JOURNAL OF APPROXIMATION THEORY 77, 212–227 (1994)

Integration and Approximation of Multivariate Functions: Average Case Complexity with Isotropic Wiener Measure*

G. W. WASILKOWSKI†

Department of Computer Science, University of Kentucky, 40506 Lexington, Kentucky

Communicated by Allan Pinkus

Received October 5, 1992; accepted in revised form January 29, 1993

We study the average case complexity of multivariate integration and L_2 function approximation for the class $F = C([0, 1]^d)$ of continuous functions of d variables. The class F is endowed with the isotropic Wiener measure (Brownian motion in Lévy's sense). For the integration problem, the average case complexity of solving the problem to within ε is proportional to $\varepsilon^{-2/(1+1/d)}$. This is a negative result since for a large number d of variables, the average case complexity is close to ε^{-2}; the latter is also achieved by the classical Monte Carlo method in the randomized worst case setting. Furthermore, $\Theta(\varepsilon^{-2})$ is the highest possible average case complexity among *all* probability measures with finite expectation of $\|f\|_{L_2}^2$. Thus, for large d, the average case complexity of the integration problem with isotropic Wiener measure behaves as the worst possible average complexity. For the function approximation problem, the complexity is even higher since it is proportional to ε^{-2d}.

These two negative results are in a sharp contrast to (H. Woźniakowski, *Bull. Amer. Math. Soc.* **24**, No. 1 (1991), 185–194; *Bull. Amer. Math. Soc.*, to appear), where, for F endowed with the Wiener sheet measure, small average case complexities have been proven. Indeed, they are of order $\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2}$ and $\varepsilon^{-2}(\log \varepsilon^{-1})^{2(d-1)}$ for the integration and function approximation problems, respectively.

1. INTRODUCTION

We study the integration and function approximation problems for multivariate functions f. For the integration problem, we want to approximate the integral of f to within a specified error ε, and for the function approximation problem, we want to recover f with the L_2 error not

* An announcement of the paper appeared in *Bull. Amer. Math. Soc.* **28** (1993), 308–314.

† Partially supported by the National Science Foundation under Grant CCR-91-14042.

0021-9045/94 $6.00

exceeding ε. To solve both problems, we would like to use as small a number of function values as possible.

Both problems have been extensively studied in the literature (see, e.g., [9, 16] for hundreds of references). However, they are mainly addressed in the worst case setting. In the worst case setting, the cost and the error of an algorithm are defined by the worst performance with respect to the given class F of functions f. Not surprisingly, for a number of classes F, the integration and function approximation problems are intractable (prohibitively expensive) or even unsolvable. For instance, if F consists of continuous functions that are bounded by 1, any algorithm that uses a finite number of function values cannot approximate the integral of f nor can recover f with the worst case error less than 1. Hence, both problems are unsolvable for $\varepsilon < 1$. Assuming that functions f have bounded rth derivative in the sup–norm, the number of function values required for the worst case error not to exceed ε is of order $\varepsilon^{-d/r}$. Hence, for fixed r, it is exponential in d.

Due to intractability in the worst case setting, the average case setting is of interest. In the average case setting, the class F is equipped with a probability measure μ. The error and the cost of an algorithm are measured by the expectations with respect to μ. Then, the average case complexity (with respect to μ) is defined as the minimal expected cost needed to compute an approximation with the expected error not greater than ε.

The majority of the average case results obtained so far (see, e.g., [3–6, 9–13, 15–18, 21]) deal with scalar functions ($d = 1$). These results indicate that for a "reasonable" choice of measure μ, the integration and function approximation problems are significantly easier on the average than in the worst case setting. Thus, one could hope that the intractability (or even noncomputability) of multivariate problems in the worst case setting can be removed by switching from the worst case to the average case setting.

This hope has recently been supported by Woźniakowski, see [23, 24] who analyzes integration and function approximation for the class $F = C([0, 1]^d)$ endowed with the Wiener sheet measure μ. He proves that the average case complexities of both problems are only weakly dependent on the number of variables. Indeed, the average case complexity of computing an ε-approximation is $\Theta(\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2})$ for the integration problem, and $\Theta(\varepsilon^{-2}(\log \varepsilon^{-1})^{2(d-1)})$ for the function approximation problem.

In this paper, we study the average case complexity of the integration and function approximation problems. However, instead of the Wiener sheet measure, we endow the class $F = C([0, 1]^d)$ with the isotropic Wiener measure (or Brownian motion in Lévy's sense). We prove that the average case complexity equals $\Theta(\varepsilon^{-2/(1+1/d)})$ for the integration problem, and $\Theta(\varepsilon^{-2d})$ for the function approximation problem. Unlike for the Wiener

sheet measure, the average case complexity of the function approximation problems depends strongly on d. In particular, for large d, this problem is intractable since its complexity $\Theta(\varepsilon^{-2d})$ is exponential in d, and is huge even for a modest error demand ε. For large d, the average case complexity of the integration problem is essentially proportional to ε^{-2} which is the highest possible average case complexity of the integration problem. Indeed, for any probability measure with finite expected value of $\|f\|_{L_2}^2$, the average case complexity is bounded from above by $O(\varepsilon^{-2})$. Hence, this is again a negative result.

