*The Annals of Applied Probability*
1995, Vol. 5, No. 2, 518–540

# MULTIVARIATE INTEGRATION AND APPROXIMATION FOR RANDOM FIELDS SATISFYING SACKS–YLVISAKER CONDITIONS

BY KLAUS RITTER,[1] GRZEGORZ W. WASILKOWSKI[2] AND HENRYK WOŹNIAKOWSKI[3]

*Universität Erlangen–Nünberg, University of Kentucky and Columbia University*

We present sharp bounds on the minimal errors of linear estimators for multivariate integration and $L_2$-approximation. This is done for a random field whose covariance kernel is a tensor product of one-dimensional kernels that satisfy the Sacks–Ylvisaker regularity conditions.

**1. Introduction.** We study multivariate integration and $L_2$-approximation for random fields $Y$ that are defined on the $d$-dimensional unit cube $D = [0,1]^d$ and that have mean zero and known covariance kernel $K$. We assume that $K$ is at least continuous, and hence we may assume that $Y$ is a measurable random field whose realizations are in $L_2(D)$ with probability 1. For integration we want to estimate the integral $\int_D Y(t)\,dt$, whereas for $L_2$-approximation we want to estimate the values $Y(t)$ for all $t$, and we study the distance of the estimate and the realization of the field in the space $L_2(D)$.

For both problems we mainly consider linear estimators that use $n$ observations of the random field. These estimators are of the form

$$I_n(Y) = \sum_{i=1}^{n} Y(t_i) \cdot a_i \quad \text{and} \quad A_n(Y) = \sum_{i=1}^{n} Y(t_i) \cdot g_i$$

with sampling points $t_i \in D$ as well as real weights $a_i$ for integration and functions $g_i \in L_2(D)$ for $L_2$-approximation. In both cases we define the errors in the mean square sense

$$e(I_n, K) = \left(E\left(\int_D Y(t)\,dt - I_n(Y)\right)^2\right)^{1/2}$$

and

$$e(A_n, K) = (E\|Y - A_n(Y)\|_2^2)^{1/2} = \left(\int_D E(Y(t) - A_n(Y)(t))^2\,dt\right)^{1/2},$$

where $E$ denotes expectation.

Received September 1993; revised July 1994.
[1]Supported in part by the Deutsche Forschungsgemeitschaft (DFG).
[2]Supported by NSF Grant CCR-91-14042.
[3]Supported by NSF Grant IRI-92-12597 and AFO.
*AMS* 1991 *subject classifications*. 41A50, 41A55, 41A63, 62H12, 62G05.
*Key words and phrases*. Multivariate stochastic processes, integration, approximation, optimal linear estimators.

Each finite set of sampling points is called a design. For a fixed design it is well known that the best linear estimators are defined by orthogonal projections in the Hilbert space generated by $Y$. More precisely, let $C = (K(t_i, t_j))_{i,j}$ denote the $n \times n$ covariance matrix of $Y(t_1), \dots, Y(t_n)$, let $b = (\int_D K(t, t_i)\, dt)_i$ denote the $n \times 1$ vector of covariances of $\int_D Y(t)\, dt$ and $Y(t_i)$ and let $z(t) = (K(t, t_i))_i$ be an $n \times 1$ vector. Then any solution $a = (a_1, \dots, a_n)^T$ of $Ca = b$ defines an optimal linear estimator for integration. Analogously, any solution $g(t) = (g_1(t), \dots, g_n(t))^T$ of $Cg(t) = z(t)$ defines an optimal linear estimator for $L_2$-approximation.

The design problem of how to choose $n$ sampling points optimally, that is, with minimal error in the class of all linear estimators using $n$ observations, is in general much harder.

Equivalent formulations of the design problem are well known. See, for example, Sacks and Ylvisaker (1970b), Cambanis (1985) and Wahba (1990). In particular, it is known that for integration the error $e(I_n, K)$ coincides with the maximal error of $I_n$ in the unit ball of the reproducing kernel Hilbert space $H(K)$, that is,

$$e(I_n, K) = \sup_{\substack{f \in H(K) \\ \|f\|_K \le 1}} \left| \int_D f(t)\, dt - \sum_{i=1}^{n} f(t_i) \cdot a_i \right|. \tag{1.1}$$

Similarly, for $L_2$-approximation we have

$$e(A_n, K)^2 = \int_D \sup_{\substack{f \in H(K) \\ \|f\|_K \le 1}} \left( f(t) - \sum_{i=1}^{n} f(t_i) \cdot g_i(t) \right)^2 dt. \tag{1.2}$$

Moreover, the design problem for integration is equivalent to a design problem in a linear regression model. Finally, estimating the integral is closely related to a signal detection problem. See Cambanis and Masry (1983).

The majority of the results for design problems are for the univariate case $d = 1$. For example, the optimal $n$-point design for Brownian motion is $t_i = 2i/(2n+1)$ for integration [see Suldin (1959, 1960)] and $t_i = 3i/(3n+1)$ for $L_2$-approximation [see Lee (1986)]. The minimal error is $1/(\sqrt{3}\,(2n+1))$ for integration and $1/\sqrt{2(3n+1)}$ for $L_2$-approximation.

In a series of papers, Sacks and Ylvisaker (1966, 1968, 1970b) introduced regularity conditions for the covariance kernel $K$ and studied the design problem for weighted integration. Modifications of these conditions for the design problem for weighted integration and $L_2$-approximation are used in Wahba (1971), Hájek and Kimeldorf (1974), Speckman (1979), Eubank, Smith and Smith (1982), Benhenni and Cambanis (1992), Su and Cambanis (1993) and Müller-Gronbach (1993). The conditions assure that the stochastic process has exactly $r \in \mathbb{N}_0$ mean square derivatives. Strongly asymptotically optimal designs and sharp error bounds are obtained for weighted integration and approximation. There are no such results yet for the multivariate case $d \ge 2$.

The goal of this paper is to find sharp bounds on the minimal errors of linear estimators for the multivariate case. Obviously the minimal errors depend on

the covariance kernel $K$. For smooth $K$, the minimal errors go quickly to zero as a function of $n$; however, it may happen that for nonsmooth $K$, the minimal error is zero even for $n = 1$; see Ritter, Wasilkowski and Woźniakowski (1993). To obtain sharp bounds on the minimal errors, we need to require specific regularity conditions on $K$.

In this paper we assume that $K$ is a tensor product of one-dimensional covariance kernels $K_j$ and each $K_j$ satisfies the Sacks–Ylvisaker regularity conditions of order $r_j$. Kernels in a tensor product form are studied in a number of papers. We just mention Ylvisaker (1975), Micchelli and Wahba (1981), Sacks, Welch, Mitchell and Wynn (1989), Papageorgiou and Wasilkowski (1990), Woźniakowski (1991, 1992), Paskov (1993), and Ritter, Wasilkowski and Woźniakowski (1993).

We prove that the minimal errors for the kernel $K$ behave as the minimal errors for the kernel $Q$ of the folded Wiener sheet. For the latter, the minimal errors (modulo a multiplicative constant) recently have been found for integration by Woźniakowski (1991) and Paskov (1993), and for $L_2$-approximation by Woźniakowski (1992). Therefore, if $K_j$ satisfies the Sacks–Ylvisaker conditions of order $r_j$, then the minimal error is at most of order

$$\begin{aligned} &\text{for integration:} && n^{-\gamma-1}(\log n)^{(\nu-1)/2}, \\ &\text{for } L_2\text{-approximation:} && n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+1)}. \end{aligned}$$

Here $\gamma = \min_{1\le j\le d} r_j$ and $\nu = \#\{j\colon\ r_j = \gamma\}$.

The presented bounds are sharp for $L_2$-approximation for arbitrary $r_j$ and they are sharp for integration with $r_j = 0$ for all $j$. For integration with $r_j > 0$ for some $j$, the presented bounds are, in general, not sharp. In fact, it may even happen that the integral of the random fields is zero and the integration problem is trivial. We need an extra condition to conclude that the bounds are sharp for integration. They are sharp if the kernel $K$ satisfies additionally the boundary conditions

$$\frac{\partial^k}{\partial t^k} K_j(\cdot, t)\Big|_{t=0} = 0 \quad \text{for } k = 0, 1, \ldots, r_j - 1 \text{ and } j = 1, 2, \ldots, d.$$

When the bounds are sharp, then an optimal design for $K$ is modulo a multiplicative constant equivalent to an optimal design for $Q$. In particular, an optimal design for $L_2$-approximation can be derived from hyperbolic cross points; see Temlyakov (1987) Woźniakowski (1992), as well as Remark 5.

Our proof technique is based on the characterization of the reproducing kernel Hilbert space $H(K)$ with kernel $K$. We show that if $K_j$ satisfies the Sacks–Ylvisaker conditions of order $r_j$, then $H(K_j)$ may differ from the corresponding Sobolev space $W^{r_j+1}([0,1])$ only by a finite dimensional subspace of polynomials.

