Extremum Global Sensitivity Analysis with Least Squares Polynomials and their Ridges

Chun Yui Wong\textsuperscript{a,}\textsuperscript{*}, Pranay Seshadri\textsuperscript{a}, Geoffrey Parks\textsuperscript{a}

\textsuperscript{a}Department of Engineering, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Abstract

Global sensitivity analysis is a powerful set of ideas and computational heuristics for understanding the importance and interplay between different parameters in a data-set. Such a data-set is characterized by a set of input parameters and an output quantity of interest, where we typically assume that the inputs are independent and information on their joint density can be obtained. If the output quantity of interest is smooth and continuous, polynomial least squares approximations can be used to extract Sobol' indices.

In this paper, we build on these previously well-known ideas by studying two different aspects of this paradigm. First, we study whether sensitivity indices can be computed efficiently if one leverages a polynomial ridge approximation—a polynomial least squares fit over a subspace. We discuss a recipe that utilizes this special dependence structure to reduce the number of model evaluations required for this process. Second, we discuss two heuristics for evaluating input sensitivities when constrained near output extrema: skewness-based sensitivity indices and Monte Carlo filtering. We provide algorithms that implement the ideas discussed in this paper, codes for which are available online.

Keywords: global sensitivity analysis, ridge approximations, active subspaces, higher-order statistics, analysis of skewness

1. Introduction and motivation

Across all data-centric disciplines, sensitivity analysis is an important task. The question “which of my parameters is the most important?” can be heard daily across shop floors, office spaces and board rooms. The measure of importance is typically the conditional variance, leading to the familiar Sobol’ indices \cite{23,24} and the related total indices \cite{9}. However, these are not the only metrics available; over the years, other metrics for sensitivity analysis have been proposed, including those based on conditional skewness \cite{27} and kurtosis \cite{4}. Beyond this, Hooker \cite{10} and Liu and Owen \cite{13} introduce the notion of a superset importance, which is the sum of all Sobol’ indices based on a superset of an input set. Techniques based on partial derivative evaluations of a function have also been extensively studied \cite{1,15}, and the relationship between these elementary effects and total Sobol’ indices have also been investigated (see \cite{12}).

More recently, there has been work centered on higher-order indices (see Owen et al. \cite{17} and Geraci et al. \cite{5}). The former text is particularly interesting as the authors try to draw inference from higher-order skewness- and kurtosis-based indices. More specifically, their working analogy is that a positively skewed
distribution will attain more extreme positive values than a symmetric distribution with the same variance; distributions with a large kurtosis will also attain greater extremes of both signs [16]. Geraci et al., who introduce polynomial approximation-based ideas for computing skewness- and kurtosis-based indices, find applications of these indices in more accurate approximation of the probability density function (PDF) of the output response. In other words, higher values of a variable’s higher-order sensitivity index indicate that a higher-degree polynomial is required to capture the response’s behavior with respect to that variable—requiring more samples.

These ideas are part of a broader effort to identify which variables—and under what distributions—drive the output response to its highest levels. In this paper, we take clues from [17] and Monte Carlo filtering (see page 39 of [19]). We study known metrics based on skewness and, in the general case, we identify them to offer information on the convexity along each input direction, but not necessarily any information on extremum sensitivity, assuming monotonicity. For more general functions, we propose a Monte Carlo filtering-based strategy (see section 3).

Central to our investigation is the tractable computation of various sensitivity metrics. It is well-known that computing sensitivity indices—particularly in high-dimensional problems—can be prohibitive owing to the excessively large number of evaluations required. While access to a function’s gradients can abate this curse of dimensionality, our focus will be on the general case where access to a function’s gradients (or adjoint) is not available. Our approach is to use orthogonal polynomial least squares representations of the function. This implies that we are restricting our study to functions that can be well-approximated by globally smooth and continuous polynomials. Using orthogonal polynomials to this end has been relatively well-studied and prior work can be found in [25, 21] and the references therein.

The remainder of this paper is structured as follows. In section 2 we interpret the global polynomial approximation problem as that of computing its coefficients. For polynomials of a few dimensions, this can be done in the absence of assuming anisotropy; for higher-dimensional approximations, on the other hand, anisotropy is necessary to circumvent the curse of dimensionality. More specifically, we construct polynomial approximations over subspaces—a few linear combinations of all the parameters. From the coefficients in the projected space, we design a polynomial surrogate in the full space for determining the coefficients in the full space. These coefficients are then used for quantifying extremum sensitivity information, detailed in section 3. In section 4, we demonstrate the algorithms and ideas proposed in this paper with two numerical examples.

