Average Case Complexity of Multivariate Integration

H. WOŹNIAKOWSKI

Department of Computer Science, Columbia University
and Institute of Informatics, University of Warsaw

September, 1989

CUCS-464-89

Abstract. We study the average case complexity of multivariate integration for the class of continuous functions of d variables equipped with the classical Wiener sheet measure. To derive the average case complexity one needs to obtain optimal sample points. This optimal design problem has long been open. All known designs guaranteeing average case error ε lead to an exponential number of sample points, roughly $\Theta(\varepsilon^{-d})$. For d large this makes the problem intractable for even the fastest computers.

Yet good designs have to exist since the average case complexity is bounded by $O(\varepsilon^{-2})$ as can be proven by considering the Monte-Carlo algorithm. We just did not know how to construct them.

In this paper we prove that optimal design is closely related to discrepancy theory which has been extensively studied for many years. Of particular importance for our purpose are papers by Roth [10, 11]. This relation enables us to show that optimal sample points can be derived from Hammersley points. Extending the result of Roth [10] and using the recent result of Wasilkowski [19], we conclude that the average case complexity is $\Theta(\varepsilon^{-1}(\ln \varepsilon^{-1})^{(d-1)/2})$.

1. Introduction

The approximate computation of multivariate integrals has been extensively studied in many papers, see [5–7, 17] for hundreds of references. We assume that multivariate integrals are approximated by evaluating integrands at finitely many deterministic sample points and by performing arithmetic operations and comparisons on real numbers. Assume that the cost of one integrand evaluation is $\mathbf{c}$, and that the cost of one arithmetic operation or comparison is taken as 1. Usually $\mathbf{c} \gg 1$.

We seek the computational complexity of multivariate integration, which is defined as the *minimal cost* of approximating multivariate integrals with *error* at most ε for a given class F of integrands, see [17]. How we define cost and error depends on the *setting*.

In the worst case setting, the cost of an approximation is defined as $\sup_{f \in F}(\mathbf{c}\, n_1(f) + n_2(f))$, where $n_1(f)$ is the number of integrand evaluations and $n_2(f)$ is the number of arithmetic operations and comparisons needed to compute the approximation for $f \in F$. The error of the approximation is defined as the maximal absolute difference between the exact and approximate values of multivariate integrals over the class F. The computational complexity in the worst case setting is denoted by $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, F)$ and is known for many

This research was supported in part by the National Science Foundation under Contract IRI-89-07215

classes F. For example, if $F = W_p^{r,d}$ is the Sobolev class of real functions defined on the d-dimensional cube $D = [0,1]^d$ whose rth distributional derivatives exist and are bounded in the L_p norm by one, then for $pr > d$ we have $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, W_p^{r,d}) = \Theta(\mathrm{c}\,\varepsilon^{-d/r})$, see [7] for a recent survey. For d large relative to r, the worst case complexity is huge even for moderate ε.

In the average case setting, the cost and error of approximations are defined as the expected cost and error with respect to some probability measure on the class F. The computational complexity in the average case setting is denoted by $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F)$. In contrast to the enormous literature for the worst case setting, the average case setting for multivariate integration has been studied in relatively few papers, see [1–4, 7–9, 12–20].

We report briefly on what is known about $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, F)$. For the scalar case $d = 1$, let $F = C^r$ be the class of r times continuously differentiable functions equipped with r-fold Wiener measure. Then

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

where $\alpha = \sqrt{|B_{2r+2}|/(2r+2)!}$ and B_{2r+2} is a Bernoulli number. Furthermore, optimal sample points and how to optimally combine the integrand evaluations at these points are known, see [3, 15, 17].

For the multivariate case, $d \geq 1$, assume that $F = C_d$ is 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

$$R(s,t) \stackrel{\mathrm{def}}{=} \int_{C_d} f(s)\,f(t)\;w(df) = \min(s,t) \stackrel{\mathrm{def}}{=} \prod_{j=1}^{d} \min(s_j, t_j)$$

for any vectors $s = [s_1, \ldots, s_d]$ and $t = [t_1, \ldots, t_d]$ from D.

