JOURNAL OF COMPLEXITY 9, 291–312 (1993)

Average Case Complexity of Multivariate Integration for Smooth Functions

S. H. PASKOV

Department of Computer Science, Columbia University, New York, New York 10027

Received November 11, 1991

DEDICATED TO JOSEPH F. TRAUB ON THE OCCASION OF HIS 60TH BIRTHDAY

We study the average case complexity of multivariate integration for the class of smooth functions equipped with the folded Wiener sheet measure. The complexity is derived by reducing this problem to multivariate integration in the worst case setting but for a different space of functions. Fully constructive optimal information and an optimal algorithm are presented. Next, fully constructive almost optimal information and an almost optimal algorithm are also presented which have some advantages for practical implementation.

1. INTRODUCTION

Many papers have been devoted to the approximate computation of multivariate integrals; see TWW (1988)[1], Novak (1988), and Niederreiter (1978) for references. Multivariate integrals are approximated by algorithms that use finitely many deterministic integrand values. It is usually assumed that the cost of one integrand evaluation is $c \gg 1$, whereas the cost of arithmetic operation or comparison is taken as 1. The computational complexity of multivariate integration is defined as the minimal cost of approximating multivariate integrals with error at most ε for a given class F of integrands, e.g., TWW (1988). We will consider the average

[1] By TWW (1988) we mean Traub, Wasilkowski, and Woźniakowski (1988).

0885-064X/93 $5.00

case setting, in which cost and error are defined as the expected cost and and error with respect to some probability measure on the class F. Complexity is denoted by $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F)$.

We now present what is known about $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F)$. Let $d = 1$ and $F = C^r$ be the class of r times continuously differentiable functions equipped with the r-folded Wiener measure. Then $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F)$ is essentially equal to $c(\alpha\varepsilon^{-1})^{1/(r+1)}$, i.e.,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F) \simeq c(\alpha\varepsilon^{-1})^{1/(r+1)},$$

where $\alpha = \sqrt{|B_{2r+2}|/(2r + 2)!}$ and B_{2r+2} is a Bernoulli number. Optimal sample points and an optimal algorithm are known as well; see TWW (1988), Lee and Wasilkowski (1986), and Sacks and Ylvisaker (1970).

For the multivariate case $d \geq 1$, let $F = C_d$ be the class of real continuous functions defined on $D = [0, 1]^d$ and equipped with the classical Wiener sheet measure w. That is, w is Gaussian with mean zero and covariance kernel R_w defined as

$$R_w(t, x) \stackrel{\text{def}}{=} \int_{C_d} f(t)f(x)w(df) = \min(t, x) \stackrel{\text{def}}{=} \prod_{j=1}^{d} \min(t_j, x_j),$$

where $t = (t_1, \ldots, t_d)$, $x = (x_1, \ldots, x_d)$ for $t, x \in D$. Woźniakowski (1991) established that

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d) = \Theta(c\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2}).$$

Unfortunately, optimal sample points are not fully constructive. On the other hand, fully constructive almost optimal points and algorithm are given with an average cost of order $c\varepsilon^{-1}(\log \varepsilon^{-1})^{d-1}$.

In this paper, we consider the space $C^{r_1,\ldots,r_d}(D)$, which consists of functions which are continuously differentiable r_j-times with respect to the jth variable. Let $r = r_1 = \cdots = r_k < r_{k+1} \leq \cdots \leq r_d$. The space $C^{r_1,\ldots,r_d}(D)$ is equipped with the folded Wiener sheet measure. We show that

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C^{r_1,\ldots,r_d}(D)) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\left(\log\frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right). \quad (1.1)$$

This is derived by using a relation between the average and worst case settings. The average case problem is reduced to the worst case one for a different space of nonperiodic functions. This worst case problem has been considered by Bykovskij (1985) and Temlyakov (1990); see also Frolov (1976) where the periodic case has been studied.

We present fully constructive optimal sample points and a fully constructive optimal algorithm. For this algorithm (1.1) is attained modulo a multiplicative factor, which may depend on d and $r_1, \ldots, r_d$ but is independent of ε. The optimal sample points are essentially the same as in Bykovskij (1985) and Temlyakov (1991). The optimal algorithm is a simple linear combination of function values at these sample points. We stress that the computation of the optimal sample points, as well as their number, requires *preprocessing*. This preprocessing can be done at cost $O(\varepsilon^{-2/(r+1)}(\log \varepsilon^{-1})^{(k-1)/(r+1)})$, which may significantly exceed (1.1).

We also present a different set of fully constructive sample points which are almost optimal. The sample points can be preprocessed by using the Monte Carlo algorithm at cost $O(\varepsilon^{-1/(r+1)}(\log \varepsilon^{-1})^{(k-1)(4r^2+11r+8)/(r+1)})$. Note that now the cost of preprocessing differs from (1.1) only by a power of $\log \varepsilon^{-1}$. By the almost optimality of sample points and of the algorithm, we mean that the average error ε is achieved with average cost $\Theta(c\varepsilon^{-1/(r+1)}(\log \varepsilon^{-1})^{(k-1)(r+2)/(r+1)})$, which differs from the average case complexity by a power of $\log \varepsilon^{-1}$. These sample points are derived from *hyperbolic cross* points. Hyperbolic cross points have been studied by Temlyakov (1987) for the approximation of functions in the worst case setting. Hyperbolic cross points are also used to derive optimal or almost optimal sample points for multivariate linear problems in the average case setting studied by Woźniakowski (1992). We plan to test the almost optimal sample points and algorithm in the future.

2. Preliminaries

In this section we recall a known fact, see Wahba (1990), Micchelli and Wahba (1981), and also TWW (1988, p. 304), that the average case error of a linear algorithm that uses nonadaptive information is equal to the worst case error of the same algorithm over the unit ball of a Hilbert space. The Hilbert space is defined by the covariance kernel of the a priori measure. We now present the details.

Consider a Banach space F of real-valued functions with domain D. We assume that all functionals of the form $L_x(f) = f(x)$, $x \in D$, are continuous with respect to the norm $\|\cdot\|$ of F. Let F be equipped with a probability measure μ with mean zero and covariance kernel R_μ,

$$R_\mu(t, x) = \int_F f(t)f(x)\mu(df), \qquad t, x \in D. \tag{2.1}$$

Let

$$I(f) = \int_D f(x)dx, \qquad f \in F.$$

We first approximate $I(f)$ by linear algorithms that use function values at nonadaptive points. That is, let

$$N(f) = [f(x^1), \ldots, f(x^n)]$$

be nonadaptive information with cardinality n, where the points x^i of D are given in advance, $i = 1, \ldots, n$. Denote by

$$\varphi(N(f)) = \sum_{i=1}^{n} a_i f(x^i)$$

a linear algorithm that uses $N(f)$. The average case error is defined as

$$e^{\text{avg}}(\varphi, N, F) = \left[\int_F (I(f) - \varphi(N(f)))^2 \mu(df)\right]^{1/2}.$$

We make use of some facts about Hilbert spaces with reproducing kernels. For more details, we refer to Wahba (1990, Chap. 1) or Aronszajn (1950).

