*The Annals of Statistics*
1998, Vol. 26, No. 6, 2264–2288

# SPATIAL ADAPTION FOR PREDICTING RANDOM FUNCTIONS

BY THOMAS MÜLLER-GRONBACH AND KLAUS RITTER

*Freie Universität Berlin and Universität Passau*

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 two-stage designs that lead to asymptotically optimal methods. We show that every nonadaptive design is less efficient.

**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 $Y(t_1), \dots, Y(t_n)$ of a realization of $Y$ are used for:

1. Prediction of the integral $\int_D Y(t)\,dt$, called *integration*.
2. Prediction of $Y(t)$ for all $t \in D$, called *reconstruction*.

In this paper we study the univariate case, $D = [0, 1]$. 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 class of processes $Y$ having inhomogeneous local smoothness.

Our approach is motivated by several applications. Problems (1) and (2) 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 (1996), Bates, Buck, Riccomagno and Wynn (1996). Moreover, the random function approach is used in numerical analysis to complement the classical worst case approach; 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.

Received April 1997; revised July 1998.

*AMS* 1991 *subject classifications*. Primary 62M20, 62L03; secondary 60G15, 60G25.

*Key words and phrases.* Reconstruction, integral estimation, spatial adaption, asymptotically optimal designs, unknown smoothness, nonparametric model.

Usually a properly chosen design $(t_1, \dots, t_n) \in D^n$ 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 $Y(t_1), \dots, Y(t_{i-1})$. 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

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

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

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

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) allows for regions of different spatial variability of $Y$. We present an adaptive two-stage method which detects these regions and places additional points accordingly. We prove that our method is asymptotically optimal. Furthermore, we show that every nonadaptive method is less efficient.

The design problem, that is, the optimal choice of observation points, 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 adaption does not help. See, for example, Sacks and Ylvisaker (1966, 1970), Benhenni and Cambanis (1992), Stein (1995a) and Ritter (1996b) for integration. Reconstruction is analyzed in Speckman (1979), Su and Cambanis (1993), Müller-Gronbach (1996a, b), Ritter (1996a), and Müller-Gronbach and Ritter (1997a). 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, for example, Stein (1995b), Ritter (1996a), Ritter, Wasilkowski and Woźniakowski (1995) and Müller-Gronbach (1997).

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 $X$, that is,

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

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 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 cl $\Omega_j$ for $j \in \{+,-\}$. By $L_j$ we denote the extension of $L$ to $[0,1]^2$ which is continuous on cl $\Omega_j$ and on $[0,1]^2 \setminus \text{cl}\, \Omega_j$.

The following smoothness conditions were introduced by Sacks and Ylvisaker (1996) and thereafter studied in many papers, some of which are cited in the introduction.

(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 cl $\Omega_+$ as well as to 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 $\|\cdot\|_R$.

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 $Y$ is Hölder continuous in quadratic mean with exponent $1/2$. This degree of smoothness frequently occurs in geostatistics and computer experiments. The local Hölder constant of $Y$ is given by

$$\alpha(t) = g(t)\sqrt{f'(t)}\,. \tag{2}$$

Thus

$$\lim_{s\to t}\frac{\left(E(Y(s)-Y(t))^2\right)^{1/2}}{|s-t|^{1/2}}=\alpha(t). \tag{3}$$

See Lemma 1 for a more detailed estimate. Property (3) is too weak for analysis, in particular for proving lower bounds. Hence we use slightly stronger conditions (A)–(F).

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

$$\begin{aligned} R(s,t) &= (1-|s-t|)/2,\\ m(t) &= 2t,\\ g(t) &= 1/2+5t^2,\\ f(t) &= (\tanh(15t-5)+1)/2. \end{aligned} \tag{4}$$

The corresponding function $\alpha$ is shown in Figure 2.