Thus, the average case complexities of integration and function approximation problems are very different depending on whether μ is the Wiener sheet or isotropic Wiener measure. It is interesting to note that both measures are identical when $d=1$. They are different for $d>1$; results of [23, 24] and our results indicate how drastically different they are.

In this paper we allow nondeterministic (randomized) methods, even though typically only deterministic methods are considered in the average case setting. Formally, this could be viewed as strengthening our negative results. However, our main reason for allowing nondeterminism is to unify the presentation and to simplify some of the proofs; as will be clear later, nondeterminism is no more powerful than determinism for problems studied in this paper. In particular, the complexity bounds provided in Theorem 1 remain true when only deterministic methods are allowed.

The paper is organized as follows. Section 2 provides basic definitions and the statement of the main results. In addition to results already mentioned, it contains a result relating the average case complexities of the integration and function approximation problems for general probability measures. Section 3 provides the proof of the main result for the integration problem. The proof concerning the function approximation problem is given in Section 4. In addition to these proofs, we show that Haber's [2] modified Monte Carlo quadrature and a piecewise constant function approximation provide almost optimal algorithms.

2. Basic Concepts and Main Results

In this paper, we consider the following integration and function approximation problems for multivariate functions. Let $F=C(D)$ be the space of continuous functions $f: D \to \mathbb{R}$ where D is a bounded subset of $\mathbb{R}^d$. For simplicity, we take $D=[0,1]^d$ as a unit cube. For every $f \in F$ we wish to approximate $S(f)$, where $S: F \to G$ with

$$S(f)=\mathrm{Int}(f)=\int_D f(x)\,dx \qquad \text{and} \qquad G=\mathbb{R}$$

for the integration problem, and

$$S(f) = \mathrm{App}(f) = f \qquad \text{and} \qquad G = L_2(D)$$

for the approximation problem.

We assume that the functions f are unknown; instead, we can compute information $N(f)$ that consists of a finite number of values of f taken at some points from D. For a precise definition of N, see, e.g., [16]. Here we only stress that

$$N(f) = [f(x_1), ..., f(x_n)],$$

where the points x_i and the number n of them (called the *cardinality* of N) can be selected adaptively and/or randomly. That is, for adaptive N, x_i's depend on previously computed values $f(x_1), ..., f(x_{i-1})$, and the cardinality $n = n(f)$ varies with f based on computed values. For randomized N, the points x_i and the cardinality $n(f)$ may also depend on an outcome of a random process t. (That is, x_i is selected randomly with an arbitrary distribution that may depend on previously computed values of f; the distribution of $n(f)$ may also depend on observed values.) In such a case we sometimes write $N(f) = N_t(f)$.

An approximation $U(f)$ to $S(f)$ is computed based on $N(f)$. That is

$$U(f) = \phi(N(f)), \qquad \text{where} \quad \phi: N(F) \to G$$

is an arbitrary mapping; ϕ is called an *algorithm* that uses N. The algorithm ϕ can also be random; in such a case, we sometimes write $\phi = \phi_t$.

In the average case setting, we assume that the space F is endowed with a (Borel) probability measure μ. Then the *average error* and the *average cost*[1] of ϕ are defined respectively by

$$e^{\mathrm{avg}}(\phi, N, S, \mu) := \sqrt{E_\mu E_t(\|S(f) - \phi_t(N_t(f))\|_G^2}$$

and

$$\mathrm{cost}^{\mathrm{avg}}(\phi, N, S, \mu) := E_\mu E_t(n(f)).$$

(By E_μ and E_t we denote the expectations w.r.t. μ and t, respectively.) Of course, for *deterministic* N *and* ϕ,

$$e^{\mathrm{avg}}(\phi, N, S, \mu) = \sqrt{\int_F \|S(f) - \phi(N(f))\|_G^2 \, \mu(df)}$$

[1] We measure the cost by the expected number of function values neglecting the combinatory cost of N and of ϕ. With the exception of Theorem 2, this is without loss of generality since, as explained in a number of references (see, e.g., [16]), for Gaussian measures the same results hold for a more general definition of the average cost, provided that a single arithmetic operation is no more expensive than a function evaluation.

and

$$\text{cost}^{\text{avg}}(\phi, N, S, \mu) = \int_F n(f)\, \mu(df).$$

The *average case complexity* is the minimal average cost for solving the problem to within a preassigned error accuracy ε. That is,

$$\text{comp}^{\text{avg}}(\varepsilon, S, \mu) := \inf\{\text{cost}^{\text{avg}}(\phi, N, S, \mu) : e^{\text{avg}}(\phi, N, S, \mu) \leqslant \varepsilon\}.$$

(We stress that the infimum above is taken with respect to all randomized ϕ and N.)

In this paper, we analyze the average case complexity of the integration and function approximation problems ($S = \text{Int}$ and $S = \text{App}$) assuming that the probability μ is the *isotropic* Wiener measure. This measure is also referred to as the Brownian motion in Lévy's sense. For more detailed discussion and properties of μ, see, e.g., [1, 7, 8]. Here we only recall that μ is a zero-mean Gaussian measure with the correlation function

$$K(x, y) = \frac{\|x\| + \|y\| - \|x - y\|}{2} \qquad \forall x, y \in \mathbb{R}^d;$$

$\|x\|$ is the Euclidean norm in $\mathbb{R}^d$.