A symmetric real-valued function $R(\cdot, \cdot) : D^2 \to \mathbb{R}$ is said to be *positive definite* if for any real $\xi_1, \ldots, \xi_n$ and $y^1, \ldots, y^n \in D$,

$$\sum_{i,j=1}^{n} R(y^i, y^j)\xi_i \xi_j \geq 0.$$

Observe that the covariance kernel R_μ is positive definite. It turns out that every positive definite function $R(\cdot, \cdot)$ uniquely generates a Hilbert space H_R of functions with $R(\cdot, \cdot)$ as its reproducing kernel. That is, H_R is the completion of functions of the form

$$f(\cdot) = \sum_{i=1}^{k} a_i R(\cdot, y^i), \qquad a_i \in \mathbb{R},\ y^i \in D,$$

where k is an arbitrary integer. The inner product of H_R is defined as

$$\langle f, g \rangle_{H_R} \stackrel{\text{def}}{=} \sum_{i=1}^{k} \sum_{j=1}^{m} a_i b_j R(x^i, y^j),$$

where $f(\cdot) = \sum_{i=1}^{k} a_i R(\cdot, x^i)$, $g(\cdot) = \sum_{j=1}^{m} b_j R(\cdot, y^j)$. The basic property of H_R is that

$$f(x) = \langle f, R(\cdot, x) \rangle_{H_R}, \qquad \forall f \in H_R,\ \forall x \in D.$$

Let $H = H_{R_\mu}$ be the Hilbert space generated by the covariance kernel R_μ of the measure μ. It is known that, see Wahba (1990), Micchelli and Wahba (1981), and TWW (1988, p. 304), the average case error of (φ, N) over the class F is equal to the worst case error of (φ, N) over the unit ball of the Hilbert space H, i.e.,

$$e^{\text{avg}}(\varphi, N, F) = e^{\text{wor}}(\varphi, N, B(H)) \overset{\text{def}}{=} \sup_{f \in B(H)} |I(f) - \varphi(N(f))| \quad (2.2)$$

where $B(H) = \{f \in H : \|f\|_H \leq 1\}$.

For the reader's convenience, we provide a sketch of the proof. Using Fubini's Theorem and (2.1) we can compute the average case error of (φ, N) directly to get

$$e^{\text{avg}}(\varphi, N, F) = \int_D \int_D R_\mu(t, x)dtdx - 2 \sum_{i=1}^{n} a_i \int_D R_\mu(t, x^i)dt + \sum_{i,j=1}^{n} a_i a_j R_\mu(x^i, x^j).$$

Since $I(f) = \int_D \langle f, R_\mu(\cdot, t)\rangle_H dt$ and $\varphi(N(f)) = \langle f, \sum_{i=1}^n a_i R_\mu(\cdot, x^i)\rangle_H$, we obtain

$$\begin{aligned} e^{\text{wor}}(\varphi, N, B(H))^2 &= \sup_{f \in B(H)} \left| \left\langle f, \int_D R_\mu(\cdot, t)dt - \sum_{i=1}^{n} a_i R_\mu(\cdot, x^i) \right\rangle_H \right| \\ &= \left\| \int_D R_\mu(\cdot, t)dt - \sum_{i=1}^{n} a_i R_\mu(\cdot, x^i) \right\|_H^2 \\ &= \int_D \int_D R_\mu(t, x)dtdx - 2 \sum_{i=1}^{n} a_i \int_D R_\mu(t, x^i)dt \\ &\quad + \sum_{i,j=1}^{n} a_i a_j R_\mu(x^i, x^j), \end{aligned} \quad (2.3)$$

which proves (2.2).

3. Main Result

Consider the class $C^{\bar{r}} = C^{r_1, \ldots, r_d}(D)$ of real-valued functions on D that are r_j times continuously differentiable with respect to x_j, $j = 1, 2, \ldots, d$, where $\bar{r} = (r_1, \ldots, r_d)$ and the r_j are nonnegative integers. Let $D^{i_1, \ldots, i_d}$ denote the derivative operator

$$D^{i_1,\dots,i_d}f = \frac{\partial^{i_1+\cdots+i_d}}{\partial x_1^{i_1}\dots\partial x_d^{i_d}} f.$$

Define

$$C_0^{\bar r} \stackrel{\text{def}}{=} \{f \in C^{\bar r} : (D^{i_1,\dots,i_d}f)(t_1, \dots, t_d) = 0, \forall i_j = 0, 1 \dots, r_j,$$
$$\forall j = 1, 2, \dots, d, \text{ whenever } t_k = 0 \text{ for some } 1 \le k \le d\}.$$

The space $C_0^{\bar r}$ is equipped with the sup norm $\|f\| = \sup_{t\in D} |D^{r_1,\dots,r_d}f(t)|$. Note that $D^{r_1,\dots,r_d}$ restricted to $C_0^{\bar r}$ is a one-to-one operator. The folded Wiener sheet measure on $\mathcal{B}(C_0^{\bar r})$ is defined by

$$\mu(B) = w(D^{r_1,\dots,r_d}B), \qquad \forall B \in \mathcal{B}(C_0^{\bar r}),$$

where w is the classical Wiener sheet measure on $\mathcal{B}(C_0^{\bar 0})$; for more details, see Kuo (1975) and Adler (1981). As already mentioned in the introduction, w is Gaussian with mean zero and covariance kernel $R_w(t, x) = \min(t, x)$.

The covariance kernel R_μ of the folded Wiener sheet measure μ can be founded, for instance, in Papageorgiou and Wasilkowski (1990),

$$R_\mu(t, x) = \prod_{j=1}^{d} \int_0^1 \frac{(t_j - s_j)_+^{r_j-1}}{(r_j - 1)!} \frac{(x_j^{r_j+1} - (x_j - s_j)_+^{r_j+1})}{(r_j + 1)!} ds_j.$$

Integration by parts in each direction yields

$$R_\mu(t, x) = \int_D \prod_{j=1}^{d} \frac{(t_j - s_j)_+^{r_j}}{r_j!} \frac{(x_j - s_j)_+^{r_j}}{r_j!} ds_1 \dots ds_d. \tag{3.1}$$

Following Section 2, we explicitly describe the Hilbert space $H = H_{R_\mu}$ generated by R_μ.

Consider the space $W_2^{\overline{r+1}} \stackrel{\text{def}}{=} W_2^{r_1+1,\dots,r_d+1}(D)$, which consists of all real-valued functions f on D having the form

$$f(x) = \int_D \prod_{j=1}^{d} \frac{(x_j - s_j)_+^{r_j}}{r_j!} \varphi(s_1, \dots, s_d) ds_1 \dots ds_d, \qquad \forall x \in D, \tag{3.2}$$

where $\varphi \in L_2(D)$. Note that

$$D^{r_1+1,\dots,r_d+1}f = \varphi. \tag{3.3}$$

The space $W_2^{\overline{r+1}}$ becomes a Hilbert space when equipped with the inner product

$$\langle f, g\rangle = \int_D (D^{r_1+1,\ldots,r_d+1}f)(s)(D^{r_1+1,\ldots,r_d+1}g)(s)\, ds. \tag{3.4}$$

LEMMA 1. $H_{R_\mu} = W_2^{\overline{r+1}}$.

