{S,per}}(s,R)}\sum_{|\ell|_\infty\geq(m+1)/2}|\langle f,\varphi_\ell\rangle_{L^2}|^2\\&=R^2\Bigl(\frac{m+1}{2}\Bigr)^{-2s}.\end{aligned}$$

For $s>d/2$, we obtain, using the Cauchy–Schwarz inequality,

$$\begin{aligned}&\sup_{f\in\mathscr{F}^d_{\mathrm{S,per}}(s,R)}\|P_nf-\mathscr{I}_nf\|_{L^2}^2\\&\quad=\sup_{f\in\mathscr{F}^d_{\mathrm{S,per}}(s,R)}\sum_{|\ell|_\infty\leq(m-1)/2}\Biggl(\sum_{k\in\mathbb{Z}^d\setminus\{0\}}\langle f,\varphi_{\ell+km}\rangle_{L^2}\Biggr)^2\end{aligned}$$

$$\le \left(\sup_{f\in\mathscr{F}^d(s,R)} \sum_{|\ell|_\infty\le(m-1)/2} \sum_{k\in\mathbb{Z}^d\setminus\{0\}} |\ell+km|_\infty^{2s}\langle f,\varphi_{\ell+km}\rangle_{L^2}^2\right)$$

$$\times\left(\sup_{|\ell|_\infty\le(m-1)/2} \sum_{k\in\mathbb{Z}^d\setminus\{0\}} |\ell+km|_\infty^{-2s}\right)$$

$$= R^2 \sup_{|\ell|_\infty\le(m-1)/2} \sum_{k\in\mathbb{Z}^d\setminus\{0\}} |\ell+km|_\infty^{-2s}.$$

Noting that the grid points $\ell+km$ do not overlap and that $m^{-1}|\ell+km|_\infty \ge |k|_\infty - 1/2$ for the values of ℓ and k considered, we arrive at

$$\sup_{f\in\mathscr{F}^d_{\mathrm{S,per}}(s,R)} \|P_n f-\mathscr{I}_n f\|_{L^2}^2$$

$$\le R^2 m^{-2s} \sum_{k\in\mathbb{Z}^d\setminus\{0\}} |k-1/2|_\infty^{-2s}$$

$$= R^2 m^{-2s} \sum_{k=1}^{\infty} \big((2k+1)^d-(2k-1)^d\big)(k-1/2)^{-2s}.$$

Hence, using Theorem 2.4, we have proven the following result, which extends the scalar results of Brown and Low (1996) and, more specifically, Rohde (2004) to any dimension $d\ge 1$.

THEOREM 2.8. *For d-dimensional periodic Sobolev classes $\mathscr{F}^d_{\mathrm{S,per}}(s,R)$ with regularity $s>d/2$ and equidistant design on the cube $[0,1]^d$, the nonparametric regression experiment $\mathbb{E}_n^d$ and the Gaussian shift experiment $\mathbb{G}_n^d$ are asymptotically equivalent as $n\to\infty$. The Le Cam distance satisfies*

$$\Delta_{\mathscr{F}^d_{\mathrm{S,per}}(s,R)}(\mathbb{E}_n^d,\mathbb{G}_n^d) \lesssim \sigma^{-1}Rn^{1/2-s/d}.$$

3. Isomorphic approximation.

3.1. *General theory.* We extend the preceding framework by merely requiring an isomorphic property. Since it will suffice for the subsequent applications, we immediately specialize here to $D_n g=(g(x_1),\dots,g(x_n))$. Let $S_n\subseteq\mathscr{L}^2(\mathscr{D})$, $\dim S_n=n$, have the property

$$\forall g_n\in S_n: g_n(x_1)=\cdots=g_n(x_n)=0 \Longrightarrow g_n=0. \tag{3.1}$$

Let

$$\langle f,g\rangle_n := \frac{1}{n}\sum_{i=1}^n f(x_i)\overline{g(x_i)} \quad\text{and}\quad \langle v,g\rangle_n := \frac{1}{n}\sum_{i=1}^n v_i\overline{g(x_i)},$$

$$f,g\in\mathscr{L}^2,\ v\in\mathbb{R}^n$$

and $\|g\|_n^2 = \langle g, g\rangle_n$. In this notation, equation (3.1) is equivalent to the isomorphy of the norms $\|\bullet\|_n$ and $\|\bullet\|_{L^2}$ on S_n:

$$\exists A_n, B_n > 0\ \forall g_n \in S_n : A_n\|g_n\|_{L^2} \le \|g_n\|_n \le B_n\|g_n\|_{L^2}. \tag{3.2}$$

We choose any L^2-orthonormal basis $(\varphi_j)_{1\le j\le n}$ of S_n and introduce the linear mappings $\Pi_n, \mathscr{I}_n : \mathscr{L}^2(\mathscr{D}) \to S_n$, $\Sigma_n : S_n \to S_n$:

$$\Pi_n g := \sum_{j=1}^n \langle g, \varphi_j\rangle_n \varphi_j, \qquad \Sigma_n := \Pi_n|_{S_n} : S_n \to S_n, \qquad \mathscr{I}_n g := \Sigma_n^{-1}\Pi_n g.$$

Observe the following properties: $\langle \Sigma_n g_n, h_n\rangle_{L^2} = \langle g_n, h_n\rangle_n$ holds for $g_n, h_n \in S_n$; $\|\Sigma_n\|_{L^2\to L^2} \le B_n$ and $\|\Sigma_n^{-1}\|_{L^2\to L^2} \le A_n^{-1}$; $\mathscr{I}_n$ is a projection onto S_n (i.e., $\mathscr{I}_n\mathscr{I}_n g = \mathscr{I}_n g$, but it is not an L^2-orthogonal projection) and $\mathscr{I}_n g$ interpolates g at the points (x_i); Π_n and $\mathscr{I}_n$ are independent of the choice of basis (φ_j).

The regression experiment $\mathbb{E}_n^d$ can be transformed to a functional Gaussian shift by expanding the observations (Y_i) in the basis (φ_j),

$$Z_1 := \sum_{j=1}^n \langle Y, \varphi_j\rangle_n \varphi_j = \Pi_n f + \frac{\sigma}{\sqrt{n}}\Sigma_n^{1/2}\zeta \in S_n, \tag{3.3}$$

with Gaussian white noise $\zeta := \Sigma_n^{-1/2}(\sqrt{n}\sum_{j=1}^n \langle \varepsilon_j, \varphi_j\rangle_n \varphi_j) \sim \mathscr{N}(0, \mathrm{Id}_{S_n})$ because

$$E[\langle \zeta, g_n\rangle_{L^2}\langle \zeta, h_n\rangle_{L^2}] = \langle \Sigma_n^{-1/2} g_n, \Sigma_n^{-1/2} h_n\rangle_n = \langle g_n, h_n\rangle_{L^2}, \qquad g_n, h_n \in S_n.$$

By applying $\Sigma_n^{-1/2}$ and Σ_n^{-1}, respectively, we conclude that the regression experiment $\mathbb{E}_n^d$ is also equivalent to observing with $\zeta \sim \mathscr{N}(0, \mathrm{Id}_{S_n})$

$$Z_2 = \Sigma_n^{-1/2} Z_1 = \Sigma_n^{1/2}\mathscr{I}_n f + \frac{\sigma}{\sqrt{n}}\zeta \in S_n, \tag{3.4}$$

$$Z_3 = \Sigma_n^{-1} Z_1 = \mathscr{I}_n f + \frac{\sigma}{\sqrt{n}}\Sigma_n^{-1/2}\zeta \in S_n. \tag{3.5}$$

THEOREM 3.1. *The regression experiment $\mathbb{E}_n^d$ is equivalent to each of the experiments given by observing Z_1 in* (3.3), *Z_2 in* (3.4) *and Z_3 in* (3.5), *respectively.*

The Le Cam distance between $\mathbb{E}_n^d$ and $\mathbb{G}_n^d$ for the class $\mathscr{F}^d$ satisfies the bounds

$$\Delta_{\mathscr{F}^d}(\mathbb{E}_n^d, \mathbb{G}_n^d) \le 1 - 2\Phi\Big(-\frac{\sqrt{n}}{2\sigma}\sup_{f\in\mathscr{F}^d}\|f - \Sigma_n^{1/2}\mathscr{I}_n f\|_{L^2}\Big), \tag{3.6}$$

$$\begin{aligned}\Delta_{\mathscr{F}^d}(\mathbb{E}_n^d, \mathbb{G}_n^d) \le {}& 1 - 2\Phi\Big(-\frac{\sqrt{n}}{2\sigma}\sup_{f\in\mathscr{F}^d}\|f - \mathscr{I}_n f\|_{L^2}\Big)\\ &+ \sqrt{2}\|\Sigma_n^{-1} - \mathrm{Id}_{S_n}\|_{\mathrm{HS}},\end{aligned} \tag{3.7}$$

where $\|\bullet\|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm of an operator.