Remark 1. Although F consists of functions f that are defined only on $D \subset \mathbb{R}^d$, sometimes it is convenient to consider f as a function over the whole space $\mathbb{R}^d$. Formally, this corresponds to extending f (as a stochastic process) to the zero mean Gaussian process $\tilde{f}$ with the correlation $K(x, y)$ given above. The corresponding integration and function approximation problems are then defined by $\widetilde{\text{Int}}\,(\tilde{f}) = \text{Int}(\tilde{f}|_D)$ and $\widetilde{\text{App}}\,(\tilde{f}) = \text{App}(\tilde{f}|_D)$, respectively. In the proof of Lemma 1, when using this correspondence between f and $\tilde{f}$, we drop the $\tilde{}$-sign.

We are ready to state the main results.

THEOREM 1. *For the integration and function approximation problems,*

$$\text{comp}^{\text{avg}}(\varepsilon, \text{Int}, \mu) = \Theta(\varepsilon^{-2/(1+1/d)})$$

and

$$\text{comp}^{\text{avg}}(\varepsilon, \text{App}, \mu) = \Theta(\varepsilon^{-2d}).$$

For $d = 1$, μ equals the classical Wiener measure. Hence, for scalar functions, this theorem follows from known results; see [12, 13, 15, 19]. Therefore, in the rest of the paper, we shall consider $d > 1$.

Since the proof of the theorem provides additional results of independent interest (e.g., nearly optimal information and algorithms, and a relation between the complexities of the integration and function approximation problems for general measures) we present it in the next two sections: in Section 3 for $S = \text{Int}$ and in Section 4 for $S = \text{App}$.

As a matter of fact, a lower bound on complexity for $S = \text{App}$ is derived via a general result which, in turn, implies the following theorem.[2]

THEOREM 2. *Let ν be an arbitrary measure on F. If*

$$\text{comp}^{\text{avg}}(\varepsilon, \text{Int}, \nu) = \Omega(\varepsilon^{-p})$$

for some $p \leqslant 2$ (obviously, $p \leqslant 2$ whenever $\|f\|^2_{L_2(D)}$ has a finite (ν-)expectation), then

$$\text{comp}^{\text{avg}}(\varepsilon, \text{App}, \nu) = \Omega(\varepsilon^{-2p/(2-p)}).$$

3. INTEGRATION PROBLEM

In this section we prove Theorem 1 for $S = \text{Int}$. We also show that Haber's (see [2]) algorithm is nearly optimal.

The proof is based on the nth minimal average radius which is the minimal average error among all algorithms that use n function values.

More precisely, let Λ_n be the class of all deterministic and nonadaptive information operators N of cardinality n. That is, $N \in \Lambda_n$ means that $N(f) = [f(x_1), ..., f(x_n)]$ for some points x_i that are chosen a priori. By the *average radius* of N, we mean the minimal average error among all algorithms ϕ that use N,

$$r^{\text{avg}}(N, \text{Int}, \mu) = \inf_{\phi} e^{\text{avg}}(\phi, N, \text{Int}, \mu).$$

Then, the nth *minimal average radius* is defined by

$$r^{\text{avg}}(n, \text{Int}, \mu) = \inf_{N \in \Lambda_n} r^{\text{avg}}(N, \text{Int}, \mu). \tag{1}$$

In the next two subsections we prove that

$$r^{\text{avg}}(n, \text{Int}, \mu) = \Theta(n^{-1/2 - 1/(2d)}). \tag{2}$$

[2] This theorem is the only place in the paper where neglecting the combinatory cost in the definition of the average cost of algorithms is important; see footnote 1 and the proof of the theorem in Section 4.1.

This is sufficient to prove Theorem 1 for $S = \mathrm{Int}$. Indeed, from [19, 20] we know that for Gaussian μ and linear S, the average case complexity is fully determined by the behavior of nth minimal average radii. In particular, if they are semi-convex in n (as in (2)), then

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, S, \mu) = \Theta(n^*(\varepsilon)), \qquad \text{where} \quad n^*(\varepsilon) = \min\{n : r^{\mathrm{avg}}(n, S, \mu) \leqslant \varepsilon\}.$$

Furthermore, information $N^*_{n^*(\varepsilon)} \in \Lambda_{n^*(\varepsilon)}$ and an algorithm ϕ^* that satisfy

$$e^{\mathrm{avg}}(\phi^*, N^*_{n^*(\varepsilon)}, S, \mu) = r^{\mathrm{avg}}(n^*(\varepsilon), S, \mu)$$

are nearly optimal. In particular, randomization does not help and

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon) \leqslant \mathrm{comp}^{\mathrm{avg\text{-}det}}(\varepsilon) \leqslant \mathrm{comp}^{\mathrm{avg}}(\varepsilon) + 1/2,$$

where $\mathrm{comp}^{\mathrm{avg\text{-}det}}(\varepsilon)$ is the minimal average cost among all *deterministic algorithms* using *deterministic information* with the average error less than or equal to ε.