Proof. First we check that R_μ is a reproducing kernel for $W_2^{\overline{r+1}}$. Indeed $R_\mu(\cdot, x) \in W_2^{\overline{r+1}}$ for every $x \in D$, because $\prod_{j=1}^d (t_j - \cdot)_+^{r_j}/r_j! \in L_2(D)$. For an arbitrary function f of the form (3.2) consider

$$\langle f, R_\mu(\cdot, x)\rangle_H = \int_D (D^{r_1+1,\ldots,r_d+1}f)(s)(D^{r_1+1,\ldots,r_d+1}R_\mu(\cdot, x))(s)\, ds.$$

Applying (3.3) and (3.1), we get

$$\langle f, R_\mu(\cdot, x)\rangle_H = \int_D \varphi(s_1, \ldots, s_d) \prod_{j=1}^d \frac{(x_j - s_j)_+^{r_j}}{r_j!}\, ds_j = f(x).$$

Thus, R_μ is a reproducing kernel for the Hilbert space $W_2^{\overline{r+1}}$. Since every reproducing kernel Hilbert space determines uniquely its reproducing kernel and since $R_\mu(\cdot, x) \in W_2^{\overline{r+1}}$ for every $x \in D$, we conclude the claim of Lemma 1. ■

Let $B(W_2^{\overline{r+1}}) = \{\|f\|_H \le 1 : f \in W_2^{\overline{r+1}}\}$ be the unit ball of $W_2^{\overline{r+1}}$. From Section 2 and Lemma 1 we conclude that

$$e^{\text{avg}}(\varphi, N, C_0^{\bar r}) = e^{\text{wor}}(\varphi, N, B(W_2^{\overline{r+1}})) \tag{3.5}$$

for any linear algorithm $\varphi(N(f)) = \sum_{i=1}^n a_i f(x^i)$, that uses nonadaptive information $N(f) = [f(x^1), \ldots, f(x^n)]$. Let

$$e_n(F) \stackrel{\text{def}}{=} \inf\{e(\varphi, N, F) : N \text{ nonadaptive and of cardinality } n,\ \varphi \text{ linear}\}.$$

Then (3.5) yields

COROLLARY 1.

$$e_n^{\text{avg}}(C_0^{\bar r}) = e_n^{\text{wor}}(B(W_2^{\overline{r+1}})).$$

To find bounds on $e_n^{\text{wor}}(B(W_2^{\overline{r+1}}))$ we proceed as follows. Observe that the space $W_2^{\overline{r+1}}$ consists of smooth functions such that

$$D^{i_1,\ldots,i_d}f(t) = 0$$

for all $i_j \le r_j$ and for all t with at least one component equal to zero.

If we apply the linear transformation $x = 1 - t$ to the functions from $W_2^{\overline{r+1}}$ we obtain the space that was considered by Bykovskij (1985) and Temlyakov (1990). By using a periodization technique, see also Keng and Yuan (1981), each function of this nonperiodic space is transformed into a periodic one with the same smoothness and with the same integral over D. More precisely, let $\tilde{W}_2^{\overline{r+1}} \overset{\text{def}}{=} \tilde{W}_2^{r_1+1,\ldots,r_d+1}(D)$ be the class of all functions $r_j + 1$ times differentiable with respect to x_j which satisfy $(D^{i_1,\ldots,i_d}f)(x) = 0$ for all x belonging to the boundary of D and $0 \le i_j \le r_j + 1$, $j = 1, \ldots, d$ equipped with the norm,

$$\|f\|_{\overline{r+1}} = \sum_{\substack{0 \le i_j \le r_j + 1 \\ 1 \le j \le d}} \|D^{i_1,\ldots,i_d}f\|_2$$

where, as usual, $\|f\|_2 = (\int_D f(x)^2\,dx)^{1/2}$.

It is clear that

$$B(\tilde{W}_2^{\overline{r+1}}) \subset B(W_2^{\overline{r+1}}). \tag{3.6}$$

Define

$$\psi_r(u) = \frac{\int_0^u t^{2r}(1-t)^{2r}\,dt}{\int_0^1 t^{2r}(1-t)^{2r}\,dt} \qquad \text{for } 0 \le u \le 1. \tag{3.7}$$

Consider now arbitrary nonadaptive information $N_n(f) = [f(x^1), \ldots, f(x^n)]$ and an arbitrary linear algorithm $\varphi(N(f)) = \sum_{i=1}^n a_i f(x^i)$. Let $\psi(x) = (\psi_{r_1+1}(x_1), \ldots, \psi_{r_d+1}(x_d))$. For the pair (φ, N), we define $(\overline{\varphi}, \overline{N})$ by

$$\begin{aligned} \overline{N}_n(f) &= [f(\psi(x^1)), \ldots, f(\psi(x^n))] \\ \overline{\varphi}(\overline{N}(f)) &= \sum_{i=1}^n a_i \psi'_{r_1+1}(x_1^i) \ldots \psi'_{r_d+1}(x_d^i) f(\psi(x^i)). \end{aligned} \tag{3.8}$$

It follows from Bykovskij (1985) that there exists a constant K depending only d and r_i such that

$$e_n^{\text{wor}}(\overline{\varphi}, \overline{N}_n, B(W_2^{\overline{r+1}})) \le K e_n^{\text{wor}}(\varphi, N_n, B(\tilde{W}_2^{\overline{r+1}})). \tag{3.9}$$

Taking the infimum with respect to φ and N in (3.9),

$$e_n^{\text{wor}}(B(W_2^{\overline{r+1}})) \le K e_n^{\text{wor}}(B(\tilde{W}_2^{\overline{r+1}})).$$

Taking into account (3.6), this means that

$$e_n^{\text{wor}}(B(W_2^{\overline{r+1}})) = \Theta(e_n^{\text{wor}}(B(\tilde{W}_2^{\overline{r+1}}))). \tag{3.10}$$

Frolov (1976) and Bykovskij (1985) proved that

$$e_n^{\text{wor}}(B(\tilde{W}_2^{\overline{r+1}})) = \Theta\left(\frac{(\log n)^{(k-1)/2}}{n^{r+1}}\right), \tag{3.11}$$

where $r = r_1 = \ldots r_k < r_{k+1} \leq \ldots \leq r_d$. Corollary 1, (3.10), and (3.11) yield

COROLLARY 2.

$$e_n^{\text{avg}}(C_0^{\bar{r}}) = \Theta\left(\frac{(\log n)^{(k-1)/2}}{n^{r+1}}\right).$$

In what follows, when we use the word "optimal," we ignore a multiplicative factor which may depend on d and $r_1, \ldots, r_d$ but is independent of n.