2. Computing the coefficients of polynomial approximations

Prior to detailing techniques for computing coefficients of the aforementioned polynomials, it will be worthwhile to make more precise our notion of anisotropy. Let us consider a polynomial

$$ p = p(M^T x) \text{ where } M \in \mathbb{R}^{d \times n} \text{ and } x \in \mathbb{R}^d. $$

(1)

For the case where \( n = d \) and where \( M = I \) (the identity matrix), we have a polynomial approximation in the full \( d \)-dimensional space. When \( n < d \) and when the \( n \) columns of \( M \) are orthogonal, they span a \( n \)-dimensional subspace in \( \mathbb{R}^d \). In this case, the function \( p \) is referred to as a generalized ridge function

\[ \text{Note that the columns do not have to be orthogonal; they simply have to be linearly independent.} \]
Depending on what the weights in $M$ are, the polynomial will be anisotropic with the larger weights prescribing linear combinations along which the polynomial varies the most. For notational clarity, we add a subscript below $p$ to denote the dimension over which the polynomial is constructed: $p_d$ denotes a polynomial in the full space, while $p_n$ denotes a polynomial over a subspace. From hereon, we will assume that $n \ll d$.

2.1. Notation and main problem

Now, let us assume the existence of a smooth and continuous function $f : \mathbb{R}^d \to \mathbb{R}$ for which we wish to compute sensitivities. Provided one has access to $N$ input/output pairs $(x_i, f_i)_{i=1}^N$, and one can express $f$ as a polynomial (albeit approximately), we can state

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix} \approx \begin{pmatrix} p_d(x_1) \\ \vdots \\ p_d(x_N) \end{pmatrix} = A_d \alpha,$$

where $A_d \in \mathbb{R}^{N \times r}$ is an orthogonal polynomial design matrix in $d$ dimensions. Writing the expansion of $p_d$ in terms of its $r$ basis terms,

$$p_d(x) = \sum_{i=1}^r \alpha_i \Psi_i(x),$$

we can identify $A_d$ to be

$$A_d = \begin{bmatrix} \Psi_1(x_1) & \Psi_2(x_1) & \cdots & \Psi_r(x_1) \\ \Psi_1(x_2) & \Psi_2(x_2) & \cdots & \Psi_r(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(x_N) & \Psi_2(x_N) & \cdots & \Psi_r(x_N) \end{bmatrix}.$$  

In other words, $A_d$ is a Vandermonde-type matrix. The number of rows $N$ corresponds to the number of model evaluations, typically scaling exponentially with respect to $d$, while the number of columns $r$ corresponds to the number of unknown polynomial coefficients. Consider two scenarios for the size of $N$ vs $r$:

- $N \geq r$: This leads to a polynomial least squares formulation where one tries to solve

$$\min_{\alpha} \|A_d \alpha - f\|_2.$$  

Even with new strategies for selecting which $N$ points to use (see [20, 21]), this approach is not computationally practical for moderate to large $d$.

- $N < r$: To solve a linear system with more unknowns than equations, one has to impose some form of regularization. Solution strategies based on linear programming and basis pursuit denoising [26, 6] have been used to solve

$$\min_{\alpha} \|A_d \alpha - f\|_1.$$  

This, of course, assumes that $\alpha$ is sparse, which need not necessarily be the case, depending on the $r$ polynomial basis terms selected.
So, why the need to use a global polynomial approximation? As alluded to already, accurate polynomial approximations of \( f \) permit us to leverage prior work [25] on computing global variance-based sensitivities using only the coefficient values. Moreover, these coefficients can also be used for computing skewness-based sensitivity metrics, which may offer insight into understanding which variables push \( f \) to its extrema.

But what if \( f \), in addition to being smooth and continuous, was also anisotropic? The idea of developing global sensitivity metrics for ridge approximations has been studied previously by Constantine and Diaz [3]. Using ideas from active subspaces [2], the authors define activity scores that are based on the eigendecomposition of the average outer product of the gradient \( \nabla f \) with itself in the input space. They draw comparisons between these activity scores and total Sobol’ indices (see Theorem 4.1 in [3]). We aim to build upon their ideas and seek to approximate higher-order sensitivity information using \( M \), leveraging polynomial approximations.

Consider the ridge approximation

\[
\mathbf{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix} \approx \begin{bmatrix} p_n(M^T x_1) \\ \vdots \\ p_n(M^T x_N) \end{bmatrix} = \mathbf{A}_n \mathbf{\beta},
\]

where \( \mathbf{A}_n \in \mathbb{R}^{N \times s} \) is another orthogonal polynomial design matrix, but now in \( \mathbb{R}^n \). Since typically \( s \ll r \), we will assume that \( N \geq s \), implying that we have at least as many model evaluations as unknown coefficients in \( \mathbf{\beta} \). In other words, computing \( \mathbf{\beta} \) using regular least squares should not be a problem. The trouble, however, is that \( \mathbf{\beta} \) are the coefficients of \( p_n \), and we need access to the coefficients \( \mathbf{\alpha} \) associated with \( p_d \) in order to estimate global sensitivities.