**3. The adaptive method.** Let $\hat{Y}_n$ denote any method for reconstruction that is based on $n$ observations. Formally, $\hat{Y}_n$ 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

$$\hat{Y}_n = \phi(y_1,\dots,y_n),$$

where

$$y_1 = Y(t_1)$$

and

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

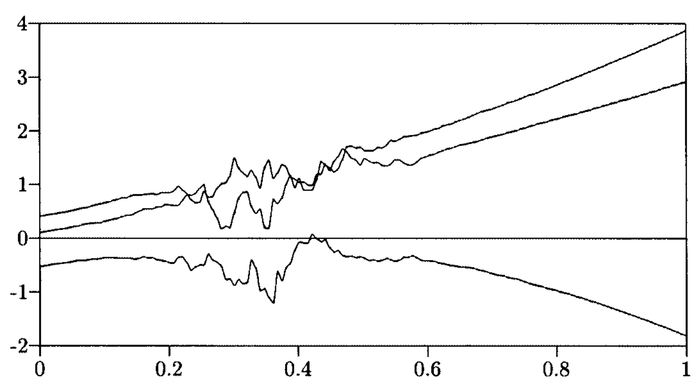


FIG. 1. *Three realizations of a process $Y$.*

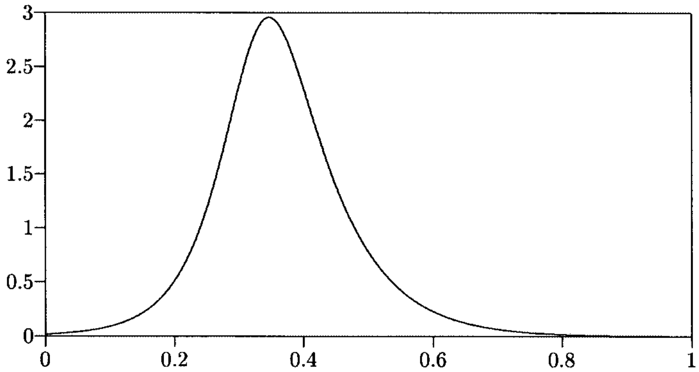


FIG. 2. *The corresponding function $\alpha$.*

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

A method is nonadaptive, by definition, if all observations could be 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 $\phi$ is irrelevant for this classification. Let us mention that the notion of adaptivity is sometimes used in a different sense; see, for example, Donoho and Johnstone (1994).

We measure the distance between a realization of $Y$ and the corresponding prediction in $L_p$-norm $\|\cdot\|_p$ on $D$. The error of $\hat{Y}_n$ is defined by

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

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 $\phi$ has range $\mathbb{R}$ and the error is defined by

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

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 $\alpha^\lambda$ of the local Hölder constant $\alpha$; see (2). We take

$$(5)\qquad \lambda = \begin{cases} (1/2 + 1/p)^{-1}, & \text{for reconstruction in } L_p\text{-norm},\\ 2/3, & \text{for integration.}\end{cases}$$

Then we select additional points adaptively with "density" proportional to the estimate $\hat{\alpha}$ of $\alpha^\lambda$. Finally, we use piecewise linear interpolation of all 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

$$0 < \delta < 1/k^2. \tag{6}$$

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

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

where $i = 1, \dots, k$ and $j = 0, \dots, k$. Additionally, we use the points $t_{0,k} = 0$ and $t_{t+1,0} = 1$. The resulting nonadaptive design

$$T = (t_{0,k}, \dots, t_{k+1,0}) \tag{7}$$

consists of $k(k+1)+2$ sample points.

Sampling $Y$ at $T$ yields data

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

which are used for estimating the function $\alpha^{\lambda}$. 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 $\alpha((2i-1)/2k)^2\delta$; see Lemma 1. Therefore, a natural choice for estimating $\alpha^{\lambda}$ at $(2i-1)/2k$ is

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

where $c_\lambda$ denotes the absolute moment of order $\lambda$ 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 $\hat{\alpha}_i(y)$ getting too small, we will use

$$\tilde{\alpha}_i(y) = \max(\hat{\alpha}_i(y), \varepsilon)$$

instead, where

$$0 < \varepsilon \le 1.$$

Estimation of $\alpha$ is also studied in Istas (1996).

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

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

together with an integer

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

which is roughly the total number of points $s_{i,j}(y)$. We estimate $\alpha^{\lambda}$ on $J_0 \cup \cdots \cup J_k$ by piecewise linear interpolation of the values $\tilde{\alpha}_i(y)$, and we use this estimate to construct an adaptive design. See Figure 3 for an example.

