Exact Cubature for a Class of Functions of Maximum Effective Dimension

Shu Tezuka [a] and Anargyros Papageorgiou [b]

[a] *Faculty of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, Fukuoka-ken, Japan 812-8581*

[b] *Department of Computer Science, Columbia University, New York, NY 10027*

Abstract

We consider high dimensional integration in a broad class of functions where all elements have maximum effective dimension. We show that there exists an exact cubature with only two points. Therefore, not only the convergence but also the worst case error of quasi-Monte Carlo need not depend on the effective dimension at all.

Key words: ANOVA, antisymmetric functions, effective dimension, high dimensional integrals

1 Introduction

About ten years ago the *effective dimension* was proposed in [1] as an indicator of the difficulty of high dimensional integration. The purpose was to explain the remarkable success of quasi-Monte Carlo (QMC) in approximating very high dimensional integrals in finance [10,16]. The authors of [1] argued that the integrands are of low effective dimension and that is why QMC is much faster than Monte Carlo (MC).

Defining a notion such as the effective dimension is an attempt to model *reality* [5]. A function, of d variables, that is equal to the sum of d functions of a single variable is one dimensional in a sense. It turns out that the effective dimension of this function is equal to 1 and, generally, QMC is quite successful in approximating the integrals of functions of a single variable. On the

Email addresses: `tezuka@math.kyushu-u.ac.jp` (Shu Tezuka), `ap@cs.columbia.edu` (Anargyros Papageorgiou).

other hand, it is known that QMC fails for certain functions that have high effective dimension. Owen [7] has pointed out that low effective dimension is not a sufficient condition for QMC integration to beat MC and additional assumptions are required, such as smoothness. So, is low effective dimension a necessary condition for QMC to beat MC or for high dimensional integration to be tractable? [1] Is it necessary for the worst case error of QMC to depend on the effective dimension?

The first question was only recently settled [17], as we will explain below. Some earlier papers, e.g., [4,9,11,12], showed that multivariate integration is tractable and that QMC converges fast for a number of classes of high dimensional integrands. Nevertheless, the impact of the arguments in [1] was great, a number of papers dealt with the relationship between the error of QMC and the effective dimension, e.g. [3,19], and a number of researchers believed the answer to the questions above was positive. For instance, the authors of [2, p. 595] state they accept the assertion in [1] and, in their opinion also, the reason for the success of QMC is the low effective dimension (in the superposition sense) of the integrands. Recently, Tezuka [17] showed a class of functions of d variables, all having maximum effective dimension equal to d, for which the QMC convergence rate is n^{-1}, where n is the number of function evaluations. Hence, QMC can beat MC and high dimensional integration is tractable for functions of high effective dimension.

In this companion paper we show that it is not only the convergence of QMC that need not depend on the effective dimension but also its worst case error, which can be zero for functions of maximum effective dimension. For any $d \geq 1$, we construct a broad class of functions (of d variables,) all having maximum effective dimension equal to d, for which the integration problem is solved exactly by a cubature with two points only. Hence, we have a QMC algorithm, in the sense that it is the average of two function evaluations at deterministic points, that has zero worst case error.

2 Background

We begin with the ANOVA (analysis of variance) decomposition of a function of d variables. Consider a square integrable function $f : [0,1]^d \to \mathbb{R}$. Let $u \subseteq \{1, 2, ..., d\}$ be a subset of the coordinates of $[0,1]^d$, let $\bar{u} = \{1, 2, ..., d\} - u$ be its complement, and denote by $|u|$ the cardinality of u. The ANOVA decomposition of f (see, e.g. [6]) is defined by

$$f(x_1, ..., x_d) = \sum_{u \subseteq \{1,2,...,d\}} \alpha_u(x_1, ..., x_d),$$

[1] See [11] for the definition of tractability.

where the sum is over all 2^d subsets of coordinates of $[0,1]^d$. The terms $\alpha_u(x_1, ..., x_d)$ are defined recursively starting with

$$\alpha_\emptyset(x_1, ..., x_d) := I(f) \equiv \int_{[0,1]^d} f(x_1, ..., x_d) dx_1 ... dx_d,$$

and for $0 < |u| < d$

$$\alpha_u(x_1, ..., x_d) := \int_{[0,1]^{d-|u|}} f(x_1, ..., x_d) \prod_{i \in \bar{u}} dx_i - \sum_{v \subset u, v \neq u} \alpha_v(x_1, ..., x_d),$$