3.1. *Lower Bound*

We show that the rhs of (2) is a lower bound on $r^{\mathrm{avg}}(n, \mathrm{Int}, \mu)$, i.e.,

LEMMA 1.

$$r^{\mathrm{avg}}(n, \mathrm{Int}, \mu) = \Omega(n^{-1/2 - 1/(2d)}).$$

Proof. Let H be the reproducing kernel Hilbert space spanned by the family $\{K(x, \cdot) : x \in \mathbb{R}^d\}$ with the corresponding inner product $\langle f, g \rangle_\mu$ determined by

$$\langle K(x, \cdot), K(y, \cdot) \rangle_\mu = K(x, y), \qquad \forall x, y \in \mathbb{R}^d.$$

The space H is critical for our analysis since, as it has been succesfully exploited for various integration problems (see, e.g., [3, 4, 11, 12, 16–18, 24]), the average radius of nonadaptive information N equals the worst case radius $r^{\mathrm{w}}(N, \mathrm{Int}, BH)$ of N defined by

$$r^{\mathrm{w}}(N, \mathrm{Int}, BH) = \inf_{\phi} \sup_{f \in BH} |\mathrm{Int}(f) - \phi(N(f))|.$$

The set BH is the unit ball in the space H. Hence, to prove the lemma, it is enough to bound from below the worst case radius of any nonadaptive information $N(f) = [f(x_1), \ldots, f(x_n)]$.

For this end, we need some characterization of the space H. Such a characterization has been provided by Molchan [8] for odd d, and later by Ciesielski [1] for arbitrary d. From Lemma 4.1 and Theorem 4.1 in [1] we

know that any real valued function f from $C_0^\infty(\mathbb{R}^d)$ that vanishes at zero belongs to H. Furthermore, for any two such functions f, g,

$$\langle f, g\rangle_\mu = a_d \langle (-\Delta)^{(d+1)/4} f, (-\Delta)^{(d+1)/4} g\rangle_{L_2(\mathbb{R}^d)},$$

where a_d is a (known) constant and Δ is the Laplace operator. For $d+1$ not divisible by 4, $(-\Delta)^{(d+1)/4}$ is understood in the generalized form (see, e.g., [14]).

Without loss of generality we assume that $(2n)^{1/d}$ is an integer. For $m = 2n$, consider m equal-size cubes that partition D. The centers of these cubes are denoted by v_i. Let

$$h = m^{-1/d} \qquad \text{and} \qquad g(y) = \psi(\|y\|^2)$$

for a nonnegative function $\psi \in C^\infty(\mathbb{R})$ with $\psi(t) = 0$ iff $t \geqslant 1/16$. For $i = 1, \ldots, m$, define the following functions:

$$f_i(x) = g((x - v_i)/h).$$

Since $\|v_i\| \geqslant h/2$, $f_i(0) = 0$. Therefore $f_i \in H$. Furthermore, the functions f_i have disjoint supports, $B_i := \operatorname{supp}(f_i) = \{x : \|x - v_i\| < h/4\}$. As a matter of fact, $\operatorname{dist}(B_i, B_j) \geqslant h/2$ for $i \neq j$. Since the information N consist of n function values at points $x_1, \ldots, x_n$, these points are not contained in at least $m - n = n$ cubes, indexed by $1, \ldots, n$. Hence, the function

$$f = \sum_{i=1}^{n} f_i$$

has zero information, $N(f) = 0$. Furthermore, $\operatorname{Int}(f) = n \operatorname{Int}(f_1) = nh^d \int_{\|y\| \leqslant 1} g(y)\, dy = C_1$ for a positive constant C_1. Therefore, the average and worst case radii of N are bounded from below by

$$r^{\mathrm{avg}}(N, \operatorname{Int}, \mu) = r^{\mathrm{w}}(N, \operatorname{Int}, BH) = \sup_{g \in BH,\, N(g) = 0} |\operatorname{Int}(g)| \geqslant \frac{C_1}{\|g\|_\mu}.$$

Thus, to complete the proof, we only need to show that

$$\|f\|_\mu \leqslant C_2 h^{-d/2 - 1/2} = O(n^{1/2 + 1/(2d)}) \tag{3}$$

for some constant C_2. In what follows, we write C to denote positive (in general different) constants which are independent of n.

We begin with the case of $d + 1 = 4k$ for an integer k. Since $(-\Delta)^{(d+1)/4} = (-\Delta)^k$ preserves disjoint supports of f_i's, f_i's are orthogonal in H. Thus, $\|f\|_\mu^2 = a_d \sum_{i=1}^n \|(-\Delta)^k f_i\|_{L_2(\mathbb{R}^d)}^2 = a_d n \|(-\Delta)^k f_1\|_{L_2(\mathbb{R}^d)}^2 = n \|f_1\|_\mu^2$. Furthermore, $\Delta^k f_1(x) = h^{-2k} \Delta^k g(y)|_{y = (x - v_1)/h}$, and a simple change of variables yields $\|f_1\|_\mu^2 = Ch^{-1}$. This proves (3) for $d + 1 = 4k$.

For $d-1=4k$, we have an analogous situation since $\langle f_i, f_j\rangle_\mu = a_d\langle(-\Delta)^{2k+1} f_i, f_j\rangle_{L_2(\mathbb{R}^d)}$ and the orthogonality of f_i's is preserved.

