

ON THE APPROXIMATION ERROR IN HIGH DIMENSIONAL MODEL REPRESENTATION

Xiaoqun Wang

Department of Mathematical Sciences
 Tsinghua University
 Beijing 100084, CHINA

ABSTRACT

Mathematical models are often described by multivariate functions, which are usually approximated by a sum of lower dimensional functions. A major problem is the approximation error introduced and the factors that affect it. This paper investigates the error of approximating a multivariate function by a sum of lower dimensional functions in the setting of high dimensional model representations. Two kinds of approximations are studied, namely, the approximation based on the ANOVA (analysis of variance) decomposition and the approximation based on the anchored decomposition. We prove new theorems for the expected errors of approximations based on anchored decomposition when the anchor is chosen randomly and establish the relationship of the expected approximation errors with the global sensitivity indices of Sobol'. The expected approximation error give indications on how good or how bad could be the approximation based on anchored decomposition and when the approximation is good or bad. The influence of the anchor on the goodness of approximation is studied. Methods for choosing good anchors are presented.

1 INTRODUCTION

Mathematical models are often described by multivariate functions $f(\mathbf{x})$, where $\mathbf{x} = (x_1, \dots, x_s)$. We study representations of a multivariate function $f(\mathbf{x})$ writing as a finite sum of lower-dimensional functions

$$f(\mathbf{x}) = f_0 + \sum_{j=1}^s f_{\{j\}}(x_j) + \sum_{1 \leq i < j \leq s} f_{\{i,j\}}(x_i, x_j) + \dots + f_{\{1, \dots, s\}}(x_1, \dots, x_s),$$

where f_0 is a constant, $f_{\{j\}}(x_j)$ represents the individual contribution of the variable x_j whereas $f_{\{i,j\}}(x_i, x_j)$ represents the cooperative effects of x_i and x_j and so on. Such decompositions are called *high dimensional model representations*, see Rabitz *et al* (1999) and Sobol' (2003).

It is argued in Rabitz *et al* (1999) that quite often in high dimensional models in practice low order interactions of input variables have the main impact on the output (usually up to the third order). In many applications (for instance, in mathematical finance), a multivariate function often has small *effective dimension*, see Caflisch *et al* and Wang and Fang (2003). Roughly, this means that the functions depend mainly on a small number of variables or is nearly a sum of functions that depend on a small number of variables. A common practice is to use

$$f(\mathbf{x}) = f_0 + \sum_{m=1}^L \sum_{i_1 < \dots < i_m} f_{\{i_1, \dots, i_m\}}(x_{i_1}, \dots, x_{i_m}) \quad (1)$$

as an approximation to $f(\mathbf{x})$ (where $L < s$).

Two decompositions, namely, the ANOVA decomposition and the anchored decomposition, are popular in practice. ANOVA decomposition has some good properties and is widely used in practice. However, the ANOVA terms and their variances are difficult to compute due to the high dimensional integrals involved (which may require Monte Carlo or quasi-Monte Carlo methods). The anchored decomposition might be less attractive in theoretical aspect, but it is much easier to obtain. Due to the complexity in computing the ANOVA terms, it is suggested to use the low order terms in anchored decomposition to approximate the model function $f(\mathbf{x})$, see Rabitz *et al* (1999). A major problem here is the approximation error and the factors that affect it. The anchored decomposition contains an arbitrary reference point (called *anchor*). The choice of a "good" anchor is almost neglected in literature. It is pointed out in Sobol' (2003) that in some cases the choice of the anchor is important in order to obtain a good approximation and a careless choice may lead to unacceptable approximation.

This paper investigates the error of approximating a multivariate function $f(\mathbf{x})$ by a sum of lower dimensional functions as in (1). We compare the approximations based on ANOVA decomposition and on anchored decomposi-

tion. For the latter one, we introduce the *expected error* of approximation as a quality measure when the anchor is chosen randomly. We prove new theorems about the expected error of the approximation and establish the relationship of the expected approximation error with the global sensitivity indices. The expected error gives indications on how good or bad could be the approximation, and when and why the approximation is good or bad.

This paper is organized as follows. In Section 2, we introduce high dimensional model representations, namely, the ANOVA decomposition and the anchored decomposition. In Section 3, we study the approximation error, with focus on the expected error of anchored approximation. In Section 4, we present methods to choose good anchors. In the last section we present conclusions.

2 HIGH DIMENSIONAL MODEL REPRESENTATION

2.1 The ANOVA decomposition

For any subset $u \subseteq \{1, \dots, s\}$, let $-u$ denote the complementary set and $|u|$ denote the cardinality of u . Consider a square-integrable function $f(\mathbf{x})$ defined on $[0, 1]^s$. We say that the representation

$$f(\mathbf{x}) = \sum_{u \subseteq \{1, \dots, s\}} f_u^{\text{anova}}(\mathbf{x}) \quad (2)$$

is an ANOVA decomposition of $f(\mathbf{x})$ if

$$\int_0^1 f_u^{\text{anova}}(\mathbf{x}) dx_j = 0 \quad \text{for } j \in u. \quad (3)$$

The condition (3) uniquely define the terms in (2). Indeed, integrating both sides of (2) over $[0, 1]^s$ and using (3), we obtain that

$$f_{\emptyset}^{\text{anova}}(\mathbf{x}) = \int_{[0, 1]^s} f(\mathbf{x}) d\mathbf{x}.$$

Integrating both sides of (2) with respect to all variables except x_j and using (3), we obtain that

$$\int_{[0, 1]^{s-1}} f(\mathbf{x}) d\mathbf{x}_{-\{j\}} = f_{\emptyset}^{\text{anova}}(\mathbf{x}) + f_{\{j\}}^{\text{anova}}(\mathbf{x}),$$

thus

$$f_{\{j\}}^{\text{anova}}(\mathbf{x}) = \int_{[0, 1]^{s-1}} f(\mathbf{x}) d\mathbf{x}_{-\{j\}} - f_{\emptyset}^{\text{anova}}(\mathbf{x}).$$

