Higher order Sobol' indices

Art B. Owen
Stanford University

Josef Dick
University of New South Wales

Su Chen
Two Sigma LLC

June 2013

Abstract

Sobol' indices measure the dependence of a high dimensional function on groups of variables defined on the unit cube $[0,1]^d$. They are based on the ANOVA decomposition of functions, which is an L^2 decomposition. In this paper we discuss generalizations of Sobol' indices which yield L^p measures of the dependence of f on subsets of variables. Our interest is in values $p > 2$ because then variable importance becomes more about reaching the extremes of f. We introduce two methods. One based on higher order moments of the ANOVA terms and another based on higher order norms of a spectral decomposition of f, including Fourier and Haar variants. Both of our generalizations have representations as integrals over $[0,1]^{kd}$ for $k \geqslant 1$, allowing direct Monte Carlo or quasi-Monte Carlo estimation. We find that they are sensitive to different aspects of f, and thus quantify different notions of variable importance.

1 Introduction

Sobol' indices (Sobol', 1990) are the standard way to measure the importance of variables and subsets of variables for a black box function defined on the unit cube $[0,1]^d$. These measures are used in applications in aerospace engineering and climate models among many others.

Sobol's indices are based on the ANOVA decomposition of $[0,1]^d$, which is an L^2 method. An aeronautics-astronautics engineering student, Gary Tang, asked us about how to construct an alternative to Sobol' indices that would identify which variables are most important when one is especially interested in the extreme values taken on by the function. In this paper we address that problem by considering alternative measures based on other criteria that place greater emphasis on extremes than L^2 does.

Perhaps the simplest way to get an index more sensitive to extremes in f is to replace the target function $f(\boldsymbol{x})$ by a transformed version such as $|f(\boldsymbol{x})|$ or $\exp(f(\boldsymbol{x}))$ or $1_{f(\boldsymbol{x})\geqslant M}$ for a threshold M and so on, followed by an application of the usual Sobol' indices. This approach will often be reasonable. In some cases though, it may complicate the problem. For example, if f is a sum of functions of one variable at a time, then f^2 involves pairwise interactions that were not present in f, and $1_{f(\boldsymbol{x})\geqslant M}$ may involve interactions of all orders. Furthermore, if f only takes two values, such as 0 or 1, (e.g., safe versus dangerous outcomes), then transforming it to take two different values does not help. As a result, we consider new generalizations.

The ANOVA can be developed as an analysis of $L^2[0,1]^d$, or as a synthesis of Fourier, Walsh or other basis expansions. Both of these methods can be used to make L^p generalizations. Additionally, the Sobol' indices satisfy some identities that can be directly generalized. These approaches coincide for $p = 2$, but they differ for $p \neq 2$.

An outline of this paper is as follows. Section 2 introduces our notation and reviews the ANOVA and Sobol' indices. Section 3 presents some related non-L_2 concepts, median polish and analysis of skewness, from the literature. One of our methods includes a crossed-effects extension of the analysis of skewness as a special case. Section 4 presents a generalization based on extending one of Sobol's identities to p'th order moments. The identity yields a representation of the index as an integral of dimension dp or lower. For even integers $p \geqslant 2$ we show that the resulting estimates are nonnegative and increase when any set of variables is replaced by a superset. Section 5 presents a generalization based on the synthesis from a Fourier expansion. When $p \geqslant 2$ is an even integer, then the resulting importance measures are sums of p'th powers of the moduli of the function's Fourier coefficients. Yet they can still be estimated directly by a high dimensional quadrature, based on an identity like one of Sobol's. That integral can be converted into one of dimension $d(p-1)$ or lower. We also provide a version based on Walsh functions, which again has nonnegativity and additivity when $p \geqslant 2$ is an even integer and also has an integral representation for quadrature. For odd p, we include a 'Dirichlet kernel trick' that produces non-negative importance measures based on L_p norms of Fourier or Walsh coefficients. That method also allows one to favor certain parts of the spectrum.

Section 6 illustrates our importance measures on test functions that are sums or products. We use such examples to confirm that our measures focus on variables that bring f towards extreme values. For product functions, and even p, our spectral measures find that the most important variables are those whose spectrum is sparsest. Our moment measure, for $p = 4$, favors variables with high kurtosis and with mean and skewness of the same sign. We look also at the important special case a rectangular spike: $f(\boldsymbol{x}) = \prod_{j=1}^d 1_{x_j \leqslant \epsilon_j}$. When f measures hitting a small region like this the variable with the smallest ϵ_j is the most important one, at least when all ϵ_j are small. Both moment and Fourier measures favor small ϵ_j. For additive functions, having no interactions, we find that the spectral measures place all their importance on singleton sets. The moment measure does this for third but not fourth moments. Section 7 has a discussion.

2 Notation

We are given a real-valued function f defined on $[0,1]^d$ for $d \geqslant 1$ and we are interested in quantifying the importance to f of various subsets of the variables in the set $\mathcal{D} = \{1, 2, \ldots, d\}$.

We make frequent use of subsets of $\mathcal{D}$ as indices. The complement of $u \subseteq \mathcal{D}$ is $u^c = \mathcal{D} - u$, or simply $-u$ when that is typographically more convenient. The cardinality of u is $|u|$. For $\boldsymbol{x} \in [0,1]^d$, the point $\boldsymbol{x}_u \in [0,1]^{|u|}$ is made up of x_j for $j \in u$ and $\mathrm{d}\boldsymbol{x}_u = \prod_{j \in u} \mathrm{d}x_j$. We use $u \subset v$ to mean that u is a proper subset of v (i.e., $u \subsetneq v$).

We often make a new point from components of two old points. If $\boldsymbol{x}, \boldsymbol{z} \in [0,1]^d$ and $u \subseteq \mathcal{D}$, then $\boldsymbol{y} \equiv \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$ is the point in $[0,1]^d$ with $y_j = x_j$ for $j \in u$ and $y_j = z_j$ for $j \notin u$.

2.1 ANOVA of $[0,1]^d$

The ANOVA decomposition represents $f(\boldsymbol{x})$ via

$$f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x}) \tag{1}$$

where the functions f_u are defined recursively by

$$f_u(\boldsymbol{x}) = \int_{[0,1]^{d-|u|}} \Bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x}) \Bigr) \,\mathrm{d}\boldsymbol{x}_{-u}. \tag{2}$$

From usual conventions, $f_\varnothing(\boldsymbol{x}) = \mu \equiv \int_{[0,1]^d} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ for all $\boldsymbol{x} \in [0,1]^d$. The function f_u only depends on x_j for $j \in u$. For $f \in L^2[0,1]^d$, these functions satisfy $\int_0^1 f_u(\boldsymbol{x}) \,\mathrm{d}x_j = 0$ when $j \in u$, from which it follows that $\int f_u(\boldsymbol{x}) f_v(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = 0$ for $u \neq v$ and that

$$\sigma^2 = \sum_{u \subseteq \mathcal{D}} \sigma_u^2 \tag{3}$$

where $\sigma^2 = \int (f(\boldsymbol{x}) - \mu)^2 \,\mathrm{d}\boldsymbol{x}$, $\sigma_\varnothing^2 = 0$ and $\sigma_u^2 = \int f_u(\boldsymbol{x})^2 \,\mathrm{d}\boldsymbol{x}$ for $u \neq \varnothing$. The name ANOVA stands for analysis of variance, as given by (3). This decomposition goes back to Hoeffding (1948).

Sobol' (1969) obtained the decomposition (1) by a different route, described next. Let ϕ_k for $k \in \mathbb{I}$ be a complete orthonormal basis for $L^2[0,1]$, where $\mathbb{I}$ is a countable index set containing a 0 element, with $\phi_0(x) = 1$, $\forall x \in [0,1]$. We can form the tensor product basis $\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{\ell=1}^d \phi_{k_\ell}(x_\ell)$, for $\boldsymbol{k} \in \mathbb{I}^d$ and then $f(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{I}^d} \beta_{\boldsymbol{k}} \phi_{\boldsymbol{k}}(\boldsymbol{x})$ where $\beta_{\boldsymbol{k}} = \int f(\boldsymbol{x}) \overline{\phi_{\boldsymbol{k}}(\boldsymbol{x})} \,\mathrm{d}\boldsymbol{x}$. Then, with $\mathbf{0}$ a vector of d zeros, and $\mathbb{I}_*$ the nonzero members of $\mathbb{I}$,

$$f_u(\boldsymbol{x}) = \sum_{\boldsymbol{k}_u \in \mathbb{I}_*^{|u|}} \beta_{\boldsymbol{k}_u:\mathbf{0}_{-u}} \phi_{\boldsymbol{k}_u:\mathbf{0}_{-u}}(\boldsymbol{x}) \tag{4}$$

recovers the functions defined at (2), and $\sigma_u^2 = \sum_{\boldsymbol{j}_u \in \mathbb{I}_*^{|u|}} \beta^2_{\boldsymbol{j}_u:\mathbf{0}_{-u}}$. Sobol' (1969) used Haar functions for his 'decomposition into summands of different dimensions' given by (4). Where Hoeffding has an analysis, Sobol' has a synthesis of variance.