We summarize the content of the paper. Basic facts concerning reproducing kernel Hilbert spaces are in Section 2. Sacks–Ylvisaker conditions for the univariate case and the characterization of the corresponding reproducing kernel Hilbert spaces are in Section 3. Main results concerning the multivariate

case are in Section 4. A brief discussion of nonlinear estimators and optimal sequential designs for Gaussian processes is in Section 5.

**2. Reproducing kernel Hilbert spaces.** In this section we recall some basic facts about reproducing kernel Hilbert spaces. See Aronszajn (1950), Parzen (1961), Vakhania, Tarieladze and Chobanyan (1987), and Wahba (1990). Then we prove a simple lemma that will be needed later.

Let $K$ be the covariance kernel of a random field $Y(t)$, $t \in D = [0,1]^d$. Equivalently, let $K$: $D^2 \to \mathbb{R}$ be a symmetric and nonnegative definite function, that is, $(K(t_i, t_j))_{i,j=1}^m \geq 0$ for any $m$ and points $t_i$ from $D$. There exists a uniquely determined Hilbert space $H(K)$ of real-valued functions on $D$ with

$$K(\cdot, t) \in H(K) \quad \text{and} \quad f(t) = \langle f, K(\cdot, t)\rangle_K \qquad \forall\, t \in D,\ \forall\, f \in H(K).$$

By $\langle \cdot, \cdot \rangle_K$ we denote the inner product in $H(K)$. The space $H(K)$ is called the Hilbert space with reproducing kernel $K$.

Any closed subspace $G \subset H(K)$ possesses a uniquely determined reproducing kernel $M$, that is, a symmetric and nonnegative definite function $M$ on $D^2$ with

$$M(\cdot, t) \in G \quad \text{and} \quad \langle g, M(\cdot, t)\rangle_M = \langle g, M(\cdot, t)\rangle_K = g(t) \qquad \forall\, t \in D,\ \forall\, g \in G.$$

The kernel $K$ is the sum of the kernels corresponding to $G$ and $G^\perp$.

For two covariance kernels $K$ and $L$ on $D^2$ we write $K \ll L$ if $cL - K$ is nonnegative definite for some positive constant $c$. Then

$$K \ll L \quad \Leftrightarrow \quad H(K) \subset H(L). \tag{2.1}$$

Moreover, $H(K) \subset H(L)$ implies that the embedding $H(K) \hookrightarrow H(L)$ is continuous.

Assume that $K$ is a continuous covariance kernel. Let $\lambda_1(K) \geq \lambda_2(K) \geq \cdots > 0$ denote the nonzero eigenvalues, repeated according to their multiplicity, of the integral operator

$$(Tf)(t) = \int_D K(s,t) f(s)\, ds$$

on $L_2(D)$. In what follows, we extend finite sequences $\lambda_i(K)$ by zeros. The minimax principle [see Weidmann (1980), Theorem 7.3] yields

$$K \ll L \quad \Rightarrow \quad \lambda_i(K) = O(\lambda_i(L)) \tag{2.2}$$

for any two continuous covariance kernels $K$ and $L$ on $D^2$. [Here and elsewhere we use big $O$, $\Theta$ and $\Omega$ notations to describe asymptotic properties of sequences. Let $\{a_i\}_i$ and $\{b_i\}_i$ be two sequences of nonnegative reals. Then $a_i = O(b_i)$ means that there is a positive constant $c$ such that $a_i \leq c\, b_i$ holds for all sufficiently large $i$. Moreover, $a_i = \Omega(b_i)$ means that $b_i = O(a_i)$, and $a_i = \Theta(b_i)$ means that $a_i = O(b_i)$ and $a_i = \Omega(b_i)$ hold.]

Let $K_j$, $j = 1, \ldots, d$, be covariance kernels on $[0,1]^2$. By the tensor product kernel $K = \bigotimes_{j=1}^{d} K_j$ we mean the covariance kernel on $D^2$ given by

$$K(s,t) = \prod_{j=1}^{d} K_j(s_j, t_j),$$

where $s = (s_1, \ldots, s_d)$ and $t = (t_1, \ldots, t_d)$. The operation $\otimes$ preserves the relation $\ll$ as shown in the following lemma.

LEMMA 1. *Let $K_j$ and $L_j$ be covariance kernels on $[0,1]^2$. If $K_j \ll L_j$ for $j = 1, \ldots, d$, then*

$$\bigotimes_{j=1}^{d} K_j \ll \bigotimes_{j=1}^{d} L_j.$$

PROOF. Let $d = 2$. It is enough to show that $K_1 \otimes K_2 \ll K_1 \otimes L_2 \ll L_1 \otimes L_2$. Let $a_k \in \mathbb{R}$, $t_k = (u_k, v_k) \in [0,1]^2$, and let $(f_i)_{i \in I}$ be an orthonormal base in $H(K_1)$. Then

$$K_1(u_k, u_j) = \langle K_1(\cdot, u_k), K_1(\cdot, u_j) \rangle_{K_1} = \sum_{i \in I} f_i(u_k) \cdot f_i(u_j).$$

Because $K_2 \ll L_2$, there exists a positive constant $c$ such that $0 \ll cL_2 - K_2$. We have

$$\begin{aligned}
&\sum_{k,j=1}^{n} a_k a_j (c \cdot K_1 \otimes L_2(t_k, t_j) - K_1 \otimes K_2(t_k, t_j)) \\
&\quad = \sum_{k,j=1}^{n} a_k a_j \cdot K_1(u_k, u_j)(cL_2(v_k, v_j) - K_2(v_k, v_j)) \\
&\quad = \sum_{i \in I} \sum_{k,j=1}^{n} a_k f_i(u_k) \cdot a_j f_i(u_j)(cL_2(v_k, v_j) - K_2(v_k, v_j)) \geq 0.
\end{aligned}$$

Hence, $K_1 \otimes K_2 \ll K_1 \otimes L_2$. The second relation $K_1 \otimes L_2 \ll L_1 \otimes L_2$ can be shown analogously.

For $d > 2$, let $K = \bigotimes_{j=1}^{d-1} K_j$ and $L = \bigotimes_{j=1}^{d-1} L_j$. By induction we can assume that $K \ll L$. Then we show $K \otimes K_d \ll L \otimes L_d$ as for $d = 2$. □

**3. Sacks–Ylvisaker conditions in the univariate case.** In this section we cite the Sacks–Ylvisaker conditions for covariance kernels $K$ in the univariate case $d = 1$. We show that these conditions determine the reproducing kernel Hilbert space uniquely up to a subspace of polynomials. We also conclude that the order of convergence of the eigenvalues $\lambda_i(K)$ to zero is uniquely determined. These consequences of the Sacks–Ylvisaker conditions will be used for the multivariate case in Section 4.

We denote one-sided limits at the diagonal in $[0,1]^2$ in the following way. Let

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

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 $[0,1]^2 \setminus \operatorname{cl} \Omega_j$. Furthermore, by $M^{(k,l)}(s,t)$ we denote a partial derivative of order $k$ with respect to $s$ and of order $l$ with respect to $t$. That is, $M^{(k,l)}(s,t) = (\partial^{k+l}/(\partial s^k \partial t^l))M(s,t)$.

We say that a covariance kernel $K$ on $[0,1]^2$ satisfies the Sacks–Ylvisaker conditions of order $r \in \mathbb{N}_0$ if the following three conditions hold:

(A) $K \in C^{r,r}([0,1]^2)$, the partial derivatives of $L = K^{(r,r)}$ up to order 2 are continuous on $\Omega_+ \cup \Omega_-$ and they are continuously extendable to $\operatorname{cl} \Omega_+$ as well as to $\operatorname{cl} \Omega_-$.

(B) The function

$$\alpha(s) = L_-^{(1,0)}(s,s) - L_+^{(1,0)}(s,s),$$

which belongs to $C^1([0,1])$ due to (A), satisfies

$$\min_{0 \le s \le 1} \alpha(s) > 0.$$

(C) $L_+^{(2,0)}(s,\cdot) \in H(L)$ for any $0 \le s \le 1$ and

$$\sup_{0 \le s \le 1} \|L_+^{(2,0)}(s,\cdot)\|_L < \infty.$$

Observe that, due to (A) and (B), any process $Y$ with covariance kernel $K$ has exactly $r$ derivatives in the mean square sense; they are denoted by $Y^{(k)}$. Furthermore, (A) implies

$$\begin{aligned} &E(Y^{(r)}(t+h) - Y^{(r)}(t))^2 \\ &\quad = L(t+h,t+h) - 2L(t+h,t) + L(t,t) \le 2c|h|, \end{aligned} \tag{3.1}$$

where $c = \sup_{s,t \in \Omega_+ \cup \Omega_-} |L^{(1,0)}(s,t)|$. Hence $Y^{(r)}$ is Lipschitz in the mean square sense. If $Y$ is a Gaussian process, then its realizations are $r$-fold continuously differentiable with probability 1. This follows from (3.1) and Adler [(1981), Theorem 3.4.1].