Only upper and lower bounds are known for $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d)$. Papageorgiou [8] established

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d) \;=\; \Omega\left(\mathrm{c}\;\varepsilon^{-1}\right) \quad \text{and} \quad \mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d) \;=\; O\left(\mathrm{c}\;\varepsilon^{-2}\right), \quad \forall d \geq 1. \qquad (1)$$

The proof of the upper bound is based on the Monte-Carlo algorithm using randomized sample points. Therefore it does not provide a constructive way to find deterministic sample points that achieve the bound $O(\mathrm{c}\,\varepsilon^{-2})$.

To obtain the average case complexity, one needs to find optimal sample points, i.e., sample points which lead to minimal average case error. It is known that optimal sample points do not form a grid. That is, if one assumes that sample points form a grid

$$\{[i_1 h_1, \ldots, i_d h_d] : \; i_j = 1, \ldots, m_j, \; j = 1, 2 \ldots, d\}$$

with $h_j = m_j^{-1}$ for some integers m_j, then Papageorgiou and Wasilkowski [8, 9], see also [20], showed that $\Theta(\varepsilon^{-d})$ grid sample points are needed to achieve the average case error ε. For $d > 2$, this cost is worse than the upper bound in (1). This proves that grid points are a poor choice of sample points. Papageorgiou [8] slightly improved this result by constructing $\Theta(\varepsilon^{-(d^2-d+1)/d})$ sample points for which the average case error is ε. For $d = 2$, these sample points reduce to those proposed by Ylvisaker [20]. Thus, for $d = 2$ we have an improved upper bound

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d) \;=\; O\left(\mathrm{c}\,\varepsilon^{-1.5}\right).$$

The average case complexity $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d)$ has been unknown due to the difficulty of finding optimal sample points. In the statistical literature this is called the optimal design problem for multivariate integration, see [20]. Micchelli and Wahba [4] conjectured that Hammersley points should lead to an optimal convergence rate $n^{-1}(\log n)^{d-1}$ which would imply that $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d)$ is $\Theta(\mathrm{c}\,\varepsilon^{-1}(\log \varepsilon^{-1})^{d-1})$. This form of the average case complexity has been also conjectured by Papageorgiou and Wasilkowski [9], based on their work on the approximation problem.

In this paper we show that optimal sample points are indeed related to Hammersley points and that the average case complexity is $\Theta(\mathrm{c}\,\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2})$. This will be done by showing that the optimal sample points problem is related to the L_2 discrepancy. The discrepancy in L_2 and other norms has been extensively studied in the literature and deep relations with number theory have been established. The reader is reffered to excellent surveys by Niederreiter [5, 6] with about 500 references. The L_2 discrepancy has been studied by Roth [10, 11] who proved in 1954 that the L_2 disrepancy of n points has to be at least of order $n^{-1}(\log n)^{(d-1)/2}$, and in 1980 that this bound is sharp.

2. Main Result

For $t = [t_1, \ldots, t_d] \in D = [0,1]^d$, define $[0,t) = [0,t_1) \times \cdots \times [0,t_d)$. Let $\chi_{[0,t)}$ be the characteristic (indicator) function of $[0,t)$. For $z_1, \ldots z_n \in D$ define

$$R_n(t; z_1, \ldots, z_n) \;= n^{-1} \sum_{k=1}^{n} \chi_{[0,t)}(z_k) \;-\; t_1 t_2 \cdots t_d$$

as the difference between the fraction of the points z_i in $[0,t)$ and the volume of $[0,t)$. The L_2 *discrepancy* of $z_1, \ldots, z_n$ is defined as the L_2 norm of the function $R(\cdot; z_1, \ldots, z_n)$. Roth [10, 11] proved that

$$\inf_{z_1,\ldots,z_n} \left(\int_D R_n^2(t; z_1, \ldots, z_n)\, dt \right)^{1/2} \;=\; \Theta\left(n^{-1}\,(\log n)^{(d-1)/2}\right). \qquad (2)$$

