

# Sensitivity analysis using anchored ANOVA expansion and high order moments computation

Kunkun Tang, Pietro Marco Congedo, Rémi Abgrall

► To cite this version:

Kunkun Tang, Pietro Marco Congedo, Rémi Abgrall. Sensitivity analysis using anchored ANOVA expansion and high order moments computation. [Research Report] RR-8531, 2014. hal-00987189

HAL Id: hal-00987189

<https://hal.inria.fr/hal-00987189>

Submitted on 5 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Sensitivity analysis using anchored ANOVA expansion and high order moments computation

Kunkun Tang, Pietro M. Congedo , Rémi Abgrall

**RESEARCH  
REPORT**

**N° 8531**

May 2014

Project-Team BACCHUS





## Sensitivity analysis using anchored ANOVA expansion and high order moments computation

Kunkun Tang<sup>\*</sup>, Pietro M. Congedo<sup>\*</sup>, Rémi Abgrall<sup>†</sup>

Project-Team BACCHUS

Research Report n° 8531 — May 2014 — 36 pages

**Abstract:** An *anchored* ANOVA method is proposed in this paper to decompose statistical moments. Compared to *standard* ANOVA with mutually orthogonal components, anchored ANOVA, with arbitrary *anchor point*, loses orthogonality if employing the same measure. However, an advantage consists in the considerably reduced number of deterministic solver's computations, which renders uncertainty quantification of real engineering problems much easier. Different from existing methods, *covariance decomposition* of output variance is used in this paper to take account of interactions between non-orthogonal components, yielding an exact variance expansion, and thus, with a suitable numerical integration method, provides a strategy that converges. This convergence is verified by studying academic tests. In particular, sensitivity problem of existing method to anchor point is analyzed via Ishigami case, and we point out covariance decomposition survives from it. *Covariance