where the integral is with respect to the variables with indices in the complement of u, and the sum is over proper subsets of u. When $u = \{1, ..., d\}$,

$$\alpha_{\{1,...,d\}}(x_1, ..., x_d) = f(x_1, ..., x_d) - \sum_{v \subset \{1,...,d\}} \alpha_v(x_1, ..., x_d),$$

where the sum is over proper subsets of $\{1, \ldots, d\}$. Each of the $\alpha_u(x_1, ..., x_d)$ is the effect of the subset $\{x_i \mid i \in u\}$ on $f(x_1, ..., x_d)$ minus the effects of its proper subsets $\{x_i \mid i \in v\}$ with $v \subset u$. The functions $\alpha_u(x_1, ..., x_d)$ have the following properties:

- Let $i \in u$. If we fix all the $x_j, j \neq i$, then

$$\int_0^1 \alpha_u(x_1, ..., x_d) dx_i = 0.$$

  Thus, when $\emptyset \neq u \subset \{1, ..., d\}$,

$$\int_{[0,1]^d} \alpha_u(x_1, ..., x_d) dx_1 ... dx_d = 0.$$

- When $u \neq v$,

$$\int_{[0,1]^d} \alpha_u(x_1, ..., x_d) \alpha_v(x_1, ..., x_d) dx_1 ... dx_d = 0.$$

  Hence, the variance $\sigma^2 = \sigma^2(f)$ of f is given by

$$\sigma^2 = \int_{[0,1]^d} (f(x_1, ..., x_d) - \alpha_\emptyset(x_1, ..., x_d))^2 dx_1 ... dx_d = \sum_{|u|>0} \sigma_u^2,$$

  where

$$\sigma_u^2 := \sigma^2(\alpha_u) = \begin{cases} 0 & \text{if } u = \emptyset, \\ \int_{[0,1]^d} \alpha_u(x_1, ..., x_d)^2 dx_1 ... dx_d & \text{otherwise} . \end{cases}$$

The definition of the effective dimension was introduced in two ways in [1]:

Truncation sense:

$$D_{trunc} := \min\{i : 1 \le i \le d \ \text{ such that } \sum_{u \subseteq \{1,2,\ldots,i\}} \sigma_u^2 \ge (1-\epsilon)\sigma^2\}.$$

Superposition sense:

$$D_{super} := \min\{i : 1 \le i \le d \ \text{ such that } \sum_{|u| \le i} \sigma_u^2 \ge (1-\epsilon)\sigma^2\}.$$

In practice the value of ϵ is often chosen to be 0.01. Nevertheless, ϵ can be set to any value in $[0,1)$ (the case $\epsilon = 1$ is not interesting.) As pointed out in [3], $(1-\epsilon)$ reflects a proportion of the variance and one can choose to study how different proportions of the variance are affected by different values of D_{trunc} or D_{super}, and vice versa.

On the other hand, as remarked by one of the referees, for fixed positive ϵ, even if the QMC error is surprisingly small for initial values of n, the asymptotic behavior of the QMC error must be the same as that for $\epsilon = 1$ and, therefore, claims concerning the impact of the effective dimension on the convergence rate of QMC are not well founded mathematically. The case $\epsilon = 0$ is different. Consider the effective dimension in the superposition sense. Then we know that the function is equal to the sum of functions of at most D_{super} variables. Assuming that D_{super} is independent of d, this corresponds to *finite-order weights* recently studied in many papers, see, e.g. [13,20]. Nevertheless, the effective dimension cannot provide a necessary condition for good QMC error bounds for arbitrary classes of functions.

3 Functions with the maximum effective dimension

Let us first introduce the following functions of d variables:

Definition 1 *Let $\phi : [0,1]^d \to \mathbb{R}$ be such that*

$$\phi(x_1, \ldots, x_d) = \prod_{i=1}^{d} r_i(x_i),$$

where $r_i : [0,1] \to \mathbb{R}$, $1 \le i \le d$, are continuous functions such that

$$\int_0^1 r_i(x_i)dx_i = 0.$$

In addition, $r_1(x)$ is antisymmetric about 1/2, i.e, for $0 \le x < 1/2$, $r_1(1-x) = -r_1(x)$.

For instance, r_1 can be the one of the functions $\sin(2j\pi x)$, $j = 1, 2, \ldots,$ and the r_i, $i > 1$, can be periodic trigonometric functions with integral zero; such functions are frequently encountered in physics.

We now define the class $\mathfrak{F}_d$ of functions in d dimensions.