The points $z_1^*, \ldots, z_n^*$ with L_2 discrepancy of order $n^{-1}(\log n)^{(d-1)/2}$ are related, as indicated below, to Hammersley points. Let $p_1, p_2, \ldots, p_{d-1}$ be the first $(d-1)$ prime numbers. Any integer $k \geq 0$ can be uniquely represented as $k = \sum_{i=0}^{\lceil \log k \rceil} a_i p_j^i$ with integers $a_i \in [0, p_{j-1}]$. The radical inverse function ϕ_{p_j} is given as $\phi_{p_j}(k) = \sum_{i=0}^{\lceil \log k \rceil} a_i p_j^{-i-1}$. The sequence $\{u_k\}$ of $(d-1)$ dimensional vectors for $k = 0, \pm 1, \pm 2, \ldots$ is defined by

$$u_k = [\phi_{p_1}(k), \phi_{p_2}(k), \ldots, \phi_{p_{d-1}}(k)], \quad k = 0, 1, \ldots, M-1,$$

with $M = (p_1 p_2 \cdots p_{d-1})^{\lceil \log n \rceil}$, and by $u_{k+M} = u_k$, $\forall k$. Then there exists a real number t^* such that the points $z_1^*, \ldots, z_n^*$ are given by

$$\{z_1^*, \ldots, z_n^*\} \;=\; \left\{ \left[(k+t^*)n^{-1}, u_k \right] \;:\; 0 \leq k + t^* < n \right\}. \tag{3}$$

For $t^* = 0$, they are Hammersley points. Thus, the points z_i^* are obtained by adding t^*/n to the first component of Hammersley points.

We are ready to show how the results on L_2 discrepancy can be used to derive optimal sample points and average case complexity. In what follows, we use the word "optimal" modulo a constant which may depend on d but is independent of ε.

Let $n = \Theta(\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2})$ be chosen such that the L_2 discrepancy of the function $R_n(\cdot, z_1^*, \ldots, z_n^*)$ is at most ε. Letting $\vec{1} = [1, 1, \ldots, 1]$, we define

$$x_k^* \;=\; \vec{1} - z_k^*, \quad k = 1, 2, \ldots, n. \tag{4}$$

We approximate the integral of f from C_d by the arithmetic mean of its values at x_k,

$$I(f) \stackrel{\text{def}}{=} \int_D f(t)\,dt \;\sim\; U(f) \stackrel{\text{def}}{=} n^{-1} \sum_{k=1}^{n} f(x_k^*), \quad \forall f \in C_d. \tag{5}$$

Clearly cost(U), the cost of computing $U(f)$, is $(c+1)n$. The average case error of U is defined as $(\int_{C_d} (I(f) - U(f))^2\, w(df))^{1/2}$. We summarize the main results as a

THEOREM.

(i) $comp^{avg}(\varepsilon, C_d) \;=\; \Theta\left(c\,\varepsilon^{-1}\,(\log \varepsilon^{-1})^{(d-1)/2}\right)$,

(ii) $x_1^*, x_2^*, \ldots, x_n^*$ *given by (4) are optimal sample points,*

(iii) U *given by (5) is optimal, i.e., the average case error of* U *is at most* ε *and*

$$cost(U) = \Theta(comp^{avg}(\varepsilon, C_d)).$$

SKETCH OF THE PROOF:

An upper bound on $\text{comp}^{\text{avg}}(\varepsilon, C_d)$ is cost(U) provided we prove that the average case error of U is at most ε. This will be done by exhibiting the indentity which relates the average case error of U with the L_2 discrepancy.