The conditions (A), (B) and (C) were, modulo small differences discussed in Remark 1, introduced and discussed in Sacks and Ylvisaker (1966, 1968, 1970a, 1970b). Various examples of kernels satisfying the Sacks–Ylvisaker conditions are known. For example, these conditions are satisfied for $r = 0$ by (i) the Brownian motion kernel $K(s,t) = \min(s,t)$, (ii) the Ornstein–Uhlenbeck kernel $K(s,t) = \exp(-|s-t|)$ and (iii) the kernel $K(s,t) = 1-|s-t|$, which corresponds to the sum $Y_1(t) + Y_2(1-t)$ of two independent Brownian motions $Y_1$ and $Y_2$.

We now discuss stochastic processes with a covariance kernel $K$ satisfying the Sacks–Ylvisaker conditions of order $r$. We show that the reproducing Hilbert space $H(K)$ is closely related to the Sobolev space $W^{r+1}([0,1])$ and

may differ from $W^{r+1}([0,1])$ only by some polynomials. Recall that

$$W^{r+1}([0,1]) = \{f \in C^r([0,1]):\ f^{(r)} \text{ abs. cont.},\ f^{(r+1)} \in L_2([0,1])\} = H(R_r),$$

where the kernel $R_r$ and the inner product are given by

$$\begin{aligned} R_r(s,t) &= \sum_{k=0}^{r} \frac{s^k \cdot t^k}{(k!)^2} + Q_r(s,t) \quad \text{with} \\ Q_r(s,t) &= \int_0^1 \frac{(s-u)_+^r (t-u)_+^r}{(r!)^2}\, du \end{aligned} \tag{3.2}$$

and

$$\langle f, g\rangle_{R_r} = \sum_{k=0}^{r} f^{(k)}(0) \cdot g^{(k)}(0) + \int_0^1 f^{(r+1)}(u) \cdot g^{(r+1)}(u)\, du;$$

see Wahba [(1990), page 8]. Moreover,

$$H(Q_r) = \{f \in H(R_r):\ f^{(k)}(0) = 0 \text{ for } k = 0, 1, \ldots, r\}.$$

Let $P_r$ denote the kernel that corresponds to the closed subspace of all functions from $H(R_r)$ vanishing at the boundary, that is,

$$H(P_r) = \{f \in H(R_r):\ f^{(k)}(0) = f^{(k)}(1) = 0 \text{ for } k = 0, 1, \ldots, r\}. \tag{3.3}$$

A formula for $P_r$ is given in Sacks and Ylvisaker [(1970a), page 2060]. For $r = 0$,

$$P_0(s,t) = \min(s,t) - s\,t.$$

We add that $K = P_r$ or $K = Q_r$ differs from $R_r$ only by a polynomial $g(s,t)$ of degree less than or equal to $r+1$ in each variable. Furthermore, $R_r^{(r,r)}(s,t) = 1 + \min(s,t)$. Therefore, $P_r$ and $Q_r$ as well as $R_r$ satisfy the Sacks–Ylvisaker conditions with $\alpha \equiv 1$ and $L_+^{(2,0)} \equiv 0$. Note also that

$$H(P_r) \subset H(Q_r) \subset H(R_r), \quad \text{that is,} \quad P_r \ll Q_r \ll R_r.$$

The function $Q_r$ is the covariance kernel of the $r$-fold integrated Brownian motion, and the kernels $R_r$ and $P_r$ correspond to the following stochastic processes. Consider a random Taylor polynomial $\sum_{k=0}^{r} X_k \cdot t^k/k!$ with independent standard normal variables $X_k$ and an $r$-fold integrated Brownian motion $Y$ independent from $(X_0, \ldots, X_r)$. Then the sum of the random Taylor polynomial and $Y$ defines a process with covariance kernel $R_r$. The difference between $Y$ and its conditional mean, given $Y^{(k)}(1)$ for $k = 0, 1, \ldots, r$, is a process with covariance kernel $P_r$.

Observe that the spaces $H(P_r)$, $H(Q_r)$ and $H(R_r)$ differ only by some polynomials. That is,

$$H(P_r) \oplus \mathbb{B}_{2r+1} = H(R_r) \quad \text{and} \quad H(Q_r) \oplus \mathbb{B}_r = H(R_r), \tag{3.4}$$

where $\mathbb{B}_i$ is the space of polynomials of degree at most $i$. We now show that the Hilbert space $H(K)$ with $K$ satisfying the Sacks–Ylvisaker conditions of

order $r$ is closely related to the spaces $H(P_r)$ and $H(R_r)$ and may differ from the space $H(R_r)$ also only by some polynomials.

THEOREM 1. *If the covariance kernel $K$ satisfies the Sacks–Ylvisaker conditions of order $r \in \mathbb{N}_0$, then*

$$H(P_r) \subset H(K) + \mathbb{B}_{r-1} \subset H(R_r), \tag{3.5}$$

*where $\mathbb{B}_{r-1}$ is the space of polynomials of degree at most $r-1$ and $\mathbb{B}_{-1} = \{0\}$. Moreover, $H(K)$ is a closed subspace of $H(R_r)$.*

PROOF. Let $L = K^{(r,r)}$ and let $M$ denote the kernel that corresponds to the closed subspace

$$H(M) = \{f \in H(L):\ f(0) = f(1) = 0\} = \operatorname{span}\{L(\cdot,0), L(\cdot,1)\}^{\perp}$$

of the space $H(L)$. We show that

$$H(M) = H(P_0). \tag{3.6}$$

Let $\beta(s) = 1/\alpha(s)$ and let $f = \sum_{i=1}^n a_i L(\cdot, t_i) \in H(M)$ with distinct $t_i$. Then (A) implies that $f \in H(P_0)$, and integration by parts yields

$$\begin{aligned}\|\beta^{1/2} f'\|_2^2 = \sum_{i,j=1}^n a_i a_j \Big( &\beta(1)\, L(1,t_i) L_+^{(1,0)}(1,t_j) - \beta(0)\, L(0,t_i) L_-^{(1,0)}(0,t_j) \\ &+ \beta(t_j)\, L(t_i,t_j)(L_-^{(1,0)}(t_j,t_j) - L_+^{(1,0)}(t_j,t_j)) \\ &- \int_0^1 L(s,t_i)(\beta'(s)\, L_+^{(1,0)}(s,t_j) + \beta(s)\, L_+^{(2,0)}(s,t_j))\, ds \Big).\end{aligned}$$

Observe that $f(0) = f(1) = \sum_{i=1}^n a_i L(0,t_i) = \sum_{i=1}^n a_i L(1,t_i) = 0$ and due to (B) we have

$$\sum_{i,j=1}^n a_i a_j\, \beta(t_j)\, L(t_i,t_j)(L_-^{(1,0)}(t_j,t_j) - L_+^{(1,0)}(t_j,t_j)) = \sum_{i,j=1}^n a_i a_j L(t_i,t_j) = \|f\|_L^2.$$

Thus

$$\|\beta^{1/2} f'\|_2^2 - \|f\|_L^2 = -\gamma_1 - \gamma_2,$$

where

$$\begin{aligned}\gamma_1 &= \sum_{i,j=1}^n a_i a_j \int_0^1 L(s,t_i)\, \beta(s)\, L_+^{(2,0)}(s,t_j)\, ds \\ &= \sum_{j=1}^n a_j \int_0^1 f(s)\, \beta(s)\, L_+^{(2,0)}(s,t_j)\, ds\end{aligned}$$

and

$$\gamma_2 = \sum_{i,j=1}^{n} a_i a_j \int_0^1 L(s,t_i)\,\beta'(s)\,L_+^{(1,0)}(s,t_j)\,ds = \int_0^1 f(s)\,\beta'(s)\,f'(s)\,ds.$$

Observe that (C) yields $L_+^{(2,0)}(s,t_j) = \langle L_+^{(2,0)}(s,\cdot), L(\cdot,t_j)\rangle_L$. Hence

$$\begin{aligned}|\gamma_1| &= \left|\int_0^1 f(s)\,\beta(s)\langle L_+^{(2,0)}(s,\cdot), f\rangle_L\,ds\right| \\ &\le c_0\,\|\beta\|_\infty\,\|f\|_L\,\|f\|_1 \le c_1\,\|f\|_L\,\|\beta^{1/2}f'\|_2\end{aligned}$$

with $c_0 = \sup_{0\le s\le 1}\|L_+^{(2,0)}(s,\cdot)\|_L < \infty$ [see (C)], and $c_1 = c_0\,\|\beta\|_\infty\,\|\alpha^{1/2}\|_\infty$. Moreover,

$$|\gamma_2| \le \|\beta'/\beta^{1/2}\|_\infty\,\|\beta^{1/2}f'\|_2\left(\int_0^1 \langle f, L(\cdot,s)\rangle_L^2\,ds\right)^{1/2} \le c_2\,\|f\|_L\,\|\beta^{1/2}f'\|_2$$

with $c_2 = \|\beta'/\beta^{1/2}\|_\infty\,\|L\|_\infty$. Therefore, we obtain

$$|\,\|\beta^{1/2}f'\|_2^2 - \|f\|_L^2\,| \le (c_1+c_2)\,\|f\|_L\,\|\beta^{1/2}f'\|_2.$$

This implies

$$[1/(c_1+c_2+1)]\,\|f\|_L \le \|\beta^{1/2}f'\|_2 \le (c_1+c_2+1)\,\|f\|_L$$

and finally

$$c_3\,\|f\|_L \le \|f'\|_2 \le c_4\,\|f\|_L$$

with suitable positive constants $c_3$ and $c_4$.