PROOF. It remains to prove the second part. The first bound (3.6) follows from the equivalence with observing Z_2, by the same arguments as for Theorem 2.4. To establish (3.7), we use the fact that the Hellinger distance between two multivariate normal distributions with the same mean satisfies

$$H^2(N(\mu,\alpha Q),N(\mu,\alpha\,\mathrm{Id}_{\mathbb{R}^n}))\le 2\|Q-\mathrm{Id}_{\mathbb{R}^n}\|_{\mathrm{HS}}^2,\qquad Q\in\mathbb{R}^{n\times n},\alpha>0,\tag{3.8}$$

which follows from, for example, Lemma 3 of Brown, Cai, Low and Zhang (2002) via the diagonalization $Q=O^\top\operatorname{diag}(\lambda_1,\dots,\lambda_n)O$ and the property $\|Q-\mathrm{Id}_{\mathbb{R}^n}\|_{\mathrm{HS}}^2=\|O(Q-\mathrm{Id}_{\mathbb{R}^n})O^\top\|_{\mathrm{HS}}^2=\sum_{i=1}^n\lambda_i^2$. Therefore, the total variation distance between the laws of Z_3 and $Z_4:=\mathscr{I}_nf+\frac{\sigma}{\sqrt{n}}\zeta$ is bounded by

$$\|\mathscr{L}(Z_3)-\mathscr{L}(Z_4)\|_{\mathrm{TV}}\le H(\mathscr{L}(Z_3),\mathscr{L}(Z_4))\le\sqrt{2}\|\Sigma_n^{-1}-\mathrm{Id}_{S_n}\|_{\mathrm{HS}}.$$

The by now standard arguments yield, with obvious notation,

$$\begin{aligned}\Delta_{\mathscr{F}^d}(\mathbb{E}_n^d,\mathbb{G}_n^d)&=\Delta_{\mathscr{F}^d}(Z_3,\mathbb{G}_n^d)\le\Delta_{\mathscr{F}^d}(Z_4,\mathbb{G}_n^d)+\Delta_{\mathscr{F}^d}(Z_4,Z_3)\\&\le 1-2\Phi\biggl(-\frac{\sqrt{n}}{2\sigma}\sup_{f\in\mathscr{F}^d}\|f-\mathscr{I}_nf\|_{L^2}\biggr)+\sqrt{2}\|\Sigma_n^{-1}-\mathrm{Id}_{S_n}\|_{\mathrm{HS}},\end{aligned}$$

as asserted. □

3.2. *Linear spline approximation.* Let us briefly explain how the approach of Carter (2006) fits into the isomorphic framework. As in Section 2.3, we consider equidistant design points $(k/m)_{k\in\{1,\dots,m\}^d}$ with $m=n^{1/d}\in\mathbb{N}$ and periodic functions on the unit cube $\mathscr{D}=[0,1]^d$. The space S_n is spanned by the periodized and tensorized linear B-splines

$$b_k(x)=b_k(x_1,\dots,x_d)=\prod_{r=1}^d\bar b(mx_r-k_r\bmod 1),$$
$$\bar b:=\mathbf{1}_{[-1/2,1/2]}*\mathbf{1}_{[-1/2,1/2]},$$

indexed by $k\in\{1,\dots,m\}^d$. For $\alpha\in(1,2]$, it is well known [cf. De Boor (2001)] that interpolation on S_n for the periodic Hölder class

$$\mathscr{F}_{\mathrm{H,per}}^d(\alpha,R):=\Bigl\{f\in C^\alpha(\mathbb{R}^d)\mid f\ \mathbb{Z}^d\text{-periodic},\ \sup_{x\neq y}|\nabla f(x)-\nabla f(y)|/|x-y|^{\alpha-1}\le R\Bigr\}$$

satisfies the estimate

$$\sup_{f\in\mathscr{F}_{\mathrm{H,per}}^d(\alpha,R)}\|f-\mathscr{I}_nf\|_{L^2([0,1]^d)}\lesssim Rn^{-\alpha/d}.\tag{3.9}$$

On the other hand, we have, for $g_n \in S_n$,

$$\|g_n\|_{L^2}^2 = \Big\| \sum_{k\in\{1,\dots,m\}^d} g_n(k/m) b_k \Big\|_{L^2}^2 = \sum_{k,\ell\in\{1,\dots,m\}^d} \langle b_k, b_\ell\rangle_{L^2} g_n(k/m) g_n(\ell/m)$$

with $\langle b_k, b_\ell\rangle_{L^2} = 0$ for $|k-\ell|_\infty > 1$ and $\langle b_k, b_\ell\rangle_{L^2} = 4^{\#\{r:k_r=\ell_r\}}/(6^d n)$ for $|k-\ell|_\infty \le 1$. Since $\sum_\ell \langle b_k, b_\ell\rangle_{L^2} = \langle b_k, 1\rangle_{L^2} = n^{-1}$, a weighted Cauchy–Schwarz inequality yields

$$\|g_n\|_{L^2}^2 \le n^{-1} \sum_{k\in\{1,\dots,m\}^d} g_n(k/m)^2 = \langle g_n, g_n\rangle_n = \langle \Sigma_n g_n, g_n\rangle_{L^2}$$

and we conclude, using the ordering of symmetric operators, that $\Sigma_n^{-1} \le \mathrm{Id}_{S_n}$.

Adding independent Gaussian noise $\eta \sim \mathscr{N}(0, \frac{\sigma^2}{n}(\mathrm{Id}_{S_n} - \Sigma_n^{-1}))$ to the observation Z_3 in (3.5), we infer that the regression experiment $\mathbb{E}_n^d$ is more informative than observing

$$Z_5 := Z_3 + \eta = \mathscr{I}_n f + \frac{\sigma}{\sqrt{n}} \tilde{\zeta} \in S_n \tag{3.10}$$

with Gaussian white noise $\tilde{\zeta} := \Sigma_n^{-1/2}\zeta + n^{1/2}\sigma^{-1}\eta \sim \mathscr{N}(0, \mathrm{Id}_{S_n})$. This randomization, together with estimate (3.9), shows that the regression experiment $\mathbb{E}_n^d$ is asymptotically at least as informative as the Gaussian experiment $\mathbb{G}_n^d$ on Hölder classes $\mathscr{F}_{\mathrm{per}}^d(\alpha, R)$ with $\alpha > d/2$ and $d \in \{1, 2, 3\}$. Together with an (easier) randomization in the other direction and a more sophisticated boundary treatment for nonperiodic function classes, this reproduces the proof in Carter (2006) for asymptotic equivalence of regression and white noise experiments in dimensions 2 and 3. For B-splines of higher order, the interpolation property $b_k(i/m) = \delta_{k,i}$ gets lost and $\Sigma_n^{-1} \le \mathrm{Id}_{S_n}$ cannot be shown, so a more refined analysis is needed. This will be accomplished in the next section for a similar approach using compactly supported wavelets.

3.3. *Wavelet multiresolution analysis.*

The construction. Let us assume an equidistant dyadic design $(k2^{-j})_{k\in\{1,\dots,2^j\}^d}$ with $n = 2^{dj}$ points for some $j \in \mathbb{N}$ and $\mathscr{D} = [0,1]^d$. We consider a wavelet multiresolution analysis $(V_j)_{j\ge 0}$ on $L^2([0,1]^d)$ obtained by means of periodization and tensor products. Let $\bar{\varphi}$ be a standard orthonormal scaling function of an r-regular multiresolution analysis for $L^2(\mathbb{R})$, that is, $(\bar{\varphi}(\bullet + k))_{k\in\mathbb{Z}}$ forms an orthonormal system in $L^2(\mathbb{R})$ and satisfies $\int \bar{\varphi} = 1$, as well as the polynomial exactness condition that $\sum_{k\in\mathbb{Z}} k^q \bar{\varphi}(x-k) - x^q$ is a polynomial of maximal degree $q-1$ for all $q = 0, \dots, R-1$ [Cohen (2000), Theorem 16.1]. We suppose that $\bar{\varphi}$ has compact support in $[-S+1, S]$, as in Daubechies' construction, so that the functions

$\varphi_{jk}:[0,1]^d \to \mathbb{R}$, $j \geq 1$, $k \in \{1,\dots,2^j\}^d$, with

$$\varphi_{jk}(x_1,\dots,x_d) := \sum_{m\in\mathbb{Z}^d} 2^{jd/2} \prod_{i=1}^d \bar{\varphi}(2^j x_i - k_i + 2^j m_i),$$

are well defined and form an orthonormal system in $L^2([0,1]^d)$ [Wojtaszczyk (1997), Proposition 2.21]. We set $S_{2^{jd}} := V_j := \operatorname{span}\{\varphi_{jk} \mid k \in \{1,\dots,2^j\}^d\}$.