LEMMA 1.

$$\int_{C_d} \left(I(f) - U(f) \right)^2 \, w(df) \;=\; \int_D R_n^2(t; z_1^*, \ldots, z_n^*)\, dt.$$

To prove Lemma 1, we use

$$\int_{C_d} I^2(f)\, w(df) = 3^{-d} \quad \text{and} \quad \int_{C_d} I(f)\, f(t)\, w(df) = \prod_{i=1}^{d} t_i(1 - t_i/2), \forall t = [t_1, \ldots, t_n].$$

From these formulas we compute the left-hand side of Lemma 1,

$$\text{LHS} = 3^{-d} - \frac{2}{n} \sum_{k=1}^{n} \prod_{i=1}^{d} x_{k,i}^*(1 - x_{k,i}^*/2) + \frac{1}{n^2} \sum_{k,p=1}^{n} \min(x_k^*, x_p^*),$$

where $x_k^* = [x_{k,1}^*, \ldots, x_{k,d}^*]$. The right-hand side of Lemma 1 can be directly computed,

$$\begin{aligned} \text{RHS} &= \int_D t_1^2 \cdots t_d^2 \, dt - \frac{2}{n} \sum_{k=1}^{n} \int_D t_1 \cdots t_d \, \chi_{[0,t)}(z_k^*)\, dt + \frac{1}{n^2} \sum_{k,p=1}^{n} \int_D \chi_{[0,t)}(z_k^*) \chi_{[0,t)}(z_p^*)\, dt \\ &= 3^{-d} - \frac{2}{n} \sum_{k=1}^{n} \prod_{i=1}^{d} \frac{1 - {z_{k,i}^*}^2}{2} + \frac{1}{n^2} \sum_{k,p=1}^{n} \prod_{i=1}^{d} \left(1 - \max\left(z_{k,i}^*, z_{p,i}^*\right)\right). \end{aligned}$$

We complete the proof of Lemma 1 by substituting $z_{k,i}^* = 1 - x_{k,i}^*$ and noting that $1 - \max(1 - a, 1 - b) = \min(a, b)$.

Due to the choice of n and the construction of z_i^*, the average case error of U is at most ε. This completes the proof of the upper bound on $\text{comp}^{\text{avg}}(\varepsilon, C_d)$.

To derive a lower bound on $\text{comp}^{\text{avg}}(\varepsilon, C_d)$, we use a proof technique from [17]. First take any nonadaptive sample points $x_1, \ldots, x_n$ from D. Since w is Gaussian and the average case error is defined in the L_2 sense, it is known that the approximation U^* with minimal average case error is the mean of the conditional probability, given $f(x_1), \ldots, f(x_n)$. Since the mean depends linearly on $f(x_i)$, U^* takes the form $U^*(f) = \sum_{k=1}^{n} c_k f(x_k)$, $\forall f \in C_d$, for some numbers c_k. Using similar calculations as in the proof of Lemma 1 we get

LEMMA 2.

$$\int_{C_d} \left(I(f) - U^*(f) \right)^2 \, w(df) \;=\; \int_d \left(\sum_{k=1}^{n} c_k \, \chi_{[0,t)}(\vec{1} - x_k) \;-\; t_1 \cdots t_d \right)^2 dt.$$

It is possible to extend the proof of Roth [10] to show that the lower bound of (2) holds for all c_k. More precisely we have

LEMMA 3. *There exists a positive number γ_d such that*

$$\int_{\bar{d}} \left(\sum_{k=1}^{n} c_k \, \chi_{[0,t)}(\vec{1} - x_k) \; - \; t_1 \cdots t_d \right)^2 dt \; \geq \; \gamma_d \, n^{-2} \, (\log n)^{d-1}$$

for all n, c_k and x_k.

From Lemmas 2 and 3 we conclude that the average case error of U^* is bounded by ε only if $n = \Omega(\varepsilon^{-1} (\log \varepsilon^{-1})^{(d-1)/2})$. Therefore the cost of approximating $I(f)$ with average case error ε must be at least $cn = \Omega(c\, \varepsilon^{-1} (\log \varepsilon^{-1})^{(d-1)/2})$.