Becuase $H(M)\cap \operatorname{span}\{L(\cdot,t)\colon 0\le t\le 1\}$ is dense in $H(M)$, we conclude that $H(M)\subset H(P_0)$ and that the norms $\|\cdot\|_L$ and $\|\cdot\|_{R_0}$ are equivalent on $H(M)$. Hence $H(M)$ is a closed subspace of $H(P_0)$.

To prove that $H(M) = H(P_0)$, we proceed as follows. Consider the subspace

$$H(M)^\perp = \operatorname{span}\{L(\cdot,0), L(\cdot,1)\}$$

of $H(L)$. Its reproducing kernel is $L - M$. From Aronszajn [(1950), page 346] we know that $L - M$ is a linear combination of

$$L(s,0)L(t,0),\ L(s,0)L(t,1),\ L(s,1)L(t,0),\ L(s,1)L(t,1).$$

Due to (A),

$$H(M)^\perp \subset C^2([0,1]) \subset H(R_0) \tag{3.7}$$

and

$$L - M \in C^{2,2}([0,1]^2). \tag{3.8}$$

Hence, $M$ satisfies (A) and (B) with $r = 0$.

Consider the integral operator $T$: $C([0,1]) \to H(M)$ given by

$$(T\varphi)(t) = \int_0^1 M(s,t)\varphi(s)\,ds$$

and let

$$g(t) = \int_0^1 L(s,t)\varphi(s)\,ds = (T\varphi)(t) + \int_0^1 (L-M)(s,t)\varphi(s)\,ds.$$

If $T\varphi = 0$, then $g \in \operatorname{span}\{L(\cdot,0), L(\cdot,1)\}$. Theorem 3.1 of Sacks and Ylvisaker (1966), applied to the kernel $L$, yields $\varphi = 0$. Hence the operator $T$ is one-to-one.

To prove that $H(M) = H(P_0)$, it is enough to show that the only $f \in H(P_0)$ with $\langle f, T\varphi\rangle_{R_0} = 0$ for all $\varphi \in C([0,1])$ is $f = 0$. Observe that

$$(T\varphi)''(t) = -\alpha(t)\varphi(t) + \int_0^1 N(s,t)\varphi(s)\,ds = \alpha(t)((U\varphi)(t) - \varphi(t)) \tag{3.9}$$

with $N = M_+^{(0,2)}$ and

$$(U\varphi)(t) = \int_0^1 [N(s,t)/\alpha(t)]\cdot\varphi(s)\,ds.$$

Moreover, (A) and (C) imply $L_+^{(0,2)}(t,t) = L_-^{(0,2)}(t,t)$ for any $0 \le t \le 1$ and, therefore, $L_+^{(0,2)} \in C([0,1]^2)$ by (A). See Sacks and Ylvisaker [(1966), page 75] for these facts. Using (3.8) we obtain $N \in C([0,1]^2)$ and, therefore, $U$ is a compact operator on $C([0,1])$. Because $T$ is injective and $T\varphi$ vanishes at zero and one, $(T\varphi)'' = 0$ implies $\varphi = 0$. Thus, 1 is not an eigenvalue of $U$.

Assume now that $f \in H(P_0)$ with $\langle f, T\varphi\rangle_{R_0} = 0$ for all $\varphi \in C([0,1])$. The Fredholm alternative and (3.9) imply the existence of $\varphi$ such that $(T\varphi)'' = -f$. Because

$$0 = \int_0^1 f'(t)(T\varphi)'(t)\,dt = -\int_0^1 f(t)(T\varphi)''(t)\,dt = \|f\|_2^2,$$

we have $f = 0$, which completes the proof of (3.6).

From (3.6) and (3.7) we conclude that

$$H(P_0) \subset H(L) \subset H(R_0).$$

The sum $H(L) = H(M) \oplus H(M)^\perp$ is closed in $H(R_0)$ because $H(M)^\perp$ is of finite dimension and $H(M)$ is closed in $H(R_0)$. This completes the proof of the Theorem for $r = 0$. We remark that the norms $\|\cdot\|_L$ and $\|\cdot\|_{R_0}$ are equivalent on $H(L)$ due to the open mapping theorem.

Consider now the case $r > 0$. Because $K \in C^{r,r}([0,1]^2)$, we know that $H(K) \subset C^r([0,1])$ and $f^{(k)}(t) = \langle f, K^{(0,k)}(\cdot,t)\rangle_K$ for any $f \in H(K)$, $k = 0,1,\ldots,r$, and $0 \le t \le 1$. See Sacks and Ylvisaker [(1970b), page 123]. For

$f \in H(K)$, define the kernel $V(s,t) = f^{(r)}(s)f^{(r)}(t)$. Then

$$\sum_{i,j=1}^{n} a_i a_j V(t_i,t_j) = \left\langle f, \sum_{i=1}^{n} a_i K^{(0,r)}(\cdot,t_i)\right\rangle_K^2 \le \|f\|_K^2 \sum_{i=1}^{n} a_i a_j L(t_i,t_j)$$

for any $a_i \in \mathbb{R}$ and $0 \le t_i \le 1$. Hence $V \ll L$ and $H(V) \subset H(L)$. Because $H(V) = \operatorname{span}\{f^{(r)}\}$, we have $f^{(r)} \in H(L)$ and

$$Uf = f^{(r)}$$

defines a linear operator from $H(K)$ to $H(L)$. The continuity of this mapping follows from the closed graph theorem. Hence, $f \in H(K)$ implies that $f \in C^r([0,1])$ and $f^{(r)} \in H(L) \subset H(R_0)$. Therefore,

$$H(K) \subset H(R_r),$$

which is, of course, equivalent to $H(K) + \mathbb{B}_{r-1} \subset H(R_r)$.

Let $U^*\colon H(L) \to H(K)$ denote the adjoint of $U$. We show that

$$U(U^*f) = f$$

for any $f \in H(L)$. It is enough to verify that

$$U(U^*L(\cdot,t)) = L(\cdot,t)$$

for any $0 \le t \le 1$. For $f \in H(K)$, we have

$$\langle f, U^*L(\cdot,t)\rangle_K = \langle f^{(r)}, L(\cdot,t)\rangle_L = f^{(r)}(t) = \langle f, K^{(0,r)}(\cdot,t)\rangle_K.$$

Hence, $U^*L(\cdot,t) = K^{(0,r)}(\cdot,t)$ and $U(U^*L(\cdot,t)) = K^{(r,r)}(\cdot,t) = L(\cdot,t)$, as claimed. This yields that $U$ is surjective and, therefore,

$$H(P_0) = H(M) \subset H(L) = \operatorname{ran} U = \{f^{(r)}\colon\ f \in H(K)\}. \tag{3.10}$$

From this it is easy to prove the first inclusion of (3.5). Indeed, take $f \in H(P_r)$. Then $f^{(r)} \in H(P_0)$ and (3.10) yields that there exists $g \in H(K)$ such that

$$f^{(r)} = g^{(r)}, \tag{3.11}$$

which implies $f = g + p$ for some polynomial $p$ of degree at most $r-1$. Therefore,

$$H(P_r) \subset H(K) + \mathbb{B}_{r-1}.$$

Clearly, $\ker U \subset \mathbb{B}_{r-1}$, and

$$\operatorname{ran} U^* = \{f \in H(K)\colon\ f = U^*(Uf)\}$$

is easy to verify. Hence, $\operatorname{ran} U^*$ is closed in $H(K)$ and

$$H(K) = \operatorname{ran} U^* \oplus \ker U.$$

To conclude that $H(K)$ is closed in $H(R_r)$ it remains to show that $\operatorname{ran} U^*$ is also closed in $H(R_r)$. Let $f_n \in \operatorname{ran} U^*$ and $f \in H(R_r)$ with $\lim_{n\to\infty} \|f_n - f\|_{R_r} = 0$. Then $f_n^{(r)} \in H(L)$ and $\lim_{n\to\infty} \|f_n^{(r)} - f^{(r)}\|_{R_0} = 0$. Because $H(L)$ is

closed in $H(R_0)$, we get $f^{(r)} \in H(L)$, and the equivalence of the norms $\|\cdot\|_L$ and $\|\cdot\|_{R_0}$ on $H(L)$ yields $\lim_{n\to\infty} \|f_n^{(r)} - f^{(r)}\|_L = 0$. The continuity of $U^*$ and of the embedding $H(K) \hookrightarrow H(R_r)$ together with $U^* f_n^{(r)} = f_n$ imply that $\lim_{n\to\infty} \|f_n - U^* f^{(r)}\|_{R_r} = 0$. Therefore, $f = U^* f^{(r)} \in \operatorname{ran} U^*$. □