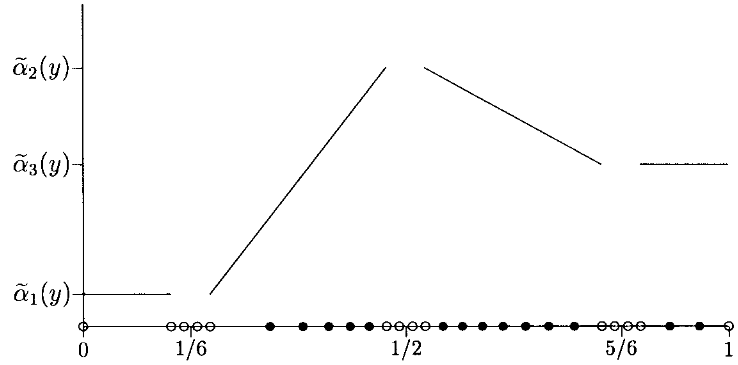


FIG. 3. *Estimate for $\alpha$, adaptive design with $k = 3$, $\delta = 1/50$ and $n = 16$.*

More precisely,

$$r_i(y) = \left\lfloor n \frac{\tilde\alpha_i(y) + \tilde\alpha_{i+1}(y)}{2\sum_{j=1}^k \tilde\alpha_j(y)} \right\rfloor$$

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 $\tilde\alpha_i(y)$ and $\tilde\alpha_{i+1}(y)$. Formally,

$$\int_{t_{i,k}}^{s_{i,j}(y)} (\tilde\alpha_i(y)(t_{i+1,0} - t) + \tilde\alpha_{i+1}(y)(t - t_{i,k}))\, dt = j \frac{(\tilde\alpha_i(y) + \tilde\alpha_{i+1}(y))(1/k - k\delta)^2}{2(r_i(y) + 1)}$$

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

$$r_0(y) = \left\lfloor n \frac{\tilde\alpha_1(y)}{2\sum_{j=1}^k \tilde\alpha_j(y)} \right\rfloor \quad \text{and} \quad r_k(y) = \left\lfloor n \frac{\tilde\alpha_k(y)}{2\sum_{j=1}^k \tilde\alpha_j(y)} \right\rfloor.$$

Therefore

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

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

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

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

Summarizing, the first stage depends on the parameters $k$ and $\delta$, and the second stage depends on the parameters $\varepsilon$ and $n$, as well as on $\lambda$, defined in

(5). The whole design consists of

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

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 $\hat{Y}_N^\lambda$.

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

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

In order to obtain asymptotic results we study sequences of methods $\hat{Y}_{N_n}^\lambda$ which are defined by sequences of the respective parameters $k_n$, $\delta_n$ and $\varepsilon_n$. 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 $\alpha^\lambda$ even for a small sample size. Hence we take

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

where

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

Hereby the maximum size $N_n$ of the adaptive design satisfies

$$N_n \approx n.$$

The parameters $\delta_n$ must satisfy

(ii) $$\delta_n \le c(1/n)$$

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 $\varepsilon_n$ according to the quality of our estimate of $\alpha^\lambda$. Due to Lemma 3, the corresponding error is of order at most $n^{-(1/2-\gamma)/2}$. Therefore we take

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

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

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

We extend the definition of $\|\cdot\|_p$ to $0 < p \le \infty$, 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 $\alpha$ be given by* (2). *If $\lambda$ is selected according to* (5) *and* (i)–(iii) *are satisfied, then*

$$e_2\left(\hat{Y}^{\lambda}_{N_n}, \mathrm{Int}\right) \approx \frac{1}{2\sqrt{3}}\|\alpha\|_{\lambda} n^{-1},$$

$$e_p\left(\hat{Y}^{\lambda}_{N_n}, L_p\right) \approx (c_p \nu_p)^{1/p}\|\alpha\|_{\lambda} n^{-1/2}, \qquad 1 \le p < \infty.$$