The key is to notice that \( p_n(M^T x) \) is simply a rearrangement of coefficients from \( p_d(x) \). In other words, we can convert the coefficients of \( p_n \) to those of \( p_d \) in a lossless manner. Namely, we can solve the square linear system

\[
\mathbf{A}_d \hat{\mathbf{\alpha}} = \mathbf{A}_n \mathbf{\beta},
\]

where we can choose \( N = r \) points of evaluation for both Vandermonde matrices arbitrarily. As long as the points of evaluation are distinct, we can invert \( \mathbf{A}_d \) to solve for \( \hat{\mathbf{\alpha}} = \mathbf{A}_d^{-1} \mathbf{A}_n \mathbf{\beta} \) using, e.g., LU factorization. Using this method, the accuracy of the estimated polynomial coefficients is now dependent solely on the accuracy of estimating the dimension-reducing subspace spanned by \( \mathbf{M} \) and the inherent error of a ridge approximation due to variations of \( f \) outside the range of \( \mathbf{M} \). Assuming the presence of strong anisotropy, the subspace can be computed with significantly fewer data than a full polynomial approximation using techniques such as active subspaces [2], polynomial variable projection [8], and minimum average variance estimation (MAVE) [29].

As an illustrative example, suppose we calculate \( \mathbf{M} \) as the leading eigenvectors of the gradient covariance matrix

\[
\mathbf{C} = \int_{\mathcal{D}} (\nabla f)(\nabla f)^T \rho(x) dx
\]

\[
= \begin{bmatrix} \mathbf{M} & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{M}^T \\ \mathbf{W}^T \end{bmatrix}
\]

\[\text{We are assuming that the } N \text{ points are selected such that } \mathbf{A}_n \text{ has a relatively low condition number.}\]
where we have assumed a PDF $\rho(x)$ and input domain $D$ for the function $f$ with mild assumptions to allow the existence of this matrix. It can be shown that the mean squared error between a function and its ridge approximation $g(M^T x) = E[f(x)|M^T x]$ is bounded [2]

$$\int_D (f(x) - g(M^T x))^2 \rho(x) dx \leq C \text{Tr}(\Lambda_2)$$

with $C$ being the Poincaré constant depending on $D$ and $\rho$ and $\text{Tr}$ the trace. Assuming now that $f(x) = p_d(x)$ and $g(M^T x) = p_n(M^T x)$, we can bound the total sum of squares difference between the coefficients with this quantity

$$\|\hat{\alpha} - \alpha\|^2 \leq C \text{Tr}(\Lambda_2).$$

With a strong ridge structure, the trailing eigenvalues—i.e. the diagonal entries in $\Lambda_2$—will be small compared to the leading ones, resulting in better approximation of the coefficients of the full space.

3. On extremum sensitivity information

We now address the second key contribution of this paper: our study of derived higher-order sensitivities to make inference on the extrema of $f$. We start with a brief overview on the computation of sensitivity indices—including higher-order indices—using orthogonal polynomial expansions. Before we proceed with this endeavour, we need to establish a system to identify sensitivity indices with their associated input variables. For this, we introduce multi-indices, which are similar to formulations in the literature such as [25].

3.1. Multi-index notation and Sobol’ indices

We associate each scalar index $i \in \mathbb{N}$ with a multi-index $i = (i_1, ..., i_d) \in \mathbb{N}_0^d$, which is a non-negative integer sequence. This association is bijective, but the order is not important. For consistency with the previous sections, we insist that the first polynomial be a constant, implying that $i = 1 \iff i = (0, ..., 0)$. As an example, with $d = 3$, one possible assignment is shown in table 1. Define a function $\text{nz}(\cdot)$ that returns the positions of the non-zero indices in a multi-index. For instance,

$$\text{nz}((0, 1, 2, 0, 4)) = \{2, 3, 5\}.$$  

The set of all possible multi-indices in the expansion is known as the index set. For instance, an index set of maximum total order $p$ is specified by

$$\left\{ i \in \mathbb{N}_0^d \bigg| \sum_{k=1}^d i_k \leq p \right\}.$$  

With this notation, we can write the polynomial expansion (4) by identifying each basis term with its degree in each input variable.

