SPATIAL ADAPTION FOR PREDICTING RANDOM FUNCTIONS

THOMAS MÜLLER-GRONBACH AND KLAUS RITTER

Abstract. We study integration and reconstruction of Gaussian random functions with inhomogeneous local smoothness. A single realization may only be observed at a finite sampling design and the correct local smoothness is unknown. We construct adaptive (sequential) designs that lead to asymptotically optimal methods. We show that any nonadaptive design will fail.

1. Introduction

Various problems of prediction from correlated data are studied in the literature. Often the underlying stochastic model is a Gaussian random function $Y(t)$, $t \in D$, where $D \subset \mathbb{R}^d$. We are interested in the case when discrete observations

$$Y(t_1), \ldots, Y(t_n)$$

of a realization of Y are used to predict a quantity which depends linearly on Y. The main examples are

(a) prediction of $Y(t)$ at an untried point $t \notin \{t_1, \ldots, t_n\}$,
(b) prediction of the integral $\int_D Y(t)\,dt$, called *integration*,
(c) prediction of $Y(t)$ for all $t \in D$, called *reconstruction*.

Given a fixed set of observation points t_i, one is interested in the best predictor with respect to some criterion, e.g., error in quadratic mean. This question is closely related to the theory of splines, see Kimeldorf and Wahba (1970), Traub, Wasilkowski, and Woźniakowski (1988), and Wahba (1990).

Even more challenging seems to be how to choose the points t_i optimally (except in (a)), if only their total number n is given. This task is also called the design problem, and it is only solved for few specific processes Y in the univariate case $D = [0, 1]$. Hence the optimality conditions are weakened and asymptotically optimal designs are sought.

In this paper we study the univariate case. We present a new framework for analyzing integration and reconstruction in case of an *unknown* mean and covariance kernel of Y. We construct an adaptive method that is asymptotically optimal for a large class of processes Y having inhomogeneous local smoothness.

Our new approach is motivated by several applications. Problems (a)–(c) arise, for instance, in geostatistics and in computer experiments, see Cressie (1993) and Hjort and Omre (1994), as well as Sacks, Welch, Mitchell, and Wynn (1989), Koehler and Owen (1995), Bates, Buck, Riccomagno, and Wynn (1996). Moreover, the random function approach is used in numerical analysis to complement the classical worst case approach,

Date: April 28, 1997.

see Novak (1988), Traub, Wasilkowski, and Woźniakowski (1988), Ritter (1996a), and Plaskota (1996).

Typically, the second order properties of Y are not known precisely in geostatistical applications or in computer experiments. Therefore *parametric* assumptions on the mean m and the covariance kernel K of Y are frequently used. The observations $Y(t_i)$ are used to estimate the parameters, and thereafter predictors are constructed on the basis of the estimated second order structure. We stress that only the observations of a single realization are at hand for parameter estimation and prediction.

Usually a properly chosen design $(t_1, \ldots, t_n) \in D^n$ is fixed in advance. Inference about K and m from the observations $Y(t_i)$ is only used in the final stage, when the prediction is generated. Any method that is based on an a priori fixed design is called *nonadaptive*; the design is called nonadaptive as well.

In this paper we analyze *adaptive* methods. The points t_i are chosen sequentially and therefore t_i may depend on the $Y(t_1), \ldots, Y(t_{i-1})$. In this way inference about m and K is already used in the observation stage. An adaptive method is based on an adaptive design, which is of the form

$$t_1,\ t_2(Y(t_1)),\ \ldots,\ t_n(Y(t_1), \ldots Y(t_{n-1}(\ldots))).$$

Furthermore we propose a *nonparametric* approach regarding K and m, and we do not choose a specific type of covariance kernel in advance. Instead, we assume that Y is of the form

$$Y(t) = m(t) + g(t) \cdot X(f(t)), \qquad t \in D. \tag{1}$$

Here m, f, and g are deterministic functions and X is a zero mean Gaussian random function. The mean m, the transformations f and g, and the covariance kernel R of X are unknown; only some smoothness conditions are assumed to hold.

The nonparametric model (1) does not predetermine the spatial variability of Y. We present an adaptive method which detects regions of different variability and places additional points accordingly. We prove that our method is asymptotically optimal. Furthermore we show that any nonadaptive method will fail.

The design problem for processes of the form (1) is analyzed in numerous papers. However, at least the functions f and g are assumed to be known, and therefore adaptation does not help. We give a brief survey of results.

It turns out that asymptotically optimal designs $(t_1^{(n)}, \ldots, t_n^{(n)})$ can be constructed as quantiles of a suitably chosen density ψ on $[0, 1]$. The optimal density ψ depends on f and g and on the smoothness of X. See, e.g., Sacks and Ylvisaker (1966, 1970), Hájek and Kimeldorf (1974), Wahba (1974), Eubank, Smith, and Smith (1982), Benhenni and Cambanis (1992), Istas and Laredo (1994), Stein (1995a), and Ritter (1996b) for integration. Reconstruction is analyzed in Speckman (1979), Su and Cambanis (1993), Müller-Gronbach (1996a, 1996b, 1996c), Ritter (1996a), and Müller-Gronbach and Ritter (1997). The (asymptotically) optimal designs are nonadaptive. Once f, g, and the smoothness of X are specified and n is selected, the design is fixed and does not depend on any observation of Y.

The above results serve as benchmarks for the nonparametric model (1) with unknown functions m, f, and g. We will demonstrate that asymptotically the same errors are achievable. To this end properly chosen adaptive designs must be used.

Much less is known in the multivariate case $D = [0,1]^d$ with $d > 1$. In fact, only order optimal designs are known, while finding the best asymptotic constants seems to be an open problem. See, e.g., Woźniakowski (1991, 1992), Wasilkowski (1993), Paskov (1993), Ritter (1996a), Ritter, Wasilkowski, and Woźniakowski (1995), and Ritter and Wasilkowski (1996).

For reasonable subclasses of designs, however, the asymptotically optimal ones are known for certain multivariate problems. Stein (1995b) studies integration for isotropic random fields and he analyzes designs which locally form parallelogram lattices. Müller-Gronbach (1997) analyzes reconstruction of random fields Y with separable covariance kernels, and he studies hyperbolic cross designs. The particular case of a non Gaussian random field with product structure is analyzed in Müller-Gronbach and Schwabe (1996), In each of these cases the “density” of a design may vary over D. The optimal “density”, which depends on f, g and on the smoothness of X, is found. The corresponding designs are nonadaptive.

In a forthcoming paper we construct adaptive hyperbolic cross designs for the nonparametric model (1) with separable covariance kernel. These designs perform asymptotically as well as the optimal nonadaptive ones that require complete knowledge of f and g.

In the following section we specify the smoothness of the random function X and the deterministic functions m, f, and g in (1). Our adaptive method is defined in Section 3, and Section 4 contains the results and some remarks. In Section 5 we compare the adaptive designs with equidistant designs by means of a simulation. Proofs are given in Section 6.

2. Smoothness Assumptions

Let X denote a zero mean Gaussian random function on

$$D = [0,1]$$

and let R denote the covariance kernel of X, i.e.,

$$R(s,t) = E(X(s) \cdot X(t)), \qquad s,t \in D.$$

Suppose that X is at least Hölder continuous in quadratic mean for some positive exponent. Then we may assume that X is a measurable random function whose realization are continuous with probability 1. The same properties hold for the random function Y defined by (1), if m, f, and g are at least continuous and $f(D) \subset D$.

In geostatistics and computer experiments quite often processes Y are considered that are precisely Hölder continuous in quadratic mean with exponent $1/2$. This means that

$$\begin{gathered}\Big(E(Y(s)-Y(t))^2\Big)^{1/2} \le c\cdot|s-t|^{1/2},\\ \liminf_{s\to t}\frac{(E(Y(s)-Y(t))^2)^{1/2}}{|s-t|^{1/2}}>0\end{gathered}\tag{2}$$

for $s,t\in D$ with a positive constant c. We will mainly analyze this case, but we discuss methods for other degrees of smoothness in Remark 4. Note that the multivariate analogue of (2) is satisfied by random fields with linear, spherical, and exponential variograms. See Cressie (1993, p. 61) for the respective definitions.

Property (2) is too weak for analysis, in particular for proving lower bounds. Hence we use slightly stronger conditions that were introduced by Sacks and Ylvisaker (1966) and thereafter studied in many papers, some of which are cited in the Introduction.

Regularity in quadratic mean of X is specified by the regularity of its covariance kernel R at the diagonal in D^2. We denote one-sided limits at the diagonal in the following way. Let

$$\Omega_+ = \{(s,t)\in\,]0,1[^2 : s>t\}, \qquad \Omega_- = \{(s,t)\in\,]0,1[^2 : s<t\},$$

and let $\operatorname{cl} A$ denote the closure of a set A. Suppose that L is a continuous function on $\Omega_+\cup\Omega_-$ such that $L|_{\Omega_j}$ is continuously extendable to $\operatorname{cl}\Omega_j$ for $j\in\{+,-\}$. By L_j we denote the extension of L to $[0,1]^2$ which is continuous on $\operatorname{cl}\Omega_j$ and on $[0,1]^2\setminus\operatorname{cl}\Omega_j$.

The covariance kernel R of X satisfies the Sacks-Ylvisaker conditions if the following three conditions hold.

(A) *R is continuous on $[0,1]^2$, the partial derivatives of R up to order two are continuous on $\Omega_+\cup\Omega_-$ and continuously extendable to $\operatorname{cl}\Omega_+$ as well as to $\operatorname{cl}\Omega_-$.*

(B) $$R_-^{(1,0)}(t,t)-R_+^{(1,0)}(t,t)=1, \qquad 0\le t\le 1.$$

(C) *$R_+^{(2,0)}(t,\cdot)\in H(R)$ for all $0\le t\le 1$ and*

$$\sup_{0\le t\le 1}\|R_+^{(2,0)}(t,\cdot)\|_R<\infty.$$

Here $H(R)$ denotes the Hilbert space with reproducing kernel R; the corresponding inner product and norm are denoted by $\langle\cdot,\cdot\rangle_R$ and $\|\cdot\|_R$. Conditions (A) and (B) imply

$$\lim_{s\to t}\frac{(E(X(s)-X(t))^2)^{1/2}}{|s-t|^{1/2}}=1 \tag{3}$$

for every $t\in[0,1]$.

Let m, g, and f denote deterministic, real-valued functions on $[0,1]$. We assume that the following conditions hold.

(D) *m is continuously differentiable,*

(E) *g is positive and continuously differentiable,*

(F) *$f([0,1])\subset[0,1]$ and f is differentiable with $f'>0$ being positive and Lipschitz continuous.*

Consider a stochastic process Y of the form (1), and assume that the conditions (A)–(F) hold. It is easily checked that the local Hölder constant for Hölder exponent 1/2 is given by

$$\lim_{s\to t} \frac{\left(E(Y(s)-Y(t))^2\right)^{1/2}}{|s-t|^{1/2}} = \alpha(t)$$

with

$$\alpha(t) = g(t)\cdot\sqrt{f'(t)}. \tag{4}$$

See Lemma 1 for a more detailed estimate.

The process Y is of inhomogeneous local smoothness, in contrast to X, see (3). An example showing three realizations of a process Y which satisfies our assumptions is given in Figure 1. Here

$$\begin{aligned} R(s,t) &= (1-|s-t|)/2,\\ m(t) &= 2t,\\ g(t) &= 1/2+5t^2,\\ f(t) &= (\tanh(15t-5)+1)/2 \end{aligned} \tag{5}$$

The corresponding function α is shown in Figure 2.