2.2 Sobol' indices and identities

The importance of variable $j \in \mathcal{D}$ is due in part to $\sigma^2_{\{j\}}$, but also due to σ_u^2 for other sets u with $j \in u$. More generally, we may be interested in the importance of a subset u of the variables.

Sobol' introduced two measures of variable subset importance, which we denote

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2, \quad \text{and} \quad \overline{\tau}_u^2 = \sum_{v \cap u \neq \varnothing} \sigma_v^2.$$

These satisfy $\underline{\tau}_u^2 \leqslant \overline{\tau}_u^2$ and $\underline{\tau}_u^2 + \overline{\tau}_{-u}^2 = \sigma^2$. Sobol' usually normalized these quantities by σ^2, yielding global sensitivity indices $\underline{\tau}_u^2/\sigma^2$ and $\overline{\tau}_u^2/\sigma^2$. We will use the unnormalized versions.

It is an elementary consequence of the ANOVA definitions that

$$\iint f(\boldsymbol{x}) f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{z} = \mu^2 + \underline{\tau}_u^2 \tag{5}$$

and

$$\frac{1}{2} \iint \big(f(\boldsymbol{x}) - f(\boldsymbol{x}_{-u}{:}\boldsymbol{z}_u)\big)^2 \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{z} = \overline{\tau}_u^2. \tag{6}$$

We write these integrals over $(\boldsymbol{x}, \boldsymbol{u}) \in [0,1]^{2d}$, although the first really only uses $2d - |u|$ components and the second uses $d + |u|$.

The great convenience of Sobol's measures is that they can be directly estimated by integration without bias. We do not need to explicitly estimate, square, integrate and sum the individual ANOVA terms. As a consequence, we can avoid numerical optimization and bias corrections.

It is computationally convenient to replace equation (5) by

$$\iint f(\boldsymbol{x}) \big(f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}) - f(\boldsymbol{z})\big) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{z} = \underline{\tau}_u^2, \tag{7}$$

because it eliminates the need to subtract an estimate of μ. Equation (7) was developed independently in Saltelli (2002) and by Mauntz (2002), and it performs better when $\underline{\tau}_u^2$ is small. For discussion and another estimator, see Owen (2012a).

2.3 Generalizations

We have three different ways to generalize the ANOVA to higher moments. First, we can generalize the original ANOVA decomposition by noticing that the integrals in it minimize a quadratic quantity, and then replacing that quadratic by a higher order moment. Second, we can generalize the Sobol' indices directly, replacing the integrals of products of pairs of function values by integrals of products of three or more function values. Third, we can generalize Sobol's synthesis.

3 Related literature

In this section we consider two non-L_2 methods from the literature. A natural approach to generalizing the ANOVA to $p \neq 2$ begins with the probabilistic interpretation of $f_u(\boldsymbol{x})$ as a conditional expectation

$$f_u(\boldsymbol{x}) = \mathbb{E}\Bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x}) \mid \boldsymbol{x}_u\Bigr).$$

For any $\boldsymbol{x}_u \in [0,1]^d$

$$f_u(\boldsymbol{x}) = \operatorname*{argmin}_m \mathbb{E}\biggl(\Bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x}) - m\Bigr)^2 \mid \boldsymbol{x}_u\biggr).$$

Just as the conditional expectation minimizes conditional variance, we may generalize the ANOVA to moments $p \geqslant 1$, via

$$f_u^{(p)}(\boldsymbol{x}) = \operatorname*{argmin}_m \mathbb{E}\biggl(\Bigl|f(\boldsymbol{x}) - \sum_{v \subset u} f_v^{(p)}(\boldsymbol{x}) - m\Bigr|^p \mid \boldsymbol{x}_u\biggr).$$

This generalization satisfies $f(\boldsymbol{x}) = \sum_u f_u^{(p)}(\boldsymbol{x})$ through the definition of $f_{\mathcal{D}}^{(p)}$, but the terms in it are not generally orthogonal. Nor do they decompose $\int |f(\boldsymbol{x})|^p \,\mathrm{d}\boldsymbol{x}$, nor do they generally integrate to 0 over x_j for $j \in u$. If $|f|$ is bounded, then there is a $p = \infty$ version corresponding to a statistic called the midrange.

It is cumbersome to minimize norms other than L_2 to define alternatives to f_u. The one example we found for this approach is the median polish method, in the next section. It uses $p = 1$, which might be expected to place *less* emphasis on extremes of f than the ANOVA, and is based on conditional medians.

3.1 Median polish

Tukey (1977) describes the median polish algorithm for a two dimensional table of numbers X_{ij}, $i = 1, \dots, I$ and $j = 1, \dots, J$. The median polish algorithm generates a decomposition

$$X_{ij} = a_i + b_j + R_{ij}.$$

Starting with $a_i = b_j = 0$ and $R_{ij} = X_{ij}$, it alternates between row steps

$$m_i \leftarrow \operatorname{median}(R_{i1}, \dots, R_{iJ}), \quad 1 \leqslant i \leqslant I$$

$$a_i \leftarrow a_i + m_i, \quad 1 \leqslant i \leqslant I$$
$$R_{ij} \leftarrow R_{ij} - m_i, \quad 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J$$

and analogous column steps. Siegel (1983) shows that the algorithm converges when all of the X_{ij} are rational numbers. While median polish will converge to a result where every row and column of R_{ij} has median 0, the result does necessarily have the L_1 minimizing values of a_i and b_j. For a table of data with an even number $I = 2k$ of rows, Siegel (1983) gets better results via the 'low median', which is the k'th smallest value, instead of the median which averages the k'th and $k+1$'st values.

In principal one could evaluate f on a grid embedded in $[0,1]^2$ and apply the median polish algorithm. While there may be reasonable ways to generalize median polish to $d > 2$, the necessity of estimating the additive components in order to measure them is computationally unattractive.

3.2 Analysis of skewness

Wang (2001) defines an analysis of skewness for problems in biology. Let X_{ij} be a measure on animal $j = 1, \dots, n_i$ from population $i = 1, \dots, I$. Here animals are nested within populations and the appropriate analysis of variance is:

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\bullet\bullet})^2 = \sum_{i=1}^{I} n_i(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^2 + \sum_{i=1}^{I}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\bullet})^2.$$

An analogous analysis of skewness is

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\bullet\bullet})^3 = \sum_{i=1}^{I} n_i(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^3 + \sum_{i=1}^{I}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\bullet})^3 + 3\sum_{i=1}^{I}(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\bullet})^2.$$

The terms above correspond to skewness of group means, skewness of observations within groups and a third term measuring the correlation of within group variance and the group mean. The relative sizes of these terms have been interpreted in terms of driven versus passive trends in evolutionary biology. The total skewness can be negative as can any of its terms.

The analysis is centered on $\bar{X}_{\bullet\bullet}$ which is not generally the minimizer of $\sum_i\sum_j |X_{ij} - m|^3$ over $m \in \mathbb{R}$. Similarly, $\bar{X}_{i\bullet}$ minimizes $\sum_j |X_{ij} - m|^2$ not $\sum_j |X_{ij} - m|^3$. In other words, this method is not based on generalizing the successive minimization property of ANOVA terms.

For functions on the unit cube, we can develop an analysis of skewness. A crossed decomposition is more appropriate than a nested one. Let $f(\boldsymbol{x}) = \mu + \sum_{u \neq \varnothing} f_u(\boldsymbol{x})$ be the ANOVA decomposition of f. Then

$$\int (f(\boldsymbol{x}) - \mu)^3 \, \mathrm{d}\boldsymbol{x} = \sum_{u\neq\varnothing}\sum_{v\neq\varnothing}\sum_{w\neq\varnothing}\int f_u(\boldsymbol{x}) f_v(\boldsymbol{x}) f_w(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

The product $f_u f_v f_w$ has mean zero if there is some index j that belongs to precisely one of the sets u, v, w. There can be more nonzero terms than nonempty subsets of $\mathcal{D}$. For example $f_{\{1,2\}}(\boldsymbol{x})f_{\{2,3\}}(\boldsymbol{x})f_{\{1,3\}}(\boldsymbol{x})$ need not integrate to zero. After eliminating the terms that must be zero, we find that $\int (f(\boldsymbol{x})-\mu)^3\,\mathrm{d}\boldsymbol{x}$ equals

$$\sum_{u\neq\varnothing}\int f_u(\boldsymbol{x})^3\,\mathrm{d}\boldsymbol{x} + \sum_{u\neq\varnothing}\sum_{\substack{v\neq u\\ v\neq\varnothing}}\sum_{z\subset u\cap v}\int f_u(\boldsymbol{x})f_v(\boldsymbol{x})f_{(u\Delta v)\cup z}\,\mathrm{d}\boldsymbol{x}.$$

For example, with $d=2$, there are 3 nonempty subsets of $\{1,2\}$ providing 27 combinations for u, v and w of which only 12 vanish, yielding

$$\begin{aligned}\int (f(\boldsymbol{x})-\mu)^3\,\mathrm{d}\boldsymbol{x} &= \int f^3_{\{1\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int f^3_{\{2\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int f^3_{\{1,2\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\quad + 3\int f_{\{1\}}(\boldsymbol{x})f^2_{\{1,2\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + 3\int f_{\{2\}}(\boldsymbol{x})f^2_{\{1,2\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\quad + 6\int f_{\{1\}}(\boldsymbol{x})f_{\{2\}}(\boldsymbol{x})f_{\{1,2\}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Our moment based method in Section 4 provide crossed decompositions for d dimensions and p'th powers. The terms are sums together into 2^d-1 effects.