Theorem 1 states that for $r = 0$ the Hilbert space $H(K)$ lies between $H(P_0)$ and $H(R_0)$. For $r \geq 1$, we have the similar inclusions for the sum of the space $H(K)$ and the class $\mathbb{B}_{r-1}$ of polynomials of degree at most $r-1$. We now show an example that demonstrates that the class $\mathbb{B}_{r-1}$ is essential and $H(P_r)$ does not have to be a subset of $H(K)$ alone.

EXAMPLE 1. Let $X$ be an arbitrary stochastic process whose kernel $M$ satisfies the Sacks–Ylvisaker condition with regularity $r \geq 1$. Define the process $Y$:

$$Y(t) = X(t) - \int_0^1 X(u)\,du.$$

Obviously, the kernel $K$ of $Y$ equals

$$K(s,t) = M(s,t) - \int_0^1 M(s,u)\,du - \int_0^1 M(t,u)\,du + \int_0^1 \int_0^1 M(u,v)\,du\,dv$$

and hence $K^{(r,r)}(s,t) = M^{(r,r)}(s,t)$. Thus, $K$ satisfies the Sacks–Ylvisaker conditions with regularity $r$.

Observe that $\int_0^1 K(s,t)\,dt = 0$ for all $0 \leq s \leq 1$. Hence, $H(K)$ consists of functions with zero integral. Because this is not true for $H(P_r)$, we have

$$H(P_r) \not\subset H(K).$$

This illustrates that, in general, the presence of $\mathbb{B}_{r-1}$ in (3.5) of Theorem 1 is crucial.

To guarantee that $H(P_r)$ is a subset of $H(K)$, we need to strengthen the Sacks–Ylvisaker conditions in the case $r \geq 1$. Assume that any $f \in H(K)$ satisfies $f^{(k)}(0) = 0$ for $k = 0, 1, \ldots, r-1$. Obviously, this is equivalent to $K^{(0,k)}(\cdot, 0) = 0$. Then, in the proof of Theorem 1, $U^*$ is given by $r$-fold integration and $\ker U = \{0\}$. From (3.11) we easily conclude that $H(P_r) \subset H(K)$. We summarize this in the following corollary.

COROLLARY 1. *If the covariance kernel $K$ satisfies the Sacks–Ylvisaker conditions of order $r \geq 1$ and*

$$K^{(0,k)}(\cdot, 0) = 0 \quad \text{for } k = 0, 1, \ldots, r-1, \tag{3.12}$$

*then*

$$H(P_r) \subset H(K) \subset H(R_r).$$

In the next corollary we study the eigenvalues $\lambda_i(K)$ of the integral operator on $L_2([0,1])$ with kernel $K$. Their order is determined by the Sacks–Ylvisaker conditions.

COROLLARY 2. *If the covariance kernel $K$ satisfies the Sacks–Ylvisaker conditions of order $r \in \mathbb{N}_0$, then*

$$\lambda_i(K) = \Theta(i^{-2r-2}).$$

PROOF. From Theorem 1 and (3.4) we know that $H(K)$ is a closed subspace of $H(R_r)$ with $H(R_r) = H(P_r) + \mathbb{B}_{2r+1} \subset H(K) + \mathbb{B}_{2r+1} \subset H(R_r)$. Therefore, $H(K)$ is of codimension $k \le 2r+2$ in $H(R_r)$. Let $K'$ denote the reproducing kernel of $H(K)$ as a subspace of $H(R_r)$, $K'(\cdot, t) \in H(K)$. Then

$$\lambda_i(K) = O(\lambda_i(R_r)) \quad \text{and} \quad \lambda_i(K') = \Theta(\lambda_i(K)) \tag{3.13}$$

due to (2.1) and (2.2).

Let $f_1, \dots, f_k$ be an orthonormal base of $H(R_r - K')$. Moreover, let $T$ and $U$ denote the integral operators on $L_2([0,1])$ with kernels $K'$ and $R_r$, respectively, and let $g_1, g_2, \dots$ be an orthonormal sequence in $L_2([0,1])$ with $Tg_i = \lambda_i(K') \cdot g_i$. Finally let $F$ and $G_i$ denote the sets of normalized functions in $L_2([0,1])$ that are orthogonal to $f_1, \dots, f_k$ or $g_1, \dots, g_{i-1}$, respectively. The minimax principle [see Weidmann (1980), Theorem 7.3], implies

$$\lambda_{i+k}(R_r) \le \sup_{f \in F \cap G_i} \langle Uf, f \rangle_2 = \sup_{f \in F \cap G_i} \langle Tf, f \rangle_2 \le \sup_{f \in G_i} \langle Tf, f \rangle_2 = \lambda_i(K'). \tag{3.14}$$

Consider the covariance kernel $Q_r$; see (3.2). From Micchelli and Wahba [(1981), page 338] we know that $\lambda_i(Q_r) = \Theta\left(i^{-2r-2}\right)$. Observe that (3.13) and (3.14) hold for $K = K' = Q_r$ with $k = r+1$. This yields that $\lambda_i(R_r) = \Theta\left(i^{-2r-2}\right)$. Applying (3.13) and (3.14) again for the original $K$, we obtain the same order for $\lambda_i(K)$. □

REMARK 1. We now discuss the conditions (A), (B) and (C) and the original conditions introduced by Sacks and Ylvisaker (1966, 1968, 1970a). Conditions (A), (B) and (C) are the same as those of Sacks and Ylvisaker for $r = 0$. For $r \ge 1$, (A), (B) and (C) are slightly more general. More precisely, it is assumed in Sacks and Ylvisaker (1970a) that the stochastic process $Y$ is the $r$-fold integral of a process $Z$ whose kernel satisfies the conditions (A), (B) and (C) with regularity zero, $r = 0$. It can be easily verified that such a class of processes is identical to the class of processes whose kernels satisfy (A), (B) and (C) with regularity $r$ and also satisfy the boundary conditions (3.12). Thus, the conditions (A), (B) and (C) together with (3.12) are equivalent to the conditions assumed in Sacks and Ylvisaker (1970a).

We add that Müller-Gronbach (1993) studied $L_2$-approximation for $r = 0$ with generalized conditions (A) and (B) together with (C). Su and Cambanis (1993) studied weighted $L_2$-approximation for $r = 0$ requiring (A) and (C) together with (B) for a nonzero and nonnegative $\alpha$. Benhenni and Cambanis

(1992) studied weighted integration under regularity conditions (A), (B) and a modified condition (C). They find the asymptotic performance of optimal linear estimators that use designs generated by positive continuous density functions.

REMARK 2. Wahba (1971), Hájek and Kimeldorf (1974) and Speckman (1979) study weighted integration and $L_p$-approximation for autoregressive processes $Y$ of order $r+1 \in \mathbb{N}$. From Hájek and Kimeldorf [(1974), page 522] it follows that $H(K) = H(Q_r)$ for the covariance kernel $K$ of $Y$. Hence the conclusions of the preceding corollaries are valid for these processes as well.

**4. Multivariate designs under Sacks–Ylvisaker conditions.** In this section we derive the main result of the paper. We consider zero mean random fields $Y(t)$, $t \in D = [0,1]^d$, with covariance kernels $K$ in a tensor product form. More precisely, we assume:

(D) $K = \bigotimes_{j=1}^d K_j$ and $K_j$ is a covariance kernel on $[0,1]^2$ that satisfies the Sacks–Ylvisaker conditions (A), (B) and (C) with regularity $r_j \in \mathbb{N}_0$, $j = 1,2,\dots,d$.

Due to (D), the kernel $K$ is at least continuous and, therefore, we may assume that $Y$ is a measurable random field whose realizations are in $L_2(D)$ with probability 1. As stated in the Introduction, we want to estimate the integral $\int_D Y(t)\,dt$ or the values $Y(t)$ for all $t \in D$ simultaneously, assuming that $Y$ may be observed at a finite number of points $t_i \in D$. We consider linear estimators $I_n(Y)$, $A_n(Y)$ and their errors $e(I_n, K)$, $e(A_n, K)$ defined as in the Introduction.