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

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

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

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

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

$$\int_0^{t_i^{(n)}} \psi(t)\, dt = \frac{i-1}{n-1}\int_0^1 \psi(t)\, dt, \qquad i = 1, \dots, n. \tag{8}$$

Then the trapezoidal rules $\hat{Y}_n$ that are based on $Y(t_1^{(n)}), \dots, Y(t_n^{(n)})$ are asymptotically optimal, that is,

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

A sequence of nonadaptive designs $(t_1^{(n)}, \dots, t_n^{(n)})$ that is defined by (8) for some fixed positive density $\psi$ 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

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

where $\lambda = (1/2 + 1/p)^{-1}$, as in (5). A regular sequence of designs together with piecewise linear interpolation is asymptotically optimal. The optimal density is given by

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

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

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

COROLLARY 1. *The methods* $\hat{Y}^{\lambda}_{N_n}$ *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 $\hat{Y}_n$, 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

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

and

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

Therefore,

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

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 $\hat{Y}^{e}_n$ denote the optimal method, either for integration or reconstruction, that is based on $t_i^{(n)} = (i-1)/(n-1)$. Then

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

and

$$\lim_{n\to\infty} \frac{e_p\left(\hat{Y}^e_{N_n}, L_p\right)}{e_p\left(\hat{Y}^\lambda_{N_n}, L_p\right)} = \frac{\|\alpha\|_p}{\|\alpha\|_{(1/2+1/p)^{-1}}}$$

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

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

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

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

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

with

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

Note that $Y^{(r)}$ is essentially a local stationary process; see Berman (1974).

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 $\alpha^\lambda$ 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. The conjecture is motivated by results for the case when $m$ is sufficiently smooth, $f(t) = t$ and $g$ is known. See Benhenni and Cambanis (1992), Stein (1995a), Ritter (1996a, b), Istas and Laredo (1997), Seleznev (1997) and Müller-Gronbach and Ritter (1997a).

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 $\varphi$ of the random field $Y$ 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 $\hat{Y}_{n,i}$ denote the best linear predictor that is based on $T_n$ and $(m_i, K_i)$. Stein shows that $\hat{Y}_{n,1}$ is asymptotically efficient, that is,

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

Here $E$ denotes expectation under the correct second-order structure $(m_0, K_0)$. 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 analyze only prediction on the basis of a fixed sequence of designs. Instead, estimating $\alpha^{\lambda}$, which is the essential ingredient of $K_0$ and constructing an asymptotically optimal adaptive design form the main part of our analysis.

Compatibility of $(m_0, K_0)$ and $(m_1, K_1)$ is rather restrictive. For instance, if

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

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.

**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 (4).

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

$$e_2\left(\hat{Y}^{e}_{N_n}, L_2\right) / e_2\left(\hat{Y}^{\lambda}_{N_n}, L_2\right)$$

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.

Figure 4 illustrates 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 $\varepsilon$ turned out to be of minor importance in this example, whence we take $\varepsilon = 0$. For every choice of $n$ and $k$ we use 50 simulations of the process $Y$ to determine approximately the error of the respective variant of the adaptive method.