Note that Corollary 1, (3.5), (3.8), and (3.9) can be used to translate optimality results from the worst case to the average case setting. More specifically, let

$$N_n^*(f) = [f(x^{1*}), \ldots, f(x^{n*})]$$

and

$$\varphi^*(N^*(f)) = \sum_{i=1}^{n} a_i^* f(x^{i*})$$

be optimal information and an optimal algorithm for integration in the worst case setting for the class $B(\tilde{W}_2^{\overline{r+1}})$, i.e., $e_n^{\text{wor}}(\varphi^*, N^*, B(\tilde{W}_2^{\overline{r+1}})) = \Theta(e_n^{\text{wor}}(B(\tilde{W}_2^{\overline{r+1}})))$ with the constant in the Θ notation depending only on d and r_j. Then

$$\begin{aligned} \overline{N}_n^*(f) &= [f(\psi(x^{1*})), \ldots, f(\psi(x^{n*}))], \\ \overline{\varphi}^*(\overline{N}^*(f)) &= \sum_{i=1}^{n} a_i \psi'_{r_1+1}(x_1^{i*}) \ldots \psi'_{r_d+1}(x_d^{i*}) f(\psi(x^{i*})) \end{aligned} \tag{3.12}$$

are optimal information and an optimal algorithm for integration in the average case setting for the class $C_0^{\bar{r}}$, i.e., $e_n^{\text{avg}}(\overline{\varphi}^*, \overline{N}^*, C_0^{\bar{r}}) = \Theta(e_n^{\text{avg}}(C_0^{\bar{r}}))$, with the constant in the Θ notation depending only on d and r_j.

We now describe optimal information and an optimal algorithm for $B(\tilde{W}_2^{\overline{r+1}})$, which are given in Frolov (1976), Bykovskij (1985), and Temlyakov (1990). Let Q be an irreducible monic polynomial of degree d with integer coefficients and d distinct real roots $\omega_1, \omega_2, \ldots, \omega_d$. An example of Q is

$$Q(x) = \prod_{j=1}^{d} (x - (2j - 1)) - 1, \qquad d > 2.$$

Let A be a $d \times d$ Vandermonde matrix

$$A = \begin{pmatrix} 1 & \ldots & 1 \\ \omega_1 & \ldots & \omega_d \\ \vdots & \ddots & \vdots \\ \omega_1^{d-1} & \ldots & \omega_d^{d-1} \end{pmatrix},$$

Consider the inverse matrix $B = A^{-1}$, which is assumed to be of the form

$$B = \frac{1}{\det(A)} \begin{pmatrix} \Omega_1^{(1)} & \ldots & \Omega_d^{(1)} \\ \vdots & \ddots & \vdots \\ \Omega_1^{(d)} & \ldots & \Omega_d^{(d)} \end{pmatrix}.$$

For integers m_i, define

$$L_1^*(m_1, \ldots, m_d) = \Omega_1^{(1)} m_1 + \ldots + \Omega_d^{(1)} m_d$$
$$\vdots$$
$$L_d^*(m_1, \ldots, m_d) = \Omega_1^{(d)} m_1 + \ldots + \Omega_d^{(d)} m_d.$$

Let q be an integer greater than 1. Then the optimal sample points are given by the set

$$\mathcal{M}_q = \left\{ \left(\frac{L_1^*(m)}{q \det(A)}, \ldots, \frac{L_d^*(m)}{q \det(A)} \right) \in D : m = (m_1, \ldots, m_d), \text{ all } m_i \text{ are integers} \right\}. \qquad (3.13)$$

Let n_q be the number of the points in $\mathcal{M}_q$. It turns out that $n_q = \Theta(q^d)$, where the constant in the Θ notation depends only on the matrix A.

Let n be an arbitrary integer. Consider first the case $n = n_q$ for some q.

Then optimal information of cardinality n and an optimal algorithm are given by

$$N_n^*(f) = [f(x^{1*}), \ldots, f(x^{n*})], \qquad \text{where } x^{i*} \in \mathcal{M}_q,$$

$$\varphi^*(N^*(f)) = \frac{1}{|\det(A)|q^d} \sum_{i=1}^{n} f(x^{i*}). \tag{3.14}$$

Consider next the case $n \neq n_q$. Then we can find q such that $n_q \le n < n_{q+1}$. Then $n_q = O(q^d)$. As optimal information of cardinality at most n and an optimal algorithm, we can take (3.14) with $n = n_q$. Note that the weights of the optimal algorithm are equal.

Clearly, the computation of the optimal sample points requires *preprocessing* since we have to compute the roots of the polynomial Q as well as the elements of the matrix B.

We now derive bounds on the average case computational complexity of multivariate integration for the class $C_0^{\bar{r}}$.

Consider first nonadaptive information. Let $\{x^{i*}\}_{i=1}^n$ be the sample points from the set $\mathcal{M}_q$ given by (3.13). For a given positive ε choose the smallest n for which the error of the algorithm given by (3.14) for the class $B(\bar{W}_2^{\overline{r+1}})$ does not exceed ε/K. Due to (3.11) and optimality of the algorithm (3.14), we have

$$n = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{1/(r+1)} \left(\log \frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right) \tag{3.15}$$

Define the sample points

$$x^{i**} = \psi(x^{i*}), \qquad i = 1, 2, \ldots, n \tag{3.16}$$

and weights

$$a_i^{**} = \psi'_{r_1+1}(x_1^{i*}) \ldots \psi'_{r_d+1}(x_d^{i*}))/|\det(A)|q^d.$$

Define the algorithm U_n^*,

$$U_n^*(f) = \sum_{i=1}^{n} a_i^{**} f(x^{i**}). \tag{3.17}$$

If n is known then the cost of computing $U_n^*(f)$ is $cn + 2n - 1$ since the sample points x^{i**} and weights a_i^{**} can be preprocessed.

We now explain how n satisfying (3.15) can be also preprocessed. Indeed, for arbitrary n it is easy to check, see also Lemma 3, that

$$[e^{\text{avg}}(U_n^*)]^2 = \int_D \prod_{j=1}^d \frac{(1-s_j)^{2(r_j+1)}}{[(r_j+1)!]^2}\,ds$$

$$- 2\sum_{i=1}^n a_i^{**} \int_D \prod_{j=1}^d \frac{(1-s_j)^{r_j+1}}{(r_j+1)!}\frac{(x_j^{i**}-s_j)_+^{r_j}}{r_j!}\,ds$$

$$+ \sum_{i,k=1}^n a_i^{**} a_k^{**} \int_D \prod_{j=1}^d \frac{(x_j^{i**}-s_j)_+^{r_j}}{r_j!}\frac{(x_j^{k**}-s_j)_+^{r_j}}{r_j!}\,ds.$$

After explicitly calculating the integrals in the above formula, it can be shown that $e^{\text{avg}}(U_n^*)$ can be computed exactly using at most $4(\sum_{j=1}^d r_j^2)n^2$ arithmetic operations. This way, we can check whether $e^{\text{avg}}(U_n^*) \le \varepsilon$. If so, we are done. If not, we can, for instance, double n and after a few steps we will find n for which $e^{\text{avg}}(U_n^*) \le \varepsilon$. Hence, we can find the proper n and the cost of this preprocessing is proportional to n^2 which is $O(\varepsilon^{-2/(r+1)}(\log \varepsilon^{-1})^{(k-1)/(r+1)})$.