We are interested in the minimal errors $\inf_{I_n} e(I_n, K)$ and $\inf_{A_n} e(A_n, K)$ in the class of all linear estimators that use $n$ observations. The order of these quantities is known for folded Wiener sheets $Y$. The folded Wiener sheets are obtained from the classical Wiener sheet by $r_j$ iterated integrations with respect to the $j$th variable. The covariance kernels of these random fields are given by $Q = \bigotimes_{j=1}^d Q_{r_j}$, where $Q_{r_j}$ is given by (3.2) with $r = r_j$. Clearly, $Q$ satisfies condition (D). We denote

$$\gamma = \min_{1 \le j \le d} r_j \quad \text{and} \quad \nu = \#\{j\colon\ r_j = \gamma\} \tag{4.1}$$

as the minimal regularity and the number of factors with the minimal regularity.

For the integration problem

$$\inf_{I_n} e(I_n, Q) = \Theta(n^{-\gamma-1}(\log n)^{(\nu-1)/2}) \tag{4.2}$$

as shown by Woźniakowski (1991) for $r_j = 0$ and Paskov (1993) for arbitrary $r_j$, $j = 1,2,\dots,d$. For the $L_2$-approximation problem

$$\inf_{A_n} e(A_n, Q) = \Theta(n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+1)}) \tag{4.3}$$

as shown by Woźniakowski (1992).

Due to the following theorem, this behavior is valid for a broad class of random fields.

THEOREM 2. *Assume that the covariance kernel $K$ satisfies condition* (D) *and the boundary conditions*

$$K_j^{(0,k)}(\cdot,0) = 0 \quad \text{for } k = 0,1,\dots,r_j - 1, \tag{4.4}$$

*for any $j = 1,2,\dots,d$ with $r_j > 0$. Then we have:*
*for integration:*

$$\inf_{I_n} e(I_n, K) = \Theta(n^{-\gamma-1}(\log n)^{(\nu-1)/2});$$

*for $L_2$-approximation:*

$$\inf_{A_n} e(A_n, K) = \Theta(n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+1)}).$$

*Here $\gamma$ and $\nu$ are given by* (4.1) *and the infima are with respect to all linear estimators $I_n$ and $A_n$ that use $n$ observations.*

PROOF. Consider the covariance kernels $R = \bigotimes_{j=1}^d R_{r_j}$ and $P = \bigotimes_{j=1}^d P_{r_j}$ on $D^2$. Here the factors $R_{r_j}$ and $P_{r_j}$ are given by (3.2) and (3.3) with $r = r_j$. Theorem 1 and Corollary 1 yield that $P_{r_j} \ll K_j \ll R_{r_j}$ due to (2.1). From Lemma 1 we get $P \ll K \ll R$. This is equivalent to

$$H(P) \subset H(K) \subset H(R).$$

Furthermore, the respective embeddings are continuous and, therefore, there exist positive constants $c_1$ and $c_2$ such that

$$\begin{aligned} c_1 \|f\|_K &\le \|f\|_P \qquad \forall\, f \in H(P) \quad \text{and} \\ \|f\|_R &\le c_2 \|f\|_K \qquad \forall\, f \in H(K). \end{aligned} \tag{4.5}$$

It is well known and used in many papers that the error $e(I_n, K)$ coincides with the maximal error of $I_n$ in the unit ball of the reproducing kernel Hilbert space $H(K)$, see (1.1). A similar relation holds for $L_2$-approximation; see (1.2). Together with (4.5) this implies

$$c_1 \cdot e(I_n, P) \le e(I_n, K) \le c_2 \cdot e(I_n, R) \tag{4.6}$$

and

$$c_1 \cdot e(A_n, P) \le e(A_n, K) \le c_2 \cdot e(A_n, R). \tag{4.7}$$

We now relate the errors in the spaces $H(P)$ and $H(R)$ by using a well known periodization technique; see, for example, Bykovskij (1985) and Temlyakov (1990). Let $Z$ be a zero mean random field with covariance kernel $R$ and let $(t)_j$ denote the $j$th component of $t \in D$.

For integration, we define the random field $\tilde{Z}$ by $\tilde{Z}(t) = \Psi'(t) \cdot Z(\Psi(t))$, where

$$\Psi(t) = (\psi_{r_1+1}((t)_1), \dots, \psi_{r_d+1}((t)_d)), \qquad \Psi'(t) = \prod_{j=1}^{d} \psi'_{r_j+1}((t)_j)$$

and

$$\psi_m(t) = \frac{\int_0^t (u(1-u))^m \, du}{\int_0^1 (u(1-u))^m \, du}$$

for $m \in \mathbb{N}$ and $t \in [0,1]$. The polynomials $\psi_m$ are strictly increasing on $[0,1]$ with $\psi_m(0) = 0$, $\psi_m(1) = 1$ and $\psi_m^{(k)}$ vanishing at $t = 0, 1$ for $k = 1, \dots, m$.

Clearly, the covariance kernel of $\tilde{Z}$ is $\tilde{R} = \bigotimes_{j=1}^{d} \tilde{R}_{r_j}$, where

$$\tilde{R}_r(s,t) = \psi'_{r+1}(s)\psi'_{r+1}(t) \cdot R_r(\psi_{r+1}(s), \psi_{r+1}(t)).$$

Observe that the integrals of $Z$ and $\tilde{Z}$ over $D$ coincide. Because

$$\int_D Z(t)\,dt - \sum_{i=1}^{n} Z(\Psi(t_i)) \cdot a_i \Psi'(t_i) = \int_D \tilde{Z}(t)\,dt - \sum_{i=1}^{n} \tilde{Z}(t_i) \cdot a_i, \tag{4.8}$$

we conclude that

$$\inf_{I_n} e(I_n, R) \le \inf_{I_n} e(I_n, \tilde{R}). \tag{4.9}$$

We now show that $\tilde{R}_r \ll P_r$. Indeed, for $f \in H(R_r)$, let $\tilde{f}(t) = \psi'_{r+1}(t) \cdot f(\psi_{r+1}(t))$. It is easy to verify that $\tilde{f} \in H(P_r)$. Moreover $f \mapsto \tilde{f}$ defines an isomorphism between $H(R_r)$ and $H(\tilde{R}_r)$. We thus conclude that $H(\tilde{R}_r) \subset H(P_r)$, which implies $\tilde{R}_r \ll P_r$ due to (2.1).

As before we thus conclude that

$$e(I_n, \tilde{R}) \le c_3 \cdot e(I_n, P)$$

with a positive constant $c_3$, which neither depends on $n$ nor $I_n$. Combining (4.6), (4.9) and the above inequality, we obtain

$$c_1 \inf_{I_n} e(I_n, P) \le \inf_{I_n} e(I_n, K) \le c_2 c_3 \inf_{I_n} e(I_n, P).$$

The foregoing inequalities hold for all $K$ satisfying (D) and (4.4). Thus, in particular, they hold for $K = Q$. This means that (4.2) also holds for $P$ and, consequently, (4.2) holds for the original $K$. This completes the proof for integration.

For $L_2$-approximation we proceed very similarly. The random field $\tilde{Z}$ is now given by $\tilde{Z}(t) = (\Psi'(t))^{1/2} \cdot Z(\Psi(t))$, where $\psi_{2r_j+2}$ instead of $\psi_{r_j+1}$ is used to define $\Psi$ and $\Psi'$. Then

$$\begin{aligned} &\int_D \left( Z(t) - \sum_{i=1}^{n} Z(\Psi(t_i)) \cdot (\Psi'(t_i)/\Psi'(t))^{1/2} g_i(\Psi^{-1}(t)) \right)^2 dt \\ &\quad = \int_D \left( \tilde{Z}(t) - \sum_{i=1}^{n} \tilde{Z}(t_i) \cdot g_i(t) \right)^2 dt. \end{aligned} \tag{4.10}$$

The covariance kernel of $\tilde{Z}$ is now $\tilde{R} = \bigotimes_{j=1}^{d} \tilde{R}_{r_j}$, where

$$\tilde{R}_r(s,t) = (\psi'_{2r+2}(s)\psi'_{2r+2}(t))^{1/2} \cdot R_r(\psi_{2r+2}(s), \psi_{2r+2}(t)).$$

As before, we can show that $\tilde{R}_r \ll P_r$. Using (4.7) we obtain

$$c_1 \inf_{A_n} e(A_n, P) \le \inf_{A_n} e(A_n, K) \le c_2\, c_4 \inf_{A_n} e(A_n, P)$$

with a suitable constant $c_4 > 0$. As before, we can set $K = Q$ to conclude that (4.3) holds for $P$ and consequently for the original $K$. The proof is complete. □

REMARK 3. Assume that the covariance kernel $K$ is a tensor product of kernels, each corresponding to an autoregressive processes of order $r_j + 1$. From Remark 2, we conclude that the error bounds given in Theorem 2 are also valid for such $K$.