Similarly, integrating both sides of (2) with respect to variables \mathbf{x}_{-u} and using (3), we obtain that

$$\int_{[0, 1]^{s-|u|}} f(\mathbf{x}) d\mathbf{x}_{-u} = \sum_{v \subset u} f_v^{\text{anova}}(\mathbf{x}), \quad (4)$$

thus we obtain the recursive formula for the ANOVA terms

$$f_u^{\text{anova}}(\mathbf{x}) = \int_{[0, 1]^{s-|u|}} f(\mathbf{x}) d\mathbf{x}_{-u} - \sum_{v \subset u} f_v^{\text{anova}}(\mathbf{x}),$$

where \subset denotes strict inclusion, \mathbf{x}_u denotes the vector comprised of the components of \mathbf{x} with j in u . The term $f_u^{\text{anova}}(\mathbf{x})$ depends only on the variable \mathbf{x}_u .

As an inverse linear relationship of (4), we have an explicit formula for the ANOVA term

$$f_u^{\text{anova}}(\mathbf{x}) = \sum_{v \subset u} (-1)^{|u|-|v|} \int_{[0, 1]^{s-|v|}} f(\mathbf{x}) d\mathbf{x}_{-v}, \quad u \subseteq \{1, \dots, s\}.$$

It follows from (3) that the ANOVA decomposition is orthogonal

$$\int_{[0, 1]^s} f_u^{\text{anova}}(\mathbf{x}) f_v^{\text{anova}}(\mathbf{x}) d\mathbf{x} = 0 \quad \text{for } u \neq v.$$

Based on this orthogonal property, we have the decomposition for the total variance

$$\sigma^2(f) = \sum_{\emptyset \neq u \subseteq \{1, \dots, s\}} \sigma_u^2(f), \quad (5)$$

where $\sigma^2(f)$ and $\sigma_u^2(f)$ are the variances of $f(\mathbf{x})$ and $f_u^{\text{anova}}(\mathbf{x})$, respectively, with \mathbf{x} being a random vector uniformly distributed on $[0, 1]^s$. Define

$$D_u = \sum_{v \subset u} \sigma_v^2(f), \quad u \subseteq \{1, \dots, s\}.$$

An inverse linear relationship of this equality is

$$\sigma_u^2(f) = \sum_{v \subset u} (-1)^{|u|-|v|} D_v, \quad u \subseteq \{1, \dots, s\}. \quad (6)$$

To measure the importance of variables, we defined the *global sensitivity indices* as

$$S_u(f) = \sigma_u^2(f) / \sigma^2(f), \quad u \subseteq \{1, \dots, s\}.$$

Global sensitivity indices can be used to identify the key variables and to fix the unessential variables, see Sobol' (2003). We may also define

$$R_\ell(f) = \frac{1}{\sigma^2(f)} \sum_{|u|=\ell} \sigma_u^2(f), \quad \ell = 1, \dots, s, \quad (7)$$

which measures the importance of all order- ℓ terms taken together. In particular, the sum of all first order global sensitivity indices $R_1(f)$ is referred as the *degree of additivity*, which measures the additive structure inherent in the function. If $R_1(f) = 1$ (or $R_1(f) \approx 1$), then the function is perfectly additive (or nearly additive). If $R_1(f) \ll 1$, then the function is dominated by higher order terms.

A good property of the ANOVA decomposition is its optimality (see Section 3). A main limitation of ANOVA decomposition is the complexity in computing the ANOVA terms, since high-dimensional integrals are involved (which may require Monte Carlo or quasi-Monte Carlo method).

2.2 The anchored decomposition

An alternative decomposition — the *anchored decomposition* — is available. Let $\mathbf{a} = (a_1, \dots, a_s) \in [0, 1]^s$ be an arbitrary point. This point is called the *anchor* or the *reference point*. Iteratively define the component terms $f_u^{\mathbf{a}}(\mathbf{x})$ as

$$f_{\emptyset}^{\mathbf{a}} = f(\mathbf{a}),$$

$$f_{\{j\}}^{\mathbf{a}}(x_j) = f(x_j, \mathbf{a}_{-\{j\}}) - f(\mathbf{a}),$$

and generally

$$f_u^{\mathbf{a}}(\mathbf{x}) = f(\mathbf{x}_u, \mathbf{a}_{-u}) - \sum_{v \subset u} f_v^{\mathbf{a}}(\mathbf{x}) \quad \text{for } \emptyset \neq u \subseteq \{1, \dots, s\},$$

where $(\mathbf{x}_u, \mathbf{a}_{-u})$ denotes an s -dimensional vector whose j th component is x_j if $j \in u$ and is a_j if $j \notin u$. For example, $(x_j, \mathbf{a}_{-\{j\}}) = (a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_s)$. Since the last term $f_{\{1, \dots, s\}}^{\mathbf{a}}(\mathbf{x})$ is determined by the difference between $f(\mathbf{x})$ and all lower order terms, we have

$$f(\mathbf{x}) = \sum_{u \subseteq \{1, \dots, s\}} f_u^{\mathbf{a}}(\mathbf{x}). \quad (8)$$

Since

$$f(\mathbf{x}_u, \mathbf{a}_{-u}) = \sum_{v \subseteq u} f_v^{\mathbf{a}}(\mathbf{x}),$$

similarly as for ANOVA terms, we have an explicit formula (inverse linear relationship) for the anchored terms

$$f_u^{\mathbf{a}}(\mathbf{x}) = \sum_{v \subseteq u} (-1)^{|u|-|v|} f(\mathbf{x}_v, \mathbf{a}_{-v}), \quad u \subseteq \{1, \dots, s\}.$$

Another explicit expression for the anchored term for functions with continuous mixed first derivatives is as follows. Let $u = \{j_1, \dots, j_m\}$, where $m = |u| > 0$, then (see

Sobol' 2003)

$$f_u^{\mathbf{a}}(\mathbf{x}) = \int_0^{h_{j_1}} \dots \int_0^{h_{j_m}} \frac{\partial^m f(\mathbf{a} + \mathbf{t})}{\partial \mathbf{x}_u} dt_{j_1} \dots dt_{j_m} \quad (9)$$

where $\mathbf{t} = (0, \dots, 0, t_{j_1}, 0, \dots, 0, t_{j_m}, 0, \dots, 0)$ and $h_i = x_i - a_i$.

