MATHEMATICS OF COMPUTATION
Volume 79, Number 270, April 2010, Pages 953–966
S 0025-5718(09)02319-9
Article electronically published on November 20, 2009

ON DECOMPOSITIONS OF MULTIVARIATE FUNCTIONS

F. Y. KUO, I. H. SLOAN, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI

ABSTRACT. We present formulas that allow us to decompose a function f of d variables into a sum of 2^d terms $f_{\mathbf{u}}$ indexed by subsets $\mathbf{u}$ of $\{1,\ldots,d\}$, where each term $f_{\mathbf{u}}$ depends only on the variables with indices in $\mathbf{u}$. The decomposition depends on the choice of d commuting projections $\{P_j\}_{j=1}^d$, where $P_j(f)$ does not depend on the variable x_j. We present an explicit formula for $f_{\mathbf{u}}$, which is new even for the ANOVA and anchored decompositions; both are special cases of the general decomposition. We show that the decomposition is minimal in the following sense: if f is expressible as a sum in which there is no term that depends on all of the variables indexed by the subset $\mathbf{z}$, then, for every choice of $\{P_j\}_{j=1}^d$, the terms $f_{\mathbf{u}} = 0$ for all subsets $\mathbf{u}$ containing $\mathbf{z}$. Furthermore, in a reproducing kernel Hilbert space setting, we give sufficient conditions for the terms $f_{\mathbf{u}}$ to be mutually orthogonal.

1. INTRODUCTION

Functions with very large numbers of variables arise nowadays in many contexts, ranging from molecular chemistry to probability and statistics and the pricing of financial options. Often there is a need to approximate such functions by a small number of terms, either concretely (as in the case of "model reduction" in molecular chemistry), or notionally (as in modern approaches to multivariate integration, which seek to exploit the existence of low-dimensional approximations without actually constructing them).

To add to the existing toolkit for the study of low-dimensional approximations, we introduce in this paper general decomposition formulas of functions of d variables into sums of 2^d terms, with each term depending on a group of variables indexed by a particular subset of $\{1,\ldots,d\}$. That is, the decomposition is of the form

$$f = \sum_{\mathbf{u}\subseteq\{1,\ldots,d\}} f_{\mathbf{u}},$$

where the summation is over all subsets of $\{1,\ldots,d\}$, and $f_{\mathbf{u}}$ depends only on the subset of variables $\{x_j : j \in \mathbf{u}\}$.

One important example of such a decomposition formula is the ANOVA decomposition; see, e.g., [2, 5, 14, 18, 21]. Another is the *anchored* decomposition; see, e.g., [5, 7, 8, 15, 19]. Both have already been discussed in Sobol′'s 1969 book [17]. They are both included here as special cases. The results obtained here for the general decomposition include not only those that are well known in special cases, such as the recursive form of the ANOVA decomposition, but also results that are new.

Received by the editor May 2, 2008 and, in revised form, October 16, 2008 and February 12, 2009.
2000 *Mathematics Subject Classification.* Primary 41A63, 41A99.

©2009 American Mathematical Society

These include an explicit expression for the general term in the decomposition, and a certain "minimal" property that applies to all decompositions in the class. This new explicit formula may turn out to be very useful for both theoretical analysis and practical computation. In its simplest form the minimal property states that if there exists a (possibly unknown) decomposition of the function in which there is no term involving one particular variable, then the decomposition (whether ANOVA or anchored or any other in the class) never unnecessarily introduces that variable. (As a simple example with $d = 2$, the function $f(x_1, x_2) := 2x_1$ can obviously be written in many non-minimal ways, such as $f(x_1, x_2) = x_1 - x_2 + (x_1 + x_2)$. Our decompositions never commit a crime of this kind.) More generally, if there is a decomposition in which no term contains all members of some subset of the variables (say x_1, x_2, x_4), then in our decomposition every term that involves all variables in that subset vanishes.

The general decomposition depends on a set of commuting projections $\{P_j\}_{j=1}^d$, where P_j acting on a multivariate function has the effect of eliminating the variable x_j in some way, while leaving all other variables unchanged. (For example, in the anchored case, P_j freezes x_j at some specified value c_j. In the ANOVA case, P_j integrates with respect to x_j.) We show that, for a given set of projections $\{P_j\}_{j=1}^d$, our decomposition is the unique decomposition of the form $f = \sum_{\mathbf{u}\subseteq\{1,\ldots,d\}} f_{\mathbf{u}}$ with the property $P_j(f_{\mathbf{u}}) = 0$ whenever $j \in \mathbf{u}$. The projections provide a useful tool for the proof of various properties, such as the minimal property mentioned above.

We also show that decompositions defined by the commuting projections are orthogonal decompositions in certain reproducing kernel Hilbert spaces. The orthogonality property is a key to the successful theoretical and practical development of integration and approximation schemes in high-dimensional reproducing kernel Hilbert spaces. Since the orthogonality is with respect to the inner product of a Hilbert space, it is clear that orthogonality can only be achieved if there is a correct association between the set of commuting projections and a particular reproducing kernel Hilbert space.

The general decomposition makes an appearance in the field of computational chemistry under the heading *high-dimensional model representation* (*HDMR*). See, e.g., the papers [11, 12, 13] and the references therein. In particular, their *cut-HDMR* is essentially our anchored decomposition. In the recent survey article [5], there is also a section discussing ANOVA-like decompositions. These decompositions are for product measures with densities, and for the Dirac measures, they become anchored decompositions.

The paper is organized as follows. The class of decompositions and their properties are presented in Section 2. Their minimal property is shown in Section 3, and orthogonality in reproducing kernel Hilbert spaces is considered in Section 4. As examples we consider two Sobolev spaces (the anchored and unanchored Sobolev spaces) both with arbitrary smoothness, and we show that the corresponding orthogonal decompositions are precisely the anchored and the ANOVA decompositions, respectively. Section 5 focuses on the computational cost of the anchored decomposition. The final section contains a short discussion about potential applications.

A more elaborate version of this paper, including a derivation of (3.2) and of the unanchored Sobolev space,[1] can be found in `www.maths.unsw.edu.au/~fkuo/pubs/`.

2. A general decomposition formula

Let $\mathcal{F}$ be a linear space of real functions $f(\boldsymbol{x}) = f(x_1, \ldots, x_d)$ defined on a domain $D \subseteq \mathbb{R}^d$. We introduce $[1..d] := \{1, \ldots, d\}$ as a shorthand notation for the set of indices from 1 to d, and we say that a function $f \in \mathcal{F}$ *does not depend on* x_j for some $j \in [1..d]$ if

$$\boldsymbol{x}_{[1..d]\setminus\{j\}} = \boldsymbol{y}_{[1..d]\setminus\{j\}} \quad \Longrightarrow \quad f(\boldsymbol{x}) = f(\boldsymbol{y}) \qquad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in D.$$

Here and elsewhere for a given $\boldsymbol{x} = (x_1, \ldots, x_d)$ we use the notation $\boldsymbol{x}_{\mathbf{u}} = (x_j)_{j\in\mathbf{u}}$.

Let $\{P_j\}_{j=1}^d$ be a set of commuting projections on $\mathcal{F}$ with the following property:

$$P_j(f) = f \text{ if } f \text{ does not depend on } x_j, \text{ and } P_j(f) \text{ does not depend on } x_j. \tag{2.1}$$

Intuitively, we can think of P_j as acting only on the jth variable x_j to produce a result that does not depend on x_j, and if P_j acts on a function that is independent of x_j it leaves the function unchanged.