Consider therefore the final case of $d=2k$. Then

$$\|f\|_\mu^2 = \sum_{i,j=1}^{n} b_{i,j} \qquad \text{with} \quad b_{i,j} = \langle f_i, f_j\rangle_\mu.$$

From [1], $b_{i,j} = C\int_{\mathbb{R}^d} \|x\|^{d+1} \hat{f}_i(x)\overline{\hat{f}_j(x)}\,dx$, and well known properties of the Fourier transform imply

$$b_{i,i} = Ch^{-1}, \qquad \forall i, \tag{4}$$

$$b_{i,j} = C\int_{\mathbb{R}^d} \|x\|^{-1} \hat{g}_i(x)\overline{\hat{f}_j(x)}\,dx \qquad \text{with} \quad g_i(x) = (-\Delta)^{k+1} f_i(x).$$

Using Lemma 1(b) on p. 117 in [14], we have for $i \neq j$

$$b_{i,j} = C\int_{\mathbb{R}^d} \overline{f_j(x)} \int_{\mathbb{R}^d} g_i(x-y)\,\|y\|^{-d+1}\,dy\,dx$$

$$= C\int_{B_j} f_j(x) \int_{B_i} g_i(y)\,\|y-x\|^{-d+1}\,dy\,dx.$$

Obviously, the inner integral equals $\int_{B_i}((-\Delta)^k f_i(y))(-\Delta\,\|y-x\|^{-d+1})\,dy$. As for the case of $d+1=4k$, the absolute value of the first term (involving f_i) is bounded from above by Ch^{-2k}. The second terms equals $C\,\|y-x\|^{-d-1}$; recall that $\|x-y\| \geq h/2$. Using the fact that f_j is also bounded, we conclude that

$$|b_{i,j}| \leq Ch^{-2k}\int_{B_i\times B_j} \|x-y\|^{-d-1}\,d(x,y),$$

where $d(x,y)$ denotes $dx\,dy$. This and (4) imply

$$\|f\|_\mu^2 \leq \sum_{i=1}^{n} b_{i,i} + \sum_{i\neq j} |b_{i,j}| \leq Ch^{-1}n + Ch^{-2k}\int_{D_h} \|x-y\|^{-d-1}\,d(x,y),$$

where $D_h = \{(x,y)\in D^2 : \|x-y\| \geq h/2\}$. Since the integral of $\|x-y\|^{-d-1}$ over D_h is bounded by Ch^{-1}, the rhs of the above inequality equals $O(h^{-2k-1})$, as claimed in (3). This completes the proof of Lemma 1. ∎

3.2. *Upper Bound*

To prove that the rhs of (2) is also an upper bound on the nth minimal average radius, we exhibit specific information N_n^{Int} of cardinality n and an algorithm ϕ_n^{Int} whose average error equals $\Theta(n^{-1/2+1/(2d)})$. Although the

information N_n^{Int} is randomized, it has a fixed cardinality. Hence, the mean value theorem implies that there exists nonadaptive deterministic information N_n^* of cardinality n with with $r^{\text{avg}}(N^*, \text{Int}, \mu) = \Theta(n^{-1/2-1(2d)})$. This suffices to complete the proof of Theorem 1. Furthermore, from the discussion at the beginning of Section 3, we conclude that the exhibited information and algorithm are almost optimal.

Let $n = p^d$ for an integer p. Partition D into n equal-size cubes U_i, $U_i = x_i + [-1/(2p), +1/(2p)]^d$. Note that the Lebesgue measure of each U_i equals $1/n$. Consider randomized information

$$N_n^{\text{Int}}(f) = [f(t_1), ..., f(t_n)], \tag{5}$$

where t_i's are independent and uniformly distributed in U_i's and an algorithm ϕ_n^{Int}, due to Haber [2],

$$\phi_n^{\text{Int}}(N_n^{\text{Int}}(f)) = \frac{1}{n}\sum_{j=1}^{n} f(t_j). \tag{6}$$

LEMMA 2. *For every n, the average error of ϕ_n^{Int} that uses N_n^{Int} equals*

$$e^{\text{avg}}(\phi_n^{\text{Int}}, N_n^{\text{Int}}, \text{Int}, \mu) = \frac{\sqrt{\int_{D\times D} \|x-y\|/2 \, d(x,y)}}{n^{1/2+1/(2d)}}. \tag{7}$$

Hence ϕ_n^{Int} and N_n^{Int} are almost optimal. Indeed, for

$$n^{\text{Int}}(\varepsilon) = \left\lceil \left(\varepsilon^{-2} \int_{D\times D} \|x-y\|/2 \, d(x,y) \right)^{1/(d+1)} \right\rceil^d,$$

$\phi_{n^{\text{Int}}(\varepsilon)}^{\text{Int}}$ has the average error not exceeding ε and its cost is proportional to $\text{comp}^{\text{avg}}(\varepsilon, \text{Int}, \mu)$.

Proof. We need only to prove (7). For brevity, let ϕ_t denote $\phi_n^{\text{Int}}(N_n^{\text{Int}}(f))$. For every f, the expectation (w.r.t. t) of the square of the error of ϕ_t equals

$$E_t((S(f) - \phi_t)^2) = \sum_{j=1}^{n} \left(n^{-1} \int_{U_j} f^2(x)\, dx - \left(\int_{U_j} f(x)\, dx \right)^2 \right).$$

Hence, taking next the expectation w.r.t. f, we get

$$\begin{aligned}
&(e^{\text{avg}}(\phi_n^{\text{Int}}, N_n^{\text{Int}}, \text{Int}, \mu))^2 \\
&\quad = \sum_{j=1}^{n} \left(n^{-1} \int_{U_j} \|x\|\, dx - \int_{U_j\times U_j} \frac{\|x\| + \|y\| - \|x-y\|}{2} d(x,y) \right) \\
&\quad = \sum_{j=1}^{n} \int_{U_j\times U_j} \frac{\|x-y\|}{2} d(x,y).
\end{aligned}$$