The decomposition of the kind (8) is considered in Rabitz *et al* (1999) under the name *cut-HDMR*. Since it is related to the anchored function spaces (see Dick *et al*. 2004), we called it anchored decomposition.

Note that for $|u| > 0$, $f_u^{\mathbf{a}}(\mathbf{x})$ vanish when any of its own variables x_j with $j \in u$ takes the value a_j , i.e.,

$$f_u^{\mathbf{a}}(\mathbf{x}) \Big|_{x_j=a_j} = 0, \quad j \in u.$$

There exists a simple relationship between the terms of ANOVA and anchored decompositions

$$f_u^{\text{anova}}(\mathbf{x}) = \int_{[0,1]^s} f_u^{\mathbf{a}}(\mathbf{x}) d\mathbf{a}.$$

In contrast to ANOVA terms, the terms in the anchored decomposition (8) are extremely simple to compute (at least for low-order terms), since they do not necessitate the computations of high-dimensional integrals. It is natural to use the truncated anchored decomposition as an approximation of $f(\mathbf{x})$. It remains open whether such an approximation is reasonably good and when it is good.

3 THE APPROXIMATION ERRORS

For an ANOVA or anchored decomposition of a function $f(\mathbf{x})$ contains 2^s terms of components. We are interested in approximating $f(\mathbf{x})$ by the sum of the terms up to order L

$$f_L(\mathbf{x}) := f_{\emptyset} + \sum_{t=1}^L \sum_{|u|=t} f_u(\mathbf{x}_u), \quad L \leq s. \quad (10)$$

This is called the ANOVA or anchored *order- L approximation*, and is denoted by $f_L^{\text{anova}}(\mathbf{x})$ or $f_L^{\mathbf{a}}(\mathbf{x})$ (with anchor \mathbf{a}), respectively. If $L = 1$, the approximation is called the (ANOVA or anchored) *additive approximation*. A major problem is the approximation error introduced and the factors that affect the approximation error. The goodness of approximation is measured by

$$e(f, f_L) = \frac{1}{\sigma^2(f)} \int_{[0,1]^s} [f(\mathbf{x}) - f_L(\mathbf{x})]^2 d\mathbf{x}.$$

3.1 The ANOVA approximation error

For the ANOVA order-zero approximation $f_0^{\text{anova}} = I(f)$, we have $e(f, f_0^{\text{anova}}) = 1$. This serves as a benchmark of the

goodness of approximation. A good approximation should have an approximation error that is much smaller than 1.

For the ANOVA additive approximation $f_1^{\text{anova}}(\mathbf{x})$, based on the orthogonal property, we have

$$e(f, f_1^{\text{anova}}) = \frac{1}{\sigma^2(f)} \sum_{|u|>1} \sigma_u^2(f) = 1 - R_1(f). \quad (11)$$

Similar result holds for the ANOVA order- L approximation

$$e(f, f_L^{\text{anova}}) = 1 - R_1(f) - \dots - R_L(f).$$

The ANOVA order- L approximation is *optimal*. More precisely, let $h(\mathbf{x})$ be any other function of \mathbf{x} expressible as a superposition of lower-dimensional functions, each of which depends on at most L components of \mathbf{x} , then

$$e(f, h) \geq e(f, f_L^{\text{anova}}). \quad (12)$$

In fact, since $f(\mathbf{x}) - f_L^{\text{anova}}(\mathbf{x})$ only contains ANOVA terms with order larger than L , while $f_L^{\text{anova}}(\mathbf{x}) - h(\mathbf{x})$ are a sum of functions with at most L components, then from (3) we have

$$\int_{[0,1]^s} [f(\mathbf{x}) - f_L^{\text{anova}}(\mathbf{x})][f_L^{\text{anova}}(\mathbf{x}) - h(\mathbf{x})] d\mathbf{x} = 0.$$

Thus

$$\begin{aligned} \int_{[0,1]^s} [f(\mathbf{x}) - h(\mathbf{x})]^2 d\mathbf{x} &= \int_{[0,1]^s} [f(\mathbf{x}) - f_L^{\text{anova}}(\mathbf{x})]^2 d\mathbf{x} \\ &\quad + \int_{[0,1]^s} [f_L^{\text{anova}}(\mathbf{x}) - h(\mathbf{x})]^2 d\mathbf{x}, \end{aligned}$$

implying the inequality (12).

3.2 The anchored approximation error

Although the ANOVA decomposition has the optimal property, it is difficult to compute the ANOVA terms. We thus turn to anchored order- L approximation

$$f_L^{\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}) + \sum_{t=1}^L \sum_{|u|=t} f_u^{\mathbf{a}}(\mathbf{x}_u), \quad L \leq s.$$

The goodness of approximation depends on the anchor and on the order L . It is obvious from (12) that for any anchor \mathbf{a} and any order L we have

$$e(f, f_L^{\mathbf{a}}) \geq e(f, f_L^{\text{anova}}).$$

We are interested in the questions: How good could be the anchored order- L approximation? How does its goodness

depend on the anchor? How do we choose an anchor to obtain a good approximation?