For $\mathbf{u} \subseteq [1..d]$, we define $P_{\mathbf{u}} := \prod_{j\in\mathbf{u}} P_j$, with $P_\emptyset := I$, the identity operator. Since the commuting projections in $P_{\mathbf{u}}$ can be applied in any order and since $P_j P_j = P_j$, we have $P_{\mathbf{u}} P_{\mathbf{v}} = P_{\mathbf{u}\cup\mathbf{v}}$ for all $\mathbf{u}, \mathbf{v} \subseteq [1..d]$.

Formally, for each $\mathbf{u} \subseteq [1..d]$ we can define $\mathcal{F}_{\mathbf{u}} := P_{[1..d]\setminus\mathbf{u}}(\mathcal{F})$. Then $\mathcal{F}_{\mathbf{u}}$ is a subspace of $\mathcal{F}$ and consists of functions which depend only on the variables listed in $\mathbf{u}$, i.e., on $\boldsymbol{x}_{\mathbf{u}} = (x_j)_{j\in\mathbf{u}}$, and which are constant with respect to the variables outside $\mathbf{u}$. For an arbitrary $g_{\mathbf{u}} \in \mathcal{F}_{\mathbf{u}}$, it is sometimes useful to ignore the uninteresting variables, by writing $g_{\mathbf{u}}(\boldsymbol{x}) = g_{\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}})$. We have $\mathcal{F}_{[1..d]} = \mathcal{F}$, and $\mathcal{F}_\emptyset$ contains only constant functions. Moreover, $\mathcal{F}_{\mathbf{v}} \subseteq \mathcal{F}_{\mathbf{u}}$ if $\mathbf{v} \subseteq \mathbf{u}$.

The commuting projections $\{P_j\}_{j=1}^d$ can be very general as long as each P_j satisfies (2.1). We have in mind two obvious examples:

- Freezing x_j at c_j:

$$P_j(f)(\boldsymbol{x}) \;=\; f(x_1, \ldots, x_{j-1}, c_j, x_{j+1}, \ldots, x_d) \quad \text{for} \quad \boldsymbol{x} \in D. \tag{2.2}$$

- With $D = [0,1]^d$, integrating out x_j:

$$P_j(f)(\boldsymbol{x}) \;=\; \int_0^1 f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_d)\,\mathrm{d}t \quad \text{for} \quad \boldsymbol{x} \in D. \tag{2.3}$$

Depending on the choice of the projections P_j, there are some implicit assumptions on the domain D and the linear space $\mathcal{F}$. In the first example we need to assume that for all $j \in [1..d]$ and all $\boldsymbol{x} \in D$ we have $(x_1, \ldots, x_{j-1}, c_j, x_{j+1}, \ldots, x_d) \in D$. In the second example, where $D = [0,1]^d$, we require f to be integrable on $[0,1]^d$. Other obvious examples include a weighted sum $P_j(f)(\boldsymbol{x}) = \sum_{m=1}^{M_j} w_{j,m} f(x_1, \ldots, x_{j-1}, c_{j,m}, x_{j+1}, \ldots, x_d)$, with $\sum_{m=1}^{M_j} w_{j,m} = 1$, and a weighted integral $P_j(f)(\boldsymbol{x}) = \int_0^1 f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_d)\,\rho_j(t)\,\mathrm{d}t$, with $\int_0^1 \rho_j(t)\,\mathrm{d}t = 1$. There is no requirement that the d projections have to be of the same form.

[1] This space is derived under the condition that its periodic subspace coincides with the weighted Korobov space of the same smoothness.

We are now ready to introduce the general decomposition formula. There are multiple ways to present the same theorem. Here we take the most elegant approach, which was brought to our attention by Griebel; see also [5, 11]. Let a_j and b_j be commuting linear operators from $\mathcal{F}$ to $\mathcal{F}$. Using the simple property

$$\prod_{j\in\mathbf{u}}(a_j+b_j) = \sum_{\mathbf{v}\subseteq\mathbf{u}}\Big(\prod_{j\in\mathbf{v}}a_j\Big)\Big(\prod_{j\in\mathbf{u}\setminus\mathbf{v}}b_j\Big), \tag{2.4}$$

we can express the identity operator I as

$$I = \prod_{j=1}^{d}[(I-P_j)+P_j] = \sum_{\mathbf{u}\subseteq[1..d]}\Big(\prod_{j\in\mathbf{u}}(I-P_j)\Big)\Big(\prod_{j\in[1..d]\setminus\mathbf{u}}P_j\Big). \tag{2.5}$$

Theorem 2.1. *Let $\{P_j\}_{j=1}^d$ be commuting projections on $\mathcal{F}$ satisfying* (2.1)*. For every function $f\in\mathcal{F}$ and each subset $\mathbf{u}\subseteq[1..d]$, we define*

$$f_{\mathbf{u}} := \Big(\prod_{j\in\mathbf{u}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(f). \tag{2.6}$$

This leads to a decomposition of f given by

$$f = \sum_{\mathbf{u}\subseteq[1..d]} f_{\mathbf{u}}, \tag{2.7}$$

where $f_{\mathbf{u}}$ depends only on variables with indices in $\mathbf{u}$.

(a) *Define the functions $g_{\mathbf{u}}$ recursively with respect to the cardinality of $\mathbf{u}\subseteq[1..d]$ by*

$$g_\emptyset := P_{[1..d]}(f), \qquad g_{\mathbf{u}} := P_{[1..d]\setminus\mathbf{u}}(f) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} g_{\mathbf{v}}.$$

Then $f_{\mathbf{u}} = g_{\mathbf{u}}$ for all $\mathbf{u}\subseteq[1..d]$.

(b) *Define the functions $h_{\mathbf{u}}$ for $\mathbf{u}\subseteq[1..d]$ by*

$$h_{\mathbf{u}} := \sum_{\mathbf{v}\subseteq\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}P_{[1..d]\setminus\mathbf{v}}(f).$$

Then $f_{\mathbf{u}} = h_{\mathbf{u}}$ for all $\mathbf{u}\subseteq[1..d]$.

(c) *Assume that $f=\sum_{\mathbf{u}\subseteq[1..d]} w_{\mathbf{u}}$, where $w_{\mathbf{u}}$ depends only on the variables with indices in the set $\mathbf{u}$ and satisfies the "annihilating" property*

$$P_j(w_{\mathbf{u}}) = 0 \quad \text{for all} \quad j\in\mathbf{u}. \tag{2.8}$$

Then $f_{\mathbf{u}} = w_{\mathbf{u}}$ for all $\mathbf{u}\subseteq[1..d]$. In other words, (2.7) *is the unique decomposition of this form with the annihilating property.*

Proof. The fact that the functions $f_{\mathbf{u}}$ defined by (2.6) yield (2.7) follows immediately from (2.5). We now proceed to prove parts (a), (b), and (c).