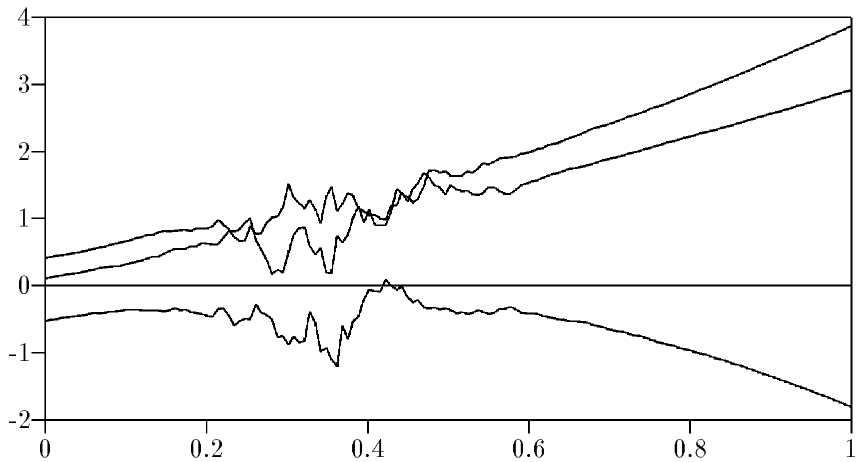

FIGURE 1. Three realizations of a process Y

3. The Adaptive Method

Let $\widehat{Y}_n$ denote any method for reconstruction that is based on n observations. Formally $\widehat{Y}_n$ is defined by a fixed point $t_1 \in D$, and by measurable mappings $\chi_i : \mathbb{R}^{i-1} \to D$ and $\phi : \mathbb{R}^n \to C(D)$. We have

$$\widehat{Y}_n = \phi(y_1,\ldots,y_n)$$

where

$$y_1 = Y(t_1)$$

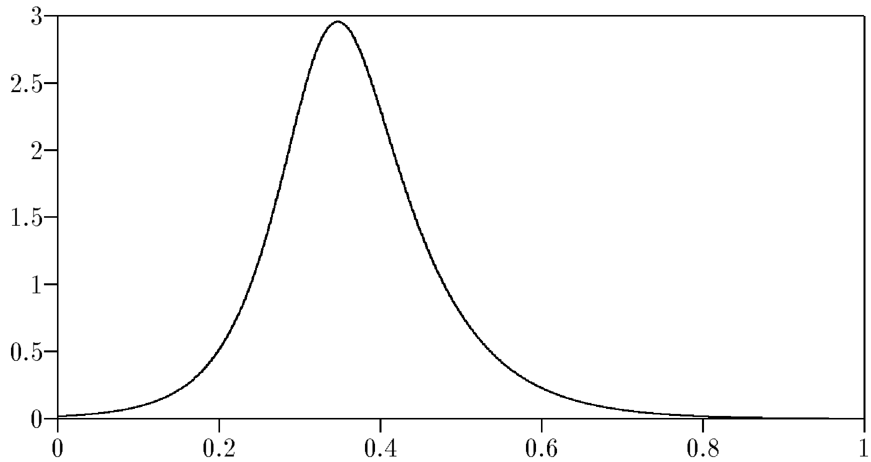

FIGURE 2. The corresponding function α

and

$$y_i = Y(\chi_i(y_1, \ldots, y_{i-1})), \qquad 2 \le i \le n.$$

The i-th observation is made at $t_i = \chi_i(y_1, \ldots, y_{i-1})$ and yields the value y_i. The method $\widehat{Y}_n$ is called nonadaptive if all mappings χ_i are constant; otherwise it is called adaptive.

A method is nonadaptive, by definition, if all observations could me made in parallel. Otherwise, if any kind of sequential observation is needed, the method is adaptive. Hence methods are classified by the way in which the data are collected. The structure of ϕ is irrelevant for this classification. Let us mention that the notion of adaptivity is sometimes used in a different sense, see, e.g., Donoho and Johnstone (1994) and Hall and Patil (1996).

Let $\|\cdot\|_p$ denote the L_p-norm on D, i.e.,

$$\|h\|_p = \left(\int_D |h(t)|^p \, dt\right)^{1/p}$$

if $1 \le p < \infty$ and

$$\|h\|_\infty = \sup_{t \in D} |h(t)|.$$

We measure the distance between a realization of Y and the corresponding prediction in L_p-norm. The error of $\widehat{Y}_n$ is defined by

$$e_q(\widehat{Y}_n, L_p) = \left(E\|Y - \widehat{Y}_n\|_p^q\right)^{1/q}$$

where $1 \le q < \infty$. For finite p it is most convenient to study the case $p = q$.

For integration the definitions are analogous, except that ϕ has range $\mathbb{R}$ and the error is defined by

$$e_q(\widehat{Y}_n, \mathrm{Int}) = \left(E\left|\int_D Y(t)\, dt - \widehat{Y}_n\right|^q\right)^{1/q}.$$

Now we describe our method for reconstruction and integration. Basically it works as follows. In the first stage, which is nonadaptive, we use a small number of observations to estimate a certain power α^λ of the local Hölder constant α, see (4). We take

$$\lambda = \begin{cases} (1/2 + 1/p)^{-1} & \text{for reconstruction in } L_p\text{-norm}, \\ 2/3 & \text{for integration.} \end{cases} \tag{6}$$

Then we select additional points adaptively with "density" proportional to the estimate $\widehat{\alpha}$ of α^λ. Finally, we use piecewise linear interpolation of the whole observations for reconstruction. Analogously, we use a trapezoidal rule for integration.

The nonadaptive part is determined by an integer

$$k \in \mathbb{N}$$

and a real number

$$0 < \delta < 1/k^2. \tag{7}$$

The corresponding sample points are clustered around the sites $(2i-1)/2k$ with distance δ within each cluster. More precisely,

$$t_{i,j} = \frac{2i-1}{2k} - \frac{k\delta}{2} + j\delta$$

where

$$i = 1, \ldots, k, \qquad j = 0, \ldots, k.$$

Additionally, we use the points

$$t_{0,k} = 0, \qquad t_{k+1,0} = 1.$$

The resulting nonadaptive design

$$T = (t_{0,k}, \ldots, t_{k+1,0}) \tag{8}$$

consists of $k \cdot (k+1) + 2$ sample points. An example is given in Figure 3.

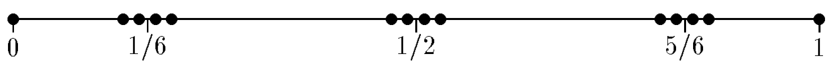

FIGURE 3. Nonadaptive design with $k = 3$, $\delta = 1/50$

Sampling Y at T yields data

$$y = (y_{0,k}, \ldots, y_{k+1,0}),$$

which are used for estimating the function α^λ. The differences $Y(t_{i,j}) - Y(t_{i,j-1})$ are normally distributed and weakly correlated with mean close to zero and second moment close to $\alpha((2i-1)/2k)^2 \cdot \delta$, see Lemma 1. Therefore, a natural choice for estimating α^λ at $(2i-1)/2k$ is

$$\widehat{\alpha}_i(y) = (c_\lambda\, \delta^{\lambda/2})^{-1} \cdot \frac{1}{k} \sum_{j=1}^{k} |y_{i,j} - y_{i,j-1}|^\lambda,$$

where c_λ denotes the absolute moment of order λ of the standard normal distribution. In fact, Lemma 3 shows that these estimates work well under the assumptions imposed on the process Y. However, taking care of $\widehat{\alpha}_i(y)$ getting to small we will use

$$\widetilde{\alpha}_i(y) = \max(\widehat{\alpha}_i(y), \varepsilon)$$

instead, where

$$0 < \varepsilon \leq 1.$$

Let us mention that the constant $(c_\lambda\, \delta^{\lambda/2}\, k)^{-1}$ is present in the definition of $\widehat{\alpha}_i$ only for mathematical convenience.

In the second stage additional points $s_{i,j}(y)$ are adaptively placed in the subintervals $J_i = [t_{i,k}, t_{i+1,0}]$. These points are determined by the values

$$\widetilde{\alpha}_1(y), \dots, \widetilde{\alpha}_k(y)$$

together with an integer

$$n \in \mathbb{N},$$

which is roughly the total number of points $s_{i,j}(y)$. We estimate α^λ on $J_0 \cup \dots \cup J_k$ by piecewise linear interpolation of the values $\widetilde{\alpha}_i(y)$, and we use this estimate to construct an adaptive design. See Figure 4 for an example.

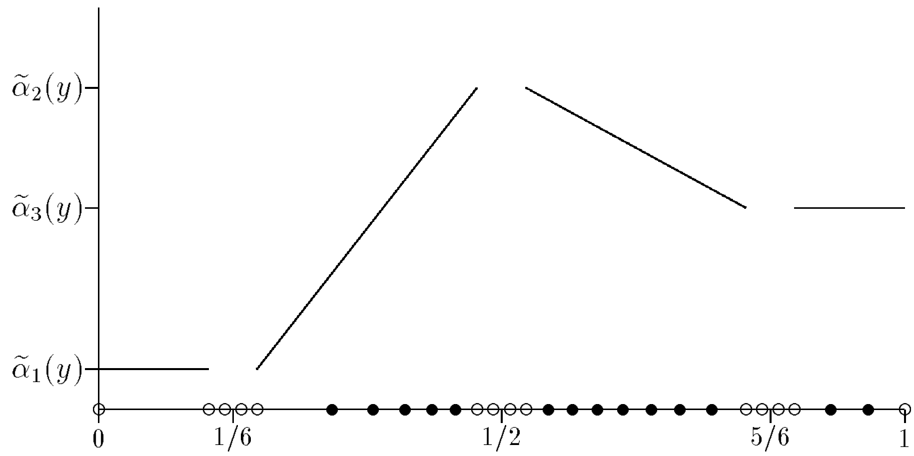

FIGURE 4. Estimate for α, adaptive design with $k = 3$, $\delta = 1/50$, and $n = 16$

More precisely,

$$r_i(y) = \left\lfloor n \cdot \frac{\widetilde{\alpha}_i(y) + \widetilde{\alpha}_{i+1}(y)}{2 \sum_{j=1}^{k} \widetilde{\alpha}_j(y)} \right\rfloor$$

is the number of points which are placed in the subinterval J_i for $i = 1, \dots, k-1$. Within each of these subintervals the spacing is proportional to the linear density with boundary values $\widetilde{\alpha}_i(y)$ and $\widetilde{\alpha}_{i+1}(y)$. Formally,

$$\int_{t_{i,k}}^{s_{i,j}(y)} (\widetilde{\alpha}_i(y)\,(t_{i+1,0} - t) + \widetilde{\alpha}_{i+1}(y)\,(t - t_{i,k}))\, dt = j \cdot \frac{(\widetilde{\alpha}_i(y) + \widetilde{\alpha}_{i+1}(y)) \cdot (1/k - k\delta)^2}{2\,(r_i(y) + 1)}$$

for $j = 1, \ldots, r_i(y)$. We use equidistant points in the subintervals J_0 and J_k. The respective numbers of points are given by

$$r_0(y) = \left\lfloor n \cdot \frac{\tilde{\alpha}_1(y)}{2 \sum_{j=1}^{k} \tilde{\alpha}_j(y)} \right\rfloor$$

and

$$r_k(y) = \left\lfloor n \cdot \frac{\tilde{\alpha}_k(y)}{2 \sum_{j=1}^{k} \tilde{\alpha}_j(y)} \right\rfloor .$$

Therefore

$$s_{0,j}(y) = j \cdot \frac{1/k - k\delta}{2\,(r_0(y)+1)}$$

with $j = 1, \ldots, r_0(y)$ and

$$s_{k,j}(y) = 1 - (r_k(y) + 1 - j) \cdot \frac{1/k - k\delta}{2\,(r_k(y)+1)}$$

with $j = 1, \ldots, r_k(y)$.

Summarizing, the first stage depends on the parameters k and δ, and the second stage depends on the parameters ε and n, as well as on λ that is defined in (6). The whole design consists of

$$N(y) \leq n + k \cdot (k+1) + 2 =: N$$

points. For reconstruction we use the piecewise linear interpolation of the whole observations from stage one and two. For integration we apply the trapezoidal rule to these data. We denote our adaptive method by $\widehat{Y}_N^\lambda$.

In the sequel we use $\approx$ to denote the strong equivalence of sequences of real numbers a_n and b_n. By definition,

$$a_n \approx b_n \qquad \text{iff} \qquad \lim_{n\to\infty} a_n/b_n = 1.$$

In order to obtain asymptotic results we study sequences of methods $\widehat{Y}_{N_n}^\lambda$ which are defined by sequences of the respective parameters k_n, δ_n, and ε_n. There are many ways to adjust the parameters such that the correct error rate is achieved. The approach taken here is based on the following considerations. Clearly, the number $k_n(k_n+1)+2$ of nonadaptive points should be small compared to the number of points chosen adaptively in the second stage. On the other hand, k_n should be sufficiently large in order to obtain a good estimate of α^λ even for a small sample size. Hence we take

$$k_n = \lceil n^{1/2-\gamma} \rceil, \tag{i}$$

where

$$0 < \gamma < 1/2.$$

Hereby the maximum size N_n of the adaptive design satisfies

$$N_n \approx n.$$

The parameters δ_n must satisfy

$$\delta_n \leq c \cdot 1/n \tag{ii}$$

with a constant $c > 0$. This guarantees that each subinterval $[t_{i,0}, t_{i,k}]$ is sufficiently small; its length is of order at most $n^{-(1/2-\gamma)}$.

Finally, we adjust the truncation parameter ε_n according to the quality of our estimate of α^λ. Due to Lemma 3 the corresponding error is of order at most $n^{-(1/2-\gamma)/2}$. Therefore we take

$$\varepsilon_n = n^{-\tau}, \tag{iii}$$

where

$$0 < \tau < (1/2 - \gamma)/2.$$

4. Results and Remarks

Let

$$\nu_p = \int_0^1 (t\,(1-t))^{p/2}\,dt = \frac{\Gamma^2(p/2+1)}{\Gamma(p+2)}$$

and recall that

$$c_p = (2\pi)^{-1/2} \int_{\mathbb{R}} |t|^p \cdot \exp(-t^2/2)\,dt.$$

We extend the definition of $\|\cdot\|_p$ to $0 < p \leq \infty$, as usual. The error of our method is given as follows.