Definition 2 *Let $\phi_k : [0,1]^d \to \mathbb{R}$, $k = 1, 2, \ldots,$ be functions that satisfy Definition 1. Assume that the ϕ_k, $k = 1, 2, \ldots,$ along with the constant function 1 are linearly independent. Let $\mathfrak{F}_d$ be the class of all functions $f : [0,1]^d \to \mathbb{R}$ such that*

$$f = c_0 + \sum_{k=1}^{\infty} c_k \phi_k,$$

with $c_k \in \mathbb{R}$, $k = 0, 1, \ldots,$ and

$$\|f\| = \left\{ \int_{[0,1]^d} f^2(x_1, ..., x_d)\, dx_1 ... dx_d \right\}^{1/2} \leq 1.$$

For example, we can construct such a class of functions by carefully selecting the functions ϕ_k among the eigenfunctions of the Laplacian, in d dimensions. Indeed, the functions $\xi_{j_1,\ldots,j_d}(x_1, \ldots, x_d) = 2^{d/2} \sin(j_1 \pi x_1) \cdots \sin(j_d \pi x_d)$, $j_i = 1, 2, \ldots,$ $i = 1, ..., d$, are orthonormal eigenfunctions of the Laplacian. The eigenfunctions that correspond to even frequencies (i.e., all the j_i even), satisfy the conditions of Definitions 1 and 2.

Let's consider the ANOVA decomposition of $f \in \mathfrak{F}_d$. We use induction on the cardinality of u. We have

$$\alpha_\emptyset(x_1, ..., x_d) = \int_{[0,1]^d} f(x_1, ..., x_d) dx_1 ... dx_d = c_0.$$

By Definition 2, for $\emptyset \neq u \subset \{1, ..., d\}$ we have

$$\begin{aligned}
\alpha_u(x_1, ..., x_d) &= \int_{[0,1]^{d-|u|}} f(x_1, ..., x_d) \prod_{i \in \bar{u}} dx_i - \sum_{v \subset u, v \neq u} \alpha_v(x_1, ..., x_d) \\
&= \int_{[0,1]^{d-|u|}} f(x_1, ..., x_d) \prod_{i \in \bar{u}} dx_i - c_0 \\
&= \int_{[0,1]^{d-|u|}} \sum_{k=1}^{\infty} c_k \phi_k(x_1, ..., x_d) \prod_{i \in \bar{u}} dx_i \\
&= \sum_{k=1}^{\infty} c_k \int_{[0,1]^{d-|u|}} \phi_k(x_1, ..., x_d) \prod_{i \in \bar{u}} dx_i = 0
\end{aligned}$$

and

$$\alpha_{\{1,...,d\}}(x_1, ..., x_d) = f(x_1, ..., x_d) - c_0.$$

Thus, we have $\sigma(f) = \sigma_{\{1,...,d\}}$.

Theorem 1 *For any non-constant function $f \in \mathfrak{F}_d$, its effective dimension, whether in the truncation or in the superposition sense, is equal to d.*

Note that the above theorem holds for every $\epsilon \in [0, 1)$ in the definition of effective dimension, i.e., the effective dimension is d regardless of the proportion of the variance that one may choose to consider.

3.1 Exact cubature with two points

We are now ready to show that although the effective dimension of all functions in the class $\mathfrak{F}_d$ is d, they can be integrated exactly by a cubature with two points.

Theorem 2 *There exists an exact cubature with two points for the integration of any function $f \in \mathfrak{F}_d$.*

Proof. By definition any function $f \in \mathfrak{F}_d$ is given by

$$f(x_1, ..., x_d) = c_0 + \sum_{k=1}^{\infty} c_k \prod_{i=1}^{d} r_i^{(k)}(x_i).$$

Take a point $(s_1, ..., s_d) \in [0, 1]^d$. Consider the following cubature with two points

$$Q_2(f) \equiv \frac{1}{2}(f(s_1, s_2, ..., s_d) + f(1 - s_1, s_2, ..., s_d))$$
$$= c_0 + \sum_{k=1}^{\infty} c_k \left(\frac{r_1^{(k)}(s_1) + r_1^{(k)}(1 - s_1)}{2} \prod_{i=2}^{d} r_i^{(k)}(s_i) \right),$$

and since each term of the infinite sum is zero the last expression is equal to c_0. Hence,

$$\int_{[0,1]^d} f(x_1, ..., x_d) dx_1 ... dx_d = c_0 = Q_2(f),$$

because $r_1^{(k)}(x), k = 1, \ldots,$ are antisymmetric from Definition 1. This completes the proof. □