4 Generalizing the Sobol' identity

Instead of generalizing the ANOVA to higher moments, we find it more convenient to directly generalize the identity (5) which yields $\mu^2+\underline{\tau}_u^2$. We are generalizing $\mu^2+\underline{\tau}_u^2$ instead of $\underline{\tau}_u^2$, because the minimizer of $\int |f(\boldsymbol{x})-m|^p\,\mathrm{d}\boldsymbol{x}$ over m, is the mean when $p=2$, but is otherwise not easy to identify.

Where (5) uses 2 points in $[0,1]^d$ with common $\boldsymbol{x}_u$, our generalization works via $p\geqslant 2$ such points. Define $\underline{\tau}^{(p)}$ via

$$\underline{\tau}_u^{(p)} + \mu^p = \int\cdots\int \prod_{k=1}^{p} f(\boldsymbol{x}_u{:}\boldsymbol{z}_{-u}^{(k)})\,\mathrm{d}\boldsymbol{x}\prod_{k=1}^{p}\mathrm{d}\boldsymbol{z}^{(k)} \tag{8}$$

where $\boldsymbol{z}^{(1)},\dots,\boldsymbol{z}^{(p)}\in[0,1]^d$. This integral is over $[0,1]^{(p+1)d}$ but only uses $|u|+p(d-|u|)$ components. For $p=2$, we get the usual Sobol' sensitivity indices (plus μ^2). The desirable property of (8) is that it is a multivariable integral and may be estimated by Monte Carlo or quasi-Monte Carlo sampling without requiring any numerical optimization.

When we seek to estimate $\underline{\tau}_u^{(p)}$ it is necessary to subtract an estimate of μ^p. One approach, generalizing an estimate studied in Janon et al. (2012) is to use

$$\widehat{\underline{\tau}}_u^{(p)} = \frac{1}{n}\sum_{i=1}^{n}\prod_{k=1}^{p} f(\boldsymbol{x}_{i,u}{:}\boldsymbol{z}_{i,-u}^{(k)}) - \hat{\mu}^p, \quad \text{where} \tag{9}$$

$$\hat{\mu} = \frac{1}{np} \sum_{i=1}^{n} \sum_{k=1}^{p} f(\boldsymbol{x}_{i,u} : \boldsymbol{z}_{i,-u}^{(k)}). \tag{10}$$

A second approach, generalizing an estimate in Mauntz (2002) and Saltelli (2002) takes

$$\widehat{\underline{\tau}}_u^{(p)} = \frac{1}{n} \sum_{i=1}^{n} \Big(\prod_{k=1}^{p} f(\boldsymbol{x}_{i,u} : \boldsymbol{z}_{i,-u}^{(k)}) - \prod_{k=1}^{p} f(\boldsymbol{z}_i^{(k)}) \Big), \tag{11}$$

a sample version of the identity

$$\underline{\tau}_u^{(p)} = \int \Bigg(\prod_{k=1}^{p} f\big(\boldsymbol{x}_u : \boldsymbol{z}_{-u}^{(k)}\big) - \prod_{k=1}^{p} f\big(\boldsymbol{z}_i^{(k)}\big) \Bigg) \,\mathrm{d}\boldsymbol{x} \prod_{k=1}^{p} \mathrm{d}\boldsymbol{z}^{(k)}.$$

Equation (11) provides unbiased estimates of $\underline{\tau}_u^{(p)}$. Even for $p = 2$ it is known that neither estimate (9) or (11) is always better than the other. For instance Owen (2012b) finds that (11) is more accurate in some examples with small $\underline{\tau}_u^2$, while (9) is better on some examples with large $\underline{\tau}_u^2$.

The most interesting cases are $p = 3$, which gives us a skewness measure for each subset of variables, and $p = 4$, the smallest even power above 2. For even integers $p \geqslant 4$ we get nonnegative measures that are increasing in u as shown below. We will use

$$\underline{f}_u(\boldsymbol{x}) = \sum_{v \subseteq u} f_v(\boldsymbol{x}) = \mathbb{E}\big(f(\boldsymbol{x}) \mid \boldsymbol{x}_u\big), \tag{12}$$

when $\boldsymbol{x} \sim \mathbf{U}[0,1]^d$.

Proposition 1. *For integer $p \geqslant 1$, $\underline{\tau}_u^{(p)} + \mu^p = \mathbb{E}\big(\underline{f}_u(\boldsymbol{x}_u)^p\big)$.*

Proof. Define $h(\boldsymbol{x}) = f(\boldsymbol{x}) - \underline{f}_u(\boldsymbol{x}_u)$ and $\boldsymbol{y}_k = \boldsymbol{x}_u : \boldsymbol{z}_{-u}^{(k)}$, for $k = 1, \dots, p$. Then $\mathbb{E}(h(\boldsymbol{y}_j) \mid \boldsymbol{x}_u) = 0$ and

$$\begin{aligned}
\mu^p + \underline{\tau}_u^{(p)} &= \mathbb{E}\Bigg(\mathbb{E}\Bigg(\prod_{k=1}^{p} (\underline{f}_u(\boldsymbol{x}_u) + h(\boldsymbol{y}_k)) \mid \boldsymbol{x}_u \Bigg) \Bigg) \\
&= \mathbb{E}\big(\mathbb{E}(\underline{f}_u(\boldsymbol{x}_u)^p \mid \boldsymbol{x}_u) \big) \\
&= \mathbb{E}\big(\underline{f}_u(\boldsymbol{x}_u)^p \big).
\end{aligned}$$

□

Theorem 1. *Let $f \in L^p[0,1]^d$ for an even integer $p \geqslant 2$. Then $\underline{\tau}_u^{(p)} \leqslant \underline{\tau}_v^{(p)}$ holds when $u \subseteq v \subseteq \mathcal{D}$.*

Proof. It suffices to consider the case where $v = u \cup \{j\}$ for $j \notin u$. Let $h(\boldsymbol{x}) = \underline{f}_v(\boldsymbol{x}_v) - \underline{f}_u(\boldsymbol{x}_u) = \sum_{w \subseteq u} f_{w \cup \{j\}}(\boldsymbol{x})$. Then by Proposition 1,

$$\mu^p + \underline{\tau}_v^{(p)} = \mathbb{E}(\underline{f}_v(\boldsymbol{x}_v)^p) = \mathbb{E}((\underline{f}_u(\boldsymbol{x}_u) + h(\boldsymbol{x}))^p)$$

$$
\begin{aligned}
&= \mathbb{E}(\mathbb{E}((\underline{f}_u(\boldsymbol{x}_u) + h(\boldsymbol{x}))^p \mid \boldsymbol{x}_u)) \\
&\geqslant \mathbb{E}(\mathbb{E}(\underline{f}_u(\boldsymbol{x}_u)^p \mid \boldsymbol{x}_u)) = \mu^p + \underline{\tau}_u^{(p)}
\end{aligned}
$$

by convexity of the function $\varphi(y) = y^p$. □

From Theorem 1, we see that $\underline{\tau}_u^{(p)}$ has some important properties for a subset importance quantity when p is an even integer. First $\underline{\tau}_u^{(p)} \geqslant \underline{\tau}_\varnothing^{(p)} = 0$, and so the importance of every subset is nonnegative. Second, increasing the number of components in a subset does not make the measure smaller. Both of these properties also hold for the measure $\underline{\tau}_u^{\varphi} = \mathbb{E}(\varphi(\underline{f}_u(\boldsymbol{x}))) - \varphi(\mu)$ for convex non-negative functions φ, but when $\varphi(y)$ is even power of y, we have a convenient estimation formula based on (8) that lets us avoid having to compute an estimate of $\underline{f}_u$.

Odd power variable measures like $\underline{\tau}_u^{(3)}$ do not have the nesting property of Theorem 1 and they can take negative values. Such negative values may be informative and interpretable. For example if $\underline{\tau}_{\{1\}}^{(3)} < 0$ while $\underline{\tau}_{\{2\}}^{(3)} > 0$ this may indicate that controlling x_1 is more important for attaining (or avoiding) very small values of f while x_2 is more important for large values of f.

5 Generalizing the synthesis

In this section we introduce a multilinear operator that allows a generalization of the synthesis approach to ANOVA. We use two different bases, Fourier and Walsh.