Theorem 1. *Let Y denote any process of the form* (1) *such that the conditions (A)–(F) hold. Let α be given by* (4). *If λ is selected according to* (6) *and (i)–(iii) are satisfied, then*

$$e_2(\widehat{Y}^\lambda_{N_n}, \mathrm{Int}) \approx \frac{1}{2\sqrt{3}} \cdot \|\alpha\|_\lambda \cdot n^{-1},$$
$$e_p(\widehat{Y}^\lambda_{N_n}, L_p) \approx (c_p\,\nu_p)^{1/p} \cdot \|\alpha\|_\lambda \cdot n^{-1/2}, \qquad 1 \leq p < \infty.$$

In the following two remarks we present results for integration and reconstruction of Y that require the mean m and the transformations f and g to be known. We impose the same kind of smoothness as required in Theorem 1.

Remark 1. The design problem for integration of an arbitrary random function of second order is equivalent to a regression design problem for a linear model with correlated errors, see Sacks and Ylvisaker (1970). Therefore the result of Sacks and Ylvisaker (1966) on regression design yields

$$\inf_{\widehat{Y}_n} e_2(\widehat{Y}_n, \mathrm{Int}) \approx \frac{1}{2\sqrt{3}} \cdot \|\alpha\|_{2/3} \cdot n^{-1}.$$

Here the infimum is over all methods that use n observations. Let

$$\psi(t) = \alpha^{2/3}(t) \cdot \|\alpha\|_{2/3}^{2/3}$$

and define the nonadaptive design $(t_1^{(n)}, \ldots, t_n^{(n)}) \in [0,1]^n$ by

$$\int_0^{t_i^{(n)}} \psi(t)\,dt = \frac{i-1}{n-1} \cdot \int_0^1 \psi(t)\,dt, \qquad i = 1, \ldots, n. \tag{9}$$

Then the trapezoidal rules $\widehat{Y}_n$ that are based on $Y(t_1^{(n)}), \ldots, Y(t_n^{(n)})$ are asymptotically optimal, i.e.,

$$e_2(\widehat{Y}_n, \mathrm{Int}) \approx \frac{1}{2\sqrt{3}} \cdot \|\alpha\|_{2/3} \cdot n^{-1}.$$

A sequence of nonadaptive designs $(t_1^{(n)}, \ldots, t_n^{(n)})$ that is defined by (9) for some fixed positive density ψ on $[0,1]$ is called a regular sequence of designs. Note that any design from a regular sequence is nonadaptive.

Remark 2. Consider the reconstruction problem in L_p-norm with $1 \le p < \infty$. The minimal error that is achievable from n observations satisfies

$$\inf_{\widehat{Y}_n} e_p(\widehat{Y}_n, L_p) \approx (c_p\, \nu_p)^{1/p} \cdot \|\alpha\|_\lambda \cdot n^{-1/2},$$

where $\lambda = (1/2+1/p)^{-1}$, as in (6). A regular sequence of designs together with piecewise linear interpolation is asymptotically optimal. The optimal density is given by

$$\psi(t) = \alpha^\lambda(t) \cdot \|\alpha\|_\lambda^\lambda.$$

See Speckman (1979) and Ritter (1996a). The case $p = 2$ is studied in Su and Cambanis (1993) and Müller-Gronbach (1996b).

For L_∞-norm we have

$$\inf_{\widehat{Y}_n} e_q(\widehat{Y}_n, L_\infty) \approx \frac{1}{\sqrt{2}} \cdot \|\alpha\|_2 \cdot (\ln n)^{1/2} n^{-1/2}.$$

Again a regular sequence of designs together with piecewise linear interpolation is asymptotically optimal. The optimal density is given by

$$\psi(t) = \alpha^2(t) \cdot \|\alpha\|_2^2.$$

See Müller-Gronbach and Ritter (1997).

Observe that the exponent λ depends continuously on $p \in [1, \infty]$. For $p = 1$ we have $\lambda = 2/3$ as for integration.

Obviously the results from Remark 1 and 2 motivate the definition of our adaptive method. Furthermore these results yield the lower bounds from Theorem 1.

Corollary 1. *The methods $\widehat{Y}_{N_n}^\lambda$ are asymptotically optimal for every process of the form* (1) *that satisfies (A)–(F).*

Hence we have a *single* sequence of methods that works well for *every* process from some large class, which is defined by smoothness properties. This optimality property cannot be achieved by any sequence of nonadaptive methods $\widehat{Y}_n$, as shown by the following Corollary.

Corollary 2. *Adaption helps for integration and reconstruction of random functions of the form* (1) *with unknown mean m and transformations f and g.*

Proof. Let $k \in \mathbb{N}$ and consider a sequence of nonadaptive methods. There exists a subinterval $I \subset [0,1]$ of length $1/k$ which contains at most $n/k+1$ observation points for infinitely many n. Take, for instance, a Brownian motion X and functions f and g such that $\alpha = 1$ on I and

$$\int_{[0,1]\setminus I} \alpha(t)\, dt = 1/k.$$

For L_2-reconstruction and $\lambda = 1$ we have

$$e_2(\widehat{Y}^{\lambda}_{N_n}, L_2) \approx (6\,n)^{-1/2} \cdot 2/k$$

and

$$\limsup_{n\to\infty} e_2(\widehat{Y}_n, L_2) \cdot (6\,n)^{1/2} \geq k^{-1/2}.$$

Therefore

$$\limsup_{n\to\infty} \frac{e_2(\widehat{Y}_n, L_2)}{e_2(\widehat{Y}^{\lambda}_{N_n}, L_2)} \geq k^{1/2}/2.$$

Clearly analogous estimates hold for L_p-reconstruction with $p \neq 2$ and for integration. □

Remark 3. If f and g are unknown in (1), then the most natural nonadaptive design is an equidistant one. Let $\widehat{Y}^e_n$ denote the optimal method, either for integration or reconstruction, that is based on $t_i^{(n)} = (i-1)/(n-1)$. Then

$$\lim_{n\to\infty} \frac{e_2(\widehat{Y}^{e}_{N_n}, \mathrm{Int})}{e_2(\widehat{Y}^{\lambda}_{N_n}, \mathrm{Int})} = \frac{\|\alpha\|_2}{\|\alpha\|_{2/3}}$$

and

$$\lim_{n\to\infty} \frac{e_p(\widehat{Y}^{e}_{N_n}, L_p)}{e_p(\widehat{Y}^{\lambda}_{N_n}, L_p)} = \frac{\|\alpha\|_p}{\|\alpha\|_{(1/2+1/p)^{-1}}}$$

for $1 \leq p < \infty$. For references concerning the error of $\widehat{Y}^e_n$ see Remarks 1 and 2. Hence equidistant designs are asymptotically optimal only if α is constant. Otherwise their performance may be arbitrarily bad for unfavorable α.

Remark 2 and results from Section 6 and Müller-Gronbach and Ritter (1997) lead to the following conjecture concerning reconstruction in L_∞-norm.

Conjecture . *Let Y denote any process of the form* (1) *such that the conditions (A)–(F) hold. Let α be given by* (4) *and take $\lambda = 2$. Then*

$$e_q(\widehat{Y}^{\lambda}_{N_n}, L_\infty) \approx \frac{1}{\sqrt{2}} \cdot \|\alpha\|_2 \cdot (\ln n)^{1/2} n^{-1/2}.$$

Remark 4. We briefly sketch how to generalize our results to other, and in particular higher degrees of smoothness. Let $r \in \mathbb{N}_0$ and $0 < \beta < 1$, and consider a zero mean Gaussian process X such that

$$\lim_{s\to t} \frac{\left(E(X^{(r)}(s) - X^{(r)}(t))^2\right)^{1/2}}{|s-t|^\beta} = 1$$

for every $t \in [0,1]$. Let Y be defined by (1) with unknown smooth functions m, g, and f. Then

$$\lim_{s \to t} \frac{\left(E(Y^{(r)}(s) - Y^{(r)}(t))^2\right)^{1/2}}{|s-t|^\beta} = \alpha(t)$$

with

$$\alpha(t) = g(t) \cdot (f'(t))^{r+\beta}.$$

We suggest the following modification of our method. Take

$$\lambda = \begin{cases} (r+\beta+1/p)^{-1} & \text{for reconstruction in } L_p\text{-norm,} \\ (r+\beta+1)^{-1} & \text{for integration} \end{cases}$$

and estimate α^λ at equally spaced points $(2i+1)/(2k)$. Use a smooth interpolation of the estimated values and construct an adaptive design with density proportional to this interpolation. We conjecture that these modifications lead to asymptotically optimal methods.

Our conjecture is motivated by results for the case when $m = 0$, $f(t) = t$, and g is known. Then, for some problems, the best regular sequence of designs is known. See Ritter (1996a) for reconstruction with $p < \infty$, $\beta = 1/2$, and $r \in \mathbb{N}_0$. For integration see Stein (1995a) for $r + \beta < 3/2$ and Benhenni and Cambanis (1992) for $\beta = 1/2$ and $r \in \mathbb{N}_0$. In the latter case these regular sequences are already asymptotically optimal, see Ritter (1996b).

Remark 5. In a series of papers Stein studies the effect of a misspecified mean m and/or covariance function K for prediction problems, see Stein (1990). Two second order structures (m_0, K_0) and (m_1, K_1) on D are called compatible if the corresponding Gaussian measures on $\mathbb{R}^D$ are mutually absolutely continuous. Stein analyzes prediction of linear functionals φ of the random field Y, e.g., $\varphi(Y) = Y(t)$ or $\varphi(Y) = \int_D Y(t)\,dt$, on the basis of a fixed sequence of designs T_n that get dense in D. The correct second order structure is assumed to be compatible to the structure (m_1, K_1) that is actually used for prediction. Let $\widehat{Y}_{n,i}$ denote the best linear predictor that is based on T_n and (m_i, K_i). Stein shows that $\widehat{Y}_{n,1}$ is asymptotically efficient, i.e.,

$$\lim_{n \to \infty} \sup_{\varphi} \frac{E(\varphi(Y) - \widehat{Y}_{n,1})^2}{E(\varphi(Y) - \widehat{Y}_{n,0})^2} = 1.$$

Here E denotes expectation under the correct second order structure (m_0, K_0). Furthermore he derives bounds on the rate of convergence.

Almost no additional assumptions on $D \subset \mathbb{R}^d$, (m_0, K_0), and T_n are needed for the above result to hold. Since (m_1, K_1) is fixed the result is not directly applicable to problems where the second order structure is estimated from observations of Y.

In this paper we make specific assumptions on the smoothness of m_0 and K_0. However, we do not only analyze prediction on the basis of a fixed sequence of designs. Instead, estimating α^λ, which is the essential ingredient of K_0, and constructing an asymptotically optimal adaptive design forms the main part of our analysis.

Compatibility of (m_0, K_0) and (m_1, K_1), or, more generally (m_1, cK_1) for some constant $c > 0$, is rather restrictive. For instance, if

$$K_i(s,t) = g_i(s)\, g_i(t) \cdot R(f_i(s), f_i(t))$$

with f_i, g_i, and R satisfying the assumptions from our paper, then compatibility of $(0, K_0)$ and $(0, K_1)$ implies

$$\alpha_1 = \alpha_2$$

for $\alpha_i = g_i\,(f_i')^{1/2}$. A proof can be based on Lemma 3. If $(0, K_1)$ is known and compatible to $(0, K_0)$ then α_1 may actually be used to define an asymptotically optimal regular sequence of designs. Wahba (1974) contains a detailed analysis of the role of compatibility for the design problem for integration.

Remark 6. Piecewise smooth random functions are studied in Gao and Wasilkowski (1992) and Cohen and d'Ales (1995).

Gao and Wasilkowski construct a Gaussian process Y with continuous derivative $Y^{(\ell-1)}$ on $[0,1]$. At an interior point z the derivative $Y^{(\ell)}$ does not exist, while all derivatives up to order $k > \ell \geq 0$ exist on $[0, z[\, \cup\,]z, 1]$. An adaptive method for locating the singularity z is constructed and analyzed. The results are applied to the integration problem for Y. It turns out that adaption helps significantly.

Cohen and d'Ales (1995) construct a process Y with discontinuities given by a Poisson process on $[0,1]$. Between successive discontinuities Y is stationary and smooth. The process may be represented as

$$Y = \sum_{i=1}^{\infty} \int_0^1 Y(t) \cdot e_i(t)\, dt \cdot e_i$$

for any orthonormal basis e_i of $L_2([0,1])$. If the first n terms are used to approximate Y in L_2-norm, then the optimal basis corresponds to the Karhunen-Loève representation of Y. If arbitrary n terms may be used, then much smaller errors can be obtained. To this end a wavelet basis is selected and the coordinates that correspond to the largest coefficients are used. In general, this approach is also called nonlinear approximation.