Consider now adaptive sample points $x_1, x_2, \ldots, x_{n(f)}$. That is, the choice of x_i may depend on the already computed $f(x_1), \ldots, f(x_{i-1})$, and the number $n(f)$ of sample points may also be adaptively chosen. For Gaussian measures adaption may help only by varying $n(f)$. We now draw on Wasilkowski's theorem, see [19], which states that adaption can help only by a multiplicative constant if the squares of the minimal average case errors r_n of n nonadaptive sample points can be bounded by two convex sequences α_n and β_n such that $\alpha_n \leq r_n \leq \beta_n$ and $\alpha_n = \Theta(\beta_n)$. In our case, $r_n = \Theta(n^{-2}(\log n)^{d-1})$ and we can take $\alpha_n = c_1 \, n^{-2}(\log n)^{d-1}$, $\beta_n = c_2 \, n^{-2}(\log n)^{d-1}$ for some positive constants c_1 and c_2. Thus, adaption does not help much and $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, C_d) = \Omega(c\, \varepsilon^{-1} (\log \varepsilon^{-1})^{(d-1)/2})$. This completes the proof of (i) and the rest follows easily from (i).

3. Final Remarks

A. The definition (4) of optimal sample points is not fully constructive due to the unspecified constant t^* in (3). It would be interesting to determine the constant t^* explicitly. On the other hand, if one takes the classical Hammersley points

$$z_k = [kn^{-1}, \phi_{p_1}(k), \ldots, \phi_{p_{d-1}}(k)], \quad k = 1, \ldots, n,$$

then Halton proved, see e.g. [5], that even the L_∞ discrepancy of the points z_k is of order $n^{-1}(\log n)^{d-1}$. Then the approximation $U(f) = n^{-1} \sum_{k=1}^{n} f(x_k)$ with $x_k = \vec{1} - z_k$ has average case error at most ε provided that $n = \Theta(\varepsilon^{-1}(\log \varepsilon^{-1})^{d-1})$.

Observe that the definition of x_k (as well as x_k^* in (4)) depends on the total number n of samples. Sometimes it is better to use an infinite sequence $\{\tilde{x}_k\}$ in which the definition of $\tilde{x}_k$ does not depend on the specific value of n. For instance $\tilde{x}_k$ can be given by

$$\tilde{x}_k = \vec{1} - [\phi_{p_1}(k), \phi_{p_2}(k), \ldots, \phi_{p_d}(k)], \quad k = 1, 2, \ldots,$$

where $p_1, \ldots, p_d$ are the first d prime numbers. The L_2 discrepancy of n points $\vec{1} - \tilde{x}_k$ is at most of order $n^{-1}(\log n)^d$, see [5], and therefore the approximation $U(f) = n^{-1} \sum_{k=1}^{n} f(\tilde{x}_k)$ has the average case error at most ε provided that $n = \Theta(\varepsilon^{-1}(\log \varepsilon^{-1})^d)$.

In both cases, we see that the use of x_k or $\tilde{x}_k$ leads to a slight increase in the average case cost. More precisely, instead of the minimal average case cost $\Theta(c\,\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2})$, we approximate multivariate integrals at average case cost $\Theta(c\,\varepsilon^{-1}(\log \varepsilon^{-1})^{d-1})$ using the sample points x_k, or $\Theta(c\,\varepsilon^{-1}(\log \varepsilon^{-1})^d)$ using the sample points $\tilde{x}_k$.