We see that functions of high effective dimension can be very easy to integrate in the worst case. Moreover, we can expand our class of functions by slightly modifying the two definitions so as to include piecewise continuous square integrable functions that may have singularities on subsets of $[0, 1]^d$ with Lebesgue measure zero. For instance, this would allow us to include functions that depend on the inverse of the standard normal distribution in some of the dimensions; such functions are common in finance. Unless the singularities are known we cannot consider deterministic algorithms that use function evaluations for the integration problem. Instead, we can use randomized algorithms. By choosing the sample points at random with uniform distribution and then applying the cubature of Theorem 2, we obtain a Monte Carlo algorithm with variance reduction that solves the integration problem exactly. On the other hand, even if one considers randomized algorithms only, why should the effective dimension, a quantity defined through the variance of a function, be the indicator of the difficulty of high dimensional integration when the error need not depend on the variance at all?

Finally, as a corollary of Theorem 2, we remark that for the class $\mathfrak{F}_d$ there exists also a one-point exact rule since $f(\frac{1}{2}, ..., \frac{1}{2}) = c_0$ because $r_1(\frac{1}{2}) = 0$, but this hardly makes any difference from a complexity point of view. Moreover, this rule cannot be used if we extend the class of functions by assuming the $r_i, 1 \le i \le d$, are integrable but not necessarily continuous functions.

4 Discussion

Global sensitivity based on ANOVA was proposed by Sobol' in 1990 [14,15] to consider the global importance of variables, of a multivariate function, on the function itself. The notion of effective dimension [1,8] is essentially the same as global sensitivity, but it is more quantitative aiming to measure how important is each subset of variables on the function. As mentioned in Introduction, the notion of effective dimensions was developed for the purpose of explaining why QMC beats MC by a wide margin for some high dimensional integration problems in finance.

We have shown that it is possible to have functions of maximum effective dimension for which the integration problem can be solved exactly. This totally contradicts what we want to imply by the effective dimension since, presumably, a high effective dimension should imply that a problem is hard to solve. Clearly a different definition of the effective dimension is needed.

Some useful insight into factors that characterize the nature of the functions is provided by considering the approximation problem. Suppose we are interested in the L_2 approximation of a function $f \in \mathfrak{F}_d$. Then the zero algorithm is optimal and its worst case error is 1, which suggests that the class of functions is too broad and needs to be restricted.

For this we follow the approach in [18, p. 121]. For simplicity let us assume that the ϕ_k, $k = 1, 2, \dots,$ is an orthonormal family of functions, i.e., $\int_{[0,1]^d} \phi_k(x_1, ..., x_d)\phi_i(x_1, ..., x_d)\, dx_1...dx_d = \delta_{k,i}$. Definitions 1 and 2 ensure that by including the constant function 1 in this family the functions remain orthonormal.

The restriction of the class of functions can be obtained through an operator of the form $Tf = \beta_0 c_0 + \sum_{k=1}^{\infty} \beta_k c_k \phi_k$, where the $\beta_k \in \mathbb{R}$, $|\beta_k| \le |\beta_{k+1}|$, $k = 0, 1, \dots$, and where the L_2 norm of Tf satisfies $\|Tf\| \le 1$. Observe that $c_0 = \int_{[0,1]^d} f(x_1, ..., x_d)\, dx_1 ... dx_d$ and $c_k = \int_{[0,1]^d} f(x_1, ..., x_d) \phi_k(x_1, ..., x_d)\, dx_1 ... dx_d$, are continuous linear functionals, $k = 1, 2, \dots$. Among all algorithms that use information about f composed of n evaluations of continuous linear functionals, the optimal algorithm that uses optimal information is $\hat{f} = c_0 + \sum_{k=1}^{n-1} c_k \phi_k$ and its worst case error is $|\beta_n|^{-1}$. The convergence of $|\beta_n|^{-1}$ as $n \to \infty$ and its dependence on d are important for the worst case error. To be able to solve the problem for any desired accuracy we need $|\beta_n| \to \infty$ as $n \to \infty$.

On the other hand, if we use function evaluations to approximate f and even if we can approximate the integrals c_k, $k = 0, \dots, n$, exactly, the worst case error will be at least $|\beta_n|^{-1}$. (Observe that in our case c_0 is the integral of f and we can approximate it exactly using two function evaluations.) Hence, the choice of the restriction operator T determines the worst case error of the optimal algorithm and thereby the problem difficulty.