Periodic approximation. Polynomial exactness and continuity of $\bar{\varphi}$ imply, for $q = 0,\dots,R-1$ and any $x \in \mathbb{R}$ [Sweldens and Piessens (1993)]

$$\sum_{m\in\mathbb{Z}} (x+m)^q \bar{\varphi}(x+m) = \int_{-\infty}^{\infty} x^q \bar{\varphi}(x)\,dx.$$

This identity is fundamental for our purposes because it implies the following fact: for $\mathbb{Z}^d$-periodic functions $h:\mathbb{R}^d \to \mathbb{R}$ that coincide with a polynomial p of maximal degree $R-1$ on $\prod_{i=1}^d [2^{-j}(k_i - S - 1), 2^{-j}(k_i + S)]$, we have

$$\begin{aligned}
\langle h, \varphi_{jk}\rangle_{L^2} &= \sum_{m\in\mathbb{Z}^d} 2^{jd/2} \int_{[0,1]^d} h(x) \prod_{i=1}^d \bar{\varphi}\big(2^j(x_i + m_i) - k_i\big)\,dx \\
&= 2^{jd/2} \int_{\mathbb{R}^d} h(x) \prod_{i=1}^d \bar{\varphi}(2^j x_i - k_i)\,dx \\
&= 2^{-jd/2} \int_{[-S-1,S]^d} p\big(2^{-j}(x+k)\big) \prod_{i=1}^d \bar{\varphi}(x_i)\,dx \\
&= 2^{-jd/2} \sum_{m\in\mathbb{Z}^d} p\big(2^{-j}(m+k)\big) \prod_{i=1}^d \bar{\varphi}(m_i) \\
&= 2^{-jd/2} \sum_{m\in\{1,\dots,2^j\}^d} h(2^{-j}m)\varphi_{jk}(2^{-j}m).
\end{aligned}$$

Hence, $\langle h, \varphi_{jk}\rangle_{L^2} = n^{1/2}\langle h, \varphi_{jk}\rangle_n$, with $n = 2^{jd}$. For any $\mathbb{Z}^d$-periodic function $g \in H^s_{\mathrm{S,per}}([0,1]^d)$ with $s \in (d/2, R)$, this local polynomial reproduction property implies by standard, but sophisticated, arguments [Cohen (2000), Theorem 30.6] that

$$\|g - \Pi_n g\|_{L^2} \lesssim 2^{-js}\|g\|_{H^s} = n^{-s/d}\|g\|_{H^s}, \tag{3.11}$$

where $\|\bullet\|_{H^s}$ denotes the standard L^2-Sobolev norm of regularity s on $[0,1]^d$. We split the bias term in (3.6) and obtain, by functional calculus,

$$\begin{aligned}
\|f - \Sigma_n^{-1/2}\Pi_n f\|_{L^2} &\leq \|f - \Pi_n f\|_{L^2} + \|\Pi_n f - \Sigma_n^{-1/2}\Pi_n f\|_{L^2} \\
&= \|f - \Pi_n f\|_{L^2} + \|H(\Sigma_n)(\mathrm{Id} - \Pi_n)\Pi_n f\|_{L^2},
\end{aligned}$$

with $H:\mathbb{R}^+ \to \mathbb{R}$, $H(x) := 1/(x + x^{1/2}) = (x^{-1/2} - 1)/(1 - x)$. Since H decreases monotonically and $H(x) \le x^{-1/2}$, we have $\|H(\Sigma_n)\|_{L^2\to L^2} \le \lambda_{\min}^{-1/2}$, $\lambda_{\min}$ being the smallest eigenvalue of Σ_n.

For $n = 2^{jd} \ge 2S - 1$, the operator Σ_n satisfies the following scaling property:

$$\begin{aligned}\langle \Sigma_n \varphi_{jk}, \varphi_{j\ell}\rangle_{L^2} &= \frac{1}{n} \sum_{\nu\in\{1,\dots,2^j\}^d} \varphi_{jk}(\nu 2^{-j})\varphi_{j\ell}(\nu 2^{-j}) \\ &= \sum_{m\in\mathbb{Z}^d} \sum_{\nu\in\{1,\dots,2^j\}^d} \prod_{a=1}^{d} \big(\bar\varphi\big((\nu - k + 2^j m)_a\big)\bar\varphi\big((\nu - \ell + 2^j m)_a\big)\big) \\ &= \prod_{a=1}^{d} \left(\sum_{b\in\mathbb{Z}} \bar\varphi(b - k_a)\bar\varphi(b - \ell_a)\right).\end{aligned}$$

Since $\bar\varphi$ has compact support, the series is just a finite sum and Σ_n has a bounded Toeplitz matrix representation in terms of (φ_{jk}). Using Fourier multipliers, it follows that $\langle \Sigma_n g_n, g_n\rangle_{L^2} \ge A_{\bar\varphi}^2 \|g_n\|_{L^2}^2$, $g_n \in S_n$, with $A_{\bar\varphi} := \inf_{u\in[0,2\pi]} |\sum_{k\in\mathbb{Z}} \bar\varphi(k)e^{\imath ku}|^d$, independently of n. Due to the compact support of $\bar\varphi$, we have $A_{\bar\varphi} > 0$ if and only if the trigonometric polynomial $\sum_{k\in\mathbb{Z}} \bar\varphi(k)e^{\imath ku}$, $u \in [0, 2\pi]$, does not vanish. It is well known [Sweldens and Piessens (1993), Lemma 3] that this is exactly the condition needed to ensure that the multiresolution analysis is also generated by an interpolating scaling function. It can be checked for standard Daubechies scaling functions, for example, by showing $|\bar\varphi(k_0)| > \sum_{k'\neq k_0} |\bar\varphi(k')|$ for some $k_0 \in \mathbb{Z}$. Moreover, gaining more flexibility by considering the shifted spaces based on $\bar\varphi_\tau = \bar\varphi(\bullet - \tau)$, $\tau \in (0, 1)$, a wavelet multiresolution analysis will almost always satisfy $A_{\bar\varphi_\tau} > 0$ for some value of τ [cf. Sweldens and Piessens (1993) and the references therein].

We arrive at

$$\|f - \Sigma_n^{-1/2}\Pi_n f\|_{L^2} \le \|f - \Pi_n f\|_{L^2} + A_{\bar\varphi}^{-1/2}\|(\mathrm{Id} - \Pi_n)\Pi_n f\|_{L^2}.$$

Because of $\|\Pi_n f\|_{H^s} \to \|f\|_{H^s}$ [Cohen (2000), Theorem 30.7], we derive from (3.11) the uniform estimate over $f \in \mathscr{F}^d_{\mathrm{S,per}}(s, R)$,

$$\|f - \Sigma_n^{-1/2}\Pi_n f\|_{L^2} \le \|f - \Pi_n f\|_{L^2} + A_{\bar\varphi}^{-1/2}\|(\mathrm{Id} - \Pi_n)\Pi_n f\|_{L^2} \lesssim Rn^{-s/d}.$$

Hence, the estimate in (3.6) yields asymptotic equivalence between the regression and the white noise experiment for any class $\mathscr{F}^d_{\mathrm{S,per}}(s, R)$ with $s > d/2$.

This result provides another way to construct explicitly the transformation between the regression and the white noise setting. It has no more theoretical implications than the Fourier basis approach, but it paves the way for proving asymptotic equivalence for nonperiodic function classes.

Nonperiodic approximation. Since every φ_{jk} has support of length $2^{-j}(2S-1)$, only those functions φ_{jk} with $k_r \in \{1, \dots, S-2\} \cup \{2^j - S + 1, \dots, 2^j\}$ for some $r = 1, \dots, d$ cross the boundary and are periodized at all. Therefore, the same derivation using only interior scaling functions shows that the regression experiment $\mathbb{E}_n^d$ for the general Sobolev function class

$$\mathscr{F}_{\mathrm{S}}^d(s, R) := \{f \in H^s([0,1]^d) \mid \|f\|_{H^s} \le R\}$$

is asymptotically more informative than the restricted white noise experiment $\bar{\mathbb{G}}_n^d$ given by observing

$$dY(x) = f(x)\,dx + \frac{\sigma}{\sqrt{n}}\,dB(x), \tag{3.12}$$
$$x \in [\delta_n, 1-\delta_n]^d \text{ with } \delta_n := (2S-1)n^{-1/d}.$$

Although $\bar{\mathbb{G}}_n^d$ is a priori less informative than $\mathbb{G}_n^d$, we may use classical extrapolation, for example, the Taylor polynomial T_f^y of order $\lfloor s \rfloor$ around $y \in [\delta_n, 1-\delta_n]^d$. At the points $x \in [0,1]^d \setminus [\delta_n, 1-\delta_n]^d$, we define the extrapolation $\tilde{f}(x) = T_f^{y_x}(x)$ for a point $y_x \in [\delta_n, 1-\delta_n]^d$ with $|y_x - x|_\infty \le 2\delta_n$, selected in a measurable way, and $\tilde{f}(x) = f(x)$ otherwise. We thereby achieve

$$\left(\int_{[0,1]^d} |\tilde{f}(x) - f(x)|^2\,dx\right)^{1/2} \lesssim Rn^{-s/d}$$

such that $\Delta_{\mathscr{F}_{\mathrm{S}}^d(s,R)}(\bar{\mathbb{G}}_n^d, \mathbb{G}_n^d) \lesssim \sigma^{-1}Rn^{1/2-s/d}$. This means that $\bar{\mathbb{G}}_n^d$ and $\mathbb{G}_n^d$ are asymptotically equivalent for $s > d/2$ and we have obtained a result for function classes without a periodicity condition.