Let us stress the following conceptual difference between reconstruction of functions, as studied for instance in this paper, and nonlinear approximation. We compare methods that use the same amount of data (function values). In nonlinear approximation methods are compared that produce approximating functions of the same "size"; formally an infinite amount of data is used. Nevertheless it seems very interesting to compare the performance of adaptive methods for reconstruction and methods for nonlinear approximation.

5. Simulation Results

Our results are asymptotic, and we do not have explicit expressions or estimates for the error of the adaptive method for finite n. Therefore we use a simulation to study errors for small to moderate numbers of observations. Here we present results for L_2-reconstruction of the process Y that is defined by (5).

We compare the adaptive method $\widehat{Y}^{\lambda}_{N_n}$ with $\lambda = 1$ and the optimal method $\widehat{Y}^{e}_{N_n}$ that is based on the equidistant design $0, 1/(N_n - 1), \dots, 1$ of the same size. By

$$e_2(\widehat{Y}^{e}_{N_n}, L_2)/e_2(\widehat{Y}^{\lambda}_{N_n}, L_2).$$

we measure the efficiency of the adaptive method. The latter quantity tends to

$$\|\alpha\|_2/\|\alpha\|_1 = 1.720\dots,$$

and this asymptotic behavior is optimal among all methods, see Corollary 1 and Remark 3.

We illustrate the dependence of the efficiency on N_n, the total number of observations, and k, the number of points where the regularity of Y is estimated. We take $\delta = 10^{-3}$ for $k = 4$ and $\delta = 10^{-4}$ for $k = 3$, 7, and 10. The parameter ε is of minor importance, and we take $\varepsilon = 0$. For every choice of n and k we use 50 simulations of the process Y to approximately determine the error of the respective variant of the adaptive method.

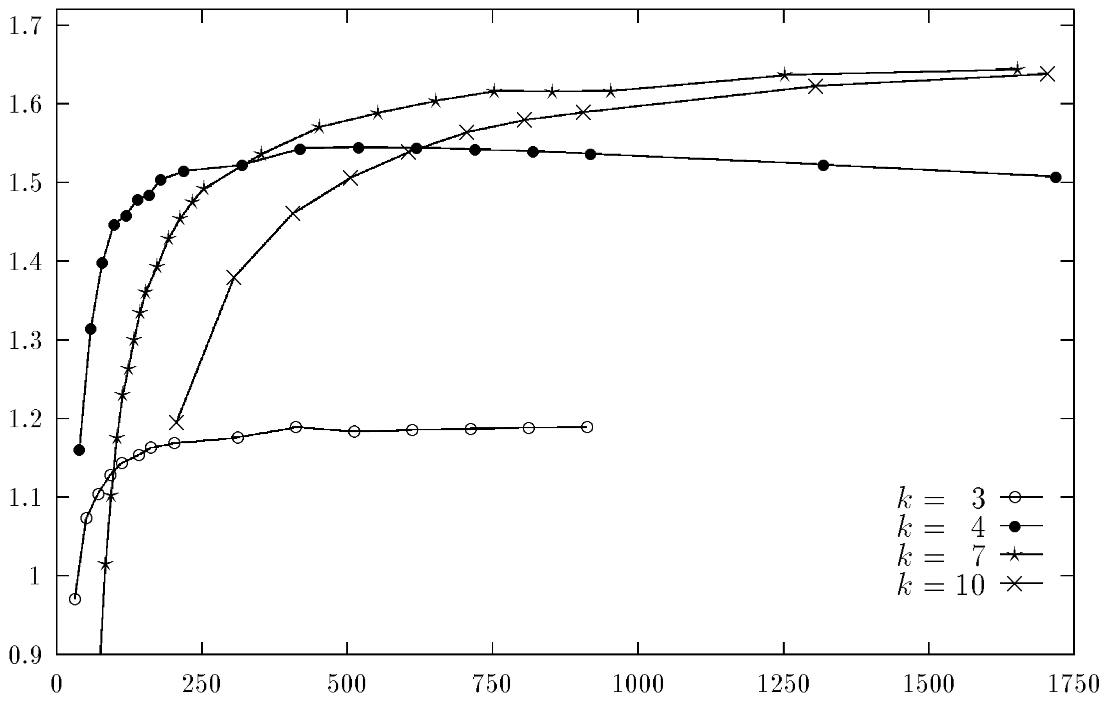

FIGURE 5. Efficiency vs. total number N_n of observations

The simulation shows that $k = 3$ leads to poor results, which get even worse for $\delta = 10^{-3}$. For $k = 4$ we have a good efficiency already for a small number of observations. Larger values of k lead to improvements only for rather large numbers of observations.

For fixed k and δ the efficiency decreases when N_n is large; actually it tends to zero like $N_n^{-1/2}$. This is caused by the error on the subintervals $[z_{i,k} - k\,\delta/2, z_{i,k} + k\,\delta/2]$, where $z_{i,k} = (2i-1)/(2k)$. Hence we have to decrease δ with increasing n. If k is still fixed, one cannot reach the efficiency $1.720\dots$. Thus we also have to increase k with increasing n.

Suppose that the estimate $\widetilde{\alpha}_i$ of α at $z_{i,k}$ is replaced by the exact value $\alpha(z_{i,k})$. Proceeding as in stage two of the adaptive method we get a nonadaptive method. Errors of these methods can be computed exactly. Further experiments have shown the that the values $\alpha(z_{i,k})$ are estimated with sufficient accuracy by the adaptive method. Therefore one might reduce the number of points in the first stage that are clustered around $z_{i,k}$.

6. Proof of Theorem 1

In the sequel we use the following notation. Let Y denote an arbitrary stochastic process and let $T = (t_1, \dots, t_n)$ denote an arbitrary n-point design. Then

$$Y_T = (Y(t_1), \dots, Y(t_n))$$

is the vector of corresponding observations. Furthermore, $\widehat{Y}_T^L$ denotes the piecewise linear interpolation of Y based on Y_T. If T is ordered increasingly and $t \in [t_{i-1}, t_i]$, then

$$\widehat{Y}_T^L(t) = \frac{t - t_{i-1}}{t_i - t_{i-1}} \cdot Y(t_i) + \frac{t_i - t}{t_i - t_{i-1}} \cdot Y(t_{i-1}).$$

We use c to denote unspecified positive constants, which only depend on R, m, g, f, and, for reconstruction, on p and q. Furthermore, we put

$$Z(t) = g(t) \cdot X(f(t)), \qquad 0 \le t \le 1,$$

where X is a zero mean Gaussian process with covariance kernel R satisfying (A)–(C). We assume that m, f, and g satisfy conditions (D)–(F). Finally,

$$K(s,t) = g(s)\, g(t) \cdot R(f(s), f(t)), \qquad 0 \le s, t \le 1,$$

denotes the covariance kernel of the process Z.

6.1. **Preliminary estimates.** We have the following estimates for increments of the process Z.

Lemma 1. *Let*

$$\Gamma = E((Z(s_2) - Z(s_1)) \cdot (Z(s_4) - Z(s_3)))$$

and

$$\Delta = \max(s_2 - s_1, s_4 - s_3),$$

where $s_1 \le s_2$ and $s_3 \le s_4$. Then

$$|\Gamma| \le c \cdot \Delta^2$$

if $s_2 \le s_3$, and

$$|\Gamma - \alpha^2(s_1) \cdot \Delta| \le c \cdot \Delta^2$$

if $s_1 = s_3$ and $s_2 = s_4$.

Proof. The partial derivatives $K^{(1,0)}$ and $K^{(1,1)}$ are continuous on $\Omega_+ \cup \Omega_-$ and continuously extendable to $\operatorname{cl}\Omega_+$ as well as to $\operatorname{cl}\Omega_-$, due to the assumptions (A), (E) and (F). Furthermore,

$$K_-^{(1,0)}(t,t) - K_+^{(1,0)}(t,t) = g^2(t) \cdot f'(t) = \alpha^2(t)$$

by assumption (B), and α^2 is Lipschitz continuous by (E) and (F). Let $s_1 < s_2$ and $s_3, s_4 \notin \,]s_1, s_2[$. Then

$$\Gamma = \int_{s_1}^{s_2} (K^{(1,0)}(u, s_4) - K^{(1,0)}(u, s_3))\, du.$$

Now assume $s_2 \leq s_3 \leq s_4$. Then

$$K^{(1,0)}(u, s_4) - K^{(1,0)}(u, s_3) = \int_{s_3}^{s_4} K^{(1,1)}(u, v)\, dv \tag{10}$$

for each $s_1 < u < s_2$, and therefore

$$|\Gamma| \leq \int_{s_1}^{s_2} \int_{s_3}^{s_4} |K^{(1,1)}(u,v)|\, dv\, du \leq c \cdot \Delta^2.$$

Next let $s_1 = s_3$ and $s_2 = s_4$. Then

$$\begin{aligned} &K^{(1,0)}(u, s_2) - K^{(1,0)}(u, s_1) \\ &\quad = K^{(1,0)}(u, s_2) - K_-^{(1,0)}(u,u) + \alpha^2(u) + K_+^{(1,0)}(u,u) - K^{(1,0)}(u, s_1) \\ &\quad = \int_u^{s_2} K^{(1,1)}(u,v)\, dv + \alpha^2(u) + \int_{s_1}^{u} K^{(1,1)}(u,v)\, dv. \end{aligned}$$

Hence

$$|\Gamma - \alpha^2(s_1) \cdot \Delta| \leq \int_{s_1}^{s_2} \int_{s_1}^{s_2} |K^{(1,1)}(u,v)|\, dv\, du + \int_{s_1}^{s_2} |\alpha^2(u) - \alpha^2(s_1)|\, du \leq c \cdot \Delta^2,$$

which completes the proof. □

As a consequence of Lemma 1 and the Lipschitz continuity of m, the process Y turns out to be Hölder continuous with exponent $1/2$ in quadratic mean, see (2).

Furthermore, Lemma 1 yields an upper bound for the error due to piecewise linear interpolation of the process Z. Observe that $K^{(1,1)}(\theta, \eta) = K^{(1,1)}(\eta, \theta)$ for $\theta \neq \eta$ and therefore $K_-^{(1,1)} = K_+^{(1,1)}$ is a continuous function on $[0,1]^2$.

Lemma 2. *Let*

$$M(u,v) = E\left((Z(u) - \widehat{Z}^L_{(s_1,s_2)}(u)) \cdot (Z(v) - \widehat{Z}^L_{(s_3,s_4)}(v))\right)$$

where $s_1 \leq s_2$ and $s_3 \leq s_4$. Define Δ as in Lemma 1, and let ω denote the modulus of continuity of $K_-^{(1,1)}$. Then

$$\left| \int_{s_1}^{s_2} \int_{s_3}^{s_4} M(u,v)\, du\, dv \right| \leq c \cdot \omega(\Delta) \cdot \Delta^4$$

if $s_2 \leq s_3$, and

$$\int_{s_1}^{s_2} \int_{s_1}^{s_2} M(u,v)\, du\, dv \leq 1/12 \cdot \Delta^3 \cdot \alpha^2(s_1) \cdot (1 + c \cdot \Delta),$$

$$\int_{s_1}^{s_2} M(u,u)^{p/2}\,du \le \nu_p \cdot \Delta^{1+p/2} \cdot \alpha^p(s_1)\cdot(1+c\cdot\Delta)$$

if $s_1 = s_3$ *and* $s_2 = s_4$.

Proof. Let $s_1 \le u \le s_2$ and $s_3 \le v \le s_4$ and put

$$U(u) = Z(u) - Z(s_1), \qquad V(v) = Z(v) - Z(s_3).$$

Clearly

$$\begin{aligned} M(u,v) &= E\Big(\Big(U(u) - \frac{u-s_1}{s_2-s_1}\,U(s_2)\Big)\cdot\Big(V(v) - \frac{v-s_3}{s_4-s_3}\,V(s_4)\Big)\Big)\\ &= E(U(u)\,V(v)) - \frac{u-s_1}{s_2-s_1}\cdot E(U(s_2)\,V(v))\\ &\quad - \frac{v-s_3}{s_4-s_3}\cdot E(U(u)\,V(s_4)) + \frac{(u-s_1)(v-s_3)}{(s_2-s_1)(s_4-s_3)}\cdot E(U(s_2)\,V(s_4)). \end{aligned}$$

First assume that $s_2 \le s_3$. Then, by (10),

$$E(U(u)\,V(v)) = (u-s_1)\,(v-s_3)\cdot K^{(1,1)}(\theta,\eta)$$

for some $\theta \in\,]s_1,s_2[$ and $\eta \in\,]s_3,s_4[$. Hence

$$|M(u,v)| \le 2\cdot\omega(\Delta)\cdot(u-s_1)\,(v-s_3)$$

which yields the first estimate.