Thus, after preprocessing of the cardinality n, the sample points x^{i**}, and the weights a_i^{**}, the average case error of U_n^* is at most ε due to (3.5) and (3.9). Therefore, we have an upper bound on the average case complexity

$$\text{comp}^{\text{avg}}(\varepsilon, C_0^{\bar{r}}) = O\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\left(\log\frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right).$$

We now prove a lower bound on the average complexity. Let $x^1, \ldots, x^n$ be any nonadaptive sample points in D. Since the problem is linear, μ is Gaussian and the average case error is defined in the L_2 sense, it is known that the algorithm U_n^* that uses nonadaptive information $N_n(f) = [f(x^1), \ldots, f(x^n)]$ and minimizes the average error, is of the form $U_n^*(f) = \sum_{i=1}^n c_i f(x^i)$ for some coefficients c_i; see TWW (1988). Using Corollary 2, we conclude that $e^{\text{avg}}(U_n^*, C_0^{\bar{r}})$ is at most ε only if

$$n = \Omega\left(\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\left(\log\frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right).$$

Therefore, the cost of approximating $I(f)$ is bounded from below by

$$cn = \Omega\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\left(\log\frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right).$$

So far, we considered only a nonadaptive choice of sample points. Let $x^1, \ldots, x^{n(f)}$ be adaptive sample points. Because a Gaussian measure is used, adaption may help only by varying cardinality $n(f)$, see TWW (1988, p. 239). On the other hand, using Wasilkowski's theorem, see TWW (1988, p. 246), it is easy to see that the squares of the radii of optimal nonadaptive information are semi-convex. Thus, varying $n(f)$ can only help by a multiplicative constant. We summarize these results in the following theorem.

THEOREM 1.

(i) $$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_0^{\bar{r}}) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\left(\log\frac{1}{\varepsilon}\right)^{(k-1)/(2(r+1))}\right),$$

(ii) $x^{1**}, \ldots, x^{n**}$ given by (3.16) are optimal sample points,

(iii) U_n^* which is given by (3.17) is optimal.

4. Hyperbolic Cross Points

We now present fully constructive almost optimal sample points and algorithm which can be *preprocessed* at cost $O(\varepsilon^{-1/(r+1)}(\log \varepsilon^{-1})^{(k-1)(4r^2+11r+8)/(r+1)})$. These almost optimal sample points are derived from hyperbolic cross points which have been studied by Temlyakov (1987). He analyzed, in particular, multivariate integration for the class of periodic functions $B(\bar{W}_2^{r+1})$ in the worst case setting. For the reader's convenience we repeat his construction here.

Assume first that $d = 1$. Define the functional

$$\sigma_\nu(f) = \frac{1}{8\nu}\sum_{i=1}^{8\nu-1} f\left(\frac{i}{8\nu}\right) \tag{4.1}$$

and define

$$\begin{aligned} \Delta\sigma_0(f) &= \sigma_1(f), \\ \Delta\sigma_0(f) &= \sigma_{2^\nu}(f) - \sigma_{2^{\nu-1}}(f) \qquad \text{for } \nu \geq 1. \end{aligned} \tag{4.2}$$

Replacing (4.1) in the right-hand side of (4.2), we obtain

$$\Delta\sigma_\nu(f) = \frac{1}{2^{\nu+3}}\sum_{i=1}^{2^{\nu+3}-1}(-1)^{\varepsilon(i+1)}f\left(\frac{i}{2^{\nu+3}}\right), \tag{4.3}$$

where $\varepsilon = 0$ if $\nu = 0$ and $\varepsilon = 1$ otherwise.

For $d \geq 1$, let

$$\Delta\sigma_{\bar{s}}(f) = \Delta\sigma_{s_1}(\Delta\sigma_{s_2} \ldots \Delta\sigma_{s_d}(f) \ldots),$$

where $\bar{s} = (s_1, \ldots, s_d)$ and the functionals $\Delta\sigma_{s_j}$ act on a function depending only on x_j, $j = 1, 2, \ldots d$.

We now give the explicit expression for $\Delta\sigma_{\bar{s}}(f)$. Making use of (4.3), it is not difficult to prove by induction on d that

$$\Delta\sigma_{\bar{s}}(f) = \sum_{i_1=1}^{2^{s_1+3}-1} \cdots \sum_{i_d=1}^{2^{s_d+3}-1} \frac{(-1)^{\varepsilon_1(i_1+1)+\cdots+\varepsilon_d(i_d+1)}}{2^{(s_1+3)+\cdots+(s_d+3)}} f\left(\frac{i_1}{2^{s_1+3}}, \ldots, \frac{i_d}{2^{s_d+3}}\right), \tag{4.4}$$

where $\varepsilon_j = 0$ if $s_j = 0$ and $\varepsilon_j = 1$ if $s_j > 0$, $j = 1, 2, \ldots, d$.

Let the vector $\overline{r' + 1}$ be such that $r + 1 = r'_1 + 1 = \cdots = r'_k + 1$ and $r + 1 < r'_j + 1 < r_j + 1$ for $j = k + 1, k + 2, \ldots, d$. For given integer m, consider the hyperbolic cross $\mathscr{X}_m$ of sample points,

$$\mathscr{X}_m = \left\{\left(\frac{i_1}{2^{s_1+3}}, \frac{i_2}{2^{s_2+3}}, \ldots, \frac{i_d}{2^{s_d+3}}\right) : i_j = 1, 2, \ldots, 2^{s_j+3} - 1, j = 1, 2, \ldots, d, \bar{s} \in S,\right\}$$

where S is the set of all nonnegative integer vectors $\bar{s}$ which satisfy the condition $(\overline{r' + 1}, \bar{s}) \leq m(r + 1)$. Let Q_m be the number of the points in $\mathscr{X}_m$. It turns out that

$$Q_m = \Theta(2^m m^{k-1}).$$

For an arbitrary number m, we define

$$\Phi_{Q_m}(f) = \sum_{(\overline{r'+1}, \bar{s}) \leq m(r+1)} \Delta\sigma_{\bar{s}}(f).$$

Now, let the number n be given. Find the number m such that $Q_m \leq n < Q_{m+1}$. Then $n = O(2^m m^{k-1})$. As the functional $\Phi_n(f)$ we take $\Phi_{Q_m}(f)$. The functional $\Phi_n(f)$ depends linearly on $f(x^i)$ and can be represented as

$$\Phi_n(f) = \sum_{x^i \in \mathscr{X}_m}^{n} b_i f(x^i). \tag{4.5}$$

We now calculate the weights b_i. Consider the point

$$x^i = \left\{\left(\frac{i_1}{2^{s_1+3}}, \frac{i_2}{2^{s_2+3}}, \ldots, \frac{i_d}{2^{s_d+3}}\right) \in \mathscr{X}_m, \qquad (4.6)$$

where $i_j \leq 2^{s_j+3} - 1$ are odd positive integers if $s_j > 0$ and $i_j = 1, 2, \ldots, 7$ if $s_j = 0$. All points from $\mathscr{X}_m$ which contribute to the weight b_i are of the form

$$\left(\frac{i_1 2^{p_1}}{2^{(s_1+p_1)+3}}, \frac{i_2 2^{p_2}}{2^{(s_2+p_2)+3}}, \ldots, \frac{i_d 2^{p_d}}{2^{(s_d+p_d)+3}}\right), \qquad (4.7)$$

where p_j are nonnegative integers satisfying $(\overline{r'+1}, \overline{s+p}) \leq m(r+1)$, or equivalently

$$(\overline{r'+1}, \bar{p}) \leq m(r+1) - (\overline{r'+1}, \bar{s}) = A(\bar{s}). \qquad (4.8)$$