A uniform error bound for anchored order- L approximation is derived in Sobol' (2003). Assume that the mixed derivatives of the form $\partial^{|\mathbf{u}|} f / \partial \mathbf{x}_u$ are piecewise continuous for $\emptyset \neq u \subseteq \{1, \dots, s\}$. Let $A_u := \sup_{\mathbf{x}} |\partial^{|\mathbf{u}|} f / \partial \mathbf{x}_u|$ and $A := \sum_{|u|>L} A_u$, then from (9) we have

$$|f(\mathbf{x}) - f_L^{\mathbf{a}}(\mathbf{x})| \leq \sum_{|u|>L} |f_u^{\mathbf{a}}(\mathbf{x})| \leq \sum_{|u|>L} A_u = A.$$

Therefore,

$$e(f, f_L^{\mathbf{a}}) \leq A^2 / \sigma^2(f).$$

A *sufficient* condition for a good anchored order- L approximation with an *arbitrary* anchor is that $A^2 \ll \sigma^2(f)$.

However, for functions which do not satisfy the conditions above the choice of anchor can be significant. To study the influence of the anchor on the goodness of anchored approximation, let the anchor \mathbf{a} be *random and uniformly distributed* over $[0, 1]^s$. We introduce the *expected error* of the anchored order- L approximation

$$\mathbf{IE}[e(f, f_L^{\mathbf{a}})].$$

This quantity measures the goodness of approximation on the average when the anchor is chosen randomly.

If the expected error $\mathbf{IE}[e(f, f_L^{\mathbf{a}})]$ is small, then we are hopefully to achieve small approximation error with a random choice of the anchor. Indeed, based on Markov inequality, for $b > 0$ we have

$$\Pr(e(f, f_L^{\mathbf{a}}) < b \mathbf{IE}[e(f, f_L^{\mathbf{a}})]) \geq 1 - \frac{1}{b}. \quad (13)$$

For example, if $\mathbf{IE}[e(f, f_L^{\mathbf{a}})] = 1/100$, then

$$\Pr\left(e(f, f_L^{\mathbf{a}}) < \frac{1}{10}\right) \geq \frac{9}{10}.$$

On the other hand, if the expected error is large, then there exists some choices of anchor such that the error of the corresponding approximation is at least as large as the expected error. Moreover, the probability of large error of anchored approximation depends on the variance $D := \text{Var}(e(f, f_L^{\mathbf{a}}))$. Based on Chebyshev inequality, we have

$$\Pr\left(|e(f, f_L^{\mathbf{a}}) - \mathbf{IE}[e(f, f_L^{\mathbf{a}})]| < \varepsilon \sqrt{D}\right) \geq 1 - \frac{1}{\varepsilon^2}.$$

It follows that

$$\Pr\left(e(f, f_L^{\mathbf{a}}) > \mathbf{IE}[e(f, f_L^{\mathbf{a}})] - \varepsilon \sqrt{D}\right) \geq 1 - \frac{1}{\varepsilon^2}. \quad (14)$$

Thus if the expected error $\mathbb{E}[e(f, f_L^{\mathbf{a}})]$ is large and the variance D is small, then there is a large probability that the error of the anchored approximation is large. For example, if $\mathbb{E}[e(f, f_L^{\mathbf{a}})] = 1 + c$ with $c > 0$ and $3\sqrt{D} < c$, then

$$\Pr(e(f, f_L^{\mathbf{a}}) > 1) \geq \Pr\left(e(f, f_L^{\mathbf{a}}) > \mathbb{E}[e(f, f_L^{\mathbf{a}})] - \varepsilon\sqrt{D}\right) \geq \frac{8}{9}.$$

3.3 The expected error of anchored order-zero approximation

We first consider the simplest case, namely, the anchored order-zero approximation

$$f_0^{\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}).$$

Since

$$\begin{aligned} e(f, f_0^{\mathbf{a}}) &= \frac{1}{\sigma^2(f)} \int_{[0,1]^s} |f(\mathbf{x}) - f(\mathbf{a})|^2 d\mathbf{x} \\ &= 1 + \frac{1}{\sigma^2(f)} [I(f) - f(\mathbf{a})]^2, \end{aligned} \quad (15)$$

the expected error of the anchored order-zero approximation is

$$\mathbb{E}[e(f, f_0^{\mathbf{a}})] = 1 + \frac{1}{\sigma^2(f)} \int_{[0,1]^s} [I(f) - f(\mathbf{a})]^2 d\mathbf{a} = 2,$$

which is exactly twice as the error of the ANOVA order-zero approximation. Thus there are some choices of anchors such that the corresponding approximation errors are at least as large as 2. This indicates that a random choice of anchor may result in poor anchored order-zero approximation. We point out that if f is continuous over $[0, 1]^s$, then there exists a point \mathbf{a}^* such that $f(\mathbf{a}^*) = I(f)$. Thus the anchor \mathbf{a}^* achieves the minimal error of the anchored order-zero approximation, which is the error of the ANOVA order-zero approximation. On the other hand, from (15) the maximal error of the anchored order-zero approximation is achieved when the anchor is chosen as to

$$\max_{\mathbf{a} \in [0,1]^s} |f(\mathbf{a}) - I(f)|.$$

Whether there is a large probability of large approximation error depends on the variance of $e(f, f_0^{\mathbf{a}})$, which is given by

$$D =: \text{Var}([e(f, f_0^{\mathbf{a}})]) = \frac{b_4}{\sigma^4(f)} - 1,$$

where $b_4 = \int_{[0,1]^s} [f(\mathbf{x}) - I(f)]^4 d\mathbf{x}$. This variance can be small or large. Based on (14), if $\sqrt{D} < 1/3$, then

$$\Pr(e(f, f_1^{\mathbf{a}}) > 1) \geq \frac{8}{9}.$$

3.4 The expected error of the anchored additive approximation

Now we consider the anchored additive approximation (i.e., the approximation (10) with $L = 1$). Let $f_1^{\mathbf{a}}(\mathbf{x})$ denote the anchored additive approximation with the anchor \mathbf{a} . From the definition of anchored decomposition, the first order terms are

$$f_{\{j\}}^{\mathbf{a}}(x_j) = f(x_j, \mathbf{a}_{-\{j\}}) - f(\mathbf{a}),$$

and the anchored additive approximation is

$$\begin{aligned} f_1^{\mathbf{a}}(\mathbf{x}) &= f(\mathbf{a}) + \sum_{j=1}^s f_{\{j\}}^{\mathbf{a}}(x_j) \\ &= -(s-1)f(\mathbf{a}) + \sum_{j=1}^s f(x_j, \mathbf{a}_{-\{j\}}). \end{aligned} \quad (16)$$

The next theorem shows how the expected error of the anchored additive approximation depends on the global sensitivity indices of various order.