5.1 Fourier synthesis

For $0 \leqslant j < p$ let $f_j : [0,1]^d \to \mathbb{R}$ have a Fourier expansion

$$
f_j(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{f}_j(\boldsymbol{k}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}.
$$

For any $j \in \{0, 1, \dots, p-1\}$ let its successor be $j+ \equiv j + 1 \pmod{p}$ and its predecessor be $j- \equiv j - 1 \pmod{p}$. Our multilinear operator is

$$
\langle f_0, \dots, f_{p-1} \rangle_p = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j\bigl(\{(-1)^j (\boldsymbol{x}_j - \boldsymbol{x}_{j+})\}\bigr) \, \mathrm{d}\boldsymbol{x}_0 \cdots \mathrm{d}\boldsymbol{x}_{p-1}, \tag{13}
$$

where $\{\boldsymbol{z}\} = \boldsymbol{z} - \lfloor \boldsymbol{z} \rfloor$ is the fractional part of $\boldsymbol{z}$ (componentwise).

The following result is the fundamental lemma, giving a multilinear orthogonality property of the operator on Fourier functions.

Lemma 1. *Let $p \geqslant 2$ be an integer and $\boldsymbol{k}_0, \dots, \boldsymbol{k}_{p-1} \in \mathbb{Z}^d$ and let $\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}}$. Then*

$$
\langle \phi_{\boldsymbol{k}_0}, \dots, \phi_{\boldsymbol{k}_{p-1}} \rangle_p = \begin{cases} 1, & \boldsymbol{k}_j = (-1)^j \boldsymbol{k}_0, \quad j = 1, \dots, p-1 \\ 0, & \textit{otherwise.} \end{cases}
$$

Proof. For p even we have

$$\langle \phi_{\boldsymbol{k}_0}, \dots, \phi_{\boldsymbol{k}_{p-1}} \rangle_p = \int_{[0,1]^{dp}} \mathrm{e}^{2\pi \mathrm{i} \sum_{j=0}^{p-1} \boldsymbol{k}_j \cdot \{(-1)^j (\boldsymbol{x}_j - \boldsymbol{x}_{j+})\}} \, \mathrm{d}\boldsymbol{x}_0 \cdots \mathrm{d}\boldsymbol{x}_{p-1}$$
$$= \prod_{j=0}^{p-1} \int_{[0,1]^d} \mathrm{e}^{2\pi \mathrm{i} (-1)^j (\boldsymbol{k}_j + \boldsymbol{k}_{j-}) \cdot \boldsymbol{x}_j} \, \mathrm{d}\boldsymbol{x}_j$$

and for p odd we have

$$\langle \phi_{\boldsymbol{k}_0}, \dots, \phi_{\boldsymbol{k}_{p-1}} \rangle_p = \prod_{j=1}^{p-1} \int_{[0,1]^d} \mathrm{e}^{2\pi \mathrm{i} (-1)^j (\boldsymbol{k}_j + \boldsymbol{k}_{j-}) \cdot \boldsymbol{x}_j} \, \mathrm{d}\boldsymbol{x}_j \int_{[0,1]^d} \mathrm{e}^{2\pi \mathrm{i} (\boldsymbol{k}_0 - \boldsymbol{k}_{p-1}) \boldsymbol{x}_0} \, \mathrm{d}\boldsymbol{x}_0.$$

The integrals are 1 if $\boldsymbol{k}_j = (-1)^j \boldsymbol{k}_0$ for $0 \leqslant j < p$ and 0 otherwise, which implies the result. □

The function $\langle \cdot, \dots, \cdot \rangle_p$ is symmetric and multi-linear. For integers $p \geqslant 2$ we will use

$$\sigma_p(f) \equiv \langle f, \dots, f \rangle_p$$
$$= \sum_{\boldsymbol{k}_0, \dots, \boldsymbol{k}_{p-1} \in \mathbb{Z}^d} \prod_{j=0}^{p-1} \widehat{f}(\boldsymbol{k}_j) \int_{[0,1]^{dp}} \mathrm{e}^{2\pi \mathrm{i} \sum_{j=0}^{p-1} (-1)^j \boldsymbol{k}_j \cdot (\boldsymbol{x}_j - \boldsymbol{x}_{j+})} \, \mathrm{d}\boldsymbol{x}_0 \dots \mathrm{d}\boldsymbol{x}_{p-1}$$
$$= \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{f}(\boldsymbol{k})^{\lceil p/2 \rceil} \widehat{f}(-\boldsymbol{k})^{\lfloor p/2 \rfloor}.$$

If f is a real-valued function we have $\widehat{f}(-\boldsymbol{k}) = \overline{\widehat{f}(\boldsymbol{k})}$. If p is an even integer we therefore get

$$\langle f, \dots, f \rangle_p = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \big| \widehat{f}(\boldsymbol{k}) \big|^p.$$

The ANOVA decomposition $f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x})$, has terms

$$f_u(\boldsymbol{x}) = \sum_{\boldsymbol{k}_u \in \mathbb{Z}_*^{|u|}} \widehat{f}(\boldsymbol{k}_u \mathbin{:} \boldsymbol{0}_{-u}) \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k}_u \cdot \boldsymbol{x}_u}.$$

The diagonality of the multilinear operator (13) yields a p-fold orthogonality for the ANOVA terms:

Lemma 2. *Let f be as above and let $f = \sum_u f_u$ be the ANOVA decomposition of f. Then for all $u_0, \dots, u_{p-1} \subseteq \mathcal{D}$, such that there are $i, j \in \{0, \dots, p-1\}$ with $u_i \neq u_j$, we have*

$$\langle f_{u_0}, \dots, f_{u_{p-1}} \rangle_p = 0.$$

Proof. If $u_i \neq u_j$, then $\boldsymbol{k}_{u_j} \mathbin{:} \boldsymbol{0}_{-u_j} \neq -\boldsymbol{k}_{u_i} \mathbin{:} \boldsymbol{0}_{-u_i}$ for all $\boldsymbol{k}_{u_j} \in \mathbb{Z}_*^{|u_j|}$ and $\boldsymbol{k}_{u_i} \in \mathbb{Z}_*^{|u_i|}$. Then

$$\int_{[0,1]^d} \mathrm{e}^{2\pi \mathrm{i} (\boldsymbol{k}_{u_j} : \boldsymbol{0}_{-u_j} + \boldsymbol{k}_{u_i} : \boldsymbol{0}_{-u_i}) \cdot \boldsymbol{x}} \, \mathrm{d}\boldsymbol{x} = 0.$$

The result follows now from Lemma 1. □

Lemma 3. *Let f be as above and let $f = \sum_u f_u$ be the ANOVA decomposition of f. Then we have*

$$\sigma_p(f) = \sum_u \sigma_p(f_u).$$

Proof. Recall that $\langle f_{u_0}, \ldots, f_{u_{p-1}} \rangle_p = 0$ unless $u_0 = \cdots = u_{p-1}$. Therefore expanding $\sigma_p(f) = \langle f, \cdots, f \rangle_p$ yields

$$\sum_{u_0,\ldots u_{p-1} \subseteq \mathcal{D}} \langle f_{u_0}, \ldots, f_{u_{p-1}} \rangle_p = \sum_{u \subseteq \mathcal{D}} \langle f_u, \ldots, f_u \rangle_p = \sum_{u \subseteq \mathcal{D}} \sigma_p(f_u). \qquad \square$$

The aim is to estimate $\sigma_p(f_u)$ or sums of those. We investigate this in the following. For $u \subseteq \mathcal{D}$, define $\underline{\tau}_u^{[p]}$ via

$$\underline{\tau}_u^{[p]} + \mu^p = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f\big(\{(-1)^j(\boldsymbol{x}_{u,j} - \boldsymbol{x}_{u,j+})\} {:} \boldsymbol{y}_{-u,j}\big) \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{x}_{u,j} \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{y}_{u,j}. \tag{14}$$

Here $\underline{\tau}_\varnothing^{[p]} = 0$.

Theorem 2. *Let $f \in L^p[0,1]^d$, for integer $p \geqslant 2$, with ANOVA decomposition $f = \sum_u f_u$. Then for any $u \subseteq \mathcal{D}$ we have*

$$\underline{\tau}_u^{[p]} + \mu^p = \sum_{v \subseteq u} \sigma_p(f_v).$$

Proof. Using the Fourier series representation of f and Lemma 1 we obtain

$$\underline{\tau}_u^{[p]} + \mu^p = \sum_{\boldsymbol{k}_u \in \mathbb{Z}^{|u|}} |\widehat{f}(\boldsymbol{k}_u {:} \boldsymbol{0}_{-u})|^p = \sum_{v \subseteq u} \sigma_p(f_v). \qquad \square$$

Theorem 2 shows that the importance measures $\underline{\tau}_u^{[p]}$ are sums of contributions $\sigma_p(f_v)$ from $v \subseteq u$. This generalizes a property of the ANOVA to $p \geqslant 2$.