Now assume that $s_1 = s_3$, $s_2 = s_4$, and $u \le v$. Then, by Lemma 1,

$$\begin{aligned} &|E(U(u)\,U(v)) - \alpha^2(s_1)\cdot(u-s_1)|\\ &\quad\le |E(U^2(u)) - \alpha^2(s_1)\cdot(u-s_1)| + |E(U(u)\,(Z(v)-Z(u))|\\ &\quad\le c\cdot\Delta^2. \end{aligned}$$

Hence

$$\begin{aligned} M(u,v) &\le \alpha^2(s_1)\cdot\left((u-s_1) - 2\cdot\frac{(u-s_1)(v-s_1)}{\Delta} + \frac{(u-s_1)(v-s_1)}{\Delta}\right) + c\cdot\Delta^2\\ &\le \alpha^2(s_1)\cdot\left(\frac{(s_2-v)(u-s_1)}{\Delta} + c\cdot\Delta^2\right), \end{aligned}$$

and the last two estimates from Lemma 2 follow. □

6.2. **Properties of the adaptive design.** Henceforth let T denote the nonadaptive design that is defined by (8). Recall that the estimate of the function α^λ on the subinterval $J_i = [t_{i,k}, t_{i+1,0}]$ depends on $\widehat{\alpha}_i(Y_T)$ and $\widehat{\alpha}_{i+1}(Y_T)$, where

$$\widehat{\alpha}_0(y) = \widehat{\alpha}_1(y), \qquad \widehat{\alpha}_{k+1}(y) = \widehat{\alpha}_k(y).$$

Lemma 3. *Let* $\theta \in J_{i-1}\cup J_i$. *Then*

$$\left(E\left(|\widehat{\alpha}_i(Y_T) - \alpha^\lambda(\theta)|^q\right)\right)^{1/q} \le c\cdot 1/k^{1/2}$$

for any $q \ge 1$.

Proof. Fix i and put $a = \alpha^2(\theta)/c_\lambda^{2/\lambda}$ as well as

$$V_j = (c_\lambda^{1/\lambda}\,\delta^{1/2})^{-1}\cdot(Y(t_{i,j}) - Y(t_{i,j-1}))$$

for $j = 1, \ldots, k$. Clearly

$$\widehat{\alpha}_i(Y_T) = \frac{1}{k}\sum_{j=1}^{k}|V_j|^\lambda. \tag{11}$$

Recall that $\alpha^2 = g^2 \cdot f'$ is positive and Lipschitz continuous, due to the assumptions on f and g. Lemma 1 and (7) yield

$$\begin{aligned}
|\operatorname{Var}(V_j) - a| &= (c_\lambda^{2/\lambda}\,\delta)^{-1}\cdot|E(Z(t_{i,j}) - Z(t_{i,j-1}))^2 - \alpha^2(\theta)\cdot\delta| \\
&\le c/\delta\cdot\Big|E(Z(t_{i,j}) - Z(t_{i,j-1}))^2 - \alpha^2(t_{i,j-1})\cdot\delta\Big| + c\cdot|\alpha^2(\theta) - \alpha^2(t_{i,j-1})| \\
&\le c\cdot\delta + c\cdot 1/k \le c\cdot 1/k
\end{aligned} \tag{12}$$

and

$$\begin{aligned}
|\operatorname{Cov}(V_j, V_\ell)| &= c/\delta\cdot\Big|E\Big((Z(t_{i,j}) - Z(t_{i,j-1}))\cdot(Z(t_{i,\ell}) - Z(t_{i,\ell-1}))\Big)\Big| \\
&\le c\cdot\delta \le c\cdot 1/k^2
\end{aligned} \tag{13}$$

if $j \neq \ell$.

Let $V = (V_1, \ldots, V_k)$ and put

$$C = \operatorname{Cov}(V), \qquad B = \operatorname{diag}(\operatorname{Var}(V_1), \ldots, \operatorname{Var}(V_k)).$$

Then (12) and (13) yield

$$\begin{aligned}
|x^\top C\,x - x^\top B\,x| = |\sum_{\ell\neq j} x_\ell\, x_j\ \operatorname{Cov}(V_\ell, V_j)| &\le c\cdot 1/k^2\cdot\sum_{\ell\neq j}|x_\ell|\,|x_j| \\
&\le c\cdot 1/k\cdot\sum_{j=1}^{k} x_j^2 \le c\cdot 1/k\cdot x^\top B\,x
\end{aligned} \tag{14}$$

for all $x = (x_1, \ldots, x_k)^\top \in \mathbb{R}^k$. For matrices $A, B \in \mathbb{R}^{k\times k}$ we write $A \ll B$ if $B - A$ is positive semidefinite. By (14) we have

$$(1 - \tilde{c}/k)\cdot B \ll C \ll (1 + \tilde{c}/k)\cdot B \tag{15}$$

with some positive constant $\tilde{c}$. Let

$$\widetilde{V} = (\widetilde{V}_1, \ldots, \widetilde{V}_k) \sim N(E(V), (1 + \tilde{c}/k)\cdot B),$$

Furthermore let $G(\mu, \Sigma)$ denote the expectation of

$$x \mapsto \left|\frac{1}{k}\cdot\sum_{j=1}^{k}\Big(|x_j|^\lambda - E(|\widetilde{V}_j|^\lambda)\Big)\right|^q$$

with respect to the normal distribution $N(\mu, \Sigma)$ on $\mathbb{R}^k$. Using (11) and (15) we get

$$\begin{aligned} E\Big(\Big|\widehat{\alpha}_i(Y_T) - 1/k \cdot \sum_{j=1}^{k} E(|\widetilde{V}_j|^\lambda)\Big|^q\Big) &= G(E(V), C) \\ &\leq (\det((1+\tilde{c}/k)\cdot B)/\det(C))^{1/2} \cdot G(E(V), (1+\tilde{c}/k)\cdot B) \\ &\leq ((1+\tilde{c}/k)/(1-\tilde{c}/k))^{k/2} \cdot E\Big(\Big|1/k \cdot \sum_{j=1}^{k} (|\widetilde{V}_j|^\lambda - E(|\widetilde{V}_j|^\lambda))\Big|^q\Big). \end{aligned}$$

Applying a moment inequality of Dharmadhikari and Jogdeo (1969), we have

$$\begin{aligned} E\Big(\Big|1/k \cdot \sum_{j=1}^{k} (|\widetilde{V}_j|^\lambda - E(|\widetilde{V}_j|^\lambda))\Big|^q\Big) &\leq c \cdot k^{-q/2-1} \cdot \sum_{j=1}^{k} E\left(\left||\widetilde{V}_j|^\lambda - E(|\widetilde{V}_j|^\lambda)\right|^q\right) \\ &\leq c \cdot 1/k^{q/2}, \end{aligned}$$

and therefore

$$E\Big(\Big|\widehat{\alpha}_i(Y_T) - 1/k \cdot \sum_{j=1}^{k} E(|\widetilde{V}_j|^\lambda)\Big|^q\Big) \leq c \cdot 1/k^{q/2}.$$

It remains to show that

$$|E(1/k \cdot \sum_{j=1}^{k} |\widetilde{V}_j|^\lambda) - \alpha^\lambda(\theta)| \leq c \cdot 1/k^{1/2}. \tag{16}$$

Let $F(\mu, \sigma^2)$ denote the expectation of $x \mapsto |x|^\lambda - \alpha^\lambda(\theta)$ with respect to the normal distribution $N(\mu, \sigma^2)$. Then

$$E(|\widetilde{V}_j|^\lambda) - \alpha^\lambda(\theta) = F(E(V_j), \mathrm{Var}(\widetilde{V}_j)),$$

and F is continuously differentiable on $\mathbb{R} \times]0, \infty[$. Using (D) we get

$$|E(V_j)| = (c_\lambda^{1/\lambda}\, \delta^{1/2})^{-1} \cdot |m(t_{i,j}) - m(t_{i,j-1})| \leq c \cdot \delta^{1/2} \leq c \cdot 1/k.$$

Together with (12) this implies that $(E(V_j), \mathrm{Var}(\widetilde{V}_j))$ belongs to some compact and convex subset of $\mathbb{R} \times]0, \infty[$ if k is sufficiently large. Moreover,

$$|F(E(V_j), \mathrm{Var}(\widetilde{V}_j)) - F(0, \mathrm{Var}(\widetilde{V}_j))| \leq c \cdot 1/k.$$

Lemma 1 and the Lipschitz continuity of α^2 imply

$$\begin{aligned} |F(0, \mathrm{Var}(\widetilde{V}_j))| &= |(c_\lambda\, \delta^{\lambda/2})^{-1} \cdot (1+\tilde{c}/k)^{\lambda/2} \cdot E(|Z(t_{i,j}) - Z(t_{i,j-1})|^\lambda) - \alpha^\lambda(\theta)| \\ &= \left|\Big(1/\delta \cdot (1+\tilde{c}/k) \cdot E(Z(t_{i,j}) - Z(t_{i,j-1}))^2\Big)^{\lambda/2} - (\alpha^2(\theta))^{\lambda/2}\right| \\ &\leq c \cdot \left|(1+\tilde{c}/k) \cdot 1/\delta \cdot E(Z(t_{i,j}) - Z(t_{i,j-1}))^2 - \alpha^2(\theta)\right| \\ &\leq c \cdot 1/k. \end{aligned}$$

Combining the last two estimates we obtain

$$\left|E\left(|\widetilde{V}_j|^\lambda\right) - \alpha^\lambda(\theta)\right| \leq c \cdot 1/k,$$

which proves (16). $\square$

In the sequel we derive asymptotic estimates, taking into account the properties (i)–(iii) of the parameters k_n, δ_n, and ε_n. The corresponding notation T_n, $\widehat{\alpha}_{i,n}$, $J_{i,n}$, etc., is canonical. We suppress the dependence on n as long as no asymptotics are involved.

We prove that asymptotically the adaptive designs behave like the regular sequence generated by α^λ. More precisely, consider the n-point design $(t^*_{1,n}, \dots, t^*_{n,n})$ from this regular sequence. The length of each subinterval $[t^*_{j,n}, t^*_{j+1,n}] \subset J_{i,n}$ is approximately

$$\rho_{i,n} = \frac{1}{n \max_{\theta \in J_{i,n}} \alpha^\lambda(\theta)} \int_0^1 \alpha^\lambda(t)\, dt.$$

In fact, this also holds true for the adaptive points placed in $J_{i,n}$. However, we only need the following upper bound.

Put

$$s_{i,0}(y) = t_{i,k}, \qquad s_{i,r_i(y)+1}(y) = t_{i+1,0}$$

and let

$$\Delta_i^{\max}(y) = \max_{1 \le j \le r_i(y)+1} (s_{i,j}(y) - s_{i,j-1}(y))$$

denote the maximum distance between two consecutive points in J_i.

Lemma 4. *For any $q \ge 1$,*

$$\limsup_{n\to\infty} \max_{0 \le i \le k_n} E\left(\Delta_{i,n}^{\max}(Y_{T_n})/\rho_{i,n}\right)^q \le 1.$$

Proof. First we show that

$$(E(\Delta_i^{\max}(Y_T))^q)^{1/q} \le c/(n\,\varepsilon). \tag{17}$$

By definition,

$$\begin{aligned} s_{i,j}(y) - s_{i,j-1}(y) &\le \frac{1}{\min(\widetilde{\alpha}_i(y), \widetilde{\alpha}_{i+1}(y))} \cdot \frac{(\widetilde{\alpha}_i(y) + \widetilde{\alpha}_{i+1}(y)) \cdot (1/k - k\delta)}{2(r_i(y)+1)} \\ &\le \frac{1}{\min(\widetilde{\alpha}_i(y), \widetilde{\alpha}_{i+1}(y))} \cdot \frac{1}{n\,k} \sum_{j=1}^k \widetilde{\alpha}_j(y) \end{aligned} \tag{18}$$

for every $i = 1, \dots, k-1$ and $j = 1, \dots, r_i(y)+1$. Since $\widetilde{\alpha}_i(y) > \varepsilon$ for $i = 1, \dots, k$, we get

$$s_{i,j}(y) - s_{i,j-1}(y) \le \frac{1}{n\,\varepsilon\,k} \sum_{j=1}^k \widetilde{\alpha}_j(y).$$

This upper bound also holds in the cases $i = 0$ and $i = k$. It thus follows

$$(E(\Delta_i^{\max}(Y_T))^q)^{1/q} \le \frac{1}{n\,\varepsilon\,k} \sum_{j=1}^k (E(\widetilde{\alpha}_j(Y_T))^q)^{1/q}.$$

Furthermore

$$(E(\widehat{\alpha}_i(Y_T)^q))^{1/q} \le \frac{1}{k} \sum_{j=1}^k \left(E(|V_j|^{\lambda q})\right)^{1/q} \le c,$$

see (11). Since $\widetilde{\alpha} - \widehat{\alpha} \le \varepsilon$, we have

$$\sum_{j=1}^{k} \left(E(\widetilde{\alpha}_j(Y_T)^q)\right)^{1/q} \le k \cdot (\varepsilon + c) \le c \cdot k,$$

which proves (17).