Theorem 1 *Assume that the anchor \mathbf{a} is random and is uniformly distributed over $[0, 1]^s$. Then the expected error of the anchored additive approximation is*

$$\mathbb{E}[e(f, f_1^{\mathbf{a}})] = b_2 R_2(f) + \cdots + b_s R_s(f), \quad (17)$$

where

$$b_\ell = \ell^2 - \ell + 2, \quad \ell = 2, \dots, s,$$

and $R_\ell(f)$ is the sum of global sensitivity indices of order ℓ defined in (7).

Remark On the average the anchored additive approximation eliminates the variance contributions due to the first order terms. The formula (17) should be compared with the error of ANOVA additive approximation (11).

Proof We introduce several notations. For any subset $u \subseteq \{1, \dots, s\}$, let

$$D_u := \sum_{v \subseteq u} \sigma_v^2(f), \quad (18)$$

where σ_v^2 denotes the variance of the ANOVA term $f_v^{(\text{anova})}(\mathbf{x})$. Moreover, let $-\{j\}$ and $-\{i, j\}$ denote complementary set of $\{j\}$ and $\{i, j\}$, respectively.

From (10) and (16), the expected error of the anchored additive approximation is

$$e(f, f_1^{\mathbf{a}}) = \frac{1}{\sigma^2(f)} \int_{[0,1]^s} [f(\mathbf{x}) + (s-1)f(\mathbf{a}) - \sum_{j=1}^s f(x_j, \mathbf{a}_{-\{j\}})]^2 d\mathbf{x}. \quad (19)$$

Expanding the square inside the integral on the right hand side of (19) and then integrating both side with respect to the anchor \mathbf{a} over $[0, 1]^s$, we have (after some calculations)

$$\begin{aligned} \mathbb{E}[e(f, f_1^{\mathbf{a}})] &= \frac{1}{\sigma^2(f)} [(s^2 - s + 2)\sigma^2(f) \\ &- 2 \sum_{j=1}^s \sigma_{\{j\}}^2(f) - 2(s-1) \sum_{j=1}^s D_{-\{j\}} + 2 \sum_{i < j} D_{-\{i,j\}}], \end{aligned}$$

where we have used $\int_{[0,1]^{2s}} f(\mathbf{x})f(\mathbf{a})d\mathbf{x}d\mathbf{a} = [I(f)]^2$ and the following formula (see Sobol' 2001)

$$D_u = \int_{[0,1]^{2s-|u|}} f(\mathbf{x})f(\mathbf{x}_u, \mathbf{a}_{-u})d\mathbf{x}d\mathbf{a}_{-u} - [I(f)]^2.$$

In the derivation of (20), we have used this formula for $u = \{j\}, u = -\{j\}$ and $u = -\{i, j\}$, respectively; moreover, the number of $[I(f)]^2$ is exactly as required.

Since the total variance $\sigma^2(f)$ and the quantities $D_{-\{j\}}$ and $D_{-\{i,j\}}$ are all the sums of some variance terms (see (5) and (18)), we may write (20) as

$$\begin{aligned} \mathbb{E}[e(f, f_1^{\mathbf{a}})] &= \frac{1}{\sigma^2(f)} [b_1 \sum_{|u|=1} \sigma_u^2(f) \\ &+ b_2 \sum_{|u|=2} \sigma_u^2(f) + \dots + b_s \sum_{|u|=s} \sigma_u^2(f)]. \end{aligned}$$

The fact that the variance terms $\sigma_u^2(f)$ with the same order $|u| = \ell$ have the same coefficients is based on the symmetry. It remains to determine the coefficients b_ℓ . To determine b_1 , we just count the number of $\sigma_{\{1\}}^2$ contained in each sum on the right hand side of (20). We have

$$b_1 = (s^2 - s + 2) - 2 - 2(s-1)^2 + (s-1)(s-2) = 0.$$

The coefficient of $\sigma_{\{1\}}^2(f)$ and $\sigma_{\{j\}}^2(f)$ is exactly zero. Similarly, by counting the number of $\sigma_{\{1,2\}}^2(f)$ contained in each sum on the right hand side of (20), we have

$$b_2 = (s^2 - s + 2) - 2(s-1)(s-2) + (s-2)(s-3) = 4.$$

In general, we have for $\ell = 2, \dots, s$

$$\begin{aligned} b_\ell &= (s^2 - s + 2) - 2(s-1)(s-\ell) + (s-\ell)(s-\ell-1) \\ &= \ell^2 - \ell + 2. \end{aligned}$$

Thus we have

$$\mathbb{E}[e(f, f_1^{\mathbf{a}})] = b_2 R_2(f) + \dots + b_s R_s(f).$$

The theorem is proven. ■

Corollary 2 Assume that the anchor \mathbf{a} is random and is uniformly distributed over $[0, 1]^s$. Then we have

$$4e(f, f_1^{\text{anova}}) \leq \mathbb{E}[e(f, f_1^{\mathbf{a}})] \leq (s^2 - s + 2)e(f, f_1^{\text{anova}}),$$

where $e(f, f_1^{\text{anova}})$ is the error of the ANOVA additive approximation given in (11).