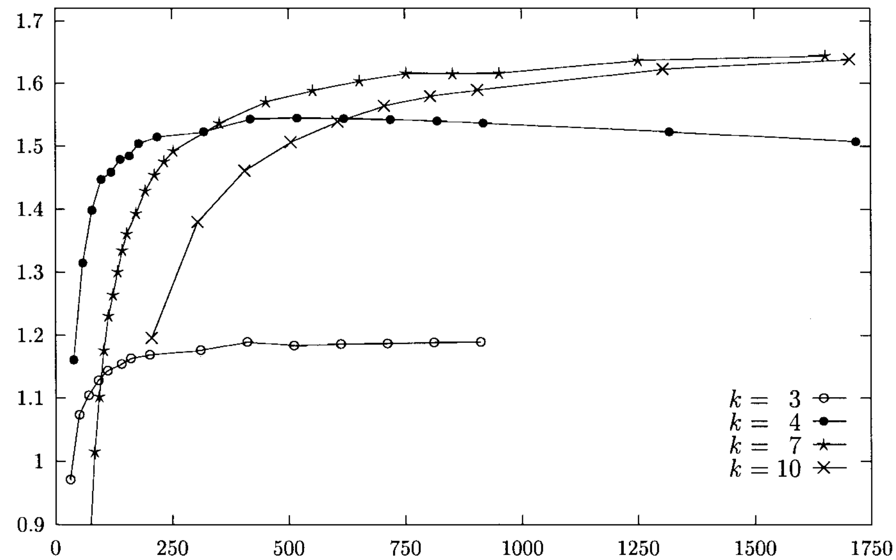


FIG. 4. *Efficiency versus total number $N_n$ of observations.*

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 $\delta$, 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 $\delta$ with increasing $n$. If $k$ is still fixed, one cannot reach the efficiency $1.720\ldots$ . Thus we also have to increase $k$ with increasing $n$.

Suppose that the estimate $\tilde{\alpha}_i$ of $\alpha$ 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 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 $T = (t_1, \ldots, t_n)$ denote an arbitrary $n$-point design. Then

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

is the vector of corresponding observations. Furthermore, $\hat{Y}_T^L$ denotes the piecewise linear interpolation of $Y$ based on $Y_T$.

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

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

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,

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

denotes the covariance kernel of the process $Z$.

6.1. *Preliminary estimates.* Using the assumptions (A)–(F), we obtain the following estimates for increments of the process $Z$.

LEMMA 1. *Let*

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

*and*

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

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

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

*if* $s_2 \le s_3$, *and*

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

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

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. Furthermore, Lemma 1 yields an upper bound for the error due to piecewise linear interpolation of the process $Z$.

LEMMA 2. *Let*

$$M(u,v) = E\Big(\Big(Z(u) - \hat{Z}^L_{(s_1, s_2)}(u)\Big)\Big(Z(v) - \hat{Z}^L_{(s_3, s_4)}(v)\Big)\Big),$$

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

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

*if* $s_2 \le s_3$, *and*

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

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

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

For proofs of Lemmas 1 and 2, we refer to Müller-Gronbach and Ritter (1997b).

6.2. *Properties of the adaptive design.* Henceforth let $T$ denote the nonadaptive design that is defined by (7). Recall that the estimate of the function $\alpha^\lambda$ on the subinterval $J_i = [t_{i,k}, t_{i+1,0}]$ depends on $\hat{\alpha}_i(Y_T)$ and $\hat{\alpha}_{i+1}(Y_T)$, where

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

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

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

*for any $q \ge 1$.*

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

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

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

$$\hat{\alpha}_i(Y_T) = \frac{1}{k}\sum_{j=1}^{k} |V_j|^\lambda. \tag{9}$$

Lemma 1 and (6) yield

$$\begin{aligned} \left|\mathrm{Var}(V_j) - a\right| &\le c/\delta \left|E\left(Z(t_{i,j}) - Z(t_{i,j-1})\right)^2 - \alpha^2(t_{i,j-1})\delta\right| \\ &\quad + c\left|\alpha^2(\theta) - \alpha^2(t_{i,j-1})\right| \\ &\le c(\delta + c)(1/k) \le c(1/k) \end{aligned} \tag{10}$$

and

$$\left|\mathrm{Cov}(V_j, V_l)\right| \le c\delta \le c(1/k^2) \tag{11}$$

if $j \ne l$.

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

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

Then (10) and (11) yield

$$|x^\top Cx - x^\top Bx| \le c(1/k)\, x^\top Bx \tag{12}$$

for all $x = (x_1, \dots, x_k)^\top \in \mathbb{R}^k$. For matrices $A, B \in \mathbb{R}^{k\times k}$ we write $A \ll B$ if $B - A$ is positive semidefinite. By (12) we have

$$(1 - \tilde{c}/k)B \ll C \ll (1 + \tilde{c}/k)B \tag{13}$$

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

$$\tilde{V} = \left(\tilde{V}_1, \dots, \tilde{V}_k\right) \sim N(E(V), (1 + \tilde{c}/k)B),$$

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

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

with respect to the normal distribution $N(\mu, \Sigma)$ on $\mathbb{R}^k$. Using (9) and (13), we get