REMARK 4. Upper bounds on the minimal errors can be obtained for covariance kernels $K = \bigotimes_{j=1}^{d} K_j$, which satisfy *only* condition (A) with regularity $r_j \in \mathbb{N}_0$. Indeed, from Ritter, Wasilkowski and Woźniakowski [(1993), Theorem 2] and (3.1), we obtain for integration

$$\inf_{I_n} e(I_n, K) = O(n^{-\gamma-1}(\log n)^{(\nu-1)(\gamma+3/2)}),$$

and for $L_2$-approximation,

$$\inf_{A_n} e(A_n, K) = O(n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+3/2)}).$$

Observe that these upper bounds differ from the bounds given in Theorem 2 only by logarithmic factors.

Obviously, if only (A) holds, then the $r_j$ are not uniquely defined and the foregoing upper bounds need not be sharp. We now show that even if we add condition (B) and the boundary conditions (4.4), then still the upper bounds need not be sharp.

We provide a univariate example for integration. Let $X(t)$, $t \in [0,1]$, be a process with covariance kernel $Q_r$ [see (3.2)], and define

$$Y(t) = X(t) - a(t) \int_0^1 X(u)\, du,$$

where

$$a(t) = [1/c] \int_0^1 Q_r(s,t)\, ds \quad \text{and} \quad c = \int_0^1 \int_0^1 Q_r(s,t)\, ds\, dt.$$

Because

$$Q_r^{(0,r)}(s,t) = \int_0^{\min(s,t)} (s-u)^r/r!\, du,$$

we have $a \in C^\infty([0,1])$. Moreover, $a(0) = \cdots = a^{(r)}(0) = 0$. Recall that $Q_r$ satisfies the Sacks–Ylvisaker condition of order $r$ and, therefore, the kernel

$$K(s,t) = Q_r(s,t) - c \cdot a(s)\, a(t)$$

of $Y$ satisfies (A) and (B) together with the boundary conditions (4.4). However, $\int_0^1 Y(t)\,dt = 0$ because $\int_0^1 a(t)\,dt = 1$ and, therefore, $e(I_1, K) = 0$ for $I_1(Y) = 0$.

This shows that we cannot eliminate condition (C) if we want to have sharp bounds on the minimal errors.

We now discuss covariance kernels $K$ that satisfy $(D)$ but violate the boundary conditions (4.4). The proof of Theorem 2 shows that the error bounds are sharp also for the kernel $R = \bigotimes_{j=1}^d R_{r_j}$. Because $K \ll R$, the bounds of Theorem 2 now become upper bounds for $K$. As follows from Example 1, they need not be sharp for the integration problem. However, they are sharp for the integration problem with $r_j = 0$ for $j = 1, 2 \ldots, d$, and for the $L_2$-approximation problem for arbitrary $r_j$'s. More precisely we have the following theorem.

THEOREM 3. *Assume that the covariance kernel $K$ satisfies condition* (D). *Then we have*
*for integration:*

$$\inf_{I_n} e(I_n, K) = O(n^{-\gamma-1}(\log n)^{(\nu-1)/2});$$

*for $L_2$-approximation:*

$$\inf_{A_n} e(A_n, K) = \Theta(n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+1)}).$$

*The bound for integration is sharp if $r_1 = \cdots = r_d = 0$.*

PROOF. It remains to show that

$$\inf_{A_n} e(A_n, K) = \Omega(n^{-\gamma-1/2}(\log n)^{(\nu-1)(\gamma+1)}). \tag{4.11}$$

From Micchelli and Wahba [(1981), Theorem 1], we know that

$$\inf_{A_n} e(A_n, K)^2 \geq \sum_{i=n+1}^{\infty} \lambda_i(K). \tag{4.12}$$

Here $\lambda_i(K)$ denote the ordered eigenvalues of the integral operator on $L_2(D)$ with kernel $K$. Because $K = \bigotimes_{j=1}^d K_j$, the set $\{\lambda_i(K)\colon i \in \mathbb{N}\}$ is equal to the set of all products $\prod_{j=1}^d \lambda_{i_j}(K_j)$ with $i_1, \ldots, i_d \in \mathbb{N}$. Due to Corollary 2, we have $\lambda_i(K_j) = \Theta(i^{-2r_j-2})$, and, as shown by Papageorgiou and Wasilkowski [(1990), Theorem 2.1], this implies

$$\sum_{i=n+1}^{\infty} \lambda_i(K) = \Theta(n^{-2\gamma-1}(\log n)^{2(\nu-1)(\gamma+1)}). \tag{4.13}$$

Hence, (4.11) follows from (4.12) and (4.13). □

We add that for $d = 1$ and $r = 0$, Müller-Gronbach (1993) studied the question whether the lower bound (4.12) is sharp in the class of all linear estimators using samples $Y(t_i)$.

Let $L_2(Y)$ denote the space of square integrable random variables that is generated by the random field $Y(t)$, $t \in D$. It is known that the bound (4.12) is attained on the larger class of linear estimators that may use samples of arbitrary elements from $L_2(Y)$. More precisely, let $g_1, g_2, \ldots$ be an orthonormal sequence in $L_2(D)$ such that

$$\int_D K(s,t) g_i(s)\, ds = \lambda_i(K) \cdot g_i(t).$$

Then an optimal linear estimator $B_n$ is given by

$$B_n(Y) = \sum_{i=1}^{n} \int_D Y(t) g_i(t)\, dt \cdot g_i \quad \text{and} \quad e(B_n, K) = \left( \sum_{i=n+1}^{\infty} \lambda_i(K) \right)^{1/2}$$

From the proof of Theorem 3 we immediately conclude the following corollary.

COROLLARY 3. *Assume that the covariance kernel $K$ satisfies condition* (D). *Then for $L_2$-approximation, optimal sampling of $Y(t)$ is, modulo a constant, as powerful as sampling of elements from $L_2(Y)$.*

REMARK 5. We now briefly discuss optimal designs. These designs yield optimal estimators $I_n$ and $A_n$, that is, estimators with minimal error $\inf_{I_n} e(I_n, K)$ and $\inf_{A_n} e(A_n, K)$, respectively. Unfortunately, optimal designs are not known in many cases and, therefore, the notion of optimality is relaxed. Recall that sequences $\{a_n\}_n$ and $\{b_n\}_n$ of positive real numbers are called strongly asymptotically equivalent if $\lim_{n\to\infty} a_n/b_n = 1$ and weakly asymptotically equivalent if $c_1 \le a_n/b_n \le c_2, \forall\ n$, with positive constants $c_1, c_2$. By strongly (weakly) asymptotically optimal designs, we mean designs whose errors are strongly (weakly) asymptotically equivalent to the minimal errors.

The concept of relaxed optimality dates back to Sacks and Ylvisaker (1966), and strongly asymptotically optimal designs are known for many stochastic processes in the univariate case. In the multivariate case, $d > 1$, only weakly asymptotically optimal designs are known.

In particular, for the Wiener sheet $Y$, the design problem for integration is equivalent to minimizing the $L_2$-discrepancy of $n$-point sets. See Woźniakowski (1991). For the latter problem, only weak asymptotic optimality results are known.

Assume now that the covariance kernel $K$ satisfies condition (D) together with the boundary conditions

$$K_j^{(0,k)}(\cdot, t) = 0 \quad \text{for } k = 0, 1, \ldots, r_j \text{ and for } t = 0, 1 \tag{4.14}$$

for any $j = 1, 2, \ldots, d$. In this case, we have $H(K) = \bigotimes_{j=1}^{d} H(P_{r_j})$.

For integration, Frolov (1976) has constructed weakly asymptotically optimal designs; see also Bykovskij (1985) and Temlyakov (1990). For $L_2$-approximation, hyperbolic cross points are weakly asymptotically optimal

designs; see Temlyakov (1987) and Woźniakowski (1992), and see also the end of this remark. An application of the respective mapping $\Psi$ [see (4.8) and (4.10)], to these designs yields weakly asymptotically optimal designs for any kernel which satisfies (D) but not necessarily (4.14). Paskov (1993) discussed practical implementation of these designs.

We add that grid points are a very poor design for the covariance kernels $K$ satisfying the condition (D) in the multivariate case; see Papageorgiou and Wasilkowski (1990).