It is clear from (2.6) that $f_\emptyset = P_{[1..d]}(f) = g_\emptyset$. For $\mathbf{u}\neq\emptyset$ we can rewrite (2.6) as

$$\begin{aligned} f_{\mathbf{u}} &= P_{[1..d]\setminus\mathbf{u}}(f) - \Big(I - \prod_{j\in\mathbf{u}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(f) \\ &= P_{[1..d]\setminus\mathbf{u}}(f) - \Big(\prod_{j\in\mathbf{u}}[(I-P_j)+P_j] - \prod_{j\in\mathbf{u}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(f). \end{aligned}$$

We use (2.4) for the first product and obtain

$$\begin{aligned} f_{\mathbf{u}} &= P_{[1..d]\setminus\mathbf{u}}(f) - \Big(\sum_{\mathbf{v}\subsetneq\mathbf{u}}\Big(\prod_{j\in\mathbf{v}}(I-P_j)\Big)P_{\mathbf{u}\setminus\mathbf{v}}\Big)P_{[1..d]\setminus\mathbf{u}}(f) \\ &= P_{[1..d]\setminus\mathbf{u}}(f) - \sum_{\mathbf{v}\subsetneq\mathbf{u}}\Big(\prod_{j\in\mathbf{v}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{v}}(f) \;=\; P_{[1..d]\setminus\mathbf{u}}(f) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v}}, \end{aligned}$$

where we used the definition (2.6) with $\mathbf{u}$ replaced by $\mathbf{v}$. This proves that $f_{\mathbf{u}}$ satisfies the same recurrence as $g_{\mathbf{u}}$. Since $f_\emptyset = g_\emptyset$, we conclude that $f_{\mathbf{u}} = g_{\mathbf{u}}$ for all $\mathbf{u}\subseteq[1..d]$.

To prove that $f_{\mathbf{u}} = h_{\mathbf{u}}$ for all $\mathbf{u}\subseteq[1..d]$, we apply (2.4) to the first product in (2.6):

$$\begin{aligned} f_{\mathbf{u}} &= \Big(\sum_{\mathbf{v}\subseteq\mathbf{u}}\Big(\prod_{j\in\mathbf{v}} I\Big)\Big(\prod_{j\in\mathbf{u}\setminus\mathbf{v}}(-P_j)\Big)\Big)P_{[1..d]\setminus\mathbf{u}}(f) \\ &= \sum_{\mathbf{v}\subseteq\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}P_{[1..d]\setminus\mathbf{v}}(f) \;=\; h_{\mathbf{u}}. \end{aligned}$$

Finally, for $f = \sum_{\mathbf{v}\subseteq[1..d]} w_{\mathbf{v}}$ as given in (c), we have from (2.6) that

$$f_{\mathbf{u}} \;=\; \sum_{\mathbf{v}\subseteq[1..d]}\Big(\prod_{j\in\mathbf{u}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(w_{\mathbf{v}}).$$

If $\mathbf{v}\neq\mathbf{u}$, then either there exists an index $k\in\mathbf{v}$ and $k\notin\mathbf{u}$, or there exists an index $k\in\mathbf{u}$ and $k\notin\mathbf{v}$. In the former case, we have from (2.8) that $P_k(w_{\mathbf{v}}) = 0$; in the latter case, we have $(I-P_k)(w_{\mathbf{v}}) = 0$. Hence only the $\mathbf{v}=\mathbf{u}$ term survives, and we have

$$f_{\mathbf{u}} \;=\; \Big(\prod_{j\in\mathbf{u}}(I-P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(w_{\mathbf{u}}) \;=\; w_{\mathbf{u}}.$$

This completes the proof. □

It is clear from the theorem that, instead of defining the decomposition through (2.6), we could use any of the parts (a), (b), or (c) as the definition for $f_{\mathbf{u}}$.

We now present two specific examples of the decomposition formula: the first requires integrals of f, while the second requires only function evaluations of f.

Example 2.2 (ANOVA Decomposition)**.** Taking $D=[0,1]^d$ and $\mathcal{F}$ as the space of square integrable functions, and choosing each P_j according to (2.3), we recover the well-known ANOVA *decomposition* (analysis of variance); see, e.g., [2, 5, 14, 18, 21]. This has the form

$$f \;=\; \sum_{\mathbf{u}\subseteq[1,d]} f_{\mathbf{u},*}\,, \tag{2.9}$$

with $f_{\mathbf{u},*}$ depending only on the variables listed in $\mathbf{u}$ and satisfying

(a) $f_{\emptyset,*} \;=\; \displaystyle\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},\;\; f_{\mathbf{u},*}(\boldsymbol{x}_{\mathbf{u}}) \;=\; \int_{[0,1]^{d-|\mathbf{u}|}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{[1..d]\setminus\mathbf{u}} - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v},*}(\boldsymbol{x}_{\mathbf{v}}),$

(b) $f_{\mathbf{u},*}(\boldsymbol{x}_{\mathbf{u}}) \;=\; \displaystyle\sum_{\mathbf{v}\subseteq\mathbf{u}}(-1)^{|\mathbf{u}|-|\mathbf{v}|}\int_{[0,1]^{d-|\mathbf{v}|}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{[1..d]\setminus\mathbf{v}},$

(c) $\displaystyle\int_0^1 f_{\mathbf{u},*}(\boldsymbol{x}_{\mathbf{u}})\,\mathrm{d}x_j \;=\; 0 \quad \text{for all} \quad j\in\mathbf{u}.$

Here $\mathrm{d}\boldsymbol{x}_{[1..d]\setminus\mathbf{u}}$ denotes $\prod_{j\in[1..d]\setminus\mathbf{u}} \mathrm{d}x_j$. For $\mathbf{u} = [1..d]$, the integral over the empty set is replaced by $f(\boldsymbol{x})$. In the literature, the recursive form (a) of the ANOVA terms is typically used as the definition of ANOVA decomposition. However, Sobol′ used the annihilating property (c) as the definition of the ANOVA terms; see, e.g., [18]. The explicit form (b) obtained here is new. The ANOVA decomposition might be difficult to use in practice because it requires integrals of f which are typically unavailable.

Example 2.3 (Anchored Decomposition)**.** Let $\boldsymbol{c} \in D$, and let $(\boldsymbol{x};\boldsymbol{c})_{\mathbf{u}}$ denote the d-dimensional vector whose jth component is x_j if $j \in \mathbf{u}$ and c_j if $j \notin \mathbf{u}$. Assuming that $(\boldsymbol{x};\boldsymbol{c})_{\mathbf{u}} \in D$ whenever $\boldsymbol{x} \in D$, and choosing each P_j according to (2.2), we obtain the *anchored decomposition with respect to the anchor* $\boldsymbol{c}$. This has the form

$$f = \sum_{\mathbf{u}\subseteq[1..d]} f_{\mathbf{u},\boldsymbol{c}}\,, \tag{2.10}$$

with $f_{\mathbf{u},\boldsymbol{c}}$ depending only on the variables listed in $\mathbf{u}$ and satisfying

(a) $f_{\emptyset,\boldsymbol{c}} = f(\boldsymbol{c}), \quad f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_{\mathbf{u}}) = f((\boldsymbol{x};\boldsymbol{c})_{\mathbf{u}}) - \sum_{\mathbf{v}\subsetneq\mathbf{u}} f_{\mathbf{v},\boldsymbol{c}}(\boldsymbol{x}_{\mathbf{v}}),$

(b) $f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_{\mathbf{u}}) = \sum_{\mathbf{v}\subseteq\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} f((\boldsymbol{x};\boldsymbol{c})_{\mathbf{v}}),$

(c) $f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_{\mathbf{u}}) = 0$ whenever $x_j = c_j$ for all $j \in \mathbf{u}$.