Note that $\int_{U_j \times U_j} \|x-y\|\, d(x,y) = n^{-2-1/d} \int_{D\times D} \|x-y\|\, d(x,y)$. This completes the proof. ∎

Remark 2. In the worst case setting with $F = C[0,1]^d$, Haber's modified Monte Carlo algorithm ϕ_n^{Int} and the classical Monte Carlo algorithm $n^{-1}\sum_{i=1}^n f(t_i)$ (with t_i's uniformly distributed in D) have (modulo constants) the same errors that are proportional to $1/\sqrt{n}$. It can be verified that the average error of the classical Monte Carlo algorithm remains proportional to $1/\sqrt{n}$; it equals $\sqrt{n^{-1}\int_{D\times D} \|x-y\|/2\, d(x,y)}$. Hence, it is precisely $n^{1/(2d)}$ times larger than the average error of the modified Monte Carlo algorithm.

4. Function Approximation Problem

As for the integration problem, the proof of Theorem 1 for $S = \mathrm{App}$ is based on showing that

$$r^{\mathrm{avg}}(n, \mathrm{App}, \mu) = \Theta(n^{-1/(2d)}) \tag{8}$$

with the nth *minimal average radius* $r^{\mathrm{avg}}(n, \mathrm{App}, \mu)$ for the function approximation problem defined in a similar way as for the integration problem in (1).

In Subsection 4.1, we prove that the rhs of (8) is a lower bound on $r^{\mathrm{avg}}(n, \mathrm{App}, \mu)$. Our proof is an immediate consequence of a more general result, Theorem 3, which in particular implies Theorem 2. In Subsection 4.2, we prove that the rhs of (8) is an upper bound on $r^{\mathrm{avg}}(n, \mathrm{App}, \mu)$; actually we show that a piecewise constant function approximation provides an almost optimal algorithm.

4.1. *Lower Bound*

We have defined the nth minimal average radius by restricting the class of information to nonadaptive information only. For Gaussian measures (such as μ), such a restriction is without loss of generality since adaptive information of fixed cardinality does not help; see [22]. Furthermore, as mentioned in Section 3, the average case complexity of a problem with linear S is fully determined by the behavior of the nth minimal average radii; see [19].

However, for an arbitrary probability measure ν, adaptive information (with or without varying cardinality, deterministic or randomized) might be more powerful than nonadaptive information. Therefore, for the purpose of the following theorem, we need a more general notion of the

minimal average radius, called cth *minimal average adaptive radius.* It is defined by

$$r_{\mathrm{ad}}^{\mathrm{avg}}(c, S, \nu) = \inf\{r^{\mathrm{avg}}(N, S, \nu) : E(n(f)) \leqslant c\}$$

for a positive number c. Recall that $E(n(f)) = E_\mu E_t(n(f))$ is the expected cardinality of N.

THEOREM 3. *For an arbitrary probability measure ν on F, the minimal (ν-) average radii for the integration and function approximation problems satisfy*

$$r^{\mathrm{avg}}(n, \mathrm{App}, \nu) \geqslant \sqrt{k}\, r^{\mathrm{avg}}(n+k, \mathrm{Int}, \nu)$$
$$r_{\mathrm{ad}}^{\mathrm{avg}}(c, \mathrm{App}, \nu) \geqslant \sqrt{k}\, r_{\mathrm{ad}}^{\mathrm{avg}}(c+k, \mathrm{Int}, \nu),$$

for all n, c, and k.

Proof. Choose any k and n. Consider an nth optimal information N_n^* and the optimal error algorithm ϕ^* for the function approximation problem (without loss of generality we assume that they exist). That is,

$$r^{\mathrm{avg}}(n, \mathrm{App}, \nu) = r^{\mathrm{avg}}(N_n^*, \mathrm{App}, \nu) = e^{\mathrm{avg}}(\phi^*, N_n^*, \mathrm{App}, \nu).$$

For brevity, f^* denotes $\phi^*(N_n^*(f))$.

For the integration problem, consider the following N and ψ. The information $N(f)$ consist of n function values provided by $N_n^*(f)$ plus k function values at randomly chosen points $t_i \in D$. The points are chosen independently with uniform distribution on D. The algorithm $\psi(N(f))$ is then the integral $\mathrm{Int}(f^*)$ plus the classical Monte Carlo method $\sum_{i=1}^{k} k^{-1}(f(t_i) - f^*(t_i))$ applied to the integrand $f - f^*$. Then, the square of the average error of ψ equals

$$\begin{aligned}(e^{\mathrm{avg}}(\psi, N, \mathrm{Int}, \nu))^2 &= \frac{1}{k}\int_F [\|f-f^*\|_{L_2(D)}^2 - (\mathrm{Int}(f-f^*))^2]\, \nu(df) \\ &\leqslant \frac{1}{k}\int_F \|f-f^*\|_{L_2(D)}^2\, \nu(df) \\ &= \frac{1}{k}\,(r^{\mathrm{avg}}(N_n^*, \mathrm{App}, \nu))^2.\end{aligned}$$

Since N has cardinality $n+k$ and since randomization does not help on the average, the average error of ψ is greater than or equal to the $(n+k)$th minimal average radius for the integration problem. This completes the proof of the first inequality.