THEOREM 3.2. *For general d-dimensional Sobolev classes $\mathscr{F}_{\mathrm{S}}^d(s,R)$ with regularity $s > d/2$ and equidistant design on the cube $[0,1]^d$, the nonparametric regression experiment $\mathbb{E}_n^d$ and the Gaussian white noise experiment $\mathbb{G}_n^d$ are asymptotically equivalent as $n \to \infty$. The Le Cam distance satisfies*

$$\Delta_{\mathscr{F}_{\mathrm{S}}^d(s,R)}(\mathbb{E}_n^d, \mathbb{G}_n^d) \lesssim \sigma^{-1}Rn^{1/2-s/d}.$$

Discussion. The property that a wavelet estimator based on an equidistant regression model and a corresponding estimator based on a white noise model are asymptotically close is well known [see, e.g., Donoho and Johnstone (1999) and Johnstone and Silverman (2004)]. Interestingly, both papers show identical asymptotics of the L^2-risk for standard estimators uniformly over balls in Besov spaces $B_{p,q}^s([0,1])$ with $s > 1/p$ or $s = p = 1$. Since $B_{p,q}^s$ embeds in the Sobolev space H^σ for $s > \sigma$ and $s - 1/p > \sigma - 1/2$, Theorem 3.2 provides, more generally, asymptotic equivalence for Besov classes with $s > 1/p$ and $p < 2$. The counterexample in Brown and Low (1996) shows, however, that for $s \le 1/2$ and all $p \in [1, \infty]$, asymptotic equivalence breaks down. Similarly, if $\psi \in B^1_{1,1}$ is a function with support in $(0, 1)$ and $\|\psi\|_{L^2} = 1$, then $\psi_n(x) := \psi(nx)$ has support in $(0, 1/n)$, L^2-norm $\|\psi_n\|_{L^2} = n^{-1/2}$ and Besov norm $\|\psi_n\|_{B^1_{1,1}} \sim 1$. Hence, testing the signal $f = 0$ versus $f = \psi_n$ has nontrivial power in the white noise model $\mathbb{G}^1_n$, while both signals generate exactly the same observations in the regression model $\mathbb{E}^1_n$. We conclude that $\mathbb{G}^1_n$ and $\mathbb{E}^1_n$ are not asymptotically equivalent on Besov classes with $s = 1$, $p = 1$. An intriguing example for the important class of bounded variation functions is given by $\psi_n(x) = \sqrt{2}\mathbf{1}_{[1/4n,3/4n]}(x)$. Asymptotic equivalence between Gaussian regression and white noise is indeed an L^2-theory and we cannot gain by measuring smoothness in an L^p-sense, $p \neq 2$.

Let us also mention that the (asymptotically negligible) loss in information due to neglecting boundary coefficients in the construction seems unavoidable. The wavelets on an interval [Cohen, Daubechies and Vial (1993)] use nonorthogonal boundary corrections and can therefore not be used, while the Coiflet approach of Johnstone and Silverman (2004) also involves some information loss at the boundary (cf. their remark on dimensions before Proposition 2).

4. Random design.

4.1. *The general idea.* Denote by $U([0, 1]^d)$ the uniform distribution on the cube $\mathscr{D} = [0, 1]^d$.

DEFINITION 4.1. Let $\mathbb{E}^d_{n,r}$ be the compound experiment obtained by observing independent random design points $X_i \sim U([0, 1]^d)$, $i = 1, \dots, n$, and the regression

$$Y_i = f(X_i) + \sigma\varepsilon_i, \qquad i = 1, \dots, n,$$

for $n \in \mathbb{N}$ and $f : [0, 1]^d \to \mathbb{R}$ in some class $\mathscr{F}^d \subseteq \mathscr{L}^2([0, 1]^d)$ and with i.i.d. random variables $\varepsilon_i \sim \mathscr{N}(0, 1)$, independent of the design.

We place ourselves in the isomorphic setting, that is, we are given an $L^2([0, 1]^d)$-orthonormal basis $(\varphi_j)_{j\geq 1}$ and we set $S_n = \operatorname{span}(\varphi_1, \dots, \varphi_n)$. For the moment, we merely assume that S_n is chosen to satisfy the isomorphic condition (3.1), given the random design points $(X_i)_{1\leq i\leq n}$. Later, certain parts will rely on fine properties of the Fourier basis. Conditionally on the design, the regression experiment is equivalent to observing

$$Z_1 := \sum_{j=1}^{n} \langle Y, \varphi_j\rangle_n \varphi_j = \Pi_n f + \frac{\sigma}{\sqrt{n}} \Sigma_n^{1/2} \zeta \in S_n,$$

with white noise $\zeta \sim N(0, \mathrm{Id}_{S_n})$. Let us briefly comment on why the foregoing approaches using Z_2 in (3.4) or Z_3 in (3.5) will not succeed here. For

$Z_2 = \Sigma_n^{-1/2} Z_1$, we need to have $\|(\Sigma_n^{1/2} - \mathrm{Id})\mathscr{I}_n f\|_{L^2}$ and $\|\mathscr{I}_n f - f\|_{L^2}$ of smaller order than $n^{-1/2}$. The second property can be ensured for Sobolev classes of regularity $s > d/2$ as before. The first property, however, will not hold. By empirical process theory, we have, for $g_1, g_2 \in S_n$, approximately $\langle \Sigma_n g_1, g_2\rangle_{L^2} = \langle g_1, g_2\rangle_n \approx \langle g_1, g_2\rangle_{L^2} + n^{-1/2} \int g_1 g_2 \, dB^0$ with a Brownian bridge B^0. By the linearization $(1+h)^{1/2} - 1 \approx h/2$ and taking expectation with respect to the random design, we find

$$E[\|(\Sigma_n^{1/2} - \mathrm{Id})\mathscr{I}_n f\|_{L^2}^2] \sim E\left[\sum_{j=1}^{n} \left| n^{-1/2} \int (\mathscr{I}_n f)\varphi_j \, dB^0 \right|^2\right]$$
$$\sim n^{-1} \sum_{j=1}^{n} \int |\varphi_j|^2 |\mathscr{I}_n f|^2.$$

Hence, in the mean over the random design, this term does not tend to zero. When considering $Z_3 = \Sigma_n^{-1} Z_1$, we would need $\|\Sigma_n^{-1} - \mathrm{Id}_{S_n}\|_{\mathrm{HS}} \to 0$ [cf. (3.7)], but the mean over this term is, by the same approximations, of order n. The main defect in these approaches is that we do not take advantage of the regularity of f.

The new idea generalizes the two-level procedure of Brown et al. (2002) and Brown et al. (2004) and can be interpreted as a localization approach, as in Nussbaum (1996). We choose an intermediate level $n_0 < n$ and split $S_n = S_{n_0} + U_{n_0}^n$ with the $\| \bullet \|_n$-orthogonal complement $U_{n_0}^n$ of S_{n_0} in S_n. On the low-frequency space S_{n_0}, we use the empirical orthogonal projection $P_{n_0}^n Y$ of the data onto S_{n_0}. This construction is analogous to Z_3 in (3.5), and the heteroskedasticity in the noise term will become asymptotically negligible provided $n_0 = o(n^{1/2})$.

On the high-frequency part $U_{n_0}^n$ of S_n, we transform to a Gaussian shift with white noise, which is independent of the noise in S_{n_0}, in the spirit of Z_2 in (3.4). In order to take advantage of the regularity of f, however, we do not use the standard square root operator $\Sigma_n^{-1/2}$ to whiten the noise, but the adjoint T^* of an operator $T : S_n \to S_n$ which has an upper triangular matrix representation in the basis (φ_j) and satisfies $TT^* = \Sigma_n^{-1}$ (as in the Cholesky decomposition). Since T^* is a unitary transformation of $\Sigma_n^{-1/2}$, the noise part remains white. Due to the triangular structure, the signal coefficients $\langle T^* \Pi_n f, \varphi_j\rangle_{L^2} = \langle T^{-1} \mathscr{I}_n f, \varphi_j\rangle_{L^2}$ do not involve the (usually large) coefficients $\langle \mathscr{I}_n f, \varphi_k\rangle_{L^2}$ for indices k smaller than j. Moreover, for the Fourier basis, the other off-diagonal matrix entries of T^{-1} are centred and uncorrelated. The deviations in the diagonal entries grow with the frequencies, but are exactly counterbalanced by the decay of the Fourier coefficients for Sobolev function classes. Provided $n_0 \to \infty$, this high-frequency transformation will imply asymptotic equivalence.