The recursive form (a) of the anchored decomposition has already appeared in earlier literature; see, e.g., [7, 8]. The explicit form (b) is new. This is attractive because it involves only function evaluations of f with respect to the anchor $\boldsymbol{c}$; see Section 5 for a discussion on computational cost. We add that, in some specific setting, the terms $f_{\mathbf{u},\boldsymbol{c}}$ from the anchored decomposition can also be expressed as integrals of the mixed first partial derivatives of f; see [7, Proposition 1]. This is useful for some theoretical error analysis.

3. The decomposition is minimal

In this section, we prove that the decomposition from Theorem 2.1 with respect to any choice of commuting projections $\{P_j\}_{j=1}^d$ is minimal in the sense of *never introducing unnecessary terms*.

In general, there are infinitely many ways to express a function f of d variables $\boldsymbol{x} = (x_1, \ldots, x_d)$ in the form

$$f = \sum_{\mathbf{u}\subseteq[1..d]} t_{\mathbf{u}}, \tag{3.1}$$

with $t_{\mathbf{u}}$ depending only on the variables listed in $\mathbf{u}$, i.e., on $\boldsymbol{x}_{\mathbf{u}} = (x_j)_{j\in\mathbf{u}}$. With respect to a particular decomposition (3.1), it is common to refer to the terms $\{t_{\mathbf{u}} : |\mathbf{u}| = \ell\}$ collectively as the "order-ℓ terms". Similarly, in the discussion below, we shall refer to the terms $\{t_{\mathbf{u}} : \mathbf{z} \subseteq \mathbf{u} \subseteq [1..d]\}$ for a given $\mathbf{z} \subseteq [1..d]$ as the "super-$\mathbf{z}$ terms". In other words, the super-$\mathbf{z}$ terms are all the terms $t_{\mathbf{u}}$ for which $\mathbf{u}$ contains $\mathbf{z}$.

The minimal property of the decomposition from Theorem 2.1 is stated below.

Theorem 3.1. *Let f be a function from $\mathcal{F}$ and let $\mathbf{z}$ be a given subset of* $[1..d]$. *Assume that there exists a decomposition* (3.1) *of f in which all super-$\mathbf{z}$ terms are*

zero, i.e.,

$$t_{\mathbf{u}} = 0 \text{ for all } \mathbf{u} \text{ containing } \mathbf{z}.$$

Then the decomposition from Theorem 2.1 *also has all super-*$\mathbf{z}$ *terms equal to zero, i.e.,*

$$f_{\mathbf{u}} = 0 \text{ for all } \mathbf{u} \text{ containing } \mathbf{z}.$$

Proof. Suppose that $f = \sum_{\mathbf{w}\subseteq[1..d]} t_{\mathbf{w}}$ with $t_{\mathbf{w}}$ depending only on the variables listed in $\mathbf{w}$, and with $t_{\mathbf{w}} = 0$ for all $\mathbf{w}$ containing $\mathbf{z}$. For each $\mathbf{u}$ satisfying $\mathbf{z} \subseteq \mathbf{u} \subseteq [1..d]$, we want to prove that $f_{\mathbf{u}} = 0$ in the decomposition in Theorem 2.1. We see from (2.6) that

$$f_{\mathbf{u}} = \Big(\prod_{j\in\mathbf{u}}(I - P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(f) = \sum_{\mathbf{w}\subseteq[1..d]}\Big(\prod_{j\in\mathbf{u}}(I - P_j)\Big)P_{[1..d]\setminus\mathbf{u}}(t_{\mathbf{w}}).$$

Note that for $\mathbf{u} \subseteq \mathbf{w}$ we have $\mathbf{z} \subseteq \mathbf{w}$, and it follows from the assumption on $t_{\mathbf{w}}$ that $t_{\mathbf{w}} = 0$. For $\mathbf{u} \not\subseteq \mathbf{w}$, there exists an index $k \in \mathbf{u}$ and $k \notin \mathbf{w}$, and for this index k we have $(I - P_k)(t_{\mathbf{w}}) = 0$ since $t_{\mathbf{w}}$ does not depend on x_k. Hence we conclude that $f_{\mathbf{u}} = 0$. □

From Theorem 3.1 we see that if for different sets $\mathbf{z}$ there exist different decompositions of f in which all super-$\mathbf{z}$ terms are zero, then the decomposition from Theorem 2.1 has all super-$\mathbf{z}$ terms equal to zero for *all* of these different sets $\mathbf{z}$. In other words, the decomposition from Theorem 2.1 uses lower order terms as much as possible and does not introduce unnecessary higher order terms. Furthermore, this holds for all possible choices of commuting projections $\{P_j\}_{j=1}^d$ satisfying (2.1). Indeed, if the decomposition with respect to one choice of operators P_j has all super-$\mathbf{z}$ terms equal to zero, then the decompositions with respect to *all* choices of operators P_j have all super-$\mathbf{z}$ terms equal to zero.

This minimal property has a strong link with the following notions of dimension:

- We say that a function of d variables has *cutoff dimension* k iff it depends only on the first k variables, and k is the smallest number for which this holds.
- We say that a function of d variables is *of order* q iff it can be written as a sum of functions each depending on at most q of the d variables, and q is the smallest number for which this holds.

These two notions correspond respectively to the more familiar concepts of *truncation dimension* and *superposition dimension*, see e.g., [2, 14, 21], except that the latter two allow small contributions from other groups of variables. Obviously the cutoff dimension of f is the smallest k for which a decomposition $f = \sum_{\mathbf{u}\subseteq[1..k]} t_{\mathbf{u}}$ exists. Similarly, the order of f is the smallest q for which a decomposition $f = \sum_{\mathbf{u}\subseteq[1..d],\,|\mathbf{u}|\le q} t_{\mathbf{u}}$ holds. Note that the cutoff dimension and the order of a given function are both uniquely defined and that the definition is not related to any particular decomposition method. Theorem 3.1 implies that the decomposition from Theorem 2.1 is minimal with respect to both cutoff dimension and order:

- A function f of d variables has cutoff dimension $k < d$ *iff* k is the smallest number such that $f_{\mathbf{u}} = 0$ for all subsets $\mathbf{u} \subseteq [1..d]$ for which $\mathbf{u} \not\subseteq [1..k]$.
- A function f of d variables is of order $q < d$ *iff* q is the smallest number such that $f_{\mathbf{u}} = 0$ for all subsets $\mathbf{u} \subseteq [1..d]$ with $|\mathbf{u}| > q$.