$$p_d(x) = \sum_{i=1}^r \alpha_i \Psi_i(x), \quad \Psi_i(x) = \prod_{k=1}^d \psi_{i_k}(x_k),$$

where the univariate basis polynomials $\psi(x)$ are orthonormal to each other with respect to the marginal input PDFs $\rho_j(x_j)$. This implies that the multivariate basis polynomials $\Psi$ also form an orthonormal system.
under $\rho(x) = \prod_j \rho_j(x_j)$, and are suitable as basis functions in a Sobol’ decomposition [23], of which we omit the details. In essence, we can consider the decomposition of the total output variance $\sigma^2$ via the following expression

$$
\sigma^2 = \mathbb{E} \left[ (p_d(x) - \mathbb{E} p_d(x))^2 \right] = \int_D \left( \sum_{i=2}^{r} \alpha_i \Psi_i(x) \right)^2 \rho(x) dx. \tag{16}
$$

Owing to orthonormality, this expansion allows us to compute the Sobol’ indices straightforwardly given the expansion coefficients. For instance, the Sobol’ index for a set of variables $S$ is given by

$$
\sigma_S^2 = \sum_{i \in S} \alpha_i^2. \tag{17}
$$

A measure of the importance of an input variable is the total Sobol’ index [24], which is the sum of all Sobol’ indices that contain the variable concerned. For instance, for variable $x_i$,

$$
\sigma_i^2 = \sum_{S} \sigma_S^2. \tag{18}
$$

3.2. Skewness-based indices

We are now ready to define skewness-based sensitivity indices. The definition is similar to (16), where we used the Sobol’ functional ANOVA decomposition and linked it with the polynomial expansion. Applying the same method here yields the following for the global skewness $\gamma$

$$
\gamma = \mathbb{E} \left[ \frac{(p_d(x) - \mathbb{E} p_d(x))^3}{\sigma^3} \right] = \frac{1}{\sigma^3} \int_D \left( \sum_{i=2}^{r} \alpha_i \Psi_i(x) \right)^3 \rho(x) dx. \tag{19}
$$

However, unlike the case for variance-based indices, the resultant indices cannot be expressed in a form as simple as (17). Here, we cite the full result from [5] for the skewness index of a set of variables $S$. Letting $p$ be the corresponding multi-index to $p$ and so on, we define the following sets for a given set of variables $S$

$$
S^1 = \{ p : \text{nz}(p) = S \},
$$

$$
S^2 = \{ (p, q) : \text{nz}(p) \cup \text{nz}(q) = S \},
$$

$$
S^3 = \{ (p, q, r) : \text{nz}(p) \cup \text{nz}(q) \cup \text{nz}(r) = S \}.
$$

It can be shown that the skewness index for $S$ is

$$
t_S = \frac{1}{\sigma^3} \gamma \left( \sum_{p \in S^1} \alpha_p^3 \mathbb{E} [\Psi_p^3] + 3 \sum_{(p, q) \in S^2} \alpha_p^2 \alpha_q \mathbb{E} [\Psi_p^2 \Psi_q] + 6 \sum_{(p, q, r) \in S^3} \alpha_p \alpha_q \alpha_r \mathbb{E} [\Psi_p \Psi_q \Psi_r] \right). \tag{20}
$$

Table 1: One possible assignment between scalar and multi-indices with $d = 3$. 

| i | i |
|---|---|
| 1 | (0,0,0) |
| 2 | (0,0,1) |
| 3 | (0,1,0) |
| 4 | (0,0,2) |
| : | : |

6
The total skewness index for $x_i$ is defined analogously,

$$
\hat{t}_i = \sum_{i \in S} t_S. \tag{21}
$$

From (20) we see how the computation of skewness indices is more involved than computing Sobol’ indices. In particular, we must be careful to selectively compute the integral terms to avoid multiple complete loops over all of the indices, which is computationally intractable. In the appendix we provide an efficient algorithm for computing skewness indices that has been implemented in the open-source polynomial-based uncertainty quantification software package Effective Quadratures [22], available at https://www.effective-quadratures.org/.

So, how useful are skewness indices for inference related to output extrema? A salient feature of skewness indices is the possibility of them being negative. While Sobol’ indices must be non-negative, skewness indices do not have this restriction—they only have to sum up to unity. In Owen et al. [17], the authors state that the sign of skewness indices may be interpretable as the measure of importance of variables near extrema of the output—with a large positive skewness index indicating importance near a maximum and large negative skewness index indicating importance near a minimum. The question is whether the signed nature of skewness indices helps differentiate between importance near maxima or minima. We find that the utility of the signs is not clear without further assumptions on the function.

At the heart of sensitivity analysis—and indeed the broader area of uncertainty quantification—is the task of characterizing the statistical moments of the inputs relative to the output. With this goal in mind, we state a theorem from van Zwet [30] that links the skewness of a random variable to the skewness of a convex transformation of the random variable:

**Theorem 3.1** (from [30]). Suppose that $g : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable univariate monotonic convex function of a random variable $X$, and all moments up to the third order exist. Then, the skewness of $g(X)$ is larger than or equal to the skewness of $X$. If the function is concave, the skewness of $g(X)$ is smaller than or equal to the skewness of $X$.