If the variances of ANOVA terms higher than L are zeros ($L < s$), then we have

$$4e(f, f_1^{\text{anova}}) \leq \mathbb{E}[e(f, f_1^{\mathbf{a}})] \leq (L^2 - L + 2)e(f, f_1^{\text{anova}}).$$

Theorem 1 and Corollary 2 show what determines the goodness of the anchored approximation, and answer how good and how bad could be the anchored additive approximation and when the approximation is good or bad. On the average the anchored additive approximation eliminates the variances due to the first order terms. This property is similar with that of the ANOVA additive approximation. Thus if $R_1(f)$ dominates and if $R_2(f), \dots, R_s(f)$ are all very small (especially when the higher order ones are very small), then the expected error of the anchored additive approximation can be small. Based on (13), it is likely to get good anchored additive approximation.

Note that the situation of very small higher order indices is not rare in practice. For example in financial applications the first order terms dominate the function and the higher order ones are negligible (Wang and Sloan 2005). In some cases, the dominance of the first order terms may not be encountered, but we may increase this dominance by a proper use of dimension reduction techniques (see Sobol' and Kucherenko 2005 and Wang 2006).

However, comparing with the error of the ANOVA additive approximation, the expected error of anchored additive approximation is at least 4 times large as the error of the ANOVA additive approximation. Moreover, the expected error of anchored additive approximation has a stronger dependence on higher order global sensitivity indices than on lower order ones. Thus the expected error of anchored additive approximation can easily become very large. Even when the error of the ANOVA additive approximation is small, the expected error of anchored additive approximation can be large. For example, if $s = 100, R_1(f) = 0.999, R_2(f) = R_3(f) = \dots = R_{s-1}(f) = 0, R_s(f) = 0.001$, then the error of the ANOVA additive approximation is 0.001, which is small. But the expected error of anchored additive approximation

is

$$\mathbb{E}[e(f, f_1^{\mathbf{a}})] = b_s R_s(f) = 9.902,$$

which is unacceptable. In this case, there are some choices of anchors such that the corresponding anchored additive approximations are poor. The probability of large error of anchored additive approximation depends on the variance of $e(f, f_1^{\mathbf{a}})$ (see the arguments at the end of Section 3.2).

For an anchored order- L approximation with $L > 1$, there is no simple expression for the expected approximation error in general. However, for the anchored highest order approximation and for functions with separate variables, it is still possible to derive a simple expression for the expected error of the anchored approximation.

3.5 The highest order approximation for multiplicative functions

Consider functions of the multiplicative form

$$f(\mathbf{x}) = \prod_{j=1}^s g_j(x_j), \quad (20)$$

where all $g_j(x_j)$ are square integrable. Denote

$$\mu_j := \int_0^1 g_j(x) dx, \quad \lambda_j^2 := \int_0^1 (g_j(x) - \mu_j)^2 dx.$$

For a function $f(\mathbf{x})$ of the form (20), both the ANOVA decomposition and the anchored decomposition can be found easily. Indeed, for all subset $u \subseteq \{1, \dots, s\}$, the ANOVA terms are

$$f_u^{\text{anova}}(\mathbf{x}) = \prod_{j \in u} [g_j(x_j) - \mu_j] \prod_{j \notin u} \mu_j,$$

and the terms in anchored decomposition are

$$f_u^{\mathbf{a}}(\mathbf{x}) = \prod_{j \in u} [g_j(x_j) - g_j(a_j)] \prod_{j \notin u} g_j(a_j),$$

where $\mathbf{a} = (a_1, \dots, a_s) \in [0, 1]^s$ is the anchor. Clearly, if the anchor \mathbf{a} is chosen such that

$$g_j(a_j) = \mu_j, \quad j = 1, \dots, s, \quad (21)$$

then the anchored decomposition coincides with the ANOVA decomposition. This indicates that for function of the form (20), if each $g_j(x_j)$ is continuous, then it is always possible to make the anchored decomposition being the same as the ANOVA decomposition by properly choosing the anchor.

The variance of the ANOVA term $f_u^{\text{anova}}(\mathbf{x})$ is

$$\sigma_u^2(f) = \prod_{j \in u} \lambda_j^2 \prod_{j \notin u} \mu_j^2,$$

and the total variance is

$$\sigma^2(f) = \prod_{j=1}^s (\lambda_j^2 + \mu_j^2) - \prod_{j=1}^s \mu_j^2.$$

Now we consider the highest order approximation (i.e., the approximation (10) with $L = s - 1$). The error of the ANOVA highest order approximation is

$$e(f, f_{s-1}^{\text{anova}}) = R_s(f) = \frac{1}{\sigma^2(f)} \prod_{j=1}^s \lambda_j^2,$$

where $R_\ell(f)$ is the sum of the global sensitivity indices of order ℓ . The quantity $e(f, f_{s-1}^{\text{anova}})$ is also the minimal error of the anchored highest order approximation, when the anchor \mathbf{a} is chosen to satisfy (21).

For an arbitrary anchor $\mathbf{a} \in [0, 1]^s$, the error of the anchored highest order approximation is

$$e(f, f_{s-1}^{\mathbf{a}}) = \frac{1}{\sigma^2(f)} \int_{[0,1]^2} |f(\mathbf{x}) - f_{s-1}^{\mathbf{a}}(\mathbf{x})|^2 d\mathbf{x},$$

where $f_{s-1}^{\mathbf{a}}(\mathbf{x})$ is the anchored order- $(s-1)$ approximation. Since

$$f(\mathbf{x}) - f_{s-1}^{\mathbf{a}}(\mathbf{x}) = f_{\{1, \dots, s\}}^{\mathbf{a}}(\mathbf{x}) = \prod_{j=1}^s [g_j(x_j) - g_j(a_j)],$$

it follows that

$$\begin{aligned} e(f, f_{s-1}^{\mathbf{a}}) &= \frac{1}{\sigma^2(f)} \int_{[0,1]^s} \prod_{j=1}^s [g_j(x_j) - g_j(a_j)]^2 d\mathbf{x} \\ &= \frac{1}{\sigma^2(f)} \prod_{j=1}^s [\lambda_j^2 + (g_j(a_j) - \mu_j)^2]. \end{aligned} \quad (22)$$