Theorem 2 can be generalized in the following way. Let $f_0, \ldots, f_{p-1}$ be functions in $L^p[0,1]^d$ for integer $p \geqslant 2$ with Fourier coefficients $\widehat{f}_j(\boldsymbol{k})$, and $\mu_j = \int f_j(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$. Next we set

$$\begin{aligned} &\underline{\tau}_u^{[p]}(f_0, \ldots, f_{p-1}) + \prod_{j=0}^{p-1} \mu_j \\ &\equiv \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j\big((-1)^j\{\boldsymbol{x}_{u,j} - \boldsymbol{x}_{u,j+}\} {:} \boldsymbol{y}_{-u,j}\big) \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{x}_{u,j} \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{y}_{-u,j}. \end{aligned}$$

Then

$$\underline{\tau}_u^{[p]}(f_0, \ldots, f_{p-1}) + \prod_{j=0}^{p-1} \mu_j = \sum_{\boldsymbol{k}_u \in \mathbb{Z}^{|u|}} \prod_{j=0}^{p-1} \widehat{f}_j((-1)^j \boldsymbol{k}_u {:} \boldsymbol{0}_{-u}).$$

5.2 Walsh synthesis

Here we replace the Fourier functions by the Walsh functions in an integer base $b \geqslant 2$. For $b = 2$, the coefficients of Walsh functions are real values. The index set is $\mathbb{I} = \mathbb{N}_0$ and then $\mathbb{I}_* = \mathbb{N}$.

For a non-negative integer k with base b representation

$$k = \kappa_{a-1} b^{a-1} + \cdots + \kappa_1 b + \kappa_0,$$

with $\kappa_i \in \{0, 1, \ldots, b-1\}$, we define the Walsh function $\mathrm{wal}_k : [0,1) \to \{z \in \mathbb{C} : |z| = 1\}$ by

$$\mathrm{wal}_k(x) := \mathrm{e}^{2\pi \mathrm{i}(x_1 \kappa_0 + \cdots + x_a \kappa_{a-1})/b},$$

for $x \in [0,1)$ with base b representation $x = x_1 b^{-1} + x_2 b^{-2} + \cdots$ (unique in the sense that infinitely many of the x_i must be different from $b-1$).

For dimension $s \geqslant 2$, $\boldsymbol{x} = (x_1, \ldots, x_s) \in [0,1)^s$ and $\boldsymbol{k} = (k_1, \ldots, k_s) \in \mathbb{N}_0^s$ we define $\mathrm{wal}_{\boldsymbol{k}} : [0,1)^s \to \{z \in \mathbb{C} : |z| = 1\}$ by

$$\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) := \prod_{j=1}^{s} \mathrm{wal}_{k_j}(x_j).$$

For more information on Walsh functions see Chrestenson (1955); Fine (1949); Walsh (1923).

Let $f_j : [0,1]^d \to \mathbb{R}$ have a Walsh series expansion of the form

$$f_j(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} \widehat{f}_{\mathrm{wal}}(\boldsymbol{k}) \mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}).$$

For $x, y \in \{z \in \mathbb{R} : z \geqslant 0\}$ with base b expansion $x = \sum_{i=w}^{-\infty} x_i b^i$ and $y = \sum_{i=w}^{-\infty} y_i b^i$ (unique in the sense that infinitely many of the x_i and y_i must be different from $b-1$) we set $x \ominus y = \sum_{i=w}^{-\infty} z_i b^i$ where $z_i = x_i - y_i \pmod b$ and $z_i \in \{0, \ldots, b-1\}$. For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ we define the operation $\ominus$ componentwise. Similarly we set $\boldsymbol{x} \oplus \boldsymbol{y}$ where we change the definition of z_i to $z_i = x_i + y_i \pmod b$. We define $\ominus \boldsymbol{x} = \boldsymbol{0} \ominus \boldsymbol{x}$ and $(\ominus 1)^j \boldsymbol{x} = \boldsymbol{x}$ if j is even and $\ominus \boldsymbol{x}$ otherwise.

We now define

$$\langle f_0, \ldots, f_{p-1} \rangle_{p,\mathrm{wal}} = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j\big((\ominus 1)^j (\boldsymbol{x}_j \ominus \boldsymbol{x}_{j+})\big) \, \mathrm{d}\boldsymbol{x}_0 \cdots \mathrm{d}\boldsymbol{x}_{p-1}.$$

With this definition we also have the fundamental lemma for the Walsh system.

Lemma 4. *Let $\boldsymbol{k}_0, \ldots, \boldsymbol{k}_{p-1} \in \mathbb{N}_0^d$. Then*

$$\langle \mathrm{wal}_{\boldsymbol{k}_0}, \ldots, \mathrm{wal}_{\boldsymbol{k}_{p-1}} \rangle_{p,\mathrm{wal}} = \begin{cases} 1, & \boldsymbol{k}_j = \ominus^j \boldsymbol{k}_0, \quad j = 1, \ldots, p-1 \\ 0, & \textit{otherwise.} \end{cases}$$

All the remaining results and definitions can therefore be obtained in an analogous manner. In particular for functions f with ANOVA decomposition $f = \sum_u f_u$ we have

$$\sigma_{p,\mathrm{wal}}(f) = \sum_{u \subseteq \mathcal{D}} \sigma_{p,\mathrm{wal}}(f_u),$$

where

$$\sigma_{p,\mathrm{wal}}(f) \equiv \langle f, \dots, f \rangle_{p,\mathrm{wal}} = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} [\widehat{f}_{\mathrm{wal}}(\boldsymbol{k})]^{\lceil p/2 \rceil} [\widehat{f}_{\mathrm{wal}}(\ominus \boldsymbol{k})]^{\lfloor p/2 \rfloor}.$$

If p is even and f a real-valued function, then we get

$$\sigma_{p,\mathrm{wal}}(f) \equiv \langle f, \dots, f \rangle_{p,\mathrm{wal}} = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} \left| \widehat{f}_{\mathrm{wal}}(\boldsymbol{k}) \right|^p.$$

Lemma 5. *Let f be as above and let $f = \sum_u f_u$ be the ANOVA decomposition of f. Then we have*

$$\sigma_{p,\mathrm{wal}}(f) = \sum_u \sigma_{p,\mathrm{wal}}(f_u).$$

The proof of Lemma 5 follows by the same arguments as the proof of Lemma 3.

We may estimate $\sigma_{p,\mathrm{wal}}(f_u)$ or their sums in the same way we did for their Fourier analogues $\sigma_p(f_u)$. For $u \subseteq \mathcal{D}$, define $\underline{\tau}^{[p]}_{u,\mathrm{wal}}$ via

$$\underline{\tau}^{[p]}_{u,\mathrm{wal}} + \mu^p_{\mathrm{wal}} = \int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f\big(\{(\ominus 1)^j (\boldsymbol{x}_{u,j} \ominus \boldsymbol{x}_{u,j+})\}{:}\boldsymbol{y}_{-u,j}\big) \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{x}_{u,j} \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{y}_{u,j}.$$

Here $\underline{\tau}^{[p]}_{\varnothing,\mathrm{wal}} = 0$.

Theorem 3. *Let $f \in L^p[0,1]^d$, for integer $p \geqslant 2$, with ANOVA decomposition $f = \sum_u f_u$. Then for any $u \subseteq \mathcal{D}$ we have*

$$\underline{\tau}^{[p]}_{u,\mathrm{wal}} + \mu^p_{\mathrm{wal}} = \sum_{v \subseteq u} \sigma_{p,\mathrm{wal}}(f_v).$$

The proof of this result follows along the same lines as the proof of Theorem 2.

In general, for $p > 2$, $\sigma_p(f)$ and $\sigma_{p,\mathrm{wal}}(f)$ are different and the Walsh measure will depend on the base b that was used. Parseval's identity implies that $\sigma_2(f) = \sigma_{2,\mathrm{wal}}(f)$.

5.3 Change of variable and dimension reduction

Our p-fold inner products are defined through a pd dimensional integral. But they are equivalent to a $(p-1)d$ dimensional integral.