Thus, if we consider additive functions of the form $f(x) = \sum_{i=1}^d f_i(x_i)$ where each $f_i$ only depends on one variable $x_i$, it is straightforward to conclude that skewness indices are intimately related to the convexity of the subfunctions $f_i$. For instance, if $f_1(x_1)$ is convex, the input marginal PDF of $x_1$ is symmetric (zero skewness), and the overall function $f(x)$ has a positive skewness, then the skewness index of $x_1$ will be positive, and $f_1(x_1)$ will contribute to skewing the output PDF towards the left.

Theorem 3.1 offers partial justification to the interpretation that skewness indices give local sensitivity measures near output extrema mentioned earlier under some assumptions—for example, the function must be additive, monotonic, and twice differentiable. This is because a monotonic convex function gives a steeper slope near maxima; and a monotonic concave function gives a steeper slope near minima. However, in general, the second derivative is not always indicative of importance, and it is difficult to make any rigorous statements about the sign of skewness indices. Instead, we will empirically examine the effectiveness of this information through a numerical example in section 4.2.

### 3.3. Monte Carlo filtering with a polynomial surrogate

Monte Carlo filtering (MCF) [19, p. 38] is a heuristic for determining the input variables that are important in driving the output to certain ranges. The input space is partitioned into two regions—one where the corresponding outputs fall in the range of interest $B$, and one where this does not happen $\bar{B}$. The importance
of a variable \(x_i\) is determined by the difference between the conditional PDFs of \(\phi(x_i|B)\) and \(\phi(x_i|\bar{B})\). If these distributions are significantly different, the variable \(x_i\) is deemed important in distinguishing whether the output falls into our range of interest or not. Statistical tests for the difference between two distributions include the Smirnov two-sample test [19, p. 185] for general distributions. Approaches based on comparison of statistics such as the median [28] and the moments [14] have also been proposed.

An approach based on MCF can be used for extremum sensitivity analysis by partitioning the input space between regions leading to output extrema and regions which do not. We can determine whether an input is important near an extremum by examining its conditional variance within the region of interest. The motivation behind using the variance is the notion that it indicates the variability of the input within the selected region—a large variance near an extremum indicates indifference to the input variable concerned in the vicinity of this output extremum. To quantify this concept, we introduce the variance reduction for an input variable \(x_i\),

\[
I_i = 1 - \frac{\nu_i^2}{\sigma_i^2} = 1 - \frac{\text{Var}[x_i|f(x) \text{ near extrema}]}{\text{Var}[x_i]}. \tag{22}
\]

The quantity \(\nu_i^2\) is the restricted variance, indicating the importance of the variable near an extremum value. To calculate the restricted variance, we use rejection sampling to generate a batch of samples that give extreme output values, and calculate their sample variance as an estimate. The cost of this sampling procedure is greatly reduced by an appropriate surrogate model, such as the multivariate polynomial approximation proposed in section 2. We compare the MCF-based heuristic with inference from skewness indices in section 4.2.

4. Numerical examples

4.1. Ridge approximation of polynomial coefficients

To demonstrate the method of approximating polynomial coefficients using a ridge approximation, we study the estimation of Sobol’ indices of the following approximate ridge polynomial

\[
f(x) = f_1(U^T x) + sf_2(W^T x), \tag{23}
\]

where

\[
f_1(u) = u_1^2 - 0.1u_2^2 - 2u_1^2u_2^2 + 2u_1u_2, \quad u \in \mathbb{R}^2, \tag{24}
\]

\[
f_2(w) = w_1^2 - 0.1w_2^2 - 2w_1^2w_2^2 + 2w_1w_2, \quad w \in \mathbb{R}^5, \tag{25}
\]

and the orthogonal columns of \(U, W\) span, respectively, the column and left null space of

\[
\begin{bmatrix}
2 & 3 & 1 & -4 & 1 & 0.1 & 0.3 \\
-3 & 1 & -2 & 1 & -1 & -0.3 & 0.1
\end{bmatrix}^T. \tag{26}
\]

In addition, \(s\) is a noise parameter that represents the deviation of \(f(x)\) from an exact ridge function within the subspace of \(U\). We compare three methods of evaluating the sensitivity indices, namely

- Through polynomial ridge approximation: using MAVE [29], we estimate a dimension-reducing subspace for the function of interest. Then, using the same samples, we fit a low-dimensional polynomial ridge over the reduced space \(p_d\), whose coefficients we use to calculate the coefficients in the full space via (9).
Computing the integrals defining the Sobol’ indices via quasi-Monte Carlo with a Sobol’ sequence, as specified in [24]. We adapt the open source Python library SALib [7] for this method.