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

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

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

and therefore

$$E\left( \left| \hat{\alpha}_i(Y_T) - (1/k) \sum_{j=1}^{k} E\left(|\tilde{V}_j|^\lambda\right) \right|^q \right) \leq c\left(1/k^{q/2}\right).$$

It remains to show that

$$\left| E\left( (1/k) \sum_{j=1}^{k} |\tilde{V}_j|^\lambda \right) - \alpha^\lambda(\theta) \right| \leq c\left(1/k^{1/2}\right). \tag{14}$$

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

$$E\left(|\tilde{V}_j|^\lambda\right) - \alpha^\lambda(\theta) = F\left(E(V_j), \operatorname{Var}\left(\tilde{V}_j\right)\right),$$

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

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

Together with (10) this implies that $(E(V_j), \operatorname{Var}(\tilde{V}_j))$ belongs to some compact and convex subset of $\mathbb{R} \times (0, \infty)$ if $k$ is sufficiently large. Moreover,

$$\left| F\left(E(V_j), \operatorname{Var}\left(\tilde{V}_j\right)\right) - F\left(0, \operatorname{Var}\left(\tilde{V}_j\right)\right) \right| \leq c(1/k).$$

Lemma 1 and the Lipschitz continuity of $\alpha^2$ imply

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

Combining the last two estimates we obtain

$$
\left|E\Big(|\tilde{V}_j|^\lambda\Big) - \alpha^\lambda(\theta)\right| \le c(1/k),
$$

which proves (14). □

In the sequel we derive asymptotic estimates, taking into account the properties (i)–(iii) of the parameters $k_n$, $\delta_n$ and $\varepsilon_n$. The corresponding notation $T_n$, $\hat{\alpha}_{i,n}$, $J_{i,n}$ and so on, 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 $\alpha^\lambda$. 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

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

denote the maximum distance between two consecutive points in $J_i$.

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

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

PROOF. We have

$$
\big(E(\Delta_i^{\max}(Y_T))^q\big)^{1/q} \le c/(n\varepsilon); \tag{15}
$$

see Müller-Gronbach and Ritter (1997b).

Let $\theta \in J_i$ for $1 \le i \le k-1$. Then

$$\alpha^\lambda(\theta)\Delta_i^{\max}(y) \le \left(\left|\alpha^\lambda(\theta) - \hat{\alpha}_i(y)\right| + \left|\alpha^\lambda(\theta) - \hat{\alpha}_{i+1}(y)\right|\right)\Delta_i^{\max}(y) + \frac{1}{nk}\sum_{j=1}^{k}\tilde{\alpha}_j(y).$$

Clearly, the same estimate holds for $i = 0$ and $i = k$. Lemma 3 and (15) yield

$$\left(E\left(\left|\alpha^\lambda(\theta) - \hat{\alpha}_l(Y_T)\right|\Delta_i^{\max}(Y_T)\right)^q\right)^{1/q} \le c(1/k^{1/2})(1/(n\varepsilon))$$

for $l = i+1$. Furthermore,

$$\left(E\left|\frac{1}{k}\sum_{j=1}^{k}\tilde{\alpha}_j(Y_T)\right|^q\right)^{1/q}$$

$$\le \varepsilon + \frac{1}{k}\sum_{j=1}^{k}\left(E\left|\hat{\alpha}_j(Y_T) - \alpha^\lambda\left(\frac{2j-1}{2k}\right)\right|^q\right)^{1/q} + \frac{1}{k}\sum_{j=1}^{k}\alpha^\lambda\left(\frac{2j-1}{2k}\right)$$

$$\le \varepsilon + c\frac{1}{k^{1/2}} + \frac{1}{k}\sum_{j=1}^{k}\alpha^\lambda\left(\frac{2j-1}{2k}\right).$$

Therefore

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

The result now follows, observing that $k_n^{1/2}\varepsilon_n$ tends 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

$$U_n(t) = E\left(Z(t)|Z_{T_n}\right)$$

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 $W\colon [0,1]^2 \to \mathbb{R}$ by

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

if

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

and $W(s,t) = 0$ otherwise. Next, let

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

and put

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

Finally, put

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

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

Observing

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

the following representation of the process $U$ can easily be derived from equation (3.2) in Müller-Gronbach (1996b).