Let $\theta \in J_i$ and put $M_i(y) = \min(\widehat{\alpha}_i(y), \widehat{\alpha}_{i+1}(y))$. If $1 \le i \le k-1$ then

$$\begin{aligned} \alpha^\lambda(\theta)\, \Delta_i^{\max}(y) &= (\alpha^\lambda(\theta) - M_i(y)) \cdot \Delta_i^{\max}(y) + M_i(y) \cdot \Delta_i^{\max}(y) \\ &\le (|\alpha^\lambda(\theta) - \widehat{\alpha}_i(y)| + |\alpha^\lambda(\theta) - \widehat{\alpha}_{i+1}(y)|) \cdot \Delta_i^{\max}(y) + \frac{1}{n\,k} \sum_{j=1}^{k} \widetilde{\alpha}_j(y), \end{aligned}$$

see (18). Clearly, the same estimate holds for $i = 0$ and $i = k$. Lemma 3 and (17) yield

$$\begin{aligned} &\Big(E(|\alpha^\lambda(\theta) - \widehat{\alpha}_\ell(Y_T)| \cdot \Delta_i^{\max}(Y_T))^q\Big)^{1/q} \\ &\qquad \le \Big(E(|\alpha^\lambda(\theta) - \widehat{\alpha}_\ell(Y_T)|^{2q})\Big)^{1/2q} \cdot \Big(E(\Delta_i^{\max}(Y_T))^{2q})\Big)^{1/2q} \\ &\qquad \le c \cdot 1/k^{1/2} \cdot 1/(n\,\varepsilon) \end{aligned}$$

for $\ell = i, i+1$. Furthermore,

$$\begin{aligned} &\left(E\left|\frac{1}{k}\sum_{j=1}^{k} \widetilde{\alpha}_j(Y_T)\right|^q\right)^{1/q} \\ &\qquad \le \varepsilon + \frac{1}{k}\sum_{j=1}^{k} \Big(E|\widehat{\alpha}_j(Y_T) - \alpha^\lambda((2j-1)/2k)|^q\Big)^{1/q} + \frac{1}{k}\sum_{j=1}^{k} \alpha^\lambda((2j-1)/2k) \\ &\qquad \le \varepsilon + c \cdot 1/k^{1/2} + \frac{1}{k}\sum_{j=1}^{k} \alpha^\lambda((2j-1)/2k). \end{aligned}$$

Therefore

$$\Big(E|\alpha^\lambda(\theta)\, \Delta_i^{\max}(Y_T)|^q\Big)^{1/q} \le \frac{1}{n} \cdot \Big(\varepsilon + \frac{c}{k^{1/2} \cdot \varepsilon} + \frac{1}{k}\sum_{j=1}^{k} \alpha^\lambda((2j-1)/2k)\Big).$$

The result now follows, observing that ε_n tends to zero while k_n and $k_n^{1/2} \cdot \varepsilon_n$ tend to infinity. □

6.3. **The conditional mean of Z.** For the proof of Theorem 1 we will use the decomposition

$$Z(t) = U_n(t) + V_n(t), \qquad 0 \le t \le 1,$$

where

$$U_n(t) = E(Z(t) \mid Z_{T_n})$$

is the optimal approximation of the process Z on the basis of Z_{T_n}, and

$$V_n = Z - U_n$$

denotes the corresponding error process. As previously we suppress the dependence on n as long as no asymptotics are involved.

Put

$$t_{i,k+1} = t_{i+1,0}$$

for $0 \le i \le k$ and define the function $W : [0,1]^2 \to \mathbb{R}$ by

$$W(s,t) = \frac{(f(t_{i,j+1}) - \max(s,t)) \cdot (\min(s,t) - f(t_{i,j}))}{f(t_{i,j+1}) - f(t_{i,j})}$$

if

$$s,t \in [f(t_{i,j}), f(t_{i,j+1})]$$

and $W(s,t) = 0$ otherwise. Clearly, W is the covariance kernel of a collection of independent Brownian bridges on the subintervals $[f(t_{i,j}), f(t_{i,j+1})]$. Next, let

$$\varphi_{i,j}(t) = \int_0^1 R_+^{(2,0)}(s, f(t_{i,j})) \cdot W(s, f(t))\, ds, \qquad 0 \le t \le 1,$$

and put

$$\varphi = (\varphi_{0,k}, \ldots, \varphi_{k+1,0})^\top.$$

Finally, put

$$f(T) = (f(t_{0,k}), \ldots, f(t_{k+1,0})),$$

and assume that the covariance matrix $R_{f(T)}$ of $X_{f(T)} = (X(f(t_{0,k})), \ldots, X(f(t_{k+1,0})))$ is positive definite.

Observing

$$U(t) = g(t) \cdot E(X(f(t)) \,|\, X_{f(T)}),$$

the following representation of the process U can easily be derived from equation (3.2) in Müller-Gronbach (1996c). It turns out that $U(t)$ is basically given by piecewise linear interpolation, up to the remainder $\varphi^\top(t)\, R_{f(T)}^{-1} X_{f(T)}$.

Lemma 5. *Let $\mu : \mathbb{R}^{k\cdot(k+1)+1} \times [0,1] \to \mathbb{R}$ be defined by*

$$\mu(x,t) = g(t) \cdot \left(\frac{f(t_{i,j+1}) - f(t)}{f(t_{i,j+1}) - f(t_{i,j})} \cdot x_{i,j} + \frac{f(t) - f(t_{i,j})}{f(t_{i,j+1}) - f(t_{i,j})} \cdot x_{i,j+1} - \varphi^\top(t)\, R_{f(T)}^{-1}\, x \right)$$

for $t \in [t_{i,j}, t_{i,j+1}]$ and $x = (x_{0,k}, \ldots, x_{k+1,0})^\top$ with $x_{i,k+1} = x_{i+1,0}$. Then

$$U(t) = \mu(X_{f(T)}, t), \qquad 0 \le t \le 1.$$

Next we analyze the smoothness of $\mu(x,\cdot)$ on each of the intervals $[t_{i,k}, t_{i+1,0}]$.

Lemma 6. *Let $t_{i,k} \le s \le t \le t_{i+1,0}$ and put $\Delta = t - s$. Then*

$$|\mu^{(0,1)}(x,s) - \mu^{(0,1)}(x,t)| \\ \le c \cdot (\omega(\Delta) + k^2 \cdot \Delta) \cdot (|x_{i+1,0}| + |x_{i,k}| + k^{-2} \cdot (x^\top R_{f(T)}^{-1}\, x)^{1/2}),$$

where ω denotes the modulus of continuity of g'.

Proof. Let
$$\varphi' = (\varphi'_{0,k}, \dots, \varphi'_{k+1,0})^\top,$$
where
$$\varphi'_{i,j}(t) = \int_0^1 R_+^{(2,0)}(s, f(t_{i,j})) \cdot W^{(0,1)}(s, f(t)) \cdot f'(t)\, ds, \qquad 0 \le t \le 1.$$
Recall that $H(R)$ denotes the Hilbert space with reproducing kernel R. Let P denote the orthogonal projection of $H(R)$ onto the subspace spanned by the functions $R(\cdot, f(t_{i,j}))$.

We have
$$\begin{aligned}
&|(\varphi'(t))^\top R_{f(T)}^{-1}\, \varphi'(t)| \\
&\quad = f'(t)^2 \cdot \left| \int_0^1 \int_0^1 \langle P R_+^{(2,0)}(u,\cdot), R_+^{(2,0)}(v,\cdot)\rangle_R \cdot W^{(0,1)}(u, f(t)) \cdot W^{(0,1)}(v, f(t))\, du\, dv \right| \\
&\quad \le c \cdot \sup_{0\le t\le 1} \|R_+^{(2,0)}(t,\cdot)\|_R^2 \cdot \left(\int_{f(t_{i,k})}^{f(t_{i+1,0})} |W^{(0,1)}(u, f(t))|\, du \right)^2 .
\end{aligned}$$
Hence
$$|(\varphi'(t))^\top R_{f(T)}^{-1}\, \varphi'(t)| \le c \cdot 1/k^2 \tag{19}$$
by (C) and
$$\int_{f(t_{i,k})}^{f(t_{i+1,0})} |W^{(0,1)}(u, f(t))|\, du \le c \cdot 1/k,$$
which follows from (F). Similarly,
$$|\varphi^\top(t)\, R_{f(T)}^{-1}\, \varphi(t)| \le c \cdot \left(\int_{f(t_{i,k})}^{f(t_{i+1,0})} W(u, f(t))\, du \right)^2 \le c \cdot 1/k^4, \tag{20}$$
and
$$\begin{aligned}
&|(\varphi'(s) - \varphi'(t))^\top R_{f(T)}^{-1}\, (\varphi'(s) - \varphi'(t))| \\
&\quad \le c \cdot \left(\int_{f(t_{i,k})}^{f(t_{i+1,0})} |W^{(0,1)}(u, f(s)) \cdot f'(s) - W^{(0,1)}(u, f(t)) \cdot f'(t)|\, du \right)^2 \\
&\quad \le c \cdot \Delta^2.
\end{aligned} \tag{21}$$
Assume that $g = 1$. Put
$$B(x) = |x_{i+1,0}| + |x_{i,k}| + k^{-2} \cdot (x^\top R_{f(T)}^{-1}\, x)^{1/2}. \tag{22}$$
Using (21) we get
$$\begin{aligned}
&|\mu^{(0,1)}(x,s) - \mu^{(0,1)}(x,t)| \\
&\quad = \left| \frac{f'(s) - f'(t)}{f(t_{i+1,0}) - f(t_{i,k})} \cdot (x_{i+1,0} - x_{i,k}) - (\varphi'(s) - \varphi'(t))^\top R_{f(T)}^{-1}\, x \right| \\
&\quad \le c \cdot k^2 \cdot \Delta \cdot B(x)
\end{aligned}$$
as claimed. Furthermore, (19) implies
$$|\mu^{(0,1)}(x,t)| \le c \cdot k \cdot B(x)$$

and therefore

$$|\mu(x,s)-\mu(x,t)| \le c\cdot k\cdot \Delta\cdot B(x).$$

By (20),

$$|\mu(x,t_{i,k})| \le c\cdot B(x).$$

In the general case we put

$$\tilde{\mu}(x,t)=\mu(x,t)/g(t)$$

and use

$$\begin{aligned}
&|\mu^{(0,1)}(x,s)-\mu^{(0,1)}(x,t)| \\
&\quad \le |g'(s)-g'(t)|\cdot|\tilde{\mu}(x,s)|+|g'(t)|\cdot|\tilde{\mu}(x,s)-\tilde{\mu}(x,t)| \\
&\qquad +|g(s)-g(t)|\cdot|\tilde{\mu}^{(0,1)}(x,s)|+|g(t)|\cdot|\tilde{\mu}^{(0,1)}(x,s)-\tilde{\mu}^{(0,1)}(x,t)|.
\end{aligned}$$

□

6.4. **Upper bounds for reconstruction.** Let

$$S(y)=(T_n, s_{0,1}(y),\ldots,s_{k,r_k(y)}(y))$$

denote the adaptive design that is defined in Section 3. Here T_n denotes the nonadaptive part, see (8), which consists of points from the set

$$I_n=\bigcup_{i=1}^{k_n}\left[\frac{2i-1}{2k_n}-\frac{k_n\delta_n}{2},\ \frac{2i-1}{2k_n}+\frac{k_n\delta_n}{2}\right], \tag{23}$$

together with the endpoints 0 and 1. The adaptively chosen points $s_{i,j}(y)$ belong to $[0,1]\setminus I_n$. For L_p-reconstruction the adaptive method $\widehat{Y}^{\lambda}_{N_n}$ is defined by

$$\widehat{Y}^{\lambda}_{N_n}=\widehat{Y}^{L}_{S(Y_{T_n})},$$

where $\lambda=(1/2+1/p)^{-1}$ according to (6).

On I_n the piecewise linear interpolation $\widehat{Y}^{L}_{S(Y_{T_n})}$ coincides with $\widehat{Y}^{L}_{T_n}$, and the respective error is small. Namely,

$$\begin{aligned}
&\left(\int_{I_n} E\left(|Y(t)-\widehat{Y}^{L}_{T_n}(t)|^p\right)\,dt\right)^{1/p} \\
&\quad\le \left(\int_{I_n}|m(t)-\widehat{m}^{L}_{T_n}(t)|^p\,dt\right)^{1/p}+\left(\int_{I_n}E\left(|Z(t)-\widehat{Z}^{L}_{T_n}(t)|^p\right)\,dt\right)^{1/p} \\
&\quad\le c\cdot(k_n^2\,\delta_n\cdot\delta_n^p)^{1/p}+c\cdot\left(\int_{I_n}\left(E\left(Z(t)-\widehat{Z}^{L}_{T_n}(t)\right)^2\right)^{p/2}\,dt\right)^{1/p} \\
&\quad\le c\cdot(k_n^2\,\delta_n)^{1/p}\cdot(\delta_n+\delta_n^{1/2}) \\
&\quad\le c\cdot n^{-(2\gamma/p+1/2)}
\end{aligned}$$

for every $1\le p<\infty$. Here we have used (D), Lemma 2, and properties (i) and (ii). Hence

$$\limsup_{n\to\infty}\ n^{1/2}\cdot\int_{I_n}\left(E\left(|Y(t)-\widehat{Y}^{L}_{S(Y_{T_n})}(t)|^p\right)\,dt\right)^{1/p}=0.$$

It remains to study the error on $[0,1]\setminus I_n$, and the upper bound in Theorem 1 follows from

$$\limsup_{n\to\infty} n^{1/2}\cdot\left(\int_{[0,1]\setminus I_n} E\left(|Y(t)-\widehat{Y}^L_{S(Y_{T_n})}(t)|^p\right)\,dt\right)^{1/p} \le (c_p\,\nu_p)^{1/p}\cdot\|\alpha\|_\lambda. \qquad (24)$$

In the proof of (24) we use the decomposition $Y = m+U_n+V_n$ that was introduced in Section 6.3. First we analyze piecewise linear interpolation of U_n, based on the adaptive design $S(Y_{T_n})$.