For the fixed $\bar{p}$ the coefficient of the point (4.7) in (4.4) is

$$\frac{(-1)^{\varepsilon_1(i_1 2^{p_1}+1)+\cdots+\varepsilon_d(i_d 2^{p_d}+1)}}{2^{(s_1+p_1+3)+\cdots+(s_d+p_d+3)}}. \qquad (4.9)$$

We remind the reader that $\varepsilon_j = 0$ if $s_j + p_j = 0$ and $\varepsilon_j = 1$ if $s_j + p_j > 0$, $j = 1, 2, \ldots, d$. Clearly, if $p_j = 0$ then $\varepsilon_j(i_j 2^{p_j} + 1)$ is an even number since either ε_j is 0 or i_j is odd. If $p_j \geq 1$ then $\varepsilon_j = 1$ and so, we have that $\varepsilon_j(i_j 2^{p_j} + 1)$ is an odd number. Let

$$\alpha(\bar{p}) = \#\{p_j : p_j > 0, j = 1, \ldots, d\}$$

be the number of positive elements of $\bar{p}$. Then, the coefficient (4.9) takes the form

$$\frac{(-1)^{\alpha(\bar{p})}}{2^{(s_1+3)+\cdots+(s_d+3)+p_1+\cdots+p_d}}.$$

Thus we conclude that

$$b_i = \frac{1}{2^{(s_1+3)+\cdots+(s_d+3)}} \sum_{(\overline{r'+1}, \bar{p}) \leq A(\bar{s})} \frac{(-1)^{\alpha(\bar{p})}}{2^{p_1+\cdots+p_d}}. \qquad (4.10)$$

Consider the system of inequalities

$$(\overline{r'+1}, \bar{p}) \leq A(\bar{s})$$
$$p_1 + p_2 + \cdots + p_d = \nu, \qquad (4.11)$$

where ν is an integer. Denote by $\tau_1(\nu)$ the number of integer non-negative solutions of (4.11) with an even number of positive coordinates. Analogously, denote by $\tau_2(\nu)$ the number of solutions of (4.11) with an odd number of positive coordinates. From the inequality $r + 1 \le r_j + 1$, $j = 1, \ldots, d$, it easily follows that if $\bar{p}$ satisfies (4.8), then $\bar{p}$ satisfies $\sum_{j=1}^{d} p_j \le m - \sum_{j=1}^{d} s_j = B(\bar{s})$. Then (4.10) can be rewritten as

$$b_i = \frac{1}{2^{(s_1+3)+\cdots+(s_d+3)}} \sum_{\nu=0}^{B(\bar{s})} \frac{\tau_1(\nu) - \tau_2(\nu)}{2^\nu}. \tag{4.12}$$

It is easy to construct an algorithm which calculates the numbers $\tau_1(\nu)$ and $\tau_2(\nu)$. For instance, we can generate all $\bar{p}$ such that $\sum_{j=1}^{d} p_j = \nu$ for $0 \le \nu \le B(\bar{s})$, and check which of them satisfy the inequality in (4.11); if so, we increase $\tau_1(\nu)$ or $\tau_2(\nu)$ by 1, respectively. Note that, in general, the coordinates of $\overline{r' + 1}$ might be real numbers.

We now give an explicit expressions for b_i under the assumption $r = r_1 = \cdots = r_d$. In this case the inequality in (4.11) takes the form $\sum_{j=1}^{d} p_j \le m - \sum_{j=1}^{d} s_j$. Clearly, if the equation in (4.11) is satisfied with $0 \le \nu \le m - \sum_{j=1}^{d} s_j$ then we can ignore the inequality in (4.11). Thus, in our case $\tau_1(\nu)$ ($\tau_2(\nu)$) is the number of solutions of

$$\sum_{i=1}^{d} p_i = \nu \tag{4.13}$$

with even (odd) number of positive coordinates. It is well known that the number of integer non-negative solutions of (4.13) is $\binom{\nu+d-1}{d-1}$. Let us assume that exactly j solutions of (4.13) are positive, i.e.,

$$p_{i_1} + p_{i_2} + \cdots + p_{i_j} = \nu. \tag{4.14}$$

By subtracting j from both sides, we can find that the number of integer positive solutions of (4.14) is $\binom{\nu-1}{j-1}$. Since we can choose any j indices from d, the number of solutions of (4.13) with exactly j positive coordinates is $\binom{\nu-1}{j-1}\binom{d}{j}$. Then we get

$$b_i = \frac{1}{2^{(s_1+3)+\cdots+(s_d+3)}} \left(1 + \sum_{\nu=1}^{B(\bar{s})} \frac{1}{2^\nu} \sum_{j=1}^{\nu} (-1)^j \binom{\nu-1}{j-1}\binom{d}{j}\right) \tag{4.15}$$

with the convention $\binom{d-1}{\nu} = 0$ for $\nu \ge d$.

We now simplify (4.15). We make use of the following equality

$$\sum_{\nu=j}^{\beta} \frac{1}{2^\nu} \binom{\nu-1}{j-1} = \frac{1}{2^\beta} \sum_{\nu=j}^{\beta} \binom{\beta}{\nu}, \tag{4.16}$$

where $\nu \leq \beta$ and β is an integer. This can be proved by induction on β and using β times the well known equality $\binom{\beta}{\nu} + \binom{\beta}{\nu+1} = \binom{\beta+1}{\nu+1}$. We omit the proof.

After changing indices of summation, (4.15) takes the form

$$b_i = \frac{1}{2^{(s_1+3)+\cdots+(s_d+3)}}\left(1 + \sum_{j=1}^{B(\bar{s})} (-1)^j \binom{d}{j} \sum_{\nu=j}^{B(\bar{s})} \frac{1}{2^\nu} \binom{\nu-1}{j-1}\right). \tag{4.17}$$

Let us use (4.16) and again change the indices of summation. We get

$$b_i = \frac{1}{2^{(s_1+3)+\cdots+(s_d+3)+B(\bar{s})}} \sum_{\nu=0}^{B(\bar{s})} \binom{B(\bar{s})}{\nu} \sum_{j=0}^{\nu} (-1)^j \binom{d}{j}. \tag{4.18}$$

We now employ the following equality

$$\sum_{j=0}^{\nu} (-1)^j \binom{d}{j} = (-1)^\nu \binom{d-1}{\nu}$$

which can be found, for example, in Gradshteyn and Ryzhik (1980, p. 3). Thus we have proven

LEMMA 2. *Assume that* $r = r_1 = \cdots = r_d$. *Then the coefficient of the point* x^i *given by* (4.6) *is*

$$b_i = \frac{1}{2^{m+3d}} \sum_{\nu=0}^{B(\bar{s})} (-1)^\nu \binom{B(\bar{s})}{\nu}\binom{d-1}{\nu}, \tag{4.19}$$

where $B(\bar{s}) = m - \sum_{j=1}^{d} s_j$ *and with the convention* $\binom{d-1}{\nu} = 0$ *for* $\nu \geq d$.