The order of f may be defined as the smallest value of q for which $f = F_q$, with the quantity F_q defined by

$$F_q := \sum_{\substack{\mathbf{u}\subseteq[1..d] \\ |\mathbf{u}|\le q}} f_{\mathbf{u}}, \qquad 0 \le q \le d.$$

Using simple algebraic manipulations and induction, it can be proved that $F_d = f$ and

$$F_q = \sum_{j=0}^{q} (-1)^{q-j} \binom{d-1-j}{q-j} \sum_{\substack{\mathbf{u}\subseteq[1..d] \\ |\mathbf{u}|=j}} P_{[1..d]\setminus\mathbf{u}}(f) \qquad \text{for} \quad q < d. \tag{3.2}$$

4. Orthogonality

In this section, we identify a *reproducing kernel Hilbert space* setting and conditions for which the decomposition is precisely an orthogonal decomposition. It turns out that the setting is very general and covers many interesting function spaces, including certain well-known Sobolev spaces. For the theory of reproducing kernels, we refer the readers to [1], and in particular, to the section on "sum of reproducing kernels". We mention here only that a reproducing kernel $K : D \times D \to \mathbb{R}$ corresponding to the reproducing kernel Hilbert space $\mathcal{H}(K)$ satisfies $K(\boldsymbol{x}, \cdot) \in \mathcal{H}(K)$ for all $\boldsymbol{x} \in D$ and $K(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in D$ and has the reproducing property

$$\langle f, K(\boldsymbol{x}, \cdot)\rangle_{\mathcal{H}(K)} = f(\boldsymbol{x}) \qquad \text{for all} \quad f \in \mathcal{H}(K) \quad \text{and} \quad \boldsymbol{x} \in D,$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}(K)}$ denotes the inner product in $\mathcal{H}(K)$. Additionally, $\mathcal{H}(K)$ is the completion of $\operatorname{span}\{K(\cdot, \boldsymbol{y}) : \boldsymbol{y} \in D\}$.

Following [24], we consider a general reproducing kernel Hilbert space $\mathcal{H}(K_d)$ of real functions defined on $D \subseteq \mathbb{R}^d$ with a reproducing kernel of the form

$$K_d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\mathbf{u}\subseteq[1..d]} K_{d,\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}}), \qquad K_{d,\emptyset} = 1. \tag{4.1}$$

For each $\mathbf{u} \subseteq [1..d]$, $K_{d,\mathbf{u}}$ is the reproducing kernel for a Hilbert space $\mathcal{H}(K_{d,\mathbf{u}})$ of real functions defined on $D_{\mathbf{u}} := \{\boldsymbol{x}_{\mathbf{u}} : \boldsymbol{x} \in D\}$. Functions in $\mathcal{H}(K_{d,\mathbf{u}})$ depend only on variables listed in $\mathbf{u}$. In particular, the space $\mathcal{H}(K_{d,\emptyset}) = \operatorname{span}\{1\}$ contains just constant functions. We allow for more generality than [24] by *not* assuming that $K_{d,\mathbf{u}}$ is of a product form.

Functions from $\mathcal{H}(K_d)$ are sums of functions from $\mathcal{H}(K_{d,\mathbf{u}})$,

$$f = \sum_{\mathbf{u}\subseteq[1..d]} f_{\mathbf{u}} \qquad \text{for some} \qquad f_{\mathbf{u}} \in \mathcal{H}(K_{d,\mathbf{u}}). \tag{4.2}$$

The representation (4.2) is generally *not* unique, because some nonzero functions may belong to $\mathcal{H}(K_{d,\mathbf{u}})$ for different subsets $\mathbf{u}$. If $\|\cdot\|_{\mathcal{H}(K_d)}$ denotes the norm in $\mathcal{H}(K_d)$ and $\|\cdot\|_{\mathcal{H}(K_{d,\mathbf{u}})}$ denotes the norm in $\mathcal{H}(K_{d,\mathbf{u}})$ for each $\mathbf{u} \subseteq [1..d]$, then we have

$$\begin{aligned} &\|f\|^2_{\mathcal{H}(K_d)} \\ &\quad = \inf\Bigg\{ \sum_{\mathbf{u}\subseteq[1..d]} \|f_{\mathbf{u}}\|^2_{\mathcal{H}(K_{d,\mathbf{u}})} \ : \ f_{\mathbf{u}} \in \mathcal{H}(K_{d,\mathbf{u}}) \text{ such that } f = \sum_{\mathbf{u}\subseteq[1..d]} f_{\mathbf{u}} \Bigg\}; \end{aligned} \tag{4.3}$$

see [1]. The representation (4.2) is unique iff

$$\mathcal{H}(K_{d,\mathbf{u}}) \cap \mathcal{H}(K_{d,\mathbf{v}}) = \{0\} \quad \text{for all subsets} \quad \mathbf{u} \neq \mathbf{v}. \tag{4.4}$$

When (4.4) holds, the space $\mathcal{H}(K_d)$ is the direct and orthogonal sum of the spaces $\mathcal{H}(K_{d,\mathbf{u}})$, and we have

$$\|f\|^2_{\mathcal{H}(K_d)} = \sum_{\mathbf{u}\subseteq[1..d]} \|f_{\mathbf{u}}\|^2_{\mathcal{H}(K_{d,\mathbf{u}})}.$$

Now we relate reproducing kernel Hilbert spaces constructed in this manner to the decompositions introduced in Section 2.

Theorem 4.1. *Let $\mathcal{H}(K_d)$ be a reproducing kernel Hilbert space of real functions defined on $D \subseteq \mathbb{R}^d$ which is constructed from reproducing kernel Hilbert spaces $\mathcal{H}(K_{d,\mathbf{u}})$ for $\mathbf{u} \subseteq [1..d]$ as in (4.1) and (4.3). Let $\{P_j\}_{j=1}^d$ be a set of commuting projections on $\mathcal{H}(K_d)$ satisfying (2.1) and such that for all nonempty $\mathbf{u} \subseteq [1..d]$ and all $\boldsymbol{y}_{\mathbf{u}} \in D_{\mathbf{u}}$ we have*

$$P_j(K_{d,\mathbf{u}}(\cdot, \boldsymbol{y}_{\mathbf{u}})) = 0 \qquad \textit{for all} \quad j \in \mathbf{u}. \tag{4.5}$$

Then the representation (4.2) is unique and is an orthogonal decomposition that coincides with the decomposition from Theorem 2.1.

Proof. Take a representation (4.2) of f. Since $\operatorname{span}\{K_{d,\mathbf{u}}(\cdot, \boldsymbol{y}_{\mathbf{u}}) : \boldsymbol{y}_{\mathbf{u}} \in D_{\mathbf{u}}\}$ is dense in $\mathcal{H}(K_{d,\mathbf{u}})$ and the projections P_j are continuous, (4.5) immediately implies that $P_j(f_{\mathbf{u}}) = 0$ if $j \in \mathbf{u}$. Hence Theorem 2.1(c) implies that (4.2) coincides with (2.7), and because of uniqueness, (4.4) then gives orthogonality of the subspaces $\mathcal{H}(K_{d,\mathbf{u}})$ in $\mathcal{H}(K_d)$. □

For the anchored decomposition, see Example 2.3, the condition (4.5) requires

$$\begin{aligned} &K_{d,\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}, \cdot) = 0 \quad \text{whenever} \quad x_j = c_j \\ &\text{for all } j \in \mathbf{u} \text{ and for all nonempty } \mathbf{u} \subseteq [1..d]. \end{aligned} \tag{4.6}$$

In the case of the ANOVA decomposition, see Example 2.2, the condition (4.5) requires

$$\int_0^1 K_{d,\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}, \cdot)\, \mathrm{d}x_j = 0 \quad \text{for all } j \in \mathbf{u} \text{ and for all nonempty } \mathbf{u} \subseteq [1..d]. \tag{4.7}$$

We now give examples of reproducing kernel Hilbert spaces satisfying (4.6) and (4.7).