Lemma 6. *For integers $p \geqslant 2$ and $d \geqslant 1$, let $f_0, f_1, \dots, f_{p-1} \in L^p[0,1)^d$. Then*

$$\int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j(\{(-1)^j(\boldsymbol{x}_j - \boldsymbol{x}_{j+})\}) \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{x}_j$$

$$= \int_{[0,1]^{dp-d}} f_0(\boldsymbol{y}_0) f_1(\boldsymbol{y}_1) \cdots f_{p-2}(\boldsymbol{y}_{p-2}) f_{p-1}\big(\{\boldsymbol{y}_0 - \boldsymbol{y}_1 + \cdots + (-1)^{p-2}\boldsymbol{y}_{p-2}\}\big) \prod_{j=0}^{p-2} \mathrm{d}\boldsymbol{y}_j.$$

Proof. We prove it for $p = 4$; the general case uses the same argument. For $\boldsymbol{x}_0, \dots, \boldsymbol{x}_3 \in [0,1)^d$ let $\boldsymbol{y}_0, \dots, \boldsymbol{y}_3$ be defined by

$$\begin{pmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_0 \\ \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix} \ \mathsf{mod}\ 1$$

where both the matrix multiplication and the modulus are taken componentwise. This transformation has Jacobian 1 almost everywhere. To simplify the integrals, we extend each f_j to a periodic function on $\mathbb{R}^d$, allowing us to remove the $\{\cdots\}$ operation. Making the change of variable,

$$\begin{aligned} &\iiiint f_0(\boldsymbol{x}_0 - \boldsymbol{x}_1) f_1(\boldsymbol{x}_2 - \boldsymbol{x}_1) f_2(\boldsymbol{x}_2 - \boldsymbol{x}_3) f_3(\boldsymbol{x}_0 - \boldsymbol{x}_3)\, \mathrm{d}\boldsymbol{x}_0\, \mathrm{d}\boldsymbol{x}_1\, \mathrm{d}\boldsymbol{x}_2\, \mathrm{d}\boldsymbol{x}_3 \\ &= \iiiint f_0(\boldsymbol{y}_0) f_1(\boldsymbol{y}_1) f_2(\boldsymbol{y}_2) f_3(\boldsymbol{y}_0 - \boldsymbol{y}_1 + \boldsymbol{y}_2)\, \mathrm{d}\boldsymbol{x}_0\, \mathrm{d}\boldsymbol{x}_1\, \mathrm{d}\boldsymbol{x}_2\, \mathrm{d}\boldsymbol{y}_3 \\ &= \iiint f_0(\boldsymbol{y}_0) f_1(\boldsymbol{y}_1) f_2(\boldsymbol{y}_2) f_3(\boldsymbol{y}_0 - \boldsymbol{y}_1 + \boldsymbol{y}_2)\, \mathrm{d}\boldsymbol{y}_0\, \mathrm{d}\boldsymbol{y}_1\, \mathrm{d}\boldsymbol{y}_2. \end{aligned}$$ □

Lemma 6 also applies for the Walsh case. We have

$$\int_{[0,1]^{dp}} \prod_{j=0}^{p-1} f_j(\{(\ominus 1)^j(\boldsymbol{x}_j \ominus \boldsymbol{x}_{j+})\}) \prod_{j=0}^{p-1} \mathrm{d}\boldsymbol{x}_j$$

$$= \int_{[0,1]^{dp-d}} f_0(\boldsymbol{y}_0) f_1(\boldsymbol{y}_1) \cdots f_{p-2}(\boldsymbol{y}_{p-2}) f_{p-1}\big(\{\boldsymbol{y}_0 \ominus \boldsymbol{y}_1 \oplus \cdots \oplus (\ominus 1)^{p-1}\boldsymbol{y}_{p-1}\}\big) \prod_{j=0}^{p-2} \mathrm{d}\boldsymbol{y}_j.$$

5.4 Weighted coefficients

The quantity $\langle f, f, \dots, f, g\rangle_{p+1}$ is also of interest for special choices of the function g. The result is to give weighted sums of powers of the Fourier (or Walsh) coefficients. We take p to be an odd integer and g to be a weighting function.

Of particular interest is the Dirichlet kernel

$$D_N(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \{-N,\dots,N\}^d} \mathrm{e}^{2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}} = \prod_{j=1}^{d} \frac{\sin(2\pi(N+1/2)x_j)}{\sin(\pi x_j)}.$$

If $x_j = 0$ or 1 we set $\sin(2\pi(N+1/2)x_j)/\sin(\pi x_j) := 2N+1$. For an odd integer $p > 1$ we have

$$\langle f, \dots, f, D_N\rangle_p = \sum_{\boldsymbol{k}\in\{-N,\dots,N\}^d} \bigl|\widehat{f}(\boldsymbol{k})\bigr|^{p-1}.$$

The result is a non-negative importance measure for f apart from its very highest spatial frequencies. Further, for $\boldsymbol{m}\in\mathbb{Z}^d$ we have

$$\langle f, \dots, f, D_N \mathrm{e}^{2\pi\mathrm{i}\boldsymbol{m}\cdot\boldsymbol{x}}\rangle_p = \sum_{\boldsymbol{k}\in\{-N,\dots,N\}^d} \bigl|\widehat{f}(\boldsymbol{k}+\boldsymbol{m})\bigr|^{p-1}.$$

The Dirichlet kernel for the Walsh system in base b is

$$D_{m,\mathrm{wal}}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\{0,\dots,b^m-1\}^d} \mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^d 1_{[0,b^{-m})}(x_j).$$

Thus for odd integer $p > 1$ we have

$$\langle f, \dots, f, D_{m,\mathrm{wal}}\rangle_{p,\mathrm{wal}} = \sum_{\boldsymbol{k}\in\{0,\dots,b^m-1\}^d} \bigl|\widehat{f}_{\mathrm{wal}}(\boldsymbol{k})\bigr|^{p-1}.$$

Further, for $\boldsymbol{a}\in\mathbb{N}_0^d$ we have

$$\langle f, \dots, f, D_{m,\mathrm{wal}}\mathrm{wal}_{\boldsymbol{a}}\rangle_{p,\mathrm{wal}} = \sum_{\boldsymbol{k}\in\{0,\dots,b^m-1\}^d} \bigl|\widehat{f}_{\mathrm{wal}}(\boldsymbol{k}\oplus\boldsymbol{a})\bigr|^{p-1}.$$

6 Special case functions

Here we consider some simple functional forms for which our analysis can be carried out in closed form. The first ones are functions of product form, including rectangular spikes. We will see the effects of third and fourth moments on the $\underline{\tau}_u^{(p)}$ and the effects of spectral sparsity on $\underline{\tau}_u^{[p]}$. The second are additive functions where we will see the spectral method does not introduce any apparent interactions.

The original Sobol' indices relate to variance components via a Moebius relation

$$\sigma_u^2 = \sum_{v\subseteq u} (-1)^{|u-v|}\underline{\tau}_v^2,$$

for $u \neq \varnothing$. Recalling that $\underline{\tau}_u^{(p)}$, $\underline{\tau}_u^{[p]}$ and $\underline{\tau}_{u,\mathrm{wal}}^{[p]}$ are generalizations of $\mu^2 + \underline{\tau}_u^2$, we can define analogues of variance components via

$$\sigma_u^{(p)} = \sum_{v\subseteq u} (-1)^{|u-v|}\underline{\tau}_v^{(p)}, \tag{15}$$

$$\sigma_u^{[p]} = \sum_{v\subseteq u} (-1)^{|u-v|}\underline{\tau}_v^{[p]}, \quad \text{and} \tag{16}$$

$$\sigma_{u,\mathrm{wal}}^{[p]} = \sum_{v\subseteq u} (-1)^{|u-v|}\underline{\tau}_{v,\mathrm{wal}}^{[p]}, \tag{17}$$

for $u \neq \varnothing$. We also have $\sigma_\varnothing^{(p)} = \sigma_\varnothing^{[p]} = \sigma_{\varnothing,\mathrm{wal}}^{[p]} = 0$.

6.1 Product functions

Product functions are frequently used as examples for sensitivity measures. A notable example is Sobol' (1993). Throughout this subsection we suppose that

$$f(\boldsymbol{x}) = \prod_{j=1}^{d} h_j(x_j) \equiv \prod_{j=1}^{d} (\mu_j + \tau_j g_j(x_j)) \tag{18}$$

for real-valued functions g_j and h_j defined on $[0,1]$. The functions g_j satisfy $\int_0^1 g_j(x)\,\mathrm{d}x = 0$ and $\int_0^1 g_j(x)^2\,\mathrm{d}x = 1$.

The ANOVA components of a product function are $\sigma_u^2 = \prod_{j\in u} \tau_j^2 \prod_{j\notin u} \mu_j^2$ for $u \neq \varnothing$. For a product function $\mu^2 + \underline{\tau}_u^2 = \prod_{j\in u}(\mu_j^2 + \tau_j^2) \prod_{j\notin u} \mu_j^2$. An important subset of variables must include any j with $\mu_j = 0$. When $\mu \neq 0$ we may write

$$\mu^2 + \underline{\tau}_u^2 = \mu^2 \prod_{j\in u} (1 + \tau_j^2/\mu_j^2)$$

and then see that coefficients of variation $\upsilon_j = \tau_j/\mu_j$ govern importance.

We need $\int_0^1 |f(x)|^p\,\mathrm{d}x < \infty$ to make the importance measures finite. We will use $\gamma_j = \int_0^1 g_j^3(x)\,\mathrm{d}x$ and $\kappa_j = \int_0^1 g_j^4(x)\,\mathrm{d}x$ which we assume are finite. If $x \sim \mathbf{U}(0,1)$, then γ_j is the skewness of $g_j(x)$ and $\kappa_j - 3$ is the kurtosis.