Fitting a polynomial on the full input space via least squares. Note that this requires at least 330 function evaluations, since we are using an isotropic polynomial basis of maximum total order 4.

The Sobol’ indices $\sigma_{13}$ and $\sigma_{24}$ are computed using the methods above. To benchmark these results, we also compute the quantities using integration via Gauss quadrature on a quadratic tensor grid of Legendre polynomials with 2187 points. The fact that $f(x)$ is a polynomial allows this integration to be exact. Figure 1 shows the results of this comparison for $s = 0$ and $s = 0.1$, respectively. It shows that leveraging the (approximate) ridge structure of the quantity of interest, we are able to compute the polynomial coefficients on the full space with higher accuracy than quasi-Monte Carlo. However, when we are allowed to evaluate the function more than 330 times, regression on the full space using a total order grid is superior. Ridge regression is also shown to be quite robust to small deviations outside of $U$, especially when the number of evaluations is small.

4.2. Higher-order inference

We empirically study the utility of skewness indices in extrema sensitivity analysis via an example. The example concerns a model of a piston adapted from [11], which describes the cycle time of the piston as a function of seven parameters. These parameters are shown in table 2. The input variables are independently, uniformly distributed over their support.
The cycle time in seconds is

\[ \tau(x) = 2\pi \sqrt{\frac{M}{k} + S^2 \frac{P_0 V_0 T_a}{T_0 V_0^2}} , \quad (27) \]

where

\[ V = \frac{S}{2k} \left( \sqrt{A^2 + 4k \frac{P_0 V_0}{T_0} T_a - A} \right) \quad (28) \]

and

\[ A = P_0 S + 19.62 M - \frac{k V_0}{S} . \quad (29) \]

We perform sensitivity analysis on this model by examining both the Sobol’ indices and skewness indices, using a seven-dimensional Legendre polynomial basis of maximum total order 5. We approximate the total sensitivity indices by summing skewness indices up to the third order (skewness indices involving up to three variables). The results are shown in table 3, in which the indices with large magnitudes are shown in bold.

Note that our first-order Sobol’ indices results agree well with those of Owen et al. [17] after normalization. The same cannot be said for the first-order skewness indices. This is most likely due to differences in the definitions of the skewness-based indices—Owen et al. generalize the Sobol’ identity to higher powers, whereas we follow the approach of Geraci et al. [5] to generalize the Sobol’ decomposition to higher powers. However, with regard to the ranking of the magnitudes and the signs, we can draw an agreement.

Now, we study the utility of the signs of the skewness indices to indicate importance near extrema, by comparing the results of the MCF-based approach detailed in section 3.3 for both the output maximum and minimum. To evaluate the restricted variance for each variable, we use rejection sampling, selecting the top 5% and bottom 5% of the output values, and calculating the sample variance of the input points that correspond to these extreme output values. The overall variance can be calculated analytically—\( \sigma_i^2 = R_i^2 / 12 \).
where $R_i$ is the range of the corresponding variable. The variance reduction of each variable is shown in table 4, and the marginal PDFs for the input variables near the extrema are shown in fig. 2.

First, we note that the global skewness is positive for the piston cycle time. For $S$, both the first order and total skewness indices are positive, indicating importance near the maximum. This corroborates with the MCF results in table 4, although the fact that $S$ is also rather important near the minimum is not reflected, since it is dominated by the effect of $V_0$.

For $V_0$, the first-order index is negative, while the total skewness index is positive. We can interpret this as the fact that terms involving $V_0$ alone strongly contribute to importance near the minimum, while accounting for every term involving $V_0$ results in a net effect of importance near the maximum. Table 4 indicates that while $V_0$ is the most important near the minimum, it is also significant near the maximum. To obtain a full interpretation one need to consider the skewness indices at different orders.

We also remark that while $M$ and $k$ are also significant near the maximum, their effects are only reflected in the total skewness indices. The piston model provides an example where the signs of the total skewness indices are correlated to the importance of a variable near either maxima or minima, but it is unclear how general this relationship is.

