RESEARCH ANNOUNCEMENTS

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 28, Number 2, April 1993

INTEGRATION AND APPROXIMATION OF MULTIVARIATE FUNCTIONS: AVERAGE CASE COMPLEXITY WITH ISOTROPIC WIENER MEASURE

G. W. WASILKOWSKI

ABSTRACT. 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 Levy's sense). Furthermore, for both problems, only function values are used as data.

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 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, [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, no algorithm that uses a finite number of function values can approximate the integral of f, nor can it 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

Received by the editors December 19, 1991.
1991 *Mathematics Subject Classification.* Primary 41A50, 41A55, 41A63, 65D15, 65D30.
This research was supported in part by the National Science Foundation under Grant CCR-91-14042.

0273-0979/93 $1.00 + $.25 per page

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, [3–6, 9–18, 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 Levy'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 the Wiener sheet measure, the average-case complexity of the function approximation problem 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.

The paper is organized as follows. Section 2 provides basic definitions. The main results are presented in §3. In addition to results already mentioned, §3 discusses optimality of Haber's [2] modified Monte Carlo quadrature and of a piecewise constant function approximation. It also contains a result relating the average-case complexities of the integration and function approximation problems for general probability measures. In this paper we omit all proofs because of their substantial length.

2. Basic definitions

In this paper we consider the following integration and function approximation problems for multivariate functions. Let $F = C(D)$ be the space of 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

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

$$S(f) = \mathrm{App}(f) = f \quad \text{and} \quad G = L_2(D) \qquad \text{for the approximation problem.}$$

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 [16]. Here we stress only that

$$N(f) = [f(x_1), \ldots, f(x_n)],$$

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 $f(x_1), \ldots, f(x_{i-1})$, 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,

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

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

$$e^{\mathrm{avg}}(\phi, N, S, \mu) := \sqrt{\mathrm{E}_\mu \mathrm{E}_t(\|S(f) - \phi_t(N_t(f))\|_G^2)},$$
$$\mathrm{cost}^{\mathrm{avg}}(\phi, N, S, \mu) := \mathrm{E}_\mu \mathrm{E}_t(n(f)).$$

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

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

[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 3, 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.

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

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

(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 ($S = \mathrm{Int}$ and $S = \mathrm{App}$) with the assumption that the probability μ is the *isotropic* Wiener measure. This measure is also referred to as the Brownian motion in Levy's sense. For more detailed discussion and properties of μ (see [1, 7, 8]). Here we only recall that μ is a zero-mean Gaussian measure with the correlation function

$$K(x, y) = \frac{\|x\| + \|y\| - \|x - y\|}{2} \quad \forall x, y \in \mathbb{R}^d, \qquad \|x\|^2 = \sum_{i=1}^{d} x_i^2.$$

3. Main results

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

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

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, \mathrm{App}, \mu) = \Theta\left(\varepsilon^{-2d}\right). \tag{2}$$

For $d = 1$, μ equals the classical Wiener measure. Hence, for scalar functions this theorem follows from known results (see [12, 13, 19]).

We now exhibit algorithms and information that are almost optimal. Let $n = p^d$. Partition D into n equal-sized cubes U_i, $U_i = x_i + [-1/(2p), +1/(2p)]^d$, each centered at x_i. For the integration problem, consider the following randomized information and algorithm due to Haber (see [2]):

$$N_n^{\mathrm{Int}}(f) = [f(t_1), \ldots, f(t_n)] \quad \text{and} \quad \phi_n^{\mathrm{Int}}(N_n^{\mathrm{Int}}(f)) = \frac{1}{n}\sum_{j=1}^{n} f(t_j), \tag{3}$$

where t_i's are uniformly distributed in U_i's. For the function approximation problem, consider

$$N_n^{\mathrm{App}}(f) = [f(x_1), \ldots, f(x_n)] \quad \text{and} \quad \phi_n^{\mathrm{App}}(N_n^{\mathrm{App}}(f)) = \sum_{i=1}^{n} g_i(\cdot) f(x_i), \tag{4}$$

with g_i being the indicator function for the set U_i.