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

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

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

$$\begin{aligned} &\left| \mu^{(0,1)}(x,s) - \mu^{(0,1)}(x,t) \right| \\ &\quad \le c\big(\omega(\Delta) + k^2\Delta\big)\Big( |x_{i+1,0}| + |x_{i,k}| + k^{-2}\big(x^\top R_{f(T)}^{-1} x\big)^{1/2} \Big), \end{aligned}$$

*where* $\omega$ *denotes the modulus of continuity of* $g'$.

See Müller-Gronbach and Ritter (1997b) for the proof.

6.4. *Upper bounds for reconstruction.* Let

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

denote the adaptive design that is defined in Section 3. Here $T_n$ denotes the nonadaptive part [see (7)], which consists of points from the set

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

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 $\hat{Y}^{\lambda}_{N_n}$ is defined by

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

where $\lambda = (1/2 + 1/p)^{-1}$ according to (5).

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

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

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

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

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

$$\limsup_{n \to \infty} n^{1/2} \left( \int_{[0,1] \setminus I_n} E\left( \left| Y(t) - \hat{Y}^{L}_{S(Y_{T_n})}(t) \right|^p \right) dt \right)^{1/p} \le (c_p \nu_p)^{1/p} \|\alpha\|_{\lambda}. \tag{16}$$

In the proof of (16) we use the decomposition $Y = m + U_n + V_n$ that was introduced in Section 6.3. First we analyze the piecewise linear interpolation of $U_n$, based on the adaptive design $S(Y_{T_n})$.

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

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

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

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

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

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

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

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

for every $t \in [t_{i,k}, t_{i+1,0}]$, where

$$B(x) = |x_{i+1,0}| + |x_{i,k}| + k^{-2}\big(x^\top R^{-1}_{f(T)} x\big)^{1/2}$$

and

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

We thus conclude that

$$\begin{aligned}
&\Big(E\Big(\Big|U(t) - \hat{U}^L_{S(Y_T)}(t)\Big|^p\Big)\Big)^{1/p} \\
&\qquad \le c\Big(\Big(E\big(A(X_{f(T)})\big)^{2p}\Big)^{1/2p}\Big)\Big(\Big(E\big(B(X_{f(T)})\big)^{2p}\Big)^{1/2p}\Big).
\end{aligned}$$

Clearly

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

and

$$\Big(E\big(X^\top_{f(T)} R^{-1}_{f(T)} X_{f(T)}\big)^p\Big)^{1/2p} \le c(k(k+1)+2) \le ck^2.$$

Hence

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

Furthermore, by Lemma 4,

$$\begin{aligned}
\Big(E\big(A(X_{f(T)})\big)^{2p}\Big)^{1/2p} &\le \omega(1/k)c(1/n) + k^2 c(1/n^2) \\
&\le c(1/n)\big(\omega(1/k) + 1/n^{2\gamma}\big).
\end{aligned}$$

Summarizing, we have

$$\Big(E\Big(\Big|U(t) - \hat{U}^L_{S(Y_T)}(t)\Big|^p\Big)\Big)^{1/p} \le c(1/n)\big(\omega(1/k) + 1/n^{2\gamma}\big),$$

which completes the proof. □

Next, we turn to the error process $V_n$.

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

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

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

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

for each $0 \le t \le 1$. Clearly,

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

such that

$$\begin{aligned}\int_{t_{i,k}}^{t_{i+1,0}} & E\left(\left|V(t) - \hat{V}^L_{S(Y_T)}(t)\right|^p\right) dt \\ & \le \int_{\mathbb{R}^{k(k+1)+1}} \int_{t_{i,k}}^{t_{i+1,0}} E\left(\left|Z(t) - \hat{Z}^L_{S(y)}(t)\right|^p\right) dt\, P(dy).\end{aligned} \tag{17}$$

Fix $y$ and let

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

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

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

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

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

and therefore

$$\sum_{i=0}^{k} \int_{t_{i,k}}^{t_{i+1,0}} E\Big(\big|V(t) - \hat{V}^L_{S(Y_T)}(t)\big|^p\Big)\,dt$$
$$\le c_p \nu_p \frac{1}{k} \sum_{i=0}^{k} \alpha^p(\theta_i) E\big(\Delta_i^{\max}(Y_T)\big)^{p/2} + c \max_{0\le i\le k} E\big(\Delta^{\max}(Y_T)\big)^{1+p/2}.$$

by (17).