Let the anchor \mathbf{a} be random and uniformly distributed over $[0, 1]^s$. Then from (22) the expected error of the anchored highest order approximation is

$$\mathbb{E}[e(f, f_{s-1}^{\mathbf{a}})] = \frac{1}{\sigma^2(f)} \prod_{j=1}^s 2\lambda_j^2 = 2^s R_s(f) = 2^s e(f, f_{s-1}^{\text{anova}}).$$

Once more, on the average the anchored highest order approximation eliminates the variance contributions due to ANOVA terms up to order $s - 1$. Due to the factor 2^s , the expected error of the anchored highest order approximation can be much larger than the error of the ANOVA highest order approximation. The variance of $e(f, f_{s-1}^{\mathbf{a}})$ determines

whether we have large or small probability of approximation error which is close to the expected value. From (22) the variance of $e(f, f_{s-1}^{\mathbf{a}})$ is

$$\begin{aligned} D &:= \text{Var} [e(f, f_{s-1}^{\mathbf{a}})] \\ &= \frac{1}{\sigma^4(f)} \left[\prod_{j=1}^s [3\lambda_j^4 + b_j] - \prod_{j=1}^s 4\lambda_j^4 \right], \end{aligned}$$

where $b_j = \int_0^1 [g_j(x) - \mu_j]^4 dx$. Based on the Chebyshev inequality, we have for any $\varepsilon > 0$

$$\Pr \left(|e(f, f_{s-1}^{\mathbf{a}}) - \mathbb{E}[e(f, f_{s-1}^{\mathbf{a}})]| < \varepsilon \sqrt{D} \right) \geq 1 - \frac{1}{\varepsilon^2}.$$

To see how bad the anchored highest order approximation could be, consider an extreme case. From (22), if the anchor $\mathbf{a} = (a_1, \dots, a_s)$ is chosen such that a_j solves the optimization problem

$$\max_{a_j \in [0,1]} (g_j(a_j) - \mu_j)^2 := c_j^2, \quad j = 1, \dots, s, \quad (23)$$

then the error of the anchored highest order approximation achieves the maximal value. The maximal error is

$$\max_{\mathbf{a} \in [0,1]^s} e(f, f_{s-1}^{\mathbf{a}}) = \frac{1}{\sigma^2(f)} \prod_{j=1}^s [\lambda_j^2 + c_j^2] = GR_s(f),$$

where $G := \prod_{j=1}^s [1 + c_j^2 \lambda_j^{-2}]$. From (23), it follows that $c_j^2 \geq \lambda_j^2$, thus $G \geq 2^s$.

Summarizing the results above we have the following.

Theorem 3 *Assume that the function $f(\mathbf{x})$ has the multiplicative form (20). Let the anchor \mathbf{a} be random and uniformly distributed over $[0, 1]^s$, then the expected error of the anchored highest order approximation is*

$$\mathbb{E} [e(f, f_{s-1}^{\mathbf{a}})] = 2^s R_s(f) = 2^s e(f, f_{s-1}^{\text{anova}}).$$

The minimal error of the anchored highest order approximation is

$$\min_{\mathbf{a} \in [0,1]^s} e(f, f_{s-1}^{\mathbf{a}}) = \frac{1}{\sigma^2(f)} \prod_{j=1}^s \lambda_j^2 = R_s(f) = e(f, f_{s-1}^{\text{anova}}).$$

which is achieved when the anchor is chosen to satisfy (21).

The maximal error of the anchored highest order approximation is

$$\max_{\mathbf{a} \in [0,1]^s} e(f, f_{s-1}^{\mathbf{a}}) = \frac{1}{\sigma^2(f)} \prod_{j=1}^s [\lambda_j^2 + c_j^2] = G e(f, f_{s-1}^{\text{anova}}).$$

which is achieved when the anchor is chosen to solve (23), where $G = \prod_{j=1}^s [1 + c_j^2 \lambda_j^{-2}]$.

From Theorem 3, for any anchor $\mathbf{a} \in [0, 1]^s$, we have

$$e(f, f_{s-1}^{\text{anova}}) \leq e(f, f_{s-1}^{\mathbf{a}}) \leq G e(f, f_{s-1}^{\text{anova}}).$$

Both the equalities hold with some particular choices of the anchors. The error of anchored approximation $e(f, f_{s-1}^{\mathbf{a}})$ varies from $R_s(f)$ to $GR_s(f)$, depending on the anchor (with the expected error to be $2^s R_s(f)$). Whether the expected error is small (say, less than 1) depends on whether the error of the ANOVA highest order approximation is very small (say less than $1/2^s$). Clearly, if $G e(f, f_{s-1}^{\text{anova}})$ is small, then any choice of anchor leads to small error of anchored highest order approximation; if $e(f, f_{s-1}^{\text{anova}})$ is large, then any choice of anchor leads to large error of anchored highest order approximation. In other cases, the expected error and the variance determine the probability of large or small approximation error.

4 FINDING A GOOD ANCHOR

For an *arbitrary* anchor $\mathbf{a} \in [0, 1]^s$, the anchored decomposition (8) is always an *exact* representation of $f(\mathbf{x})$, but the quality of the order- L approximation depends on the anchor and the approximation error can be quite different. In some cases the choice of the anchor becomes important. A “bad” choice of the anchor may lead to a large approximation error. We hope to choose a suitable anchor such that the error of the anchored order- L approximation is as small as possible. This leads to a general principle: finding an anchor \mathbf{a} such that the error of the anchored order- L approximation is minimized, i.e., solving the optimization problem

$$\min_{\mathbf{a} \in [0,1]^s} e(f, f_L^{\mathbf{a}}).$$

This general principle includes the method of Sobol’ (2003) as a special case.