B. We have shown that the minimal number of sample points $n(\varepsilon)$ to guarantee average case error ε is of order $\varepsilon^{-1}(\log \varepsilon^{-1})^{(d-1)/2}$ with the constant in the Θ notation dependent only on d. An open problem is to ascertain if there exists a constant α_d and, if so, to find α_d such that

$$n(\varepsilon) = \alpha_d\, \varepsilon^{-1} \left(\log \varepsilon^{-1}\right)^{(d-1)/2} (1 + o(1)) \text{ as } \varepsilon \to 0.$$

C. It would be interesting to extend the results of this paper for smoother classes of functions equipped with folded Wiener sheet measures.

D. Do similar relations to the one we have utilized here between discrepancy and multivariate integration, hold for other problems such as approximation of functions of d variables?

4. Acknowledgments

I am grateful to J. F. Traub for valuable comments and help during the preparation of this paper. I also appreciate remarks from M. Kon, G. W. Wasilkowski and A. G. Werschulz.

REFERENCES

1. Kimeldorf, G.S. and Wahba, G., *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Stat. **41** (1970), 495–502.
2. Kimeldorf, G.S. and Wahba, G., *Spline functions and stochastic processes*, Sankhya Ser. A **32** (1970), 173–180.
3. Lee, D. and Wasilkowski, G. W., *Approximation of linear functionals on a Banach space with a Gaussian measure*, J. Complexity **2** (1986), 12–43.
4. Micchelli, C.A. and Wahba, G., *Design problems for optimal surface interpolation*, in Approximation Theory and Applications (1981.), 329–347 (Z. Ziegler, ed.), Academic Press, New York.
5. Niederreiter, H., *Quasi-Monte Carlo methods and pseudo-random numbers*, Bull. Amer. Math. Soc. **84** (1978), 957–1041.
6. Niederreiter, H., *Quasi-Monte Carlo methods for multidimensional numerical integration*, in Numerical Integration III, International Series of Numerical Mathematics **85**, 157–171 (H. Braß and G. Hämmerlin eds.), Birhäuser Verlag, Basel, 1988.
7. Novak, E., "Deterministic and Stochastic Error Bounds in Numerical Analysis," Lectures Notes in Math. Springer Verlag, Berlin, 1988.
8. Papageorgiou, A., *On Average Case Complexity*, "Ph. D. thesis, Dept. of Computer Science, Columbia University," 1989.
9. Papageorgiou, A. and Wasilkowski, G. W., *On the average complexity of multivariate problems*, to appear in J. Complexity (1986).
10. Roth, K.F., *On irregularities of distribution*, Mathematika **1** (1954), 73–79.
11. Roth, K.F., *On irregularities of distribution, IV*, Acta Arith. **37** (1980), 67–75.
12. Sacks, J. and Ylvisaker, D., *Designs for regression with correlated errors*, Ann. Math. Stat. **37** (1966), 68–89.
13. Sacks, J. and Ylvisaker, D., *Designs for regression problems with correlated errors; many parameters*, Ann. Math. Stat. **39** (1968), 49–69.
14. Sacks, J. and Ylvisaker, D., *Designs for regression problems with correlated errors III*, Ann. Math. Stat. **41** (1970), 2057–2074.
15. Sacks, J. and Ylvisaker, D., *Statistical design and integral approximation*, Proc. 12th Bienn. Semin. Can. Math. Cong. (1970), 115–136.
16. Suldin, *Wiener measure and its applications to approximation methods, I and II, (in Russian)*, Izv. Vyssh. Ucheb. Zaved. Mat. **13, 18** (1959, 1960), 145–158, 165–179.
17. Traub, J.F., Wasilkowski, G.W. and Woźniakowski, H., "Information-based Complexity," Academic Press, New York, 1988.
18. Wahba, G., *On the regression design problem of Sacks and Ylvisaker*, Ann. Math, Stat. **42** (1971), 1035–1043.
19. Wasilkowski, G.W., *Information of varying cardinality*, J. Complexity **2** (1986), 204–228.
20. Ylvisaker, D., *Designs on random fields*, in A survey of statistical design and linear models (1975), 593–607 (J. Srivastava, ed.), North-Holland.