Theorem 2. *For every n, the average errors of ϕ_n^{Int} and ϕ_n^{App} are respectively equal to*

$$\frac{\sqrt{\int_D \int_D \|x - y\|/2 \, dx dy}}{n^{1/2+1/(2d)}} \quad \text{and} \quad \frac{\sqrt{\int_D \|x\|/2 \, dx}}{n^{1/(2d)}}.$$

These algorithms are almost optimal. Indeed, for $n^{\mathrm{Int}}(\varepsilon)$ *and* $n^{\mathrm{App}}(\varepsilon)$ *given by*

$$n^{\mathrm{Int}}(\varepsilon) = \left\lceil \left(\varepsilon^{-2} \int_D \int_D \|x - y\|/2 \, dx dy \right)^{1/(d+1)} \right\rceil^d,$$

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

$\phi^{\mathrm{Int}}_{n^{\mathrm{Int}}(\varepsilon)}$ *and* $\phi^{\mathrm{App}}_{n^{\mathrm{Int}}(\varepsilon)}$ *have the average errors less than or equal to* ε *and their costs are proportional to* (1) *and* (2), *respectively.*

Remark 1. In the worst-case setting with $F = C[0, 1]^d$, Haber's modified Monte Carlo algorithm ϕ^{Int}_n 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 proven that the average error of the classical Monte Carlo algorithm equals $\sqrt{n^{-1} \int_D \int_D \|x - y\|/2 \, dx dy}$. Thus, it is precisely $n^{1/(2d)}$ times larger than the average error of ϕ^{Int}_n.

Remark 2. Although the information N^{Int}_n is randomized, the cardinality n is fixed. Thus, the mean value theorem implies the existence of *deterministic* $N^*_n(f) = [f(x^*_1), \ldots, f(x^*_n)]$ such that $\phi^*_n(N^*_n(f)) = n^{-1} \sum_{i=1}^n f(x^*_i)$ has the average error not exceeding the average error of ϕ^{Int}_n. (We do not know the location of the points x^*_i; we only know that $x^*_i \in U_i$, $1 \le i \le n$.) This and the fact that the algorithm and information given in (4) are deterministic imply that randomization does not help for both problems. The lack of power of randomization holds for more general problems. Indeed, randomization does not help for linear S and Gaussian μ (see [20, 22]).

The final theorem relates the average-case complexities of the integration and function approximation problems for an arbitrary (Borel) probability measure.

Theorem 3. *Let* ν *be an arbitrary probability measure on* F. *If*

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

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

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, \mathrm{App}, \nu) = \Omega\left(\varepsilon^{-p(1+p/(2-p))}\right).$$

Remark 3. This theorem can easily be extended in a number of ways. For instance, it holds when the function approximation problem is considered with the $L_2(D)$-norm replaced by $\|f - f^*\|^2 = \int_D w(x)(f(x) - f^*(x))^2 \, dx$, a weighted norm, for some weight $w \ge 0$ and the integration problem defined by $\mathrm{Int}(f) = \int_D f(x) \sqrt{w(x)} \, dx$.

It also holds in 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 [16]). Hence, again, in the worst-case setting with randomization, integration is an easier problem than is the 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, [9, 16]).

ACKNOWLEDGMENT

The author wishes to thank David R. Adams, Klaus Ritter, and Henryk Woźniakowski for valuable suggestions.

REFERENCES

1. Z. Ciesielski, *On Levy's Brownian motion with several-dimensional time*, Lecture Notes in Math., vol. 472, Springer-Verlag, New York, 1975, pp. 29–56.
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. Stat. **41** (1970), 495–502.
4. ———, *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. Levy, *Processus stochastique et mouvement Brownien*, Gauthier-Villars, Paris, 1948.
8. G. M. Molchan, *On some problems concerning Brownian motion in Levy's sense*, Theory Probab. Appl. **12** (1967), 682–690.
9. E. Novak, *Deterministic and stochastic error bounds in numerical analysis*, Lecture Notes in Math., vol. 1349, Springer-Verlag, 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. Stat. **41** (1970), 2057–2074.
12. ———, *Statistical design and integral approximation*, Proc. 12th Bienn. Sem. Canad. Math. Congr., Montreal, 1970, pp. 115–136.
13. P. Speckman, *L_p approximation of autoregressive Gaussian process*, report, Department of Statistics, University of Oregon, Eugene, OR, 1979.
14. A. V. Suldin, *Wiener measure and its applications to approximation methods*. I, Izv. Vyssh. Ucheb. Zaved. Mat. **13** (1959), 145–158. (Russian)
15. ———, *Wiener measure and its applications to approximation methods*. II, Izv. Vyssh. Ucheb. Zaved. Mat. **18** (1960), 165–179. (Russian)
16. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-based complexity*, Academic Press, New York, 1988.
17. G. Wahba, *On the regression design problem of Sacks and Ylvisaker*, Ann. Math. Stat. **42** (1971), 1035–1043.
18. ———, *Spline models for observational data*, SIAM-NSF Regional Conf. Ser. in Appl. Math., vol. 59, SIAM, Philadelphia, PA, 1990.
19. G. W. Wasilkowski, *Information of varying cardinality*, J. Complexity **2** (1986), 204–228.
20. ———, *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** (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. (N.S.) **24** (1991), 185–194.

24. ———, *Average case complexity of linear multivariate problems*, (to appear in Bull. Amer. Math. Soc. (N.S.)).

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

E-mail address: greg@ms.uky.edu