Method A. For $L = 0$, the anchored order-zero approximation is $f_0^{\mathbf{a}} = f(\mathbf{a})$. In this case, we need to find the minimizer for

$$\min_{\mathbf{a} \in [0,1]^s} \int_{[0,1]^s} [f(\mathbf{x}) - f(\mathbf{a})]^2 d\mathbf{x}. \quad (24)$$

Since

$$\begin{aligned} &\int_{[0,1]^s} [f(\mathbf{x}) - f(\mathbf{a})]^2 d\mathbf{x} = \\ &= \int_{[0,1]^s} [f(\mathbf{x}) - I(f)]^2 d\mathbf{x} + [I(f) - f(\mathbf{a})]^2, \end{aligned}$$

The first term on the right-hand side is independent of the anchor, thus the optimization problem (24) is equivalent to

$$\min_{\mathbf{a} \in [0,1]^s} |f(\mathbf{a}) - I(f)|.$$

That is to say, the anchor \mathbf{a}^* should be chosen such that

$$f(\mathbf{a}^*) \approx I(f).$$

This provides a justification for the suggestion of Sobol' (2003). The searching process can be proceeded as follows:

(i) Generate a low discrepancy point set $\mathcal{P}_n := \{\mathbf{x}_j \in [0, 1]^s, j = 1, \dots, n\}$ and calculate a rough estimate for $I(f)$, i.e., $\hat{I}(f) := \frac{1}{n} \sum_{j=1}^n f(\mathbf{x}_j)$.

(ii) Select a point \mathbf{a}^* from \mathcal{P}_n as the anchor, such that

$$|f(\mathbf{a}^*) - \hat{I}(f)| = \min_{1 \leq j \leq n} |f(\mathbf{x}_j) - \hat{I}(f)|.$$

Method B. For an order $L > 0$ (say $L = 1$ or 2), we find the anchor \mathbf{a} such that the error of the anchored order- L approximation is minimized, i.e., to find the minimizer

$$\min_{\mathbf{a} \in [0,1]^s} e(f, f_L^{\mathbf{a}}).$$

This can be achieved similarly as in Method A:

(i) Generate a low discrepancy point set $\mathcal{P}_n := \{\mathbf{x}_j \in [0, 1]^s, j = 1, \dots, n\}$ and calculate the error of the anchored order- L approximation $e(f, f_L^{\mathbf{a}})$. The involved integrals in $e(f, f_L^{\mathbf{a}})$ can be calculated numerically by quasi-Monte Carlo.

(ii) Select a point \mathbf{a}^* from the set \mathcal{P}_n , such that the error of the anchored order- L approximation is minimized, i.e.,

$$e(f, f_L^{\mathbf{a}^*}) = \min_{\mathbf{a} \in \mathcal{P}_n} e(f, f_L^{\mathbf{a}}).$$

The comparisons of the errors of anchored approximations with good or bad choices of anchors for security pricing problems are presented in Wang (2007). It turns out that with a good choice of anchor, the anchored *additive* approximation is quite satisfactory. Such a good anchored additive approximation can be used to construct good importance density or control variate to increase the efficiency of quasi-Monte Carlo methods for security pricing problems (see Wang 2007).

5 CONCLUDING REMARKS

The approximation of a function by a sum of lower dimensional functions is an important problem in the theory and applications of high dimensional model representation. We studied and compared the error of approximating a function by the sum of the low order ANOVA or anchored terms

in the settings of ANOVA decomposition and anchored decomposition. We studied the expected error of the anchored approximation when the anchor is chosen randomly. We proved new theorems about the expected error of the anchored additive approximation and the anchored highest order approximation. In particular, we show that on the average the anchored additive approximation eliminates the variance contributions due to the first order terms. These theorems indicate the usefulness of an anchored approximation, but it should be used with care. We presented procedures to find a good anchor.

ACKNOWLEDGMENTS

The support of the National Science Foundation of China is gratefully acknowledged.

REFERENCES

- Cafflisch, R. E., W. Morokoff, A. Owen. 1997. Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance*, **1**, 27-46.
- Dick J., I. H. Sloan, X. Wang and H. Woźniakowski. 2004. Liberating the weights. *J. Complexity*, **20**, 593-623.
- Rabitz, H., O. F. Alis, J. Shorter, K. Shim. 1999. Efficient input-output model representation. *Comput. Phys. Commun.*, **117**, 11-20.
- Sobol', I. M. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulation*, **55**, 271-280.
- Sobol', I. M. 2003. Theorems and examples on high dimensional model representation. *Reliability Engineering and System Safety*. **79**, 187-193.
- Sobol', I. M., S. S. Kucherenko. 2005. On global sensitivity analysis of quasi-Monte Carlo algorithms. *Monte Carlo Methods Appl.*, **11**, 83-92.
- Wang, X. 2006. On the effects of dimension reduction techniques on some high-dimensional problems in finance. *Oper. Res.*, **54**, 1063-1078.
- Wang, X. 2007. High dimensional model representations in quasi-Monte carlo methods for security pricing. Working paper.
- Wang, X., K.-T. Fang. 2003. The effective dimension and quasi-Monte Carlo integration. *J. Complexity*, **19**, 101-124.
- Wang, X., I. H. Sloan. 2005. Why are high-dimensional finance problems often of low effective dimension?. *SIAM J. Sci. Comput.*, **27**, 159-183.

AUTHOR BIOGRAPHY

XIAOQUN WANG is a Professor of Department of Mathematical Sciences, Tsinghua University, Beijing,

Wang

China. His research interests include computational finance and financial engineering, Monte Carlo and quasi-Monte Carlo methods, information-based complexity. His email address is <xwang@math.tsinghua.edu.cn>.