Example 4.2 (Anchored Sobolev spaces)**.** Let $D = [0,1]^d$. The weighted Sobolev space anchored at $\boldsymbol{c} \in D$ with smoothness parameter $r \geq 1$ has a reproducing kernel of the form (4.1), with $K_{d,\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}}, \boldsymbol{y}_{\mathbf{u}}) = \gamma_{d,\mathbf{u}} \prod_{j\in\mathbf{u}} \eta_j(x_j, y_j)$, where

$$\eta_j(x,y) = \begin{cases} \displaystyle\int_{c_j}^{\min(x,y)} \frac{(x-t)^{r-1}\,(y-t)^{r-1}}{((r-1)!)^2}\,\mathrm{d}t + \sum_{i=1}^{r-1} \frac{(x-c_j)^i\,(y-c_j)^i}{(i!)^2} & \text{if } x,y > c_j, \\ \displaystyle\int_{\max(x,y)}^{c_j} \frac{(t-x)^{r-1}\,(t-y)^{r-1}}{((r-1)!)^2}\,\mathrm{d}t + \sum_{i=1}^{r-1} \frac{(c_j-x)^i\,(c_j-y)^i}{(i!)^2} & \text{if } x,y < c_j, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly the kernel is continuous, and $\eta_j(c_j,\cdot)=0$ for all $j=1,2,\ldots,d$. Thus the condition (4.6) holds, and the anchored decomposition (2.10) is precisely the orthogonal decomposition in the anchored Sobolev spaces for all $r\ge 1$. The numbers $\gamma_{d,\mathbf{u}}>0$ are referred to as *weights*. They model the relative importance between various groups of variables by weighting the norms. For $r=1$, we can express the norm solely in terms of f as

$$\|f\|^2_{\mathcal{H}(K_d)} \;=\; f^2(\boldsymbol{c}) + \sum_{\emptyset\ne\mathbf{u}\subseteq[1..d]} \frac{1}{\gamma_{d,\mathbf{u}}} \int_{[0,1]^{|\mathbf{u}|}} \left(\frac{\partial^{|\mathbf{u}|} f((\boldsymbol{x};\boldsymbol{c})_{\mathbf{u}})}{\partial \boldsymbol{x}_{\mathbf{u}}}\right)^2 \mathrm{d}\boldsymbol{x}_{\mathbf{u}}.$$

The norm for higher smoothness $r>1$ is difficult to express explicitly when $d>1$. For $d=1$ and $r\ge 1$, it takes the following form:

$$\|f\|^2_{\mathcal{H}(K_1)} \;=\; f^2(c) + \frac{1}{\gamma}\left(\sum_{i=1}^{r-1}\left[f^{(i)}(c)\right]^2 + \int_0^1 \left[f^{(r)}(x)\right]^2 \mathrm{d}x\right).$$

Special cases of anchored Sobolev spaces have been considered in many papers (specializing to $\boldsymbol{c}=\mathbf{0}$ or $\boldsymbol{c}=\mathbf{1}$, and/or $r=1$, and/or $\gamma_{d,\mathbf{u}}=1$); see, e.g., [16] and the references therein, and [20, Chapter 1.2]. See also [23] for extensions.

Example 4.3 (Unanchored Sobolev spaces)**.** Let $D=[0,1]^d$. The weighted and unanchored Sobolev space with smoothness parameter $r\ge 1$ has a reproducing kernel of the form (4.1), with $K_{d,\mathbf{u}}(\boldsymbol{x}_{\mathbf{u}},\boldsymbol{y}_{\mathbf{u}})=\gamma_{d,\mathbf{u}}\prod_{j\in\mathbf{u}}\eta(x_j,y_j)$, where

$$\eta(x,y) \;=\; \frac{B_{2r}(|x-y|)}{(-1)^{r+1}(2r)!} + \sum_{i=1}^{r}\frac{B_i(x)\,B_i(y)}{(i!)^2}.$$

Here B_i denotes the Bernoulli polynomial of degree i, which has the property $\int_0^1 B_i(x)\,\mathrm{d}x = 0$. It is easy to see that $\int_0^1 \eta(x,\cdot)\,\mathrm{d}x = 0$. Thus the condition (4.7) holds, and the ANOVA decomposition (2.9) is precisely the orthogonal decomposition in this unanchored Sobolev space. It is a striking and sometimes useful consequence that the terms in the ANOVA decomposition are orthogonal not only in $\mathcal{L}_2([0,1]^d)$, but also in the unanchored Sobolev spaces for every value of r. For $r=1$, we have

$$\begin{aligned}\|f\|^2_{\mathcal{H}(K_d)} \;=\; &\left(\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}x\right)^2 \\ &+ \sum_{\emptyset\ne\mathbf{u}\subseteq[1..d]} \frac{1}{\gamma_{d,\mathbf{u}}}\int_{[0,1]^{|\mathbf{u}|}}\left(\int_{[0,1]^{d-|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} f(\boldsymbol{x})}{\partial \boldsymbol{x}_{\mathbf{u}}}\,\mathrm{d}\boldsymbol{x}_{[1..d]\setminus\mathbf{u}}\right)^2 \mathrm{d}\boldsymbol{x}_{\mathbf{u}}.\end{aligned}$$

For $d=1$ and general $r\ge 1$, we have

$$\|f\|^2_{\mathcal{H}(K_1)} \;=\; \left(\int_0^1 f(x)\,\mathrm{d}x\right)^2 + \frac{1}{\gamma}\left[\sum_{i=1}^{r-1}\left(\int_0^1 f^{(i)}(x)\,\mathrm{d}x\right)^2 + \int_0^1\left[f^{(r)}(x)\right]^2\mathrm{d}x\right].$$

Special cases of unanchored Sobolev spaces have been considered in many papers (specializing to $r=1$ and/or $\gamma_{d,\mathbf{u}}=1$); see, e.g., [4] and the references therein, as well as [3] and [20, Chapter 10.2]. In particular, it was already shown in [4] that the ANOVA and the orthogonal decompositions for $r=1$ coincide.

Example 4.4 (General tensor product spaces)**.** We now show how to construct operators P_j when the space $\mathcal{H}(K_d)$ is given as the d-fold tensor product of a general reproducing kernel Hilbert space $\mathcal{H}(K_1)$. We only assume that $K_1 : D \times D \to \mathbb{R}$, with $D \subseteq \mathbb{R}$, is of the form $K_1(x, y) = 1 + \eta(x, y)$, where $\eta : D \times D \to \mathbb{R}$ is a reproducing kernel such that $\mathcal{H}(\eta)$ is separable and $1 \notin \mathcal{H}(\eta)$. Thus we have $K_d(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^d (1 + \eta(x_i, y_i)) = \sum_{\mathbf{u}\subseteq[1..d]} \prod_{i\in\mathbf{u}} \eta(x_i, y_i)$. Let $\xi_0 \equiv 1$, and let $\{\xi_k\}_{k=1}^\infty$ be a complete orthonormal system in $\mathcal{H}(\eta)$. Then $\{\xi_k\}_{k=0}^\infty$ is a complete orthonormal system in $\mathcal{H}(K_1)$. For $d \geq 1$ and $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d$ (where $\mathbb{N} = \{0, 1, \ldots\}$), let $\xi_{d,\boldsymbol{k}}(\boldsymbol{x}) = \prod_{i=1}^d \xi_{k_i}(x_i)$ for $\boldsymbol{x} \in D^d$. Then $\{\xi_{d,\boldsymbol{k}}\}_{\boldsymbol{k}\in\mathbb{N}^d}$ is a complete orthonormal system in $\mathcal{H}(K_d)$. Note that if $k_i = 0$, then $\xi_{d,\boldsymbol{k}}$ does not depend on the ith variable. Therefore, we set P_j to be orthogonal projections onto $\operatorname{span}\{\xi_{d,\boldsymbol{k}} : \boldsymbol{k} \in \mathbb{N}^d \text{ with } k_j = 0\}$, since they clearly satisfy (2.1) and (4.5).