### Table 4: Variance reduction of each input variable around the extrema. Significant values are shown in bold.

|    | $M$  | $S$  | $V_0$ | $k$  | $P_0$ | $T_a$ | $T_0$ |
|----|-----|-----|------|-----|------|------|------|
| Highest $\tau$ | 0.2281 | 0.9317 | 0.6326 | 0.6604 | 0.0131 | 0.0113 | 0.0021 |
| Lowest $\tau$  | 0.0129 | 0.8440 | 0.9594 | $-0.0017$ | 0.0156 | 0.0101 | $-0.0028$ |

5. Conclusions

In this paper we explore two related threads in sensitivity analysis. Through a polynomial-based approximation framework, we show the possibility of leveraging the ridge structure of a quantity of interest to mitigate the curse of dimensionality and reduce the number of observations required for sensitivity analysis. This polynomial-based approximation framework can therefore facilitate more efficient sensitivity analysis. In terms of extremum sensitivity analysis, it allows the efficient calculation of sensitivity indices, including skewness-based indices, or inexpensive sampling that accelerates MCF-based approaches. While we have demonstrated the utility of skewness-based indices in a numerical example, further studies are required to establish the general utility of these indices.

Acknowledgements

This work was supported by the Cambridge Trust; Jesus College, Cambridge; the Alan Turing Institute and Rolls-Royce plc.

Appendix A. Computing skewness indices

As described with (20) in section 3, the method for calculating skewness indices via polynomial expansion coefficients is more involved than calculating the Sobol’ indices, because the integrals of third-order cross terms cannot be simplified easily with orthogonality. Hence, the expectation terms within each summand
Figure 2: PDFs of the input variables to the piston model near extrema of the output. The blue line corresponds to the PDF at high output values and the orange line to the PDF at low output values.
need to be evaluated using numerical quadrature. That is, we can approximate the expectation of a function $f(x)$ using Gauss quadrature

$$E[f(x)] = \int_D f(x) \rho(x) dx \approx \sum_{q=1}^{M} f(s_q) \nu_q,$$

where $(s_q, \nu_q)_{q=1}^{M}$ are the quadrature points and weights. In algorithm 1 we describe the algorithm implemented in Effective Quadratures [22]. In general, an $O(r^3)$ loop is required, and since $r$ can scale exponentially with the input dimension, this presents a significant computational burden. Our algorithm incorporates several methods to reduce the computational complexity, namely

- Using a selection function similar to that proposed in [5] to avoid computation of terms which we know are zero;
- Scanning the index set in an $O(r)$ loop before computation to filter out indices that involve more variables than specified. This cuts the cost of computation significantly for indices involving a small number of variables.
- Using a formulation amenable to matrix operations to avoid explicit loops.

The code can be found at https://github.com/Effective-Quadratures/Effective-Quadratures.

References

[1] Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518, 2007.

[2] Paul G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*, volume 2. SIAM, 2015.

[3] Paul G. Constantine and Paul Diaz. Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety*, 162:1–13, 2017.

[4] Aronne Dell’Oca, Monica Riva, and Alberto Guadagnini. Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrology and Earth System Sciences*, 21:6219–6234, 2017.

[5] Gianluca Geraci, Pietro M. Congedo, Remi Abgrall, and Gianluca Iaccarino. High-order statistics in global sensitivity analysis: Decomposition and model reduction. *Computer Methods in Applied Mechanics and Engineering*, 301:80–115, 2016.

[6] Jerrad Hampton and Alireza Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363–386, 2015.

[7] Jon Herman and Will Usher. SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software*, January 2017.

[8] Jeffrey M. Hokanson and Paul G. Constantine. Data-Driven Polynomial Ridge Approximation Using Variable Projection. *SIAM Journal on Scientific Computing*, 40(3):A1566–A1589, 2018.
Algorithm 1 Computing conditional skewness terms

1: **Input**
2: Polynomial basis and coefficients, quadrature points and weights \((s_q, \nu_q)^{M}_{q=1}\), set of variable indices \(S\) for desired skewness index.
3: **Output**
4: Skewness index \(t_j\).
5: Compute the weighted evaluation matrix \(E_w \in \mathbb{R}^{P \times M}\) such that
6: \[
E_w = \begin{bmatrix}
\alpha_1 \Psi_1(s_1) & \alpha_1 \Psi_1(s_2) & \cdots & \alpha_1 \Psi_1(s_M) \\
\alpha_2 \Psi_2(s_1) & \alpha_2 \Psi_2(s_2) & \cdots & \alpha_2 \Psi_2(s_M) \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_P \Psi_P(s_1) & \alpha_P \Psi_P(s_2) & \cdots & \alpha_P \Psi_P(s_M)
\end{bmatrix}.
\] (A.2)
7: Sum every column to get the polynomial approximant at every point \(\tilde{f}(s_q)\):
8: \[
\tilde{f}(s_q) = \sum_{i=1}^{r} \alpha_i \Psi_i(s_q).
\] (A.3)
9: Compute the global skewness \(\gamma = \sum_{q=1}^{N} (\tilde{f}(s_q))^3 \nu_q\) via a dot product.
10: Compute the global variance \(\sigma^2 = \sum_{all} \alpha_i^2\).
11: Initialize set of valid indices \(V = \emptyset\).  \(\triangleright\) Prune indices
12: **for** \(a = 1, \ldots, P\) **do**
13: Compute \(ord = \sum_d \tilde{a}_d\)
14: **if** \(ord \leq \sum_d \tilde{J}_d\) **then**
15: \(V \leftarrow V \cup \{a\}\).
16: **end if**
17: **end for**  \(\triangleright\) Calculate first sum term
18: **for** \(a \in V\) **do**
19: Compute \(s_a = \sum_{q=1}^{N} (e_a^3)_{q} \nu_q\) via a dot product.
20: \(r_1 \leftarrow r_1 + s_a / (\sigma^{1.5} \gamma)\).
21: **end for**  \(\triangleright\) Calculate second sum term
22: **for** \(a, b \in V\) **do**
23: **if** \((a_d = 0 \text{ and } b_d \neq 0 \text{ for any } d) \text{ or } a = b\) **then**
24: Continue.
25: **end if**
26: Compute \(s_{ab} = \sum_{q=1}^{N} (e_a^2 \otimes e_b)_{q} \nu_q\) via a dot product.
27: \(r_2 \leftarrow r_2 + 3 s_{ab} / (\sigma^{1.5} \gamma)\).
28: **end for**
Set $r_3 = 0$.  