Generalizing the Fourier and Walsh syntheses

To generalize the Fourier synthesis we write $h_j(x) = \sum_{k\in\mathbb{Z}} \widehat{h}_j(k) e^{2\pi i k x}$ (in mean square) for $\widehat{h}_j(k) = \int_0^1 h_j(x) e^{-2\pi i k x}\,\mathrm{d}x$. We note that $\mu = \prod_j \mu_j$ where $\mu_j = \widehat{h}_j(0)$. Now $\widehat{f}(\boldsymbol{k}) = \prod_{j=1}^d \widehat{h}_j(k_j)$ and for even $p \geqslant 2$

$$\underline{\tau}_u^{[p]} + \mu^p = \prod_{j\notin u} |\mu_j|^p \sum_{\boldsymbol{k}_u \in \mathbb{Z}^{|u|}} \prod_{j\in u} |\widehat{h}_j(k_j)|^p = \prod_{j\notin u} |\mu_j|^p \prod_{j\in u} \Biggl(\sum_{k_j\in\mathbb{Z}} |\widehat{h}_j(k_j)|^p \Biggr).$$

Using the alternating sum (16) and simplifying, we obtain

$$\sigma_u^{[p]} = \prod_{j\notin u} |\mu_j|^p \prod_{j\in u} \Biggl(\sum_{k_j\in\mathbb{Z}_*} |\widehat{h}_j(k_j)|^p \Biggr)$$

for $u \neq \varnothing$. The effect is to change $\mathbb{Z}$ to $\mathbb{Z}_*$ in the sums.

Given two functions h_j with the same variance, the measure $\sum_{k_j\in\mathbb{Z}_*} |\widehat{h}_j(k_j)|^p$, for $p > 2$, is a measure of sparsity for the spectrum. It does not favor either high or low frequencies. To put more emphasis on high or low frequencies one could use weighted coefficients as outlined in subsection 5.4.

Analogous formulae hold for the Walsh synthesis. Now we write the factors of f as $h_j(x) = \sum_{k\in\mathbb{N}_0} \widehat{h}_{j,\mathrm{wal}}(k) \mathrm{wal}_k(x)$ for $\widehat{h}_{j,\mathrm{wal}}(k) = \int_0^1 h_j(x) \overline{\mathrm{wal}_k}(x)\,\mathrm{d}x$. Here

$\mu_{\mathrm{wal}} = \prod_j \mu_{j,\mathrm{wal}}$ where $\mu_{j,\mathrm{wal}} = \widehat{h}_{j,\mathrm{wal}}(0)$ and $\widehat{f}_{\mathrm{wal}}(\boldsymbol{k}) = \prod_{j=1}^d \widehat{h}_{j,\mathrm{wal}}(k_j)$. For even $p \geqslant 2$ the same argument that we used in the Fourier case leads to

$$\sigma^{[p]}_{u,\mathrm{wal}} = \prod_{j\notin u} |\mu_{j,\mathrm{wal}}|^p \prod_{j\in u}\biggl(\sum_{k_j\in\mathbb{N}} |\widehat{h}_{j,\mathrm{wal}}(k_j)|^p\biggr).$$

for $u \neq \varnothing$.

Generalizing the Sobol' identity

When we generalize the Sobol' identity we get

$$\underline{\tau}^{(p)}_u + \mu^p = \int \prod_{k=1}^p f(\boldsymbol{x}_u{:}\boldsymbol{z}^{(k)}_{-u})\mathrm{d}\boldsymbol{x} \prod_{k=1}^p \mathrm{d}\boldsymbol{z}^{(k)} = \prod_{j\in u}\int_0^1 h_j(x_j)^p \,\mathrm{d}x_j \prod_{j\notin u}\mu_j^p.$$

Where the Fourier synthesis had a p'th moment $\sum_{k_j\in\mathbb{Z}} |\widehat{h}_j(k_j)|^p$ of Fourier coefficients, this approach has an ordinary p'th moment $\int_0^1 h_j(x)^p\,\mathrm{d}x$. Using the alternating sum (15) we obtain

$$\sigma^{(p)}_u = \prod_{j\notin u}\mu_j^p \prod_{j\in u}\Bigl(\int_0^1 h_j(x)^p\,\mathrm{d}x - \mu_j^p\Bigr)$$

for $u \neq \varnothing$.

For the generalized Sobol' identity we can make use of the moments γ_j and κ_j of h_j. The special cases of most interest have $p = 3$ or 4. For $p = 3$

$$\int_0^1 h_j(x)^3\,\mathrm{d}x = \mu_j^3 + 3\mu_j\tau_j^2 + \gamma_j\tau_j^3$$

and so for $u \neq \varnothing$,

$$\begin{aligned}\underline{\tau}^{(3)}_u &= \prod_{j\notin u}\mu_j^3 \prod_{j\in u}\bigl(\mu_j^3 + 3\mu_j\tau_j^2 + \gamma_j\tau_j^3\bigr) - \mu^3, \quad \text{and}\\ \sigma^{(3)}_u &= \prod_{j\notin u}\mu_j^3 \prod_{j\in u}\tau_j^2\bigl(3\mu_j + \gamma_j\tau_j\bigr).\end{aligned}$$

The $\sigma^{(3)}_u$ are 'components of skewness' analogues of the components of variance σ^2_u. Some of these components may be negative. If every $\mu_j > 0$ and every $\tau_j > 0$, then a negative component of skewness arises if $3\mu_j + \gamma_j\tau_j < 0$ holds for an odd number of indices $j \in u$.

Product functions illustrate one challenge with importance measures taking negative values. The same variable x_j can drive the function towards negative values through one component $\sigma^{(3)}_u$ while driving it towards positive values through another component $\sigma^{(3)}_v$. Similarly, whether the total effect $\underline{\tau}^{(3)}_u$ is positive or negative depends on the signs of μ_j for $j \notin u$. These features make $p = 3$ hard to interpret.

For $p = 4$, we find

$$\int_0^1 h_j(x)^4 \,\mathrm{d}x = \mu_j^4 + 6\mu_j^2\tau_j^2 + 4\mu_j\gamma_j\tau_j^3 + \kappa_j\tau_j^4$$

and so for $u \neq \varnothing$,

$$\underline{\tau}_u^{(4)} = \prod_{j\notin u}\mu_j^4 \prod_{j\in u}\bigl(\mu_j^4 + 6\mu_j^2\tau_j^2 + 4\mu_j\gamma_j\tau_j^3 + \kappa_j\tau_j^4\bigr) - \mu^4, \quad \text{and}$$
$$\sigma_u^{(4)} = \prod_{j\notin u}\mu_j^4 \prod_{j\in u}\tau_j^2\bigl(6\mu_j^2 + 4\mu_j\gamma_j\tau_j + \kappa_j\tau_j^2\bigr).$$

If $j \notin u \neq \varnothing$ and $\mu_j \neq 0$, then

$$\frac{\sigma_{u\cup\{j\}}^{(4)}}{\sigma_u^{(4)}} = \upsilon_j^2\bigl(6 + 4\gamma_j\upsilon_j + \upsilon_j^2\kappa_j\bigr).$$

where $\upsilon_j = \tau_j/\mu_j$ is the j'th coefficient of variation.

A variable with a large absolute coefficient of variation $|\upsilon_j|$ tends to raise all of the $\sigma_u^{(4)}$ in which it participates just as it does for the $p = 2$ ANOVA case. Additionally a variable with large fourth moment κ_j becomes more important. Variables with large skewness γ_j become more important if γ_j has the same sign as μ_j but less important if the opposite holds. Both of these findings are intuitively reasonable when we are interested in driving $|f|$ to its largest values.

6.2 Indicators of rectangles

A special case of the product functions are indicator (characteristic) functions of hyperrectangles. These have $h_j(x) = 1$ for $x_j \in [x_{j*}, x_{j*} + \epsilon_j)$ and $h_j(x) = 0$ for $x \in [0, 1) \setminus [x_{j*}, x_{j*} + \epsilon_j)$, so that $f(\boldsymbol{x})$ is the indicator of a hyperrectangle with volume $\prod_j \epsilon_j$. For a binary function, all of the x_j have to be in their respective intervals for the function to take the high value. This means that we should expect important interactions. To model a spiky function we would have all of the ϵ_j be small. Then the most important one should be the smallest one. Here we let $\epsilon = \mu = \prod_{j=1}^d \epsilon_j$.

The generalization of Sobol's identity works entirely with moments of h_j and so without loss of generality $h_j(x) = 1$ for $x < \epsilon_j$ and is 0 otherwise. The generalization of the Walsh-based synthesis is not invariant to the interval one chooses. In this setting we prefer the Fourier-based synthesis. Shifting the interval from $[0, \epsilon_j)$ to $[x_{*j}, x_{*j} + \epsilon_j)$ for $0 \leqslant x_{0j} \leqslant 1 - \epsilon_j$ changes the phase but not the modulus of $\widehat{h}_j(k)$ leaving the importance measures unchanged when $p \geqslant 2$ is even.