The proof of the second inequality is very similar; therefore, we only indicate the difference between the two proofs. Instead of N_n^* of fixed

cardinality, we take N_c^* that satisfies: $E(n(f)) \leq c$ and $r^{\text{avg}}(N_c^*, \text{App}, \nu) = r_{\text{ad}}^{\text{avg}}(c, \text{App}, \nu)$. The corresponding information $N(f)$ for the integration problem consists of $N_c^*(f)$ plus k values at random t_i's. Since the expected cardinality of N equals the expected cardinality of N_c^* plus k, the rest of the proof easily follows. ∎

Remark 3. This theorem can easily be extended in a number of ways. For instance, consider the function approximation problem with the error $\|f-f^*\|$ measured in a weighted $L_2(D)$-norm, i.e., $\|f-f^*\|^2 = \int_D w(x)(f(x)-f^*(x))^2\,dx$ for some weight $w \geq 0$. Then repeating the proof with ψ derived from the classical Monte Carlo method applied to the integrand $(f(x)-f^*(x))\sqrt{w(x)}$ we immediately conclude that the nth minimal radius of the weighted function approximation problem is bounded from below by $\sqrt{k}$ times the $(n+k)$th minimal radius of the following weighted integration problem $\text{Int}(f) = \int_D f(x)\sqrt{w(x)}\,dx$.

Another generalization concerns the *worst case setting with randomization.* In this setting, instead of the expectation E_ν, we take the supremum w.r.t $f \in F_0$ (F_0 is a given subset of F) in the definitions of the error and cost (the expectation w.r.t. random t remains). For more detailed definitions, see e.g., [16]. Let $r_{\text{ad}}^{\text{wor-ran}}(c, S, F_0)$ denote the corresponding cth minimal worst case adaptive radius. Replacing the integrals $\int_F \nu(df)$ by $\sup_{f \in F_0}$ in the proof of Theorem 3, we immediately get $r_{\text{ad}}^{\text{wor-ran}}(c+k, \text{Int}, F_0)\sqrt{k} \leq r_{\text{ad}}^{\text{wor-ran}}(c, \text{App}, F_0)$. Hence, again, in the worst case setting with randomization, integration is an easier problem than function approximation problem. We stress that this need not be true if the worst case *deterministic* (*without randomization*) setting is considered since for a number of classes F_0 the integration and approximation problems have asymptotically the same worst case deterministic complexities; see, e.g., [9, 16].

As an immediate conclusion from Theorem 3 with $k = n$ and Lemma 1, we have the lower bound

LEMMA 3. *For the measure* μ, $r^{\text{avg}}(n, \text{App}, \mu) = \Omega(n^{-1/(2d)})$.

We end this subsection by providing

Proof of Theorem 2. Since the average cost of an algorithm that uses information N is measured by the expectation of the cardinality number $n(f)$ of N,

$$\text{comp}^{\text{avg}}(\varepsilon, S, \nu) = \inf\{c: r_{\text{ad}}^{\text{avg}}(c, S, \nu) \leq \varepsilon\}.$$

Hence, Theorem 3 with $k \simeq \text{comp}^{\text{avg}}(\varepsilon, \text{Int}, \nu)$ completes the proof. ∎

4.2. *Upper Bound*

We exhibit N_n^{App} and ϕ_n^{App} that are almost optimal.

As in Section 3.2, consider $n = p^d$ and n equal-size cubes U_i, each centered at x_i. Define

$$N_n^{\text{App}}(f) = [f(x_1), ..., f(x_n)] \qquad \text{and} \qquad \phi_n^{\text{App}}(N_n^{\text{App}}(f)) = \sum_{i=1}^{n} g_i(\cdot) f(x_i), \tag{9}$$

with g_i being the indicator function for the set U_i.

LEMMA 4. *For every n, the average error of ϕ_n^{App} that uses N_n^{App} equals*

$$e^{\text{avg}}(\phi_n^{\text{App}}, N_n^{\text{App}}, \text{App}, \mu) = \frac{\sqrt{\int_D \|x\|/2 \, dx}}{n^{1/(2d)}}. \tag{10}$$

Hence ϕ_n^{App} and N_n^{App} are almost optimal. Indeed, for

$$n^{\text{App}}(\varepsilon) = \left\lceil \varepsilon^{-2} \int_D \|x\|/2 \, dx \right\rceil^d$$

$\phi_{n^{\text{App}}(\varepsilon)}^{\text{App}}$ *has the average error not exceeding ε and its cost is proportional to* $\text{comp}^{\text{avg}}(\varepsilon, \text{App}, \mu)$.