Although the above examples have reproducing kernel of the product form, the product form is not a necessary property for the existence of a suitable set of projections $\{P_j\}_{j=1}^d$. To see this, let $\tilde{K}_{d,\mathbf{u}}(\boldsymbol{x}_\mathbf{u}, \boldsymbol{y}_\mathbf{u})$ be an arbitrary (nonproduct) reproducing kernel on $D_\mathbf{u}$. Then $K_{d,\mathbf{u}}(\boldsymbol{x}_\mathbf{u}, \boldsymbol{y}_\mathbf{u}) := (\prod_{j\in\mathbf{u}} x_j y_j)\, \tilde{K}_{d,\mathbf{u}}(\boldsymbol{x}_\mathbf{u}, \boldsymbol{y}_\mathbf{u})$ is the reproducing kernel of some Hilbert space $\mathcal{H}(K_{d,\mathbf{u}})$ which is *anchored* at $\mathbf{0}$, since every element of the space will vanish if any component of $\boldsymbol{x}_\mathbf{u}$ has the value 0. Thus Theorem 4.1 holds for $\mathcal{H}(K_{d,\mathbf{u}})$ with P_j defined by (2.2) and $c_j = 0$ for all $j = 1, 2, \ldots, d$. Although the present theory has wider scope than just product kernels $K_{d,\mathbf{u}}$, we do *not* claim that a suitable set of projections $\{P_j\}_{j=1}^d$ exists for every choice of reproducing kernel Hilbert spaces $\mathcal{H}(K_{d,\mathbf{u}})$.

5. Computational cost

The anchored decomposition is arguably more useful in practice than the ANOVA decomposition because it requires only function evaluations of f with respect to a chosen anchor $\boldsymbol{c}$. Let $\boldsymbol{x} \in D$ be given and fixed. It is clear from the explicit formula in Example 2.3(b) that for each $\mathbf{u}$, $f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_\mathbf{u})$ can be computed using $2^{|\mathbf{u}|} - 1$ additions and subtractions and $2^{|\mathbf{u}|}$ samples of f.

Suppose we want to compute $f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_\mathbf{u})$ for all $\mathbf{u}$ with $|\mathbf{u}| \leq q$. We have $A_{d,q} := \sum_{j=0}^q \binom{d}{j}$ such subsets $\mathbf{u}$. It is easy to check, see also [25], that $A_{d,q} \leq 2\, d^q$ and $A_{d,q} = (d^q/q!)(1 + o(1))$ as $d \to \infty$. We need to evaluate f at the points $(\boldsymbol{x}; \boldsymbol{c})_\mathbf{u}$ for all $\mathbf{u}$ with $|\mathbf{u}| \leq q$; that is, we use at most q components of $\boldsymbol{x}$ and the rest are replaced by the corresponding components of $\boldsymbol{c}$. The total number of such points is again $A_{d,q}$. Hence, all $f_{\mathbf{u},\boldsymbol{c}}(x_\mathbf{u})$ with $|\mathbf{u}| \leq q$ can be computed using at most $A_{d,q}$ samples of f and at most $B_{d,q} := \sum_{j=0}^q 2^j \binom{d}{j}$ arithmetic operations, where $B_{d,q} \leq 2^{q+1}\, d^q$ and $B_{d,q} = (2^q\, d^q/q!)(1 + o(1))$ as $d \to \infty$.

If instead we just want to compute $F_q(\boldsymbol{x})$, that is, the combined contribution from all terms $f_{\mathbf{u},\boldsymbol{c}}(\boldsymbol{x}_\mathbf{u})$ with $|\mathbf{u}| \leq q$, then we have from (3.2) that for $q < d$,

$$F_q(\boldsymbol{x}) = \sum_{j=0}^q \alpha_{q,j} \sum_{\substack{\mathbf{u}\subseteq[1..d] \\ |\mathbf{u}|=j}} f((\boldsymbol{x}; \boldsymbol{c})_\mathbf{u}), \qquad \alpha_{q,j} = (-1)^{q-j} \binom{d-1-j}{q-j}.$$

This would again require at most $A_{d,q}$ samples of f at the points $(\boldsymbol{x}; \boldsymbol{c})_\mathbf{u}$ for all $\mathbf{u}$ with $|\mathbf{u}| \leq q$. However, the total number of arithmetic operations will be significantly reduced. Starting with $\alpha_{0,0} = 1$, the numbers $\alpha_{i,j}$ for $i = 1, \ldots, q$

and $j = 1, \ldots, i$ can be computed using the general recurrence relations $\alpha_{i,0} = -\alpha_{i-1,0}(d-i)/i$, $\alpha_{i,j} = \alpha_{i-1,j} + \alpha_{i-1,j-1}$, and $\alpha_{i,i} = 1$. Once the numbers $\alpha_{q,j}$ are computed, we need to perform $A_{d,q} - 1$ additions and $q+1$ multiplications. For large q, the total number of arithmetic operations is then roughly a factor of 2^q smaller than $B_{d,q}$.

6. Discussion

In recent times, the success of quasi-Monte Carlo (QMC) methods for many practical problems of high dimension has been attributed to the fact that these problems have *low effective dimension*. Two notions of effective dimension have been defined through the variance allocation of the ANOVA terms: *truncation dimension* and *superposition dimension*; see, e.g., [2] for their precise definitions and [21] for known techniques to approximate them.

Although the explanation in terms of effective dimension is undoubtedly broadly true, the existing framework is somewhat unsatisfactory since, unlike the Monte Carlo (MC) error, the QMC error is *not* expressible in terms of the variance. Rather, QMC errors are typically bounded in terms of some discrepancy times the variation of the function, or more recently, in terms of the worst case error in a reproducing kernel Hilbert space times the norm of the function. Moreover, the ANOVA decomposition may not be orthogonal in this function space. Once the underlying function space is decided, a more suitable approach is to identify a set of projections that lead to orthogonal decomposition in the function space as discussed in our paper, and then capture the notion of low dimensionality through the allocation of the square norm.