4.2. *The main result.* Let us specify the transformation T concretely based on the Gram–Schmidt procedure for orthonormalization with respect to $\| \bullet \|_n$.

For $j \le n$, denote by $P_j,\ P_j^n : S_n \to S_n$ the L^2-orthogonal and $\|\bullet\|_n$-orthogonal projections onto S_j, respectively, and set $P_0^n := 0$. We obtain an $\|\bullet\|_n$-orthonormal basis (φ_j^n) of S_n via

$$\varphi_j^n := \frac{\varphi_j - P_{j-1}^n \varphi_j}{\|\varphi_j - P_{j-1}^n \varphi_j\|_n}, \qquad j = 1, \dots, n.$$

φ_j^n is then in S_j and the $\|\bullet\|_n$-orthogonality $\varphi_j^n \perp_n S_{j-1}$ holds. Defining $T : S_n \to S_n$ via $T\varphi_j := \varphi_j^n$, we see that T satisfies $\langle T\varphi_{j'}, \varphi_j\rangle_{L^2} = 0$ for $j > j'$ and is an isometry between $(S_n, \|\bullet\|_{L^2})$ and $(S_n, \|\bullet\|_n)$ such that $\Sigma_n = (TT^*)^{-1}$. The noise terms $(\langle \varepsilon, \varphi_j^n\rangle_n)_{1\le j\le n} \sim \mathscr{N}(0, n^{-1})$ are therefore independent and

$$P_{n_0}^n \varepsilon := \sum_{j=1}^{n_0} \langle \varepsilon, \varphi_j^n\rangle_n \varphi_j^n = \sum_{j=1}^{n_0} \langle \varepsilon, \varphi_j^n\rangle_n T\varphi_j \sim \mathscr{N}(0, n^{-1} T|_{S_{n_0}} T|_{S_{n_0}}^*).$$

Using $T|_{S_{n_0}} T|_{S_{n_0}}^* = \Sigma_{n_0}^{-1}$, we introduce the rescaled covariance operator $\overline{\Sigma}_n : S_n \to S_n$ via

$$\overline{\Sigma}_n g_n := \Sigma_{n_0}^{-1} P_{n_0} g_n + (\mathrm{Id}_{S_n} - P_{n_0}) g_n, \qquad g_n \in S_n.$$

The regression experiment is then transformed to observing

$$\begin{aligned} Z_r &:= \sum_{j=1}^{n_0} \langle Y, \varphi_j^n\rangle_n \varphi_j^n + \sum_{j=n_0+1}^{n} \langle Y, \varphi_j^n\rangle_n \varphi_j \in S_n \\ &= P_{n_0}^n f + T^{-1}(P_n^n - P_{n_0}^n) f + n^{-1/2}\sigma \overline{\Sigma}_n^{1/2} \zeta \in S_n \end{aligned} \tag{4.1}$$

with Gaussian white noise $\zeta \sim N(0, \mathrm{Id}_{S_n})$, conditional on the random design.

EXAMPLE 4.2. Let us consider the Haar basis. Write $I_{jk} = [2^{-j}k, 2^{-j}(k+1))$, $N_{jk} = \#\{i : X_i \in I_{jk}\}$ and $\psi_{jk} = 2^{j/2}(\mathbf{1}_{I_{j+1,2k}} - \mathbf{1}_{I_{j+1,2k+1}})$ for $j \ge 0$, $k = 0, \dots, 2^j - 1$. By construction, the transformed basis function ψ_{jk}^n has support I_{jk}, is constant on $I_{j+1,2k}$, $I_{j+1,2k+1}$ and satisfies $\langle \psi_{jk}^n, \mathbf{1}_{I_{jk}}\rangle_n = 0$, $\|\psi_{jk}^n\|_n = 1$. We infer that

$$\psi_{jk}^n = C_{jk}(N_{j,2k}^{-1}\mathbf{1}_{I_{j,2k}} - N_{j,2k+1}^{-1}\mathbf{1}_{I_{j,2k+1}}), \qquad C_{jk}^2 = nN_{j+1,2k}N_{j+1,2k+1}/N_{jk}.$$

This application of our framework has been used previously in one-dimensional constructions [Brown et al. (2002), equation (2.8)]. Because here S_n is not isomorphic for most design realizations, additional randomizations are needed.

For the following general d-dimensional theorem, we consider the construction (4.1) in terms of the Fourier basis functions $\varphi_j(x) = \exp(2\pi\iota\langle \ell(j), x\rangle)$, with an enumeration $\ell : \mathbb{N} \to \mathbb{Z}^d$ of $\mathbb{Z}^d$ satisfying $|\ell(j)|_{\ell^2} \le |\ell(j')|_{\ell^2}$ for $j \le j'$ (i.e., sorted in the order of magnitudes of the frequencies).

THEOREM 4.3. *For d-dimensional periodic Sobolev classes $\mathscr{F}^d_{\mathrm{S,per}}(s,R)$ with regularity $s > d/2$, the nonparametric regression experiment $\mathbb{E}^d_{n,r}$ with random design and the Gaussian shift experiment $\mathbb{G}^d_n$ are asymptotically equivalent as $n_0, n \to \infty$ and $n_0 = o(n^{1/2})$. The Le Cam distance satisfies*

$$\Delta_{\mathscr{F}^d_{\mathrm{S,per}}(s,R)}(\mathbb{E}^d_{n,r}, \mathbb{G}^d_n) \lesssim n^{-1/2} n_0 + \sigma^{-1} R n_0^{1/2-s/d}.$$

REMARK 4.4. The asymptotically optimal choice of n_0 is given by $n_0 \sim n^{d/(2s+d)}$, which yields a bound on the Le Cam distance of order $n^{(d-2s)/(2d+4s)}$. Note that this choice $n_0 \sim n^{d/(2s+d)}$ corresponds exactly to the optimal dimension of the approximation spaces in nonparametric regression and is also used by Gaiffas (2007) for his two-level construction of optimal confidence bands. It can be shown that even for parametric linear regression, the Le Cam distance between equidistant and random design is of order $n^{-1/2}$ and not smaller.

PROOF OF THEOREM 4.3. In order to bound the Le Cam distance for compound experiments, we use the fact that for distributions $K \otimes P$ and $K' \otimes P$, defined on $(\Omega \times \Omega', \mathscr{F} \otimes \mathscr{F}')$ by the measure P on $\mathscr{F}$ and the Markov kernels K, K' from Ω to $\mathscr{F}'$, the total variation distance can be calculated by conditioning:

$$\|K \otimes P - K' \otimes P\|_{\mathrm{TV}(\mathscr{F}\otimes\mathscr{F}')} = \int \|K(\omega, \bullet) - K'(\omega, \bullet)\|_{\mathrm{TV}(\mathscr{F}')} P(d\omega).$$

Therefore, we can first work conditionally on the design and then take expectations for (X_i). Moreover, the white noise experiment $\mathbb{G}^d_n$ is equivalent to the compound experiment of $\mathbb{G}^d_n$ and the observation of the random design points because the latter is a trivial randomization of $\mathbb{G}^d_n$.

It is a remarkable property of the Fourier basis that S_n is almost surely isomorphic [cf. Theorem 1.1 of Bass and Gröchenig (2004)]. In Proposition 4.8 below, we prove that the event

$$\Omega^n_j := \{\forall g \in S_j : \tfrac{1}{2}\|g\|_{L^2} \le \|g\|_n \le 2\|g\|_{L^2}\} \tag{4.2}$$

for $j\log(j) = o(n)$ even satisfies $P((\Omega^n_j)^{\complement}) \to 0$ with a convergence rate faster than any polynomial in n. This is much tighter with respect to the subspace dimension than what can be derived from Bass and Gröchenig (2004). In order to establish asymptotic equivalence, it therefore suffices to estimate the total variation distances on the event $\Omega^n_{n_0}$.

By (4.1), the regression experiment $\mathbb{E}^d_{n,r}$ is equivalent to observing Z_r together with the design. Introducing

$$\bar{Z}_r := P_n f + \sigma n^{-1/2} \zeta \in S_n, \tag{4.3}$$

we shall prove in a moment that (with obvious notation)

$$\Delta_{\mathscr{F}^d_{\mathrm{S,per}}(s,R)}(Z_r, \bar{Z}_r) \lesssim n^{-1/2} n_0 + \sigma^{-1} R n_0^{1/2-s/d}, \tag{4.4}$$

but then it follows that observing $\bar{Z}_r$ is equivalent to observing

$$dY(x) = P_n f(x) + \sigma n^{-1/2}\, dB(x), \qquad x \in [0,1]^d,$$

which has a total variation distance to the Gaussian shift $\mathbb{G}_n^d$ of order $\sigma^{-1} n^{1/2} \| f - P_n f\|_{L^2} \lesssim \sigma^{-1} n^{1/2-s/d} \|f\|_{H^s}$. Using the triangle inequality for the Le Cam distance between the intermediate experiments, we arrive at the bound for $\Delta_{\mathscr{F}^d_{\mathrm{S,per}}(s,R)}(\mathbb{E}^d_{n,r}, \mathbb{G}^d_n)$.