Proof. We need only to prove (10). For this end note that

$$\begin{aligned}
&(e^{\text{avg}}(\phi_n^{\text{App}}, N_n^{\text{App}}, \text{App}, \mu))^2 \\
&\quad = \int_D \left(K(x, x) - 2 \sum_{i=1}^{n} g_i(x) \, K(x, x_i) + \sum_{i,j=1}^{n} g_i(x) \, g_j(x) \, K(x_j, x_i) \right) dx \\
&\quad = A_1 - A_2,
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \int_D \left(K(x, x) - \sum_{i=1}^{n} g_i(x) \, K(x, x_i) \right) dx \\
&= \sum_{i=1}^{n} \int_{U_i} (K(x, x) - K(x, x_i)) \, dx \\
&= \sum_{i=1}^{n} \int_{U_i} \frac{\|x - x_i\| + \|x\| - \|x_i\|}{2} \, dx,
\end{aligned}$$

and

$$A_2 = \sum_{i=1}^{n} \int_D g_i(x) \left(K(x, x_i) - \sum_{j=1}^{n} g_j(x)\, K(x_j, x_i) \right) dx$$

$$= \sum_{i=1}^{n} \int_{U_i} K(x, x_i)\, dx - \sum_{i=1}^{n} \int_{U_i} K(x_i, x_i)\, dx$$

$$= \sum_{i=1}^{n} \int_{U_i} \frac{\|x\| - \|x - x_i\| - \|x_i\|}{2}\, dx,$$

(the second equation follows from the fact that g_i and g_j have disjoint supports when $i \neq j$). Therefore $(e^{\text{avg}}(\phi_n^{\text{App}}, N_n^{\text{App}}, \text{App}, \mu))^2 = A_1 - A_2 = \sum_{i=1}^{n} \int_{U_i} \|x - x_i\|\, dx = n^{-1/d} \int_D \|x\|/2\, dx$, which was to be proven. ∎

ACKNOWLEDGMENTS

I thank David R. Adams, Klaus Ritter, and Henryk Woźniakowski for valuable suggestions.

REFERENCES

1. Z. CIESIELSKI, On Lévy's Brownian motion with several-dimensional time, *in* "Lecture Notes in Mathematics," (A. Dold and B. Eckman, Eds.), Vol. 472, pp. 29–56, Springer-Verlag, New York/Berlin, 1975.
2. S. HABER, A modified Monte Carlo quadrature, *Math. Comp.* **20** (1966), 361–368.
3. G. S. KIMELDORF AND G. WAHBA, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41** (1970), 495–502.
4. G. S. KIMELDORF AND G. WAHBA, Spline functions and stochastic processes, *Sankhya Ser. A* **32** (1970), 173–180.
5. D. LEE, Approximation of linear operators on a Wiener space, *Rocky Mountain J. Math.* **16** (1986), 641–659.
6. D. LEE AND G. W. WASILKOWSKI, Approximation of linear functionals on a Banach space with a Gaussian measure, *J. Complexity* **2** (1986), 12–43.
7. P. LÉVY, "Processus Stochastique et Mouvement Brownien," Gauthier–Villars, Paris, 1948.
8. G. M. MOLCHAN, On some problems concerning Brownian motion in Lévy's sense, *Theory Probab. Appl.* **12** (1967), 682–690.
9. E. NOVAK, "Deterministic and Stochastic Error Bounds in Numerical Analysis," Lecture Notes in Mathematics, Vol. 1349, Springer-Verlag, New York/Berlin, 1988.
10. K. RITTER, Approximation and optimization on the Wiener space, *J. Complexity* **6** (1990), 337–364.
11. J. SACKS AND D. YLVISAKER, Design for regression problems with correlated errors III, *Ann. Math. Statist.* **41** (1970), 2057–2074.
12. J. SACKS AND D. YLVISAKER, Statistical design and integral approximation, *in* "Proc. 12th Bienn. Semin. Can. Math. Congr., 1970," pp. 115–136.

13. P. SPECKMAN, "L_p Approximation of Autoregressive Gaussian Process," Dept. Stat. Report, Univ. of Oregon, 1979.
14. E. M. STEIN, "Singular Integrals and Differentiability Properties of Functions," Princeton Univ. Press, Princeton, 1970.
15. A. V. SULDIN, Wiener measure and its applications to approximation methods, I, *Izv. Vyssh, Uchebn. Zaved. Mat.* **13** (1959) 145–158; II, *Izv. Vyssh. Uchebn. Zaved. Mat.* **18** (1960), 165–179. [In Russian]
16. J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, "Information-Based Complexity," New York, Academic Press, 1988.
17. G. WAHBA, On the regression design problem of Sacks and Ylvisaker, *Ann. Math. Statist.* **42** (1971), 1035–1043.
18. G. WAHBA, "Spline Models for Observational Data," SIAM–NSF Regional Conference Series in Appl. Math., Vol. 59, 1990.
19. G. W. WASILKOWSKI, Information of varying cardinality, *J. Complexity* **2** (1986), 204–228.
20. G. W. WASILKOWSKI, Randomization for continuous problems, *J. Complexity* **5** (1989), 195–218.
21. G. W. WASILKOWSKI AND F. GAO, On the power of adaptive information for functions with singularities, *Math. Comp.* **58**, No. 197 (1992), 285–304.
22. G. W. WASILKOWSKI AND H. WOŹNIAKOWSKI, Can adaption help on the average?, *Numer. Math.* **44** (1984), 169–190.
23. H. WOŹNIAKOWSKI, Average case complexity of multivariate integration, *Bull. Amer. Math. Soc.* **24**, No. 1 (1991), 185–194.
24. H. WOŹNIAKOWSKI, Average case complexity of linear multivariate problems, *Bull. Amer. Math. Soc.*, to appear.

Printed in Belgium
Uitgever: Academic Press, Inc.
Verantwoordelijke uitgever voor België:
Hubert Van Maele
Altenastraat 20, B-8310 Sint-Kruis