Finally, we briefly recall the construction of the hyperbolic cross points. Without loss of generality, we can assume that $\gamma = r_1 = \cdots = r_\nu < r_j$ for $j \geq \nu+1$. Let $\boldsymbol{k} = (k_1, \ldots, k_d)$ be such that $k_j = \gamma$, for $j \leq \nu$, and $r_1 < k_j < r_j$, for $j \geq \nu + 1$. Then, given parameter $q$, the hyperbolic cross points design consists of the points

$$\left(\frac{l_1}{2^{s_1}}, \ldots, \frac{l_d}{2^{s_d}}\right), \qquad l_j = 1, 2, \ldots, 2^{s_j} - 1 \text{ and } j = 1, \ldots, d,$$

where $s_j$'s are positive integers with $\sum_{j=1}^{d} s_j k_j \leq \gamma q$. Note that the sample size $n = n(q, \gamma, \boldsymbol{k}, d)$ depends on $q$, $\gamma$, $\boldsymbol{k}$ and $d$. It decreases with $k_j$'s. If $k_j = \gamma$ for all $j$, which corresponds to $\nu = d$, we have

$$n = \sum_{j=0}^{q-d} 2^j \binom{j+d-1}{d-1} \leq 2^{q-d+1} \binom{q-1}{d-1}.$$

Detailed analysis of hyperbolic cross points may be found in Wasilkowski and Woźniakowski (1994).

**5. Extensions for Gaussian random fields.** In this section, we briefly indicate how the results of the previous section can be extended in a number of directions. We will do this under the additional assumption that the corresponding random fields are Gaussian.

Consider a design $t_1, \ldots, t_n \in [0,1]^d$. In general, it may be reasonable to study arbitrary (measurable) estimators that are based on the samples $Y(t_1), \ldots, Y(t_n)$. For Gaussian random fields and for linear problems like integration and $L_2$-approximation, it is well known that it is sufficient to study linear estimators. The minimal error in the class of all estimators that use the given design is attained by a linear estimator.

In this paper, we also restricted our attention to nonsequential designs. Again, for Gaussian random fields this is essentially without loss of generality.

More precisely, consider first designs where the total number $n$ of samples is fixed but the sampling points are chosen adaptively. That is, $t_{i+1} = t_{i+1}(Y(t_1), \ldots, Y(t_i))$ depends on the previously observed values. We stress that there is no restriction on how $t_{i+1}$ may depend on $Y(t_j)$; actually we can also allow nondeterministic (randomized) dependence. Let $V_n$ be an arbitrary estimator (either for integration or $L_2$-approximation) that is based on the sequential design. Then [see Wasilkowski and Woźniakowski (1984)], there

exists a linear estimator $V_n^*$ based on a design with a priori fixed $n$ sampling points such that the errors in the mean square sense satisfy

$$e(V_n^*, K) \leq e(V_n, K).$$

Hence, a sequential design can be better than nonsequential designs only through an adaptive choice of the sample size. However, the gain is bounded by a multiplicative constant. To be more specific, consider an arbitrary design with, as before, adaptively chosen sample points and with the sample size $n = n(Y)$ varying according to some stopping rule. There is no restriction on the stopping rule; it can be arbitrary and even randomized. Let $E(n(Y))$ be the expected number of samples and let $V$ be an estimator that is based on the sequential design. From Wasilkowski (1986) and Theorem 2 it follows that there exists a positive constant $\alpha$ that only depends on $K$ such that, for any sequential design with $E(n(Y)) < \infty$ and any estimator $V$, there exists a nonsequential design with fixed $n^*$ sample points and a linear estimator $V_{n^*}^*$ such that

$$e(V_{n^*}^*, K) \leq \alpha\, e(V, K) \quad \text{and} \quad n^* < E(n(Y)) + 1.$$

Hence, the minimal error of sequential designs is of the same order as the minimal error of nonsequential designs, and Theorem 2 holds for sequential designs as well.

These results can be generalized even further by extending the definition of the error to include $L_p$-norms (instead of the $L_2$-norm), probabilistic setting, where instead of the expected error, one wants to minimize the probability that the error is large, and so forth. For more on this subject, the reader is referred to Traub, Wasilkowski and Woźniakowski (1988) as well as Section 4 of Ritter, Wasilkowski and Woźniakowski (1993).

**Acknowledgment.** We are grateful to an anonymous referee for valuable comments.

## REFERENCES

ADLER, S. J. (1981). *The Geometry of Random Fields*. Wiley, New York.

ARONZSAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

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

BYKOVSKIJ, V. A. (1985). On exact order of optimal quadrature formulas for spaces of functions with bounded mixed derivatives. Report of Dalnevostochnoi Center of Academy of Sciences, Vladivostok, (Russia).

CAMBANIS, S. (1985). Sampling designs for time series. In *Time Series in the Time Domain. Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) 337–367. North-Holland, Amsterdam.

CAMBANIS, S. and MASRY, E. (1983). Sampling designs for the detection of signals in noise. *IEEE Trans. Inform. Theory* **IT-29** 83–104.

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.

FROLOV, K. K. (1976). Upper error bounds for quadrature formulas on function classes. *Soviet Mat. Dokl.* **231** 818–821 (in Russian); *Soviet Math. Dokl.* **17** 1665–1669 (in English).

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

LEE, D. (1986). Approximation of linear operators on a Wiener space. *Rocky Mountain J. Math.* **16** 641–659.

MICCHELLI, C. A. and WAHBA, G. (1981). Design problems for optimal surface interpolation. In *Approximation Theory and Applications* (Z. Ziegler, ed.), 329–347. Academic Press, New York.

MÜLLER-GRONBACH, T. (1993). Optimal designs for approximating the path of a stochastic process. Preprint A-93-14, Fachbereich Mathematik, Freie Univ. Berlin.

PAPAGEORGIOU, A. and WASILKOWSKI, G. W. (1990). On the average complexity of multivariate problems. *J. Complexity* **6** 1–23.

PARZEN, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

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

RITTER, K., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1993). On multivariate integration for stochastic processes. In *Numerical Integration IV* (H. Brass and G. Hämmerlin, eds.) *International Series on Numerical Mathematics* **112** 331–347. Birkhäuser, Basel.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Designs and analysis of computer experiments. *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. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.* **39** 49–69.

SACKS, J. and YLVISAKER, D. (1970a). Design for regression problems with correlated errors III. *Ann. Math. Statist.* **41** 2057–2074.

SACKS, J. and YLVISAKER, D. (1970b). Statistical designs and integral approximation. In *Proc. 12th Biennial Seminar of the Canadian Mathematical Congress* (R. Pyke, ed.) 115–136. Canadian Math. Soc., Montreal.

SPECKMAN, P. (1979). $L_p$ approximation of autoregressive Gaussian processes. Report, Dept. Statistics, Univ. Oregon, Eugene.

SU, Y. and CAMBANIS, S. (1993). Sampling designs for estimation of a random process. *Stochastic Process. Appl.* **46** 47–89.

SULDIN, A. V. (1959, 1960). Wiener measure and its applications to approximation methods I, II. *Izv. Vyssh. Uchebn. Zaved. Mat.* **13** 145–158; **18** 165–179 (in Russian).

TEMLYAKOV, V. N. (1987). Approximate recovery of periodic functions of several variables. *Mat. USSR Sbornik* **56** 249–261.

TEMLYAKOV, V. N. (1990). On a way of obtaining lower estimates for the errors of quadrature formulas errors. *Mat. USSR Sbornik* **181** 1403–1413 (in Russian); *Math. USSR Sbornik* **71** 247–257 (1992) (in English).

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

VAKHANIA, N. N., TARIELADZE, V. I. and CHOBANYAN, S. A. (1987). *Probability Distributions on Banach Spaces*. Reidel, Dordrecht.

WAHBA, G. (1971). On the regression design problem of Sacks and Ylvisaker. *Ann. Math. Statist.* **42** 1035–1043.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

WASILKOWSKI, G. W. (1986). Information of varying cardinality. *J. Complexity* **2** 204–228.

WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1984). Can adaption help on the average? *Numer. Math.* **44** 169–190.

WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1994). Explicit cost bounds of algorithms for multivariate tensor product problems. *J. Complexity*. To appear.

WEIDMANN, J. (1980). *Linear Operators in Hilbert Spaces*. Springer, New York.

WOŹNIAKOWSKI, H. (1991). Average case complexity of multivariate integration. *Bull. Amer. Math. Soc.* **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.

YLVISAKER, D. (1975). Designs on random fields. In *A Survey of Statistical Design and Linear Models* (J. Srivastava, ed.); 593–607. North-Holland, Amsterdam.

K. RITTER
MATHEMATISCHES INSTITUT
UNIVERSITÄT ERLANGEN–NÜRNBERG
BISMARCKSTRASSE 1 1/2
91054 ERLANGEN
GERMANY

G. W. WASILKOWSKI
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40506

H. WOŹNIAKOWSKI
DEPARTMENT OF COMPUTER SCIENCE
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027
AND
INSTITUTE OF APPLIED MATHEMATICS
UNIVERSITY OF WARSAW
UL. BANACHA 2
02-097 WARSZAWA
POLAND