For the generalized Sobol' index construction we find for $u \neq \varnothing$

$$\underline{\tau}_u^{(p)} = \prod_{j\notin u}\epsilon_j^p \prod_{j\in u}\epsilon_j - \epsilon^p = \epsilon^p\Bigl(\prod_{j\in u}\epsilon_j^{-(p-1)} - 1\Bigr), \quad \text{and,}$$

$$\sigma_u^{(p)} = \prod_{j\notin u} \epsilon_j^p \prod_{j\in u} (\epsilon_j - \epsilon_j^p) = \epsilon^p \prod_{j\in u} (\epsilon_j^{-(p-1)} - 1).$$

Variables with smaller ϵ_j are more important than those with larger ϵ_j and the effect is magnified at larger p. Both $\underline{\tau}_u(p)$ and $\sigma_u^{(p)}$ are always nonnegative for integers $p \geqslant 2$ without requiring p to be even.

We now consider the Fourier synthesis for even $p \geqslant 2$. After applying some trigonometric identities, we find that the key quantity there, replacing $\epsilon_j - \epsilon_j^p$ satisfies

$$\sum_{k\in\mathbb{Z}_*} |\widehat{h}_j(k)|^p = 2\sum_{k=1}^{\infty} \Bigl(\frac{\sin(\pi k\epsilon_j)}{\pi k}\Bigr)^p \equiv T_p(\epsilon_j).$$

Thus $\sigma_u^{[p]} = \epsilon^p \prod_{j\in u} T_p(\epsilon_j)/\epsilon_j^p$. Lemma 6 gives some insight into T_p for $\epsilon_j < 1/2$ as follows. For $p = 4$, $\underline{\tau}_u^{[4]} + \mu^4 = \prod_{j\in u} Q_4(\epsilon_j)$ where for $f(x) = 1_{x<\epsilon}$ and $0 < \epsilon < 1/2$,

$$\begin{aligned}
Q_4(\epsilon) &= \int_0^1\int_0^1\int_0^1 f(y_0)f(y_1)f(y_2)f(\{y_0 - y_1 + y_2\})\,\mathrm{d}y_0\,\mathrm{d}y_1\,\mathrm{d}y_2\\
&= \int_0^\epsilon\int_0^\epsilon\int_0^\epsilon 1_{\{y_0-y_1+y_2\}<\epsilon}\,\mathrm{d}y_0\,\mathrm{d}y_1\,\mathrm{d}y_2\\
&= \frac{2}{3}\epsilon^3.
\end{aligned}$$

As a result we have the identity $T_4(\epsilon) = \frac{2}{3}\epsilon^3 - \epsilon^4$, and so for $u \neq \varnothing$,

$$\begin{aligned}
\underline{\tau}_u^{[4]} &= \prod_{j\notin u} \epsilon_j^4 \prod_{j\in u} \frac{2}{3}\epsilon_j^3 - \epsilon^4 = \epsilon^4\Bigl(\prod_{j\in u} \frac{2}{3}\epsilon_j^{-1} - 1\Bigr), \quad \text{and,}\\
\sigma_u^{[4]} &= \epsilon^4 \prod_{j\in u}\Bigl(\frac{2}{3}\epsilon_j^{-1} - 1\Bigr).
\end{aligned}$$

For even $p \geqslant 2$ we will find a quantity $Q_p(\epsilon)$ like Q_4 is a $p-1$ dimensional volume proportional to ϵ^{p-1}. As a result, the Fourier synthesis will use importance factors which grow as ϵ_j^{-1} compared to ϵ_j^{-p+1} for the moment method.

6.3 Additive functions

It frequently happens that high dimensional functions enountered in practice are very nearly additive. For example Caflisch et al. (1997) find that a 360 dimensional function motivated by a financial valuation problem is very nearly an additive function of its inputs. It is desirable that a measure of variable importance for additive functions should only give nonzero importance to singletons $u = \{j\}$.

Here we consider additive functions

$$f(\boldsymbol{x}) = \mu + \sum_{j=1}^{d} h_j(x_j) \tag{19}$$

where $\int_0^1 h_j(x)\,\mathrm{d}x = 0$, $\int_0^1 h_j(x)^2\,\mathrm{d}x = \tau_j^2$, $\int_0^1 h_j(x)^3\,\mathrm{d}x = \gamma_j$, and $\int_0^1 h_j(x)^4\,\mathrm{d}x = \kappa_j$.

For even integers $p \geqslant 2$ we find that $\sigma^{[p]}_{\{j\}} = \sum_{k\neq 0} |\widehat{h}_j(k)|^p$ and $\sigma^{[p]}_{\{j\},\mathrm{wal}} = \sum_{k\neq 0} |\widehat{h}_j(k)|^p$ are the only nonzero components.

For integer $p \geqslant 2$,

$$\underline{\tau}_u^{(p)} + \mu^p = \int \prod_{k=1}^p \Bigl[\mu + \sum_{j\in u} h_j(x_j) + \sum_{j\notin u} h_j(y_j^{(k)})\Bigr]\,\mathrm{d}\boldsymbol{x} \prod_{k=1}^p \mathrm{d}\boldsymbol{y}^{(k)}$$
$$= \int \Bigl[\mu + \sum_{j\in u} h_j(x_j)\Bigr]^p \mathrm{d}\boldsymbol{x}.$$

For $p = 3$, $\underline{\tau}_u^{(3)} + \mu^3 = \mu^3 + 3\mu\sum_{j\in u}\tau_j^2 + \sum_{j\in u}\gamma_j$, so that $\underline{\tau}_u^{(3)} = \sum_{j\in u}(\mu\tau_j^2 + \gamma_j)$. Next

$$\sigma_u^{(3)} = \sum_{v\subseteq u} (-1)^{|u-v|} \sum_{j\in v} (3\mu\tau_j^2 + \gamma_j).$$

Reversing the order of summation, we find that $\sigma_u^{(3)} = 0$ for $|u| > 2$ and otherwise

$$\sigma_{\{j\}}^{(3)} = 3\mu\tau_j^2 + \gamma_j,$$

compared to $\sigma_{\{j\}}^{(2)} = \tau_j^2$. We see that the only nonzero components of skewness for an additive function are for singletons.

The same simplification does not hold in general. For $p = 4$,

$$\underline{\tau}_u^{(4)} + \mu^4 = \mu^4 + 6\mu^2\sum_{j\in u}\tau_j^2 + 4\mu\sum_{j\in u}\gamma_j + \sum_{j\in u}\kappa_j + \sum_{j\in u}\sum_{k\in u-\{j\}}\tau_j^2\tau_k^2, \quad \text{so,}$$
$$\underline{\tau}_u^{(4)} = \sum_{j\in u}\bigl(6\mu^2\tau_j^2 + 4\mu\gamma_j + \kappa_j - \tau_j^4\bigr) + \Bigl(\sum_{j\in u}\tau_j^2\Bigr)^2.$$

As a result

$$\sigma_u^{(4)} = \begin{cases} 6\mu^2\tau_j^2 + 4\mu\gamma_j + \kappa_j - \tau_j^4, & u = \{j\} \\ 2\tau_j^2\tau_k^2, & u = \{j,k\},\ j \neq k \\ 0, & |u| > 2. \end{cases}$$

7 Discussion

We have shown that it is possible to generalize the ANOVA decomposition to higher order methods. Working directly with either Sobol's identities or with a synthesis of Fourier or Walsh terms both lead to measures that can be estimated by quadrature. For even values p the generalizations give non-negative importance measures. For odd values of p the Dirichlet kernel trick recovers non-negative importance measures for the Fourier and Walsh approaches. On test functions that we can study analytically, we see that these measures can identify variables which drive the function towards its extreme values.

Acknowledgments

We thank Gary Tang for bringing the problem to our attention. This work was supported in part by grant DMS-0906056 from the U.S. National Science Foundation. J. D. is supported in part by a Queen Elizabeth 2 Fellowship from the Australian Research Council.

References

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

Chrestenson, H. E. (1955). A class of generalized Walsh functions. *Pacific Journal of Mathematics*, 5(1):17–31.

Fine, N. J. (1949). On the Walsh functions. *Transactions of the American Mathematical Society*, 65(3):372–414.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325.

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2012). Asymptotic normality and efficiency of two Sobol' index estimators. Technical report, INRIA.

Mauntz, W. (2002). Global sensitivity analysis of general nonlinear systems. Master's thesis, Imperial College. Supervisors: C. Pantelides and S. Kucherenko.

Owen, A. B. (2012a). Better estimation of small Sobol' sensitivity indices. Technical report, Stanford University.

Owen, A. B. (2012b). Variance components and generalized Sobol' indices. *Journal of uncertainty quantification*, 1(1).

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297.

Siegel, A. F. (1983). Low median and least absolute residual analysis of two-way tables. *Journal of the American Statistical Association*, 78:371–374.

Sobol', I. M. (1969). *Multidimensional Quadrature Formulas and Haar Functions*. Nauka, Moscow. (In Russian).

Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118. (In Russian).

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Walsh, J. L. (1923). A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45(1):5–24.

Wang, S. C. (2001). Quantifying passive and driven large-scale evolutionary trends. *Evolution*, 55(5):849–858.