\[ \text{Calculate third sum term} \]

\[ \text{for } a \in V, \ b \in \{a + 1, \ldots, P\} \cap V, \ c = \{b + 1, \ldots, P\} \cap V \text{ do} \]

\[ \text{if } \bar{a}_d + \bar{b}_d + \bar{c}_d = 1 \text{ for any } d \text{ then} \]

\[ \text{Continue.} \]

\[ \text{else if } \bar{a}_d + \bar{b}_d + \bar{c}_d = 2 \text{ and } \bar{a}_d, \bar{b}_d, \bar{c}_d \text{ are pairwise distinct for any } d \text{ then} \]

\[ \text{Continue.} \]

\[ \text{end if} \]

\[ \text{Compute } s_{abc} = \sum_{q=1}^{N} (e_a \odot e_b \odot e_c)_q \nu_q \text{ via a dot product.} \]

\[ r_3 \leftarrow r_3 + 6s_{abc}/(\sigma^{1.5}) \]

\[ \text{end for} \]

\[ \text{Compute } t_S = r_1 + r_2 + r_3. \]

---

[9] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

[10] Giles Hooker. Discovering additive structure in black box functions. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 575–580. ACM, 2004.

[11] Ron Kenett, Shelemyahu Zacks, and Daniele Amberti. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*. John Wiley & Sons, 2013.

[12] Sergei Kucherenko and Ilya M. Sobol’. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017, 2009.

[13] Ruixue Liu and Art B. Owen. Estimating Mean Dimensionality of Analysis of Variance Decompositions. *Journal of the American Statistical Association*, 101(474):712–721, 2006.

[14] Marco Marozzi. Nonparametric Simultaneous Tests for Location and Scale Testing: A Comparison of Several Methods. *Communications in Statistics - Simulation and Computation*, 42(6):1298–1317, 2013.

[15] Max D Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, 1991.

[16] Art B. Owen. Personal communication, 2018.

[17] Art B. Owen, Josef Dick, and Su Chen. Higher order Sobol’ indices. *Information and Inference: A Journal of the IMA*, 3(1):59–81, 2014.

[18] Allan Pinkus. *Ridge Functions*, volume 205. Cambridge University Press, 2015.

[19] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008.

[20] Pranay Seshadri, Gianluca Iaccarino, and Tiziano Ghisu. Quadrature Strategies for Constructing Polynomial Approximations. In *Uncertainty Modeling for Engineering Applications*, pages 1–25. Springer, 2019.
[21] Pranay Seshadri, Akil Narayan, and Sankaran Mahadevan. Effectively Subsampled Quadratures for Least Squares Polynomial Approximations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1003–1023, 2017.

[22] Pranay Seshadri and Geoffrey Parks. Effective-Quadratures (EQ): Polynomials for Computational Engineering Studies. *The Journal of Open Source Software*, March 2017.

[23] Ilya M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.

[24] Ilya M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[25] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.

[26] Gary Tang and Gianluca Iaccarino. Subsampled Gauss Quadrature Nodes for Estimating Polynomial Chaos Expansions. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):423–443, 2014.

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

[28] Frank Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[29] Yingcun Xia, Howell Tong, W. K. Li, and Li-Xing Zhu. An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):363–410, 2002.

[30] Willem R. van Zwet. *Convex Transformations of Random Variables*. Mathematical Centre Tracts 7. Mathematisch Centrum, Amsterdam, 1964.