Observe that we have equal coefficients b_i when $\sum_{j=1}^{d} s_j$ is the same. Since $\sum_{j=1}^{d} s_j$ can take at most $m + 1$ distinct values, we have at most $m + 1$ distinct coefficients b_i, where, obviously, $m \leq \log n$.

We now turn to the error formula. Temlyakov (1987) proved that

$$e_n^{\text{wor}}(\Phi_n, B(\tilde{W}_2^{r+1})) = O\left(\frac{(\log n)^{(k-1)(r+2)}}{n^{r+1}}\right). \tag{4.20}$$

Consider now the algorithm

$$\bar{\Phi}_n(f) = \sum_{x^i \in \mathfrak{X}_m}^{n} \bar{b}_i f(\psi(x^i)), \tag{4.21}$$

where $\bar{b}_i = b_i \psi'_{r_1+1}(x_1^i) \ldots \psi'_{r_d+1}(x_d^i)$. From (3.5), (3.9), and (4.20) it follows that

$$e_n^{\mathrm{avg}}(\overline{\Phi}_n, C_0^{\bar{r}}) = O\left(\frac{(\log n)^{(k-1)(r+2)}}{n^{r+1}}\right). \tag{4.22}$$

Hence, if we choose

$$n = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{1/(r+1)} \left(\log \frac{1}{\varepsilon}\right)^{(k-1)(r+2)/(r+1)}\right) \tag{4.23}$$

then the average case error of $\overline{\Phi}_n$ is at most ε. Since n, $\psi(x^i)$ and $\bar{b}_i$ can be preprocessed, the average cost of Φ_n is $cn + 2n - 1$. Due to (4.23), this cost differs from the average case complexity by a power of $\log \varepsilon^{-1}$. We summarize this in the following theorem.

THEOREM 2. *The sample points $\psi(x^i)$, $x^i \in \mathscr{X}_m$, and weights $\bar{b}_i$, of (4.21), $i = 1, \ldots, n$ with n given by (4.23) are almost optimal,*

$$\mathrm{cost}^{\mathrm{avg}}(\overline{\Phi}_n) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)} \left(\log \frac{1}{\varepsilon}\right)^{(k-1)(r+2)/(r+1)}\right).$$

We now analyze the cost of preprocessing. Observe that if n is preprocessed, the rest is easy since (3.7), (4.18), and (4.12) can be used to compute $\psi(x^i)$ and b_i. Hence, we concentrate on how to preprocess n satisfying (4.23). Unfortunately, we cannot use (4.23) directly since the constant in the Θ notation is not specified. We now explain how to compute n cheaply.

Let $X = \{x^1, \ldots, x^n\}$ be the set of points in D and $A = \{a_1, \ldots, a_n\}$ be the set of n real numbers. Denote by

$$D(s; X, A) = \left(\sum_{i=1}^{n} a_i \prod_{j=1}^{d} \frac{(s_j - x_j^i)_+^{r_j}}{r_j!}\right) - \prod_{j=1}^{d} \frac{s_j^{r_j+1}}{(r_j+1)!}.$$

We add that if $r_1 = \ldots = r_d = 0$, and $a_i = 1/n$, $i = 1, 2, \ldots, n$, then L_2-norm of $D(s; X, A)$ is called L_2-discrepancy of X, see Niederreiter (1978) and Woźniakowski (1991). The following lemma holds.

LEMMA 3. *The average case error of the linear algorithm $\varphi(N(f)) = \sum_{i=1}^{n} a_i f(x^i)$ is given by*

$$e^{\mathrm{avg}}(\varphi, N, C_0^{\bar{r}}) = \|D(\cdot; X', A)\|_2,$$

where $X' = \{\bar{1} - x^1, \ldots, \bar{1} - x^n\}$.

Proof. From (2.3) and (3.1), we obtain

$$[e^{\text{avg}}(\varphi, N, C_0^{\bar{r}})]^2 = \int_D \int_D \left(\int_D \prod_{j=1}^{d} \frac{(t_j - s_j)_+^{r_j}}{r_j!} \frac{(x_j - s_j)_+^{r_j}}{r_j!} \, ds \right) dt dx$$

$$- 2\sum_{i=1}^{n} a_i \int_D \left(\int_D \prod_{j=1}^{d} \frac{(t_j - s_j)_+^{r_j}}{r_j!} \frac{(x_j^i - s_j)_+^{r_j}}{r_j!} \, ds \right) dt$$

$$+ \sum_{i,k=1}^{n} a_i a_k \int_D \prod_{j=1}^{d} \frac{(x_j^i - s_j)_+^{r_j}}{r_j!} \frac{(x_j^k - s_j)_+^{r_j}}{r_j!} \, ds.$$

After applying Fubini's theorem and simplifying the expressions, we get

$$\begin{aligned}[e^{\text{avg}}(\varphi, N, C_0^{\bar{r}})]^2 &= \int_D \prod_{j=1}^{d} \frac{(1 - s_j)^{2(r_j+1)}}{[(r_j + 1!]^2} \, ds \\ &\quad - 2\sum_{i=1}^{n} a_i \int_D \prod_{j=1}^{d} \frac{(1 - s_j)^{r_j+1}}{(r_j + 1)!} \frac{(x_j^i - s_j)_+^{r_j}}{r_j!} \, ds \\ &\quad + \sum_{i,k=1}^{n} a_i a_k \int_D \prod_{j=1}^{d} \frac{(x_j^i - s_j)_+^{r_j}}{r_j!} \frac{(x_j^k - s_j)_+^{r_j}}{r_j!} \, ds \\ &= \int_D \left(\sum_{i=1}^{n} a_i \prod_{j=1}^{d} \frac{(x_j^i - s_j)_+^{r_j}}{r_j!} - \prod_{j=1}^{d} \frac{(1 - s_j)^{r_j+1}}{(r_j + 1!} \right)^2 ds\end{aligned} \tag{4.24}$$

After changing variables $s_j' = 1 - s_j$, we obtain the desired result.

Remarks. We add in passing that the equivalent way of finding optimal sample points and weights for approximating $I(f)$ in the average case setting is finding optimal points and weights to minimize $\|D(\cdot; X, A)\|_2$. For $r_i = 0$ and $a_i = 1/n$, we then minimize the L_2-discrepancy. The latter problem was solved by Roth (1954, 1980), see also Frolov (1980). This approach is used by Woźniakowski (1991).