To obtain (4.4), we take expectations over the design and split

$$E[\|\mathscr{L}(Z_r) - \mathscr{L}(Z'_r)\|^2_{\mathrm{TV}} \mathbf{1}_{\Omega^n_{n_0}}] \lesssim I + II + III,$$

with the terms

$$I := n\sigma^{-2} E[\|(P^n_{n_0} - P_{n_0}) f\|^2_{L^2} \mathbf{1}_{\Omega^n_{n_0}}] \qquad \text{(difference in mean on } S_{n_0}\text{)},$$

$$II := E[\|\Sigma^{-1}_{n_0} - \mathrm{Id}_{S_{n_0}}\|^2_{\mathrm{HS}} \mathbf{1}_{\Omega^n_{n_0}}] \qquad \text{(heteroskedasticity on } S_{n_0}\text{)},$$

$$III := n\sigma^{-2} E\big[\big\|\big(T^{-1}(P^n_n - P^n_{n_0}) - (P_n - P_{n_0})\big) f\big\|^2_{L^2} \mathbf{1}_{\Omega^n_{n_0}}\big]$$

$$\text{(difference in mean on } S_{n_0}^{\perp_{L^2}}\text{)}.$$

Term I. Using the projection properties, we obtain, on $\Omega^n_{n_0}$, that

$$\|(P^n_{n_0} - P_{n_0}) f\|^2_{L^2} = \|P^n_{n_0}(\mathrm{Id} - P_{n_0}) f\|^2_{L^2} \le 4 \|P^n_{n_0}(\mathrm{Id} - P_{n_0}) f\|^2_n.$$

Because of $E[\langle \varphi_k, \varphi^n_j\rangle_n \overline{\langle \varphi_{k'}, \varphi^n_j\rangle_n}] = 0$ for $k \ne k'$, $k, k' > j$ by Proposition 4.5 below, an expansion in the basis (φ^n_j) yields

$$\begin{aligned} E[\|P^n_{n_0}(\mathrm{Id} - P_{n_0}) f\|^2_n] &= \sum_{j=1}^{n_0} \sum_{k=n_0+1}^{\infty} |\langle f, \varphi_k\rangle_{L^2}|^2 E[|\langle \varphi_k, \varphi^n_j\rangle_n|^2] \\ &= \sum_{k=n_0+1}^{\infty} |\langle f, \varphi_k\rangle_{L^2}|^2 E[\|P^n_{n_0} \varphi_k\|^2_n]. \end{aligned}$$

Proposition 4.9 below yields $E[\|P^n_{n_0}\varphi_k\|^2_n] \lesssim k/n$ and hence

$$I \lesssim \sigma^{-2} \sum_{k=n_0+1}^{\infty} |\langle f, \varphi_k\rangle_{L^2}|^2 k \lesssim \sigma^{-2} n_0^{1-2s/d} \|f\|^2_{H^s}.$$

Term II. Using $\|\Sigma^{-1}_{n_0}\|_{L^2 \to L^2} \le 4$ on $\Omega^n_{n_0}$, we find that

$$\begin{aligned} E[\|\Sigma^{-1}_{n_0} - \mathrm{Id}_{S_{n_0}}\|^2_{\mathrm{HS}} \mathbf{1}_{\Omega^n_{n_0}}] &\le E[\|\Sigma^{-1}_{n_0}\|_{L^2 \to L^2} \|\Sigma_{n_0} - \mathrm{Id}_{S_{n_0}}\|^2_{\mathrm{HS}} \mathbf{1}_{\Omega^n_{n_0}}] \\ &\le 4 E[\|\Sigma_{n_0} - \mathrm{Id}_{S_{n_0}}\|^2_{\mathrm{HS}}] \end{aligned}$$

$$= 4 \sum_{j,j'=1}^{n_0} E[|\langle \varphi_j, \varphi_{j'} \rangle_n - \delta_{j,j'}|^2]$$

$$\le 4n^{-1} \sum_{j,j'=1}^{n_0} \int |\varphi_j|^2 |\varphi_{j'}|^2.$$

For the Fourier basis, we obtain $II \le 4n^{-1} n_0^2$.

Term III. Let us write $f = f_0 + f_1 + f_2$ with $f_0 = P_{n_0} f$, $f_1 = (P_n - P_{n_0}) f$, $f_2 = (\mathrm{Id} - P_n) f$. The projection properties then imply that

$$\begin{aligned}
&E\big[\big\|\big(T^{-1}(P_n^n - P_{n_0}^n) - (P_n - P_{n_0})\big) f\big\|_{L^2}^2 \mathbf{1}_{\Omega_{n_0}^n}\big] \\
&\quad = E[\|T^{-1} f_1 + T^{-1} P_n^n f_2 - T^{-1} P_{n_0}^n (f_1 + f_2) - f_1\|_{L^2}^2 \mathbf{1}_{\Omega_{n_0}^n}] \\
&\quad \le 3E[\|(T^{-1} - \mathrm{Id}) f_1\|_{L^2}^2 + \|(P_n^n - P_{n_0}^n) f_2\|_n^2 + \|P_{n_0}^n f_1\|_n^2 \mathbf{1}_{\Omega_{n_0}^n}] \\
&\quad \le 3E[\|f_1\|_n^2 + \|f_1\|_{L^2}^2 - 2\,\mathrm{Re}(\langle T^{-1} f_1, f_1 \rangle_{L^2})] \\
&\qquad + 3E[\|f_2\|_n^2] + 3E[\|P_{n_0}^n f_1\|_n^2 \mathbf{1}_{\Omega_{n_0}^n}] \\
&\quad = 6E[\mathrm{Re}(\langle f_1 - T^{-1} f_1, f_1 \rangle_{L^2})] + 3\|f_2\|_{L^2}^2 + 3E[\|P_{n_0}^n f_1\|_n^2 \mathbf{1}_{\Omega_{n_0}^n}] \\
&\quad =: III_1 + III_2 + III_3.
\end{aligned}$$

The term III_2 is easily bounded by $\|f_2\|_{L^2}^2 \lesssim n^{-2s/d} \|f\|_{H^s}^2$. As in the estimate for term I, we obtain $III_3 \lesssim n^{-1} n_0^{1-2s/d} \|f\|_{H^s}^2$. For III_1, we use $E[\langle T^{-1}\varphi_j, \varphi_k \rangle_{L^2}] = 0$, $j \ne k$, by Proposition 4.5 below to conclude that

$$E[\mathrm{Re}(\langle f_1 - T^{-1} f_1, f_1 \rangle_{L^2})] = \sum_{j=n_0+1}^{n} |\langle f, \varphi_j \rangle_{L^2}|^2 E[\langle (\mathrm{Id} - T^{-1}) \varphi_j, \varphi_j \rangle_{L^2}].$$

Because of $\|\varphi_j\|_n = 1$ for the Fourier basis, we find that

$$\begin{aligned}
\langle T^{-1} \varphi_j, \varphi_j \rangle_{L^2} &= \langle \|\varphi_j - P_{j-1}^n \varphi_j\|_n \varphi_j^n + P_{j-1}^n \varphi_j, \varphi_j^n \rangle_n \\
&= \|\varphi_j - P_{j-1}^n \varphi_j\|_n \ge 1 - \|P_{j-1}^n \varphi_j\|_n^2.
\end{aligned}$$

By Proposition 4.9 below, the bound

$$\begin{aligned}
E[\mathrm{Re}(\langle f_1 - T^{-1} f_1, f_1 \rangle_{L^2})] &\le \sum_{j=n_0+1}^{n} |\langle f, \varphi_j \rangle_{L^2}|^2 E[\|P_{j-1}^n \varphi_j\|_n^2] \\
&\lesssim \sum_{j=n_0+1}^{n} \frac{j}{n} |\langle f, \varphi_j \rangle_{L^2}|^2
\end{aligned}$$

follows, which is of order $n^{-1}n_0^{1-2s/d}\|f\|_{H^s}^2$. Putting the estimates together, we have

$$\begin{aligned}III &\lesssim \sigma^{-2}(n_0^{1-2s/d}\|f\|_{H^s}^2 + n^{1-2s/d}\|f\|_{H^s}^2 + n_0^{1-2s/d}\|f\|_{H^s}^2)\\ &\lesssim \sigma^{-2}n_0^{1-2s/d}\|f\|_{H^s}^2\end{aligned}$$

and, summing, $I + II + III \lesssim \sigma^{-2}n_0^{1-2s/d}R^2 + n^{-1}n_0^2$ uniformly over $f \in \mathscr{F}_{\mathrm{S,per}}^d(s, R)$, which gives the asserted bound (4.4). □

4.3. *Technical results.* We now gather results on fine properties of the Fourier basis (φ_j) and its generated approximation spaces S_n. The setting is as in the proof of Theorem 4.3. For the value of the next proposition, notice that $\langle \varphi_{k'}, \varphi_k^n\rangle_n = \langle T^{-1}\varphi_{k'}, \varphi_k\rangle_{L^2}$.