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

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

and

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

which completes the proof. □

Proof of (16). Clearly

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

Let $\omega$ denote the modulus of continuity of $m'$. Then

$$\big|m(t) - \hat{m}^L_{S(Y_T)}(t)\big| \le \Delta_i^{\max}(Y_T)\,\omega(1/k)$$

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

$$\int_{[0,1]\setminus I_n} E\Big(\big|m(t) - \hat{m}^L_{S(Y_{T_n})}(t)\big|^p\Big)\,dt \le \max_{0\le i\le k_n} E\big(\Delta_i^{\max}(Y_{T_n})\big)^p \omega^p(1/k_n)$$
$$\le c(1/n^p)\,\omega^p(1/k_n).$$

Now (16) 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

$$\hat{Y}^{\lambda}_{N_n} = \int_0^1 \hat{Y}^L_{S(Y_{T_n})}(t)\,dt$$

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

The proof is similar to the proof of the upper bounds for reconstruction in the previous section, hence we omit it here. For a detailed analysis, see Müller-Gronbach and Ritter (1997b).

**Acknowledgments.** We thank Norbert Hofmann, Erich Novak and Oleg Seleznev 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.

BERMAN, S. M. (1974). Sojourns and extremes of Gaussian processes. *Ann. Probab.* **2** 999–1026.

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.

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

ISTAS, J. (1996). Estimating the singularity function of a Gaussian process with applications. *Scand. J. Statist.* **23** 581–595.

ISTAS, J. and LAREDO, C. (1997). Estimating functionals of a stochastic process. *Adv. in Appl. Probab.* **29** 249–270.

KOEHLER, J. R. and OWEN, A. B. (1996). Computer Experiments. In *Handbook of Statistics* **13** (S. Gosh and C. R. Rao, eds.) **13** 261–308. North Holland, Amsterdam.

MÜLLER-GRONBACH, T. (1996a). Optimal designs for approximating the path of a stochastic process. *J. Statist. Plann. Inference* **49** 371–385.

MÜLLER-GRONBACH, T. (1996b). 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. *J. Statist. Plann. Inference* **49** 371–385.

MÜLLER-GRONBACH, T. and RITTER, K. (1997a). Uniform reconstruction of Gaussian processes. *Stochastic Process. Appl.* **69** 55–70.

MÜLLER-GRONBACH, T. and RITTER, K. (1997b). Spatial adaption for predicting random functions. Freie Univ. Berlin, Fachbereich Mathematik, Preprint A-6-97.

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

PLASKOTA, L. (1996). Noisy Information and Computational Complexity. Cambridge Univ. Press.

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., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1995). Multivariate integration and approximation for random fields satisfying Sacks–Ylvisaker conditions. *Ann. Appl. Probab.* **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 *Proceedings Twelfth Biennial Seminar of the Canadian Mathematics Congress* (R. Pyke, ed.) 115–136. Canad. Math. Soc., Montreal.

SELEZNEV, O. (1997). Sampling designs for linear approximation of a random process. Technical report, Dept. Mathematical Statistics, Univ. Lund.

SPECKMAN, P. (1979). $L_p$ approximation of autoregression Gaussian processes. Technical report, Dept. Statistics, Univ. Oregon.

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. Probab.* **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 Process. Appl.* **46** 47–89.

TRAUB, J. F., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1988). *Information-Based Complexity*. Academic Press, New York.

MATHEMATISCHES INSTITUT
FREIE UNIVERSITÄT BERLIN
ARMINALLEE 2–6
14195 BERLIN
GERMANY
E-MAIL: gronbach@math.fu-berlin.de

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK
UNIVERSITÄT PASSAU
94030 PASSAU
GERMANY
E-MAIL: ritterk@fmi.uni-passau.de