We now present another way of computing the average case error with cost essentially less than $\Theta(n^2)$. The idea is to employ randomization. Since $e^{\text{avg}}(\varphi, N)^2$ is the integral of $D(\cdot; X', A)^2$ over D we can approximate it by applying the Monte Carlo algorithm,

$$MC(\vec{u}) = MC(u_1, \ldots, u_i) = \frac{1}{i} \sum_{j=1}^{i} D^2(u_j; X', A), \tag{4.25}$$

where u_j are independently and uniformly distributed over D. The cost of computing $D^2(u_j; X', A)$ is of order n, if the sample points X' and weights A are known. We choose $X' = \{\bar{1} - \psi(x^1), \ldots, \bar{1} - \psi(x^n)\}$ where $\{\psi(x^1), \ldots, \psi(x^n)\}$ are sample points of the algorithm $\bar{\Phi}_n$ given by (4.21) and weights $A = \{\bar{b}_1, \ldots, \bar{b}_n\}$ as in $\bar{\Phi}_n$ given by (4.21) with ψ_{2r} instead of ψ_r. Thus (4.25) can be computed at cost $O(in)$. We have

$$\begin{aligned} &\int_{D'} (e^{\text{avg}}(\bar{\Phi}_n)^2 - MC(\vec{u}))^2\, d\vec{u} \\ &\le \frac{1}{i} \int_D D_n^4(t; X', A)\, dt = \frac{1}{i} \|D(\cdot; X', A)\|_4^4. \end{aligned} \tag{4.26}$$

To estimate $\|D(\cdot; X', A)\|_4$ we use a known relation between $L_{p'}$-norms of $D(\cdot; X', A)^2$ and L_p-norms of the error of the algorithm $\bar{\Phi}_n$ in the space $W_p^{\overline{r+1}}$, where $1/p + 1/p' = 1$, see Temlyakov (1990, p. 1404), Temlyakov (1991), and TWW (1988). For $p = 4/3$ we have

$$\|D(\cdot; X', A)\|_4 = e_n^{\text{wor}}(\bar{\Phi}_n, B(W_{4/3}^{\overline{r+1}})). \tag{4.27}$$

Temlyakov (1987) proved that the error of the algorithm $\Phi_n(f)$ given by (4.5) in the L_p-sense ($1 \le p \le \infty$) for functions from $B(\tilde{W}_p^{\overline{r+1}})$ is given by

$$e_n^{\text{wor}}(\Phi_n, B(\tilde{W}_p^{\overline{r+1}})) = O\left(\frac{(\log n)^{(k-1)(r+2)}}{n^{r+1}}\right), \tag{4.28}$$

where the constant in the O notation may also depend on p. In particular, we are interested in the case $p = 4/3$. It follows from Temlyakov (1990), see also Bykovskij (1985), that

$$e_n^{\text{wor}}(\bar{\Phi}_n, B(W_{4/3}^{\overline{r+1}})) \le K e_n^{\text{wor}}(\Phi_n, B(\tilde{W}_{4/3}^{\overline{r+1}})). \tag{4.29}$$

Combining (4.27), (4.28), and (4.29) we get

$$\|D(\cdot; X', A)\|_4 = O\left(\frac{(\log n)^{(k-1)(r+2)}}{n^{r+1}}\right), \tag{4.30}$$

Using Corollary 2 and (4.27) we can rewrite (4.26) in the following form

$$\begin{aligned} &\int_{D'} \left(1 - \frac{MC(u_1, \ldots, u_i)}{e^{\text{avg}}(\varphi, N)^2}\right)^2 du_1 \ldots du_i \\ &= O\left(\frac{1}{i} (\log n)^{2(2r+3)(k-1)}\right). \end{aligned} \tag{4.31}$$

Hence, to get $MC(u_1, \ldots, u_i)$ and $e^{avg}(\varphi, N)^2$ of the same order with high probability, we should take

$$i = O((\log n)^{2(2r+3)(k-1)}).$$

Then the total cost of computing $MC(u_1, \ldots, u_d)$ is $O(n(\log n)^{2(2r+3)(k-1)})$. Then we check if $MC(u_1, \ldots, u_i)$ is less than ε. If so, we are done. If not, as explained before, we can double n and repeat the computations. Finally, we will find the proper n with the cost $O(n(\log n)^{2(2r+3)(k-1)})$.

Clearly, for sufficiently large n, or equivalently for sufficiently small ε, this cost is smaller than $\Theta(n^2)$.

ACKNOWLEDGMENTS

I express my thanks to H. Woźniakowski and J. Traub for valuable comments, discussions, and help during the preparation of this paper. I also appreciate remarks from B. Bojanov, A. Papageorgiou, V. Temlyakov, T. Boult, and A. Werschulz.

REFERENCES

ADLER, S. J. (1981), "The Geometry of Random Fields," Wiley Series in Prob. and Math. Stat., Wiley, New York.

ARONSZAJN, N. (1950), Theory of reproducing kernels, *Trans. AMS* **68**, 337–404.

BYKOVSKIJ, V. A. (1985), On exact order of optimal quadrature formulas for spaces of functions with bounded mixed derivatives, preprint of Dalnevostochnoi Center of Academy of Sciences, Vladivostok USSR. [in Russian]

FROLOV, K. K. (1976), An upper bound of the error of quadratures of some classes of functions (in Russian), *Dokl. Akad. Nauk SSSR* **231**, 818–821.

FROLOV, K. (1980), An upper bound of discrepancy in L_p, $2 \le p < \infty$ metric, *Dokl. Akad. Nauk SSSR* **252**, 805–807.

GRADSHTEYN, I. S., AND RYZHIK, I. (1980), "Table of Integrals, Series and Products, Academic Press, New York.

KENG, HUA LOO, AND YUAN, WANG (1981), "Applications of Number Theory to Numerical Analysis," Springer-Verlag, Berlin.

KUO, H.-H. (1975), "Gaussian Measures in Banach Spaces," Lectures Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin.

LEE, D., AND WASILKOWSKI, G. W. (1986), Approximation of linear functionals on a Banach space with a Gaussian measure, *J. Complexity* **2**, 12–43.

MICCHELLI, CH. AND WAHBA, G. (1981), Design problems for optimal surface interpolation, *in* "Approximation Theory and Applications" (Z. Zielger, Ed.), pp. 329–347, Academic Press, New York.

NIEDERREITER, H. (1978), Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84**, 957–1041.

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

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

ROTH, K. F. (1954), On irregularities of distribution, *Mathematika* **1,** 73–79.

ROTH, K. F. (1980), On irregularities of distribution, IV, *Acta Arith.* **37,** 67–75.

SACKS, J., AND YLVISAKER, D. (1970), Statistical design and integral approximation, *in* "Seminar of the Canadian Mathematical Congress on Time Series and Stochastic Processes" (R. Pyke, Ed.).

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

TEMLYAKOV, V. N. (1989), Approximation of functions with bounded mixed derivatives, *Proc. Steklov Inst. Math.* **1.**

TEMLYAKOV, V. N. (1990), On an approach of obtaining lower bounds for the error of quadratures, *Math. USSR Sbornik* **181,** 1403–1413. [in Russian]

TEMLYAKOV, V. N. (1991), Approximation of periodic functions, submitted for publication.

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

WAHBA, G. (1990), "Spline Models for Observational Data," CBMS, Vol. 59, SIAM, Philadelphia.

WOŹNIAKOWSKI, H. (1991), Average case complexity of multivariate integration, *Bull. AMS (New Series)* **24,** 185–194.

WOŹNIAKOWSKI, H. (1992), Average case complexity of linear multivariate problems, Part I: Theory, Part II: Applications, *J. Complexity* **8,** 337–372 and 373–392.