Many recent papers have considered *weighted* reproducing kernel Hilbert spaces, with the weights intended to characterize the relative importance of various groups of variables. The choice of weights can have practical as well as theoretical importance, since modern constructions of lattice rules (for a survey see [9]) make explicit use of the weights. The question of how best to choose the weights is a topic of active current research. In considering quadrature errors for QMC in the unanchored space of Example 4.3 with $r = 1$, [22] showed that the quadrature error bound of the form "worst case error $\times$ norm of f" achieves its minimum value for a given function f if the weights are given by $\gamma_{d,\mathbf{u}} = 6^{|\mathbf{u}|/2}\,\beta_{\mathbf{u}}(f)$, where

$$\beta_{\mathbf{u}}(f) := \left[\int_{[0,1]^{|\mathbf{u}|}} \left(\int_{[0,1]^{d-|\mathbf{u}|}} \frac{\partial^{|\mathbf{u}|} f}{\partial \boldsymbol{x}_{\mathbf{u}}}\, \mathrm{d}\boldsymbol{x}_{[1..d]\setminus\mathbf{u}}\right)^2 \mathrm{d}\boldsymbol{x}_{\mathbf{u}}\right]^{1/2} = \|f_{\mathbf{u},*}\|_{\mathcal{H}(K_{d,\mathbf{u}})},$$

where in the last step we used two easily verified properties of the ANOVA decomposition (2.9), namely that $\beta_{\mathbf{u}}(f_{\mathbf{v},*}) = 0$ if $\mathbf{v} \neq \mathbf{u}$ and $\beta_{\mathbf{u}}(f_{\mathbf{v},*}) = \|f_{\mathbf{u},*}\|_{\mathcal{H}(K_{d,\mathbf{u}})}$ if $\mathbf{v} = \mathbf{u}$. This opens a path to a good choice of the weights if we can estimate $\|f_{\mathbf{u},*}\|_{\mathcal{H}(K_{d,\mathbf{u}})}$, a task that should be made easier by a better understanding of the structure of the ANOVA decomposition. Similar arguments for choosing weights can be used for the anchored Sobolev space with the anchored decomposition.

The explicit decomposition formula can be useful for numerical integration and approximation of functions in high dimensions, since there might be a merit in approximating the solution corresponding to the lower order part (see, e.g., [10]). We may also use the explicit decomposition formula in a theoretical analysis where

no actual computation is required, noting that it may be easier to observe or identify certain properties of one particular decomposition term when it is expressed explicitly rather than recursively (see, e.g., [6]).

Acknowledgments

The authors thank Rob Womersley, Michael Griebel, and the anonymous associate editor for valuable suggestions. The support of the Australian Research Council under its Centres of Excellence Program is gratefully acknowledged. The first author is supported by an Australian Research Council Queen Elizabeth II Research Fellowship. The last two authors were partially supported by the National Science Foundation under Grants DMS-0609703 and DMS-0608727, respectively.

References

1. N. Aronszajn, *Theory of reproducing kernels,* Trans. Amer. Math. Soc., **68**, 337–404 (1950). MR0051437 (14:479c)
2. R. E. Caflisch, W. Morokoff, and A. Owen, *Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension*, J. Comput. Finance **1**, 27–46 (1997).
3. J. Dick and F. Pillichshammer, *Strong tractability of multivariate integration of arbitrary high order using digitally shifted polynomial lattice rules*, J. Complexity **23**, 436–453 (2007). MR2372007 (2008m:65067)
4. J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski, *Liberating the weights*, J. Complexity **20**, 593–623 (2004). MR2086942 (2005h:65008)
5. M. Griebel, *Sparse grids and related approximation schemes for higher dimensional problems*, Foundations of Computational Mathematics, Santander 2005 (L. M. Pardo, A. Pinkus, E. Süli, M. J. Todd, eds.), Cambridge University Press, London, 2006, pp. 106–161. MR2277104 (2007k:65206)
6. M. Griebel, F. Y. Kuo and I. H. Sloan, *The smoothing effect of the* ANOVA *decomposition*, submitted.
7. F. J. Hickernell, I. H. Sloan, and G. W. Wasilkowski, *On tractability of weighted integration over bounded and unbounded regions in* $\mathbb{R}^s$, Math. Comp., **73**, 1885–1901 (2004). MR2059741 (2005c:65018)
8. F. J. Hickernell, I. H. Sloan, and G. W. Wasilkowski, *The strong tractability of multivariate integration using lattice rules*, Monte Carlo and Quasi-Monte Carlo Methods 2002 (H. Niederreiter, ed.), Springer-Verlag, Berlin, 2004, pp. 259–273. MR2076938
9. F. Y. Kuo and I. H. Sloan, *Lifting the curse of dimensionality*, Notices Amer. Math. Soc. **52**, 1320–1329 (2005). MR2183869 (2006j:65061)
10. F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski, *Liberating the dimension*, submitted.
11. G. Li, J. Schoendorf, T.-S. Ho, and H. Rabitz, *Multicut-HDMR with an application to an ionospheric model*, J. Comput. Chem. **25**, 1149–1156 (2004).
12. G. Li, J. Hu, S.-W. Wang, P. G. Georgopoulos, J. Schoendorf, and H. Rabitz, *Random sampling-high dimensional model representation* (*RS-HDMR*) *and orthogonality of its different order component functions*, J. Phys. Chem. A **110**, 2474–2485 (2006).
13. G. Li and H. Rabitz, *Ratio control variate method for efficiently determining high-dimensional model representations*, J. Comput. Chem. **27**, 1112–1118 (2006).
14. R. Liu and A. Owen, *Estimating mean dimensionality of analysis of variance decompositions*, J. Amer. Statist. Assoc. **101**, 712–721 (2006). MR2281247
15. H. Rabitz, O. F. Alis, J. Shorter and K. Shim, *Efficient input-output model representation*, Comput. Phys. Commun. **117**, 11–20 (1999).
16. I. H. Sloan, X. Wang, and H. Woźniakowski, *Finite-order weights imply tractability of multivariate integration*, J. Complexity **20**, 46–74 (2004). MR2031558 (2004j:65034)
17. I. M. Sobol′, *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian). MR0422968 (54:10952)

18. I. M. Sobol′, *Sensitivity estimates for nonlinear mathematical models*, Matematicheskoe Modelirovanie, 1990, V. 2, N 1, 112–118 (in Russian), English translation in: Mathematical Modeling and Computational Experiment, 407–414 (1993). MR1335161 (96a:65010)
19. I. M. Sobol′, *Theorems and examples on high dimensional model representation*, Reliability Engrg. and System Safety **79**, 187–193 (2003).
20. G. Wahba, *Spline Models for Observational Data*, SIAM-NSF Regional Conference Series in Appl. Math., Vol. 59, SIAM, Philadelphia, 1990. MR1045442 (91g:62028)
21. X. Wang and K.-T. Fang, *Effective dimension and quasi-Monte Carlo integration*, J. Complexity **19**, 101–124 (2003). MR1966664 (2003m:62016)
22. X. Wang and I. H. Sloan, *Brownian bridge and principal component analysis*, IMA J. Numer. Anal. **27**, 631–654 (2007). MR2371825 (2009a:62248)
23. G. W. Wasilkowski and H. Woźniakowski, *Tractability of approximation and integration for weighted tensor product problems over unbounded domains*, Monte Carlo and Quasi-Monte Carlo Methods 2000, (K.-T. Fang, F. J. Hickernell, H. Niederreiter, eds.), Springer, Berlin, 2002, pp. 497–522. MR1958877 (2004b:65222)
24. G. W. Wasilkowski and H. Woźniakowski, *Finite-order weights imply tractability of linear multivariate problems*, J. Approx. Theory **130**, 57–77 (2004). MR2086810 (2005d:41048)
25. G. W. Wasilkowski and H. Woźniakowski, *Polynomial-time Algorithms for Multivariate Linear Problems with Finite-order Weights: Worst Case Setting*, Found. Comput. Math., **5**, 451–491 (2005). MR2189546 (2006i:65037)

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia

E-mail address: `f.kuo@unsw.edu.au`

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia

E-mail address: `i.sloan@unsw.edu.au`

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

E-mail address: `greg@cs.uky.edu`

Department of Computer Science, Columbia University, New York, New York 10027, and Institute of Applied Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland

E-mail address: `henryk@cs.columbia.edu`