PROPOSITION 4.5. *We have, for indices $k'', k' > k \geq 1$, $k'' \neq k'$,*

$$E[\langle \varphi_{k'}, \varphi_k^n\rangle_n] = 0 \quad \textit{and} \quad E[\langle \varphi_{k'}, \varphi_k^n\rangle_n \overline{\langle \varphi_{k''}, \varphi_k^n\rangle_n}] = 0.$$

PROOF. Since the randomness enters via P_{k-1}^n in a very intricate way, we use a symmetry argument. Define $X_i := (Y_i + \vartheta) \bmod 1$, $i = 1, \dots, n$, with $Y_i \sim U([0,1]^d)$, $\vartheta \sim U([0,1]^d)$ all independent such that $X_i \sim U([0,1]^d)$ are i.i.d. Working conditionally on ϑ, we shall keep track on the dependence on ϑ using brackets. We claim that for $k' > k$,

$$\langle \varphi_{k'}, \varphi_k^n\rangle_n[\vartheta] = e^{2\pi\iota\langle \ell(k')-\ell(k),\vartheta\rangle}\langle \varphi_{k'}, \varphi_k^n\rangle_n[0], \tag{4.5}$$

which implies the result due to

$$\int_{[0,1]^d} e^{2\pi\iota\langle \ell(k')-\ell(k),\vartheta\rangle}\,d\vartheta = 0 \quad \text{and}$$

$$\int_{[0,1]^d} e^{2\pi\iota(\langle \ell(k')-\ell(k),\vartheta\rangle - \langle \ell(k'')-\ell(k),\vartheta\rangle)}\,d\vartheta = 0.$$

For $m \in \mathbb{Z}^d$, put

$$A_m[\vartheta] := \frac{1}{n}\sum_{j=1}^n e^{2\pi\iota\langle m, X_j[\vartheta]\rangle} = \frac{1}{n}\sum_{j=1}^n e^{2\pi\iota\langle m, Y_j+\vartheta\rangle} = e^{2\pi\iota\langle m,\vartheta\rangle}A_m[0].$$

The proof of (4.5) will be performed by induction from $\kappa < k$ to k, considering tuples (κ', κ), $\kappa' > \kappa$ and (k', k), $k' > k$. Since $\ell(1) = 0$ and $\varphi_1^n = \varphi_1 = 1$, we have, for $k' > 1$ and $k = 1$,

$$\langle \varphi_{k'}, \varphi_k^n\rangle_n[\vartheta] = \frac{1}{n}\sum_{j=1}^n e^{2\pi\iota\langle \ell(k'), Y_j+\vartheta\rangle} = e^{2\pi\iota\langle \ell(k')-\ell(1),\vartheta\rangle}\langle \varphi_{k'}, \varphi_k^n\rangle_n[0].$$

Writing $c_k := \|\varphi_k - P^n_{k-1}\varphi_k\|_n^{-1}$, the induction hypothesis implies that

$$c_k^{-2}[\vartheta] = 1 - \sum_{j=1}^{k-1} |\langle \varphi_k, \varphi_j^n \rangle_n|^2[\vartheta] = c_k^{-2}[0]$$

and thus the induction step is achieved by calculating

$$\begin{aligned}
&\langle \varphi_{k'}, \varphi_k^n \rangle_n[\vartheta] \\
&\quad = \left\langle \varphi_{k'}, c_k\left(\varphi_k - \sum_{r=1}^{k-1} \langle \varphi_k, \varphi_r^n \rangle \varphi_r^n\right)\right\rangle_n [\vartheta] \\
&\quad = c_k\left(\langle \varphi_{k'}, \varphi_k \rangle_n - \sum_{r=1}^{k-1} \langle \varphi_{k'}, \varphi_r^n \rangle_n \overline{\langle \varphi_k, \varphi_r^n \rangle_n}\right)[\vartheta] \\
&\quad = c_k[\vartheta]\left(A_{\ell(k')-\ell(k)}[\vartheta] - \sum_{r=1}^{k-1} e^{2\pi\iota\langle \ell(k')-\ell(k),\vartheta\rangle} \langle \varphi_{k'}, \varphi_r^n \rangle_n[0] \overline{\langle \varphi_k, \varphi_r^n \rangle_n[0]}\right) \\
&\quad = e^{2\pi\iota\langle \ell(k')-\ell(k),\vartheta\rangle} \langle \varphi_{k'}, \varphi_k^n \rangle_n[0]. \qquad \square
\end{aligned}$$

PROPOSITION 4.6. *Suppose $g = \sum_{|\ell|_{\ell^2} \le L} \gamma_\ell e^{2\pi\iota\langle \ell, \bullet\rangle}$ is a d-dimensional trigonometric polynomial of degree L. Let $\Delta \in (0, L^{-1}]$ with $1/\Delta \in \mathbb{N}$ be given and define the cubes $C_m := \prod_{i=1}^d [(m_i - 1)\Delta, m_i\Delta)$. Then*

$$\Delta^d \sum_{m \in \{1,\dots,\Delta^{-1}\}^d} \sup_{x_m \in C_m} \big| |g(x_m)|^2 - |g(m\Delta)|^2 \big| \le \|g\|_{L^2}^2 (e^{2d\Delta L} - 1).$$

PROOF. We need multi-indices $\alpha, \beta \in \mathbb{N}_0^d$ with $\alpha! := \alpha_1! \cdots \alpha_d!$, $x^\alpha := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, $\binom{\alpha}{\beta} := \frac{\alpha!}{\beta!(\alpha-\beta)!}$ and differential operators $D^\alpha := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_1}}$. Since $|g|^2$ is real-analytic, a power series expansion gives, for any $x_m \in C_m$,

$$\begin{aligned}
&\big| |g|^2(x_m) - |g|^2(m\Delta) \big| \\
&\qquad = \left| \sum_{\alpha \in \mathbb{N}_0^d, \alpha \ne 0} D^\alpha |g|^2(m\Delta) \frac{(x_m - m\Delta)^\alpha}{\alpha!} \right| \\
&\qquad \le \sum_{\alpha \in \mathbb{N}_0^d, \alpha \ne 0} \frac{\Delta^{|\alpha|_{\ell^1}}}{\alpha!} \sum_{\beta \in \mathbb{N}_0^d, \beta \le \alpha} \binom{\alpha}{\beta} |D^\beta g(m\Delta)| |D^{\alpha-\beta} \bar{g}(m\Delta)|.
\end{aligned}$$

Together with g, any derivative is again a trigonometric polynomial of degree L and by the isometry (2.3) and Bernstein's inequality [cf. Meyer (1995), page 32], we obtain

$$\Delta^d \sum_{m \in \{1,\dots,\Delta^{-1}\}^d} |D^\alpha g|^2(m\Delta) = \|D^\alpha g\|_{L^2}^2 \le L^{2|\alpha|_{\ell^1}} \|g\|_{L^2}^2.$$

This implies, by the Cauchy–Schwarz inequality, that

$$
\begin{aligned}
\Delta^d & \sum_{m\in\{1,\dots,\Delta^{-1}\}^d} \sup_{x_m\in C_m} \big||g|^2(x_m)-|g|^2(m\Delta)\big| \\
& \le \Delta^d \sum_{\alpha\in\mathbb{N}^d,\alpha\neq 0} \frac{\Delta^{|\alpha|_{\ell^1}}}{\alpha!} \sum_{\beta\in\mathbb{N}^d,\beta\le\alpha} \binom{\alpha}{\beta} \left(\sum_m |D^\beta g(m\Delta)|^2\right)^{1/2} \\
& \qquad \times \left(\sum_m |D^{\alpha-\beta}\bar g(m\Delta)|^2\right)^{1/2} \\
& \le \|g\|_{L^2}^2 \sum_{\alpha\in\mathbb{N}^d,\alpha\neq 0} \frac{\Delta^{|\alpha|_{\ell^1}}}{\alpha!} \sum_{\beta\in\mathbb{N}^d,\beta\le\alpha} \binom{\alpha}{\beta} L^{|\beta|_{\ell^1}} L^{|\alpha-\beta|_{\ell^1}} \\
& = \|g\|_{L^2}^2 (e^{2d\Delta L}-1). \qquad \square
\end{aligned}
$$

LEMMA 4.7. *Let $Y\in\mathbb{R}^r$ follow the multinomial distribution with parameters n and $p_1=\cdots=p_r=1/r$. Then, for $n\to\infty$ and $r=r(n)$ with $r\log(r)/n\to 0$,*

$$
\begin{aligned}
\forall C>0:& \limsup_{n\to\infty} \tfrac{1}{4} r(n)^{C^2/4-1} \\
& \times P\left(\max_{1\le i\le r(n)} |Y_i - n/r(n)| > C\sqrt{n\log(r(n))/r(n)}\right) \le 1.
\end{aligned}
$$