Lemma 7. *For every $1\le p<\infty$,*

$$\lim_{n\to\infty} n\cdot\left(\int_{[0,1]\setminus I_n} E\left(|U_n(t)-\widehat{(U_n)}^L_{S(Y_{T_n})}(t)|^p\right)\,dt\right)^{1/p} = 0.$$

Proof. Let Q denote the distribution of $X_{f(T)}$ and define

$$\psi(x) = (m(t_{0,k})+g(t_{0,k})\cdot x_{0,k},\dots,m(t_{k+1,0})+g(t_{k+1,0})\cdot x_{k+1,0})$$

for $x=(x_{0,k},\dots,x_{k+1,0})$. Then $Y_T=\psi(X_{f(T)})$, and Lemma 5 yields

$$E\left(|U(t)-\widehat{U}^L_{S(Y_T)}(t)|^p\right) = \int_{\mathbb{R}^{k\cdot(k+1)+1}} |\mu(x,t)-\widehat{\mu(x,\cdot)}^L_{S(\psi(x))}(t)|^p\,Q(dx).$$

Let ω denote the modulus of continuity of g'. Put $\Delta(x)=\Delta_i^{\max}(\psi(x))$. By Lemma 6 we have

$$|\mu(x,t)-\widehat{\mu(x,\cdot)}^L_{S(\psi(x))}(t)| \le c\cdot A(x)\cdot B(x)$$

for every $t\in[t_{i,k},t_{i+1,0}]$, where $B(x)$ is defined by (22) and

$$A(x)=\Delta(x)\cdot(\omega(\Delta(x))+k^2\cdot\Delta(x)).$$

We thus conclude that

$$\left(E\left(|U(t)-\widehat{U}^L_{S(Y_T)}(t)|^p\right)\right)^{1/p} \le c\cdot\left(E\left(A(X_{f(T)})\right)^{2p}\right)^{1/2p}\cdot\left(E\left(B(X_{f(T)})\right)^{2p}\right)^{1/2p}.$$

Clearly

$$\sup_{0\le t\le1}(E|X(t)|^{2p})^{1/2p} \le c_{2p}^{1/2p}\cdot\sup_{0\le t\le1} R^{1/2}(t,t)\le c.$$

Moreover

$$X^\top_{f(T)}\,R^{-1}_{f(T)}\,X_{f(T)} \sim \chi(k\cdot(k+1)+2),$$

and therefore

$$\left(E(X^\top_{f(T)}\,R^{-1}_{f(T)}\,X_{f(T)})^p\right)^{1/2p} \le c\cdot(k\cdot(k+1)+2)\le c\cdot k^2.$$

Hence

$$\left(E\left(B(X_{f(T)})\right)^{2p}\right)^{1/2p}\le c.$$

Furthermore, by Lemma 4,

$$\left(E\left(A(X_{f(T)})\right)^{2p}\right)^{1/2p} \le \omega(1/k)\cdot c\cdot 1/n + k^2\cdot c\cdot 1/n^2 \le c\cdot 1/n\cdot(\omega(1/k)+1/n^{2\gamma}).$$

Summarizing we have

$$\left(E\left(|U(t)-\widehat{U}^L_{S(Y_T)}(t)|^p\right)\right)^{1/p} \le c\cdot 1/n\cdot(\omega(1/k)+1/n^{2\gamma}),$$

which completes the proof. $\square$

In the following we study piecewise linear interpolation of the error process V_n. According to the general definition, $\widehat{(V_n)}^L_{S(Y_{T_n})}$ is the piecewise linear interpolation of V_n, based on the adaptive design $S(Y_{T_n})$.

Lemma 8. *Let $\lambda = (1/2 + 1/p)^{-1}$. Then*

$$\limsup_{n\to\infty}\ n^{1/2} \cdot \left(\int_{[0,1]\setminus I_n} E\Big(|V_n(t) - \widehat{(V_n)}^L_{S(Y_{T_n})}(t)|^p\Big)\, dt\right)^{1/p} \leq (c_p\, \nu_p)^{1/p} \cdot \|\alpha\|_\lambda.$$

Proof. Let P denote the distribution of Y_T. By independence of V and Y_T it holds

$$E\left(|V(t) - \widehat{V}^L_{S(Y_T)}(t)|^p\right) = \int_{\mathbb{R}^{k(k+1)+1}} E\left(|V(t) - \widehat{V}^L_{S(y)}(t)|^p\right)\, P(dy)$$

for each $0 \leq t \leq 1$. Clearly

$$E\left(|V(t) - \widehat{V}^L_{S(y)}(t)|^p\right) \leq E\left(|Z(t) - \widehat{Z}^L_{S(y)}(t)|^p\right),$$

such that

$$\begin{aligned}
&\int_{t_{i,k}}^{t_{i+1,0}} E\left(|V(t) - \widehat{V}^L_{S(Y_T)}(t)|^p\right)\, dt \\
&\qquad \leq \int_{\mathbb{R}^{k(k+1)+1}} \int_{t_{i,k}}^{t_{i+1,0}} E\left(|Z(t) - \widehat{Z}^L_{S(y)}(t)|^p\right)\, dt\, P(dy). \qquad (25)
\end{aligned}$$

Fix y and let

$$\max_{t_{i,k}\leq t\leq t_{i+1,0}} \alpha(t) = \alpha(\theta_i)$$

for some $\theta_i \in [t_{i,k},\, t_{i+1,0}]$. Lemma 2 yields

$$\begin{aligned}
&\int_{s_{i,j}(y)}^{s_{i,j+1}(y)} E\left(|Z(t) - \widehat{Z}^L_{S(y)}(t)|^p\right)\, dt \\
&\quad \leq \int_{s_{i,j}(y)}^{s_{i,j+1}(y)} c_p \cdot \left(E(Z(t) - \widehat{Z}^L_{S(y)}(t))^2\right)^{p/2}\, dt \\
&\quad \leq c_p \cdot \nu_p \cdot \alpha^p(\theta_i) \cdot (s_{i,j+1}(y) - s_{i,j}(y))^{1+p/2} \cdot (1 + c \cdot (s_{i,j+1}(y) - s_{i,j}(y))) \\
&\quad \leq c_p \cdot \nu_p \cdot \alpha^p(\theta_i) \cdot (s_{i,j+1}(y) - s_{i,j}(y)) \cdot (\Delta_i^{\max}(y))^{p/2} \cdot (1 + c \cdot \Delta_i^{\max}(y)))
\end{aligned}$$

for each $j = 0, \ldots, r_i(y)$. Thus

$$\int_{t_{i,k}}^{t_{i+1,0}} E\left(|Z(t) - \widehat{Z}^L_{S(y)}(t)|^p\right)\, dt \leq c_p \cdot \nu_p \cdot \frac{1}{k}\, \alpha^p(\theta_i) \cdot (\Delta_i^{\max}(y))^{p/2} + \frac{c}{k} \cdot (\Delta_i^{\max}(y))^{1+p/2}$$

and therefore

$$\begin{aligned}
&\sum_{i=0}^{k} \int_{t_{i,k}}^{t_{i+1,0}} E\left(|V(t) - \widehat{V}^L_{S(Y_T)}(t)|^p\right)\, dt \\
&\qquad \leq c_p \cdot \nu_p \cdot \frac{1}{k} \sum_{i=0}^{k} \alpha^p(\theta_i)\, E\,(\Delta_i^{\max}(Y_T))^{p/2} + c \cdot \max_{0\leq i\leq k} E\,((\Delta_i^{\max}(Y_T))^{1+p/2}.
\end{aligned}$$

by (25).

Note that $\lambda \cdot p/2 = p^2/(2+p)$. Now Lemma 4 yields

$$\begin{aligned}\limsup_{n\to\infty} &\left(\frac{n^{p/2}}{k_n}\sum_{i=0}^{k_n} \alpha^p(\theta_{i,n})\, E\left(\Delta_{i,n}^{\max}(Y_{T_n})\right)^{p/2}\right)\\ &\le \limsup_{n\to\infty}\left(\frac{1}{k_n}\sum_{i=0}^{k_n}\alpha^{2p/(2+p)}(\theta_{i,n})\cdot \max_{0\le i\le k_n} E\left(\left(n\cdot\alpha^\lambda(\theta_{i,n})\cdot\Delta_{i,n}^{\max}(Y_{T_n})\right)^{p/2}\right)\right)\\ &\le \left(\int_0^1 \alpha^\lambda(t)\,dt\right)^{p/\lambda}\end{aligned}$$

and

$$\lim_{n\to\infty} n^{p/2}\cdot \max_{0\le i\le k_n} E\left((\Delta_{i,n}^{\max}(Y_{T_n})\right)^{1+p/2} = 0,$$

which completes the proof. □

Proof of (24). Clearly

$$Y - \widehat{Y}^L_{S(Y_{T_n})} = \left(m - \widehat{m}^L_{S(Y_{T_n})}\right) + \left(U_n - (\widehat{U_n})^L_{S(Y_{T_n})}\right) + \left(V_n - (\widehat{V_n})^L_{S(Y_{T_n})}\right).$$

Let ω denote the modulus of continuity of m'. Then

$$|m(t) - \widehat{m}^L_{S(Y_T)}(t)| \le \Delta_i^{\max}(Y_T)\cdot\omega(1/k) \tag{26}$$

for every $t \in [t_{i,k}, t_{i+1,0}]$. Lemma 4 implies

$$\begin{aligned}\int_{[0,1]\setminus I_n} E\left(|m(t) - \widehat{m}^L_{S(Y_{T_n})}(t)|^p\right)\,dt &\le \max_{0\le i\le k_n} E(\Delta_i^{\max}(Y_{T_n}))^p\cdot\omega^p(1/k_n)\\ &\le c\cdot 1/n^p\cdot\omega^p(1/k_n).\end{aligned}$$

Now (24) follows from Lemma 7 and 8. □

6.5. **Upper bounds for integration.** Our methods for integration and reconstruction in L_1-norm basically coincide. The respective designs are the same, since $\lambda = 2/3$ in both cases. Formally the method for integration is given as

$$\widehat{Y}^\lambda_{N_n} = \int_0^1 \widehat{Y}^L_{S(Y_{T_n})}(t)\,dt$$

where the right-hand side is also defined with $\lambda = 2/3$.