Depending on the choice of the ϕ_k, $k = 1, 2, \dots$, an option is to use smoothness to restrict the class of functions. One can define the β_k, $k = 1, 2, \dots$, in a way analogous to requiring that the L_2 norm of a certain derivative of f be bounded, say, by 1 (β_0 can be set to any convenient value since c_0 can be computed exactly.) If the β_k are supposed to control a certain rate of growth of f then they can depend on some interaction of $l \le d$ dimensions. In this sense, l can be considered as the effective dimension. However, it is not necessary to restrict the class of functions using smoothness to define the β_k's. Moreover, the β_k can be totally independent of d while we still have $|\beta_n| \to \infty$ as $n \to \infty$.

Therefore, the nature of the functions under consideration is characterized by Definitions 1 and 2, and by the restriction operator T (in terms of the choice of the β_k, $k = 0, 1, \dots$.) This is missed by the definition of the effective dimension, which is equal to d for an easy integration problem and regardless of the difficulty of the approximation problem.

5 Acknowledgments

We thank the anonymous referees for their remarks which we included in this paper.

References

[1] R. E. Caflisch, W. Morokoff, and A. B. Owen, Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension, *Journal of Computational Finance*, **1**, (1997), 27-46.

[2] J. Dick, I. H. Sloan, X. Wang and H. Woźniakowski, Liberating the weights, *Journal of Complexity*, **20**, 5, (2004), 593-623.

[3] F. J. Hickernell, Lattice rules: how well do they measure up?, in P. Hellekalek and G. Larcher (Eds.), *Random and Quasi-Random Point Sets*, Springer, (1998), 109-166.

[4] F. J. Hickernell and X. Wang, The error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimensions, *Mathematics of Computation,* **71**, 240, (2001), 1641-1661.

[5] F. Y. Kuo and I. H. Sloan, Lifting the curse of dimensionality, University of New South Whales, preprint, 2005.

[6] A. B. Owen, Monte Carlo Variance of scrambled net quadrature, *SIAM Journal on Numerical Analysis,* **34** (1997), 1884-1910.

[7] A. B. Owen, Necessity of low effective dimension, Stanford University, preprint, 2002.

[8] A. B. Owen, The dimension distribution, and quadrature test functions, *Statistica Sinica*, **13** (2003), 1-17.

[9] A. Papageorgiou, Sufficient conditions for fast quasi-Monte Carlo convergence, *Journal of Complexity*, **19**, 3, (2003), 332-351.

[10] S. H. Paskov, New methodologies for valuing derivatives, In M. A. H. Dempster and S. Pliska (Eds.), *Mathematics of Derivative Securities*, Cambridge University Press (1997), 545-582.

[11] I. H. Sloan and H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?, *Journal of Complexity,* **14**, 1, (1998), 1-33.

[12] I. H. Sloan and H. Woźniakowski, Tractability of multivariate integration for weighted Korobov classes, *Journal of Complexity,* **17**, 4, (2001), 697-721.

[13] I. H. Sloan, X. Wang and H. Woźniakowski, Finite-order weights imply tractability for multivariate integration, *Journal of Complexity,* **20**, 1, (2004), 46-74.

[14] I. M. Sobol', Sensitivity estimates for nonlinear mathematical models, *Mathematical Modelling and Computational Experiments*, **1** (1993), 407-414. (translation of I. M. Sobol', *Matematicheskoe Modelirovanie*, **2** (1990), 112-118. (in Russian))

[15] I. M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, **55** (2001), 271-280.

[16] S. Tezuka, Quasi-Monte Carlo: discrepancy between theory and practice, In K. T. Fang, F. J. Hickernell, and H. Niederreiter (Eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2000*, Springer, (2002), 124-140.

[17] S. Tezuka, On the necessity of low-effective dimension, *Journal of Complexity*, 21 (2005), 710-721.

[18] J. F. Traub and H. Woźniakowski, A General Theory of Optimal Algorithms, Academic Press (1980).

[19] X. Wang and I. H. Sloan, Why are high-dimensional finance problems often of low effective dimension?, *SIAM Journal on Scientific Computing,* **27**, 1, (2005), 159-183.

[20] G. W. Wasilkowski and H. Wozniakowski, Finite-order weights imply tractability of linear multivariate problems, *Journal of Approximation Theory,* **130**, 1, (2004), 57-77.