PROOF. If $X_1,\dots,X_r$ are independently Poisson(n/r)-distributed, then it is well known that the law of $(X_1,\dots,X_r)$ given $\sum_{i=1}^r X_i = n$ is multinomial with parameters n and $p_1=\cdots=p_r=1/r$. Set $A_{nr} := C\sqrt{n\log(r)/r}$. Since

$$
k\mapsto P\left(\max_{1\le i\le r} X_i - n/r > A_{nr} \,\Big|\, \sum_{i=1}^r X_i = k\right)
$$

is obviously increasing in $k\in\mathbb{N}$, we obtain

$$
P\left(\max_{1\le i\le r} X_i - n/r > A_{nr} \,\Big|\, \sum_{i=1}^r X_i = n\right) \le \frac{P(\max_{1\le i\le r} X_i - n/r > A_{nr})}{P(\sum_{i=1}^r X_i \ge n)}.
$$

As $\sum_{i=1}^r X_i$ is Poisson(n)-distributed, $P(\sum_{i=1}^r X_i \ge n)\to 1/2$ follows, so

$$
\limsup_{n\to\infty}\left(P\left(\max_{1\le i\le r} Y_i - n/r > A_{nr}\right) - 2P\left(\max_{1\le i\le r} X_i - n/r > A_{nr}\right)\right) \le 0. \tag{4.6}
$$

By the exponential moment estimate $E[e^{a(X_i-n/r)}] = e^{n(e^a-a-1)/r} \le e^{3na^2/4r}$ for $a := rA_{nr}/n \to 0$ and n large, the generalized Markov inequality yields

$$
P\left(\max_{1\le i\le r} X_i - n/r > A_{nr}\right) \le rP(X_i - n/r > A_{nr}) \le re^{3na^2/4r - aA_{nr}} = r^{1-C^2/4}.
$$

By use of (4.6) and a completely symmetric argument for $P(\max_{1\le i\le r}(n/r - X_i) > A_{nr})$, the result follows. $\square$

PROPOSITION 4.8. *For $j = j(n)$ such that $j\log(j) = o(n)$ and the event Ω_j^n in* (4.2), *we have* $\lim_{n\to\infty} n^p P((\Omega_{j(n)}^n)^{\complement}) = 0$ *for any power $p > 0$.*

PROOF. From Proposition 4.6, we derive, with $\Delta \le L := |\ell(j)|_{\ell^2}$, $1/\Delta \in \mathbb{N}$, the cubes $C_m := \prod_{i=1}^d [(m_i - 1)\Delta, m_i\Delta)$ and the occupations $N_m := \#\{i : X_i \in C_m\}$,

$$\begin{aligned}
&\left| \|g\|_{L^2}^2 - \frac{1}{n}\sum_{i=1}^n |g(X_i)|^2 \right| \\
&\quad = \frac{1}{n}\left| \sum_{m\in\{1,\dots,\Delta^{-1}\}^d} \left(\Delta^d n |g(m\Delta)|^2 - \sum_{i:X_i\in C_m} |g(X_i)|^2 \right) \right| \\
&\quad \le \frac{1}{n} \sum_{m\in\{1,\dots,\Delta^{-1}\}^d} \big(|\Delta^d n - N_m| |g(m\Delta)|^2 \\
&\qquad\qquad + N_m \sup_{x_m\in C_m} \big| |g(m\Delta)|^2 - |g(x_m)|^2 \big| \big) \\
&\quad \le \frac{\|g\|_{L^2}^2}{\Delta^d n} \max_{m\in\{1,\dots,\Delta^{-1}\}^d} \big(|\Delta^d n - N_m| + N_m (e^{2d\Delta L} - 1) \big) \\
&\quad \le \|g\|_{L^2}^2 \left(e^{2d\Delta L} \max_{m\in\{1,\dots,\Delta^{-1}\}^d} |1 - N_m/n\Delta^d| + (e^{2d\Delta L} - 1) \right).
\end{aligned}$$

By Lemma 4.7, $\max_m |1 - N_m/n\Delta^d|^2 \ge C(n\Delta^d)^{-1}\log(1/\Delta)$ has probability tending to zero with any given polynomial rate when C is chosen sufficiently large. Since $L^d \log(L) \lesssim j\log(j) = o(n)$, we can choose $\Delta = o(L^{-1})$ such that $\Delta^{-d}\log^2(1/\Delta) = o(n)$ still holds. This gives

$$\big| \|g\|_{L^2}^2 - \|g\|_n^2 \big| \le \big(Ce^{2d\Delta L}(n\Delta^d)^{-1}\log(1/\Delta) + (e^{2d\Delta L} - 1) \big) \|g\|_{L^2}^2 \le \tfrac{3}{4}\|g\|_{L^2}^2$$

for large n with probability larger than $1 - n^{-p}$. $\square$

PROPOSITION 4.9. *For $j \in \mathbb{N}$ with $j\log(j) = o(n)$, we have*

$$E[\|P_{j-1}^n \varphi_j\|_n^2] \lesssim j/n.$$

PROOF. By construction, $\|P_{j-1}^n\varphi_j\|_n^2 \le \|\varphi_j\|_n^2 = 1$ holds so that by Proposition 4.8, it suffices to find the bound for the expectation on the event Ω_j^n.

Setting $A_m := \frac{1}{n}\sum_{k=1}^m \exp(2\pi\iota\langle m, X_k\rangle)$, $m \in \mathbb{Z}^d$, we use Parseval's identity and $E[|A_m|^2] = 1/n$ for $m \neq 0$ to obtain

$$\begin{aligned}
&E[\|P_{j-1}^n \varphi_j\|_n^2 \mathbf{1}_{\Omega_j^n}] \\
&\quad = E\biggl[\sup_{g\in V_{j-1}} \frac{|\langle \varphi_j, g\rangle_n|^2}{\|g\|_n^2}\mathbf{1}_{\Omega_j^n}\biggr] \\
&\quad \le E\Biggl[\sup_{\|(c_r)\|_{\ell^2}=1}\Biggl|\frac{1}{n}\sum_{k=1}^n\sum_{r=1}^{j-1}\bar{c}_r e^{2\pi\iota\langle \ell(j)-\ell(r), X_k\rangle}\Biggr|^2 \sup_{g\in V_{j-1}}\frac{\|g\|_{L^2}^2}{\|g\|_n^2}\mathbf{1}_{\Omega_j^n}\Biggr] \\
&\quad \le 4E\Biggl[\sum_{r=1}^{j-1}\bigl|A_{\ell(j)-\ell(r)}\bigr|^2\Biggr] = \frac{4(j-1)}{n}. \qquad \square
\end{aligned}$$

REFERENCES

BASS, R. F. and GRÖCHENIG, K. (2004). Random sampling of multivariate trigonometric polynomials. *SIAM J. Math. Anal.* **36** 773–795. MR2111915

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

BROWN, L. D., CARTER, A. V., LOW, M. G. and ZHANG, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32** 2074–2097. MR2102503

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

BROWN, L. D. and ZHANG, C.-H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Ann. Statist.* **26** 279–287. MR1611772

CAI, T. and BROWN, L. D. (1999). Wavelet estimation for samples with random uniform design. *Statist. Probab. Lett.* **42** 313–321. MR1688134

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

COHEN, A. (2000). Wavelet methods in numerical analysis. In *Handbook of Numerical Analysis* (P. G. Ciarlet, ed.) **7** 417–711. North-Holland, Amsterdam. MR1804747

COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. MR1256527

DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451

DE BOOR, C. (2001). *A Practical Guide to Splines*, rev. ed. Springer, New York. MR1900298

DONOHO, D. L. and JOHNSTONE, I. M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statist. Sinica* **9** 1–32. MR1678879

GAIFFAS, S. (2007). Sharp estimation in sup norm with random design. *Statist. Probab. Lett.* **77** 782–794. MR2369683

GRAMA, I. and NUSSBAUM, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields* **111** 167–214. MR1633574

HOFFMANN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression (with discussion). *Ann. Statist.* **30** 325–396. MR1902892

JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Boundary coiflets for wavelet shrinkage in function estimation. *J. Appl. Probab.* **41A** 81–98. MR2057567

LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics. Some Basic Concepts*, 2nd ed. Springer, New York. MR1784901

MEYER, Y. (1995). *Wavelets and Operators.* Cambridge Univ. Press. MR1228209

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

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

SWELDENS, W. and PIESSENS, R. (1993). Wavelet sampling techniques. In *1993 Proceedings of the Statistical Computing Section* 20–29. Amer. Statist. Assoc.

WOJTASZCZYK, P. (1997). *A Mathematical Introduction to Wavelets.* Cambridge Univ. Press. MR1436437

INSTITUTE OF APPLIED MATHEMATICS
UNIVERSITY OF HEIDELBERG
IM NEUENHEIMER FELD 294
HEIDELBERG
GERMANY
E-MAIL: reiss@statlab.uni-heidelberg.de