As previously we separately deal with I_n and its complement in $[0,1]$, see (23). Moreover, we use the decomposition $Y = m + U_n + V_n$ on $[0,1]\setminus I_n$. Recall that $\widehat{Y}^L_{S(Y_{T_n})}$ coincides with $\widehat{Y}^L_{T_n}$ on I_n. From (D) and Lemma 2 we obtain

$$\begin{aligned}E&\left(\int_{I_n}(Y(t) - \widehat{Y}^L_{T_n}(t))\,dt\right)^2\\ &= \left(\int_{I_n}(m(t) - \widehat{m}^L_{T_n}(t))\,dt\right)^2 + \int_{I_n}\int_{I_n} E\left((Z(s) - \widehat{Z}^L_{T_n}(s))\cdot(Z(t) - \widehat{Z}^L_{T_n}(t))\right)\,ds\,dt\\ &\le c\cdot(k_n^2\,\delta_n\cdot\delta_n)^2 + c\cdot k_n^2\cdot\delta_n^3 + c\cdot k_n^4\cdot\delta_n^4\\ &\le c\cdot k_n^2\cdot\delta_n^3 \le c\cdot n^{-2-2\gamma}.\end{aligned}$$

By (26),

$$E\left(\int_{[0,1]\setminus I_n}(m(t)-\widehat{m}^L_{S(Y_{T_n})}(t))\,dt\right)^2 \le c\cdot\omega^2(1/k)\cdot\max_{0\le i\le k_n} E\left(\Delta^{\max}_{i,n}(Y_{T_n})\right)^2$$
$$\le c\cdot\omega^2(1/k)\cdot n^{-2},$$

where ω denotes the modulus of continuity of m'. Clearly the last two upper bounds tend to zero faster than n^{-2}. By Lemma 7 the same conclusion holds for the upper bound

$$E\left(\int_{[0,1]\setminus I_n}(U_n(t)-\widehat{(U_n)}^L_{S(Y_{T_n})}(t))\,dt\right)^2 \le c\cdot\int_{[0,1]\setminus I_n}E((U_n(t)-\widehat{(U_n)}^L_{S(Y_{T_n})}(t))^2\,dt.$$

It remains to show that

$$\limsup_{n\to\infty}\ n\cdot\left(E\Big(\int_{[0,1]\setminus I_n}(V_n(t)-\widehat{(V_n)}^L_{S(Y_{T_n})}(t))\,dt\Big)^2\right)^{1/2}\le\frac{1}{2\sqrt{3}}\cdot\|\alpha\|_{2/3}. \tag{27}$$

Proof of (27). Suppress the dependence on n and let P denote the distribution of Y_T. By independence of V and Y_T it holds

$$\begin{aligned}
&E\left(\int_{[0,1]\setminus I}(V(t)-\widehat{V}^L_{S(Y_T)}(t))\,dt\right)^2\\
&=\int_{\mathbb{R}^{k(k+1)+2}}\int_{[0,1]\setminus I}\int_{[0,1]\setminus I}E\left((V(s)-\widehat{V}^L_{S(y)}(s))\cdot(V(t)-\widehat{V}^L_{S(y)}(t))\right)\,ds\,dt\,P(dy)\\
&\le\int_{\mathbb{R}^{k(k+1)+2}}\int_{[0,1]\setminus I}\int_{[0,1]\setminus I}E\left((Z(s)-\widehat{Z}^L_{S(y)}(s))\cdot(Z(t)-\widehat{Z}^L_{S(y)}(t))\right)\,ds\,dt\,P(dy).
\end{aligned}$$

Fix y, let ω denote the modulus of continuity of $K_-^{(1,1)}$, and choose $\theta_i\in[t_{i,k},t_{i+1,0}]$ with

$$\max_{t_{i,k}\le t\le t_{i+1,0}}\alpha(t)=\alpha(\theta_i).$$

Lemma 2 yields

$$\begin{aligned}
&\int_{s_{i,j}(y)}^{s_{i,j+1}(y)}\int_{s_{\mu,\nu}(y)}^{s_{\mu,\nu+1}(y)}E\left((Z(s)-\widehat{Z}^L_{S(y)}(s))\cdot(Z(t)-\widehat{Z}^L_{S(y)}(t))\right)\,ds\,dt\\
&\quad\le\omega(1/k)\cdot(\max(s_{i,j+1}(y)-s_{i,j}(y),s_{\mu,\nu+1}(y)-s_{\mu,\nu}(y)))^4\\
&\quad\le\omega(1/k)\cdot((s_{i,j+1}(y)-s_{i,j}(y))\cdot(\Delta^{\max}_i(y))^3+(s_{\mu,\nu+1}(y)-s_{\mu,\nu}(y))\cdot(\Delta^{\max}_\mu(y))^3)
\end{aligned}$$

if $(\mu,\nu)\ne(i,j)$, and

$$\begin{aligned}
&\int_{s_{i,j}(y)}^{s_{i,j+1}(y)}\int_{s_{i,j}(y)}^{s_{i,j+1}(y)}E\left((Z(s)-\widehat{Z}^L_{S(y)}(s))\cdot(Z(t)-\widehat{Z}^L_{S(y)}(t))\right)\,ds\,dt\\
&\quad\le 1/12\cdot(s_{i,j+1}(y)-s_{i,j}(y))^3\cdot\alpha^2(s_{i,j})\cdot(1+c\cdot(s_{i,j+1}(y)-s_{i,j}(y)))\\
&\quad\le 1/12\cdot\alpha^2(\theta_i)\cdot(s_{i,j+1}(y)-s_{i,j}(y))\cdot(\Delta^{\max}_i(y))^2\cdot(1+c\cdot\Delta^{\max}_i(y)).
\end{aligned}$$

Recall that $r_i(y)$ is the number of points adaptively placed in the subinterval $[t_{i,k}, t_{i+1,0}]$. Hence

$$\begin{aligned}
&\int_{[0,1]\setminus I}\int_{[0,1]\setminus I} E\left((Z(s) - \widehat{Z}^L_{S(y)}(s)) \cdot (Z(t) - \widehat{Z}^L_{S(y)}(t)) \right) ds\, dt \\
&\quad \le \sum_{i=0}^{k} 1/12 \cdot \alpha^2(\theta_i) \cdot (t_{i+1,0} - t_{i,k}) \cdot (\Delta_i^{\max}(y))^2 \cdot (1 + c \cdot \Delta_i^{\max}(y)) \\
&\qquad + c \cdot \sum_{i=0}^{k} \sum_{j=0}^{r_i(y)} \omega(1/k) \cdot n \cdot (s_{i,j+1}(y) - s_{i,j}(y)) \cdot (\Delta_i^{\max}(y))^3 \\
&\quad \le \frac{1}{12\,k} \sum_{i=0}^{k} \alpha^2(\theta_i) \cdot (\Delta_i^{\max}(y))^2 + c/k \sum_{i=0}^{k} (\Delta_i^{\max}(y))^3 \\
&\qquad + c \cdot n \cdot \omega(1/k) \cdot \frac{1}{k} \sum_{i=0}^{k} (\Delta_i^{\max}(y))^3 .
\end{aligned}$$

Observing Lemma 4 we thus conclude that

$$\begin{aligned}
&E\left(\int_{[0,1]\setminus I} (V(t) - \widehat{V}^L_{S(Y_T)}(t))\, dt \right)^2 \\
&\quad \le \frac{1}{12\,k} \sum_{i=0}^{k} \alpha^2(\theta_i) \cdot E(\Delta_i^{\max}(Y_T))^2 + c/k \cdot (n \cdot \omega(1/k) + 1) \cdot \sum_{i=0}^{k} E(\Delta_i^{\max}(Y_T))^3 \\
&\quad \le \left(\frac{1}{12\,k} \sum_{i=0}^{k} \alpha^{2/3}(\theta_i) \right) \cdot \max_{0 \le i \le k} E(\alpha^{2/3}(\theta_i) \cdot \Delta_i^{\max}(Y_T))^2 + c \cdot (\omega(1/k)/n^2 + 1/n^3).
\end{aligned}$$

Clearly

$$\lim_{n\to\infty} n \cdot (\omega(1/k_n)/n^2 + 1/n^3)^{1/2} = 0$$

and

$$\lim_{n\to\infty} \left(1/k_n \sum_{i=0}^{k_n} \alpha^{2/3}(\theta_{i,n}) \right)^{1/2} = \left(\int_0^1 \alpha^{2/3}(t)\, dt \right)^{1/2}.$$

Finally, by Lemma 4,

$$\limsup_{n\to\infty}\ n \cdot \left(\max_{0 \le i \le k_n} E(\alpha^{2/3}(\theta_{i,n}) \cdot \Delta_{i,n}^{\max}(Y_{T_n}))^2 \right)^{1/2} \le \int_0^1 \alpha^{2/3}(t)\, dt,$$

whence (27) follows. □

Acknowledgments. We thank Norbert Hofmann and Erich Novak for helpful discussions.

References

Bates, R. A., Buck, R. J., Riccomagno, E. and Wynn, H. P. (1996). Experimental design and observation for large systems (with discussion). J. Roy. Statist. Soc. Ser. B **58** 77–94.

Benhenni, K. and Cambanis, S. (1992). Sampling designs for estimating integrals of stochastic processes. Ann. Statist. **20** 161–194.

Cohen, A. and d'Ales, J.-P. (1995). Nonlinear approximation of stochastic processes. In *Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, eds.) Lect. Notes in Statist. **103** 109–123. Springer, Berlin.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* Wiley, New York.

Dharmadhikari, S. W. and Jogdeo, K. (1969). Bounds on moments of certain random variables. Ann. Math. Statist. **40** 1506–1508.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaption by wavelet shrinkage. Biometrika **81** 425–455.

Eubank, R. L., Smith, P. L. and Smith, P. W. (1982). A note on optimal and asymptotically optimal designs for certain time series models. Ann. Statist. **10** 1295–1301.

Hájek, J. and Kimeldorf, G. S. (1974). Regression designs in autoregressive stochastic processes. Ann. Statist. **2** 520–527.

Hall, P. and Patil, P. (1996). Effect of threshold rules on performance of wavelet-based curve estimators. Statist. Sinica **6** 331–345.

Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics (with discussion). Scand. J. Statist. **21** 289–357.

Istas, J., and Laredo, C. (1994). Estimation d'intégrales de processus aléatoires à partir d'observations discrétisées. C. R. Acad. Sci. **319** 85–88.

Kimeldorf, G. S. and Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Statist. **41** 495–502.

J. R. Koehler and A. B. Owen, Computer Experiments, Preprint, 1995.

Müller-Gronbach, T. (1996a). Asymptotically optimal designs for approximating the path of a stochastic process with respect to the L^∞-norm. Tatra Mountains Math. Publ. **7** 87–95.

Müller-Gronbach, T. (1996b). Optimal designs for approximating the path of a stochastic process. J. Statist. Planning Inf. **49** 371-385.

Müller-Gronbach, T. (1996c). Optimal designs for approximating the path of a stochastic process with respect to a minimax criterion. Statistics **27** 279–296.

Müller-Gronbach, T. (1997). Hyperbolic cross designs for approximation of random fields. To appear in J. Statist. Planning Inf.

Müller-Gronbach, T. and Ritter, K. (1997). Uniform reconstruction of Gaussian processes. To appear in Stochastic Processes Appl.

Müller-Gronbach, T. and Schwabe, R. (1996). On optimal allocations for estimating the surface of a random field. Metrika **44** 239–258.

Novak, E. (1988). *Deterministic and Stochastic Error Bounds in Numerical Analysis.* Lect. Notes in Math. **1349**. Springer, Berlin.

Paskov, S. H. (1993). Average case complexity of multivariate integration for smooth functions. J. Complexity **9** 291–312.

Plaskota, L. (1996). *Noisy Information and Computational Complexity.* Cambridge University Press, Cambridge.

Ritter, K. (1996a). *Average Case Analysis of Numerical Problems.* Habilitationsschrift. Erlangen.

Ritter, K. (1996b). Asymptotic optimality of regular sequence designs. Ann. Statist. **24** 2081–2096.

Ritter, K. and Wasilkowski, G. W. (1996). Integration and L_2-approximation: average case setting with isotropic Wiener measure for smooth functions. To appear in Rocky Mountain J. Math.

Ritter, K., Wasilkowski, G. W. and Woźniakowski, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks-Ylvisaker conditions. Ann. Appl. Prob. **5** 518–540.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments (with discussion). Statist. Sci. **4** 409–435.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. Ann. Math. Statist. **37** 66-89.

Sacks, J. and Ylvisaker, D. (1970). Statistical design and integral approximation. In *Proc. 12th Bienn. Semin. Can. Math. Congr.* (R. Pyke, ed.) 115–136. Can. Math. Soc., Montreal.

Speckman, P. (1979). L_p approximation of autoregressive Gaussian processes. Technical Report, Department of Statistics, University of Oregon, Eugene.

Stein, M. (1990). Bounds on the efficiency of linear predictions using an incorrect covariance function. Ann. Statist. **18** 1116–1138.

Stein, M. (1995a). Predicting integrals of stochastic processes. Ann. Appl. Prob. **5** 158–170.

Stein, M. (1995b). Locally lattice sampling designs for isotropic random fields. Ann. Statist. **23** 1991–2012.

Su, Y. and Cambanis, S. (1993). Sampling designs for estimation of a random process. Stochastic Processes Appl. **46** 47–89.

Traub, J. F., Wasilkowski. G. W. and Woźniakowski, H. (1988). *Information-Based Complexity.* Academic Press, New York.

Wahba, G. (1974). Regression design for some equivalence classes of kernels. Ann. Statist. **2** 925–934.

Wahba, G. (1990). *Spline Models for Observational Data.* CBSM-NSF Regional Conf. Ser. Appl. Math. **59**. SIAM, Philadelphia.

Wasilkowski, G. W. (1993). Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure. Bull. Amer. Math. Soc. (N. S.) **28** 308–314. Full version J. Approx. Theory **77** (1994) 212–227.

Wasilkowski, G. W. and Gao, F. (1992). On the power of adaptive information for functions with singularities. Math. Comp. **58** 285–304.

Woźniakowski, H. (1991). Average case complexity of multivariate integration. Bull. Amer. Math. Soc. (N. S.) **24** 185–194.

Woźniakowski, H. (1992). Average case complexity of linear multivariate problems, Part 1: Theory, Part 2: Applications. J. Complexity **8** 337–372, 373–392.

1. Mathematisches Institut, Freie Universität Berlin, Arminallee 2–6, 14195 Berlin, Germany
E-mail address: gronbach@math.fu-berlin.de

Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstrasse 1 1/2, 91054 Erlangen, Germany
E-mail address: ritter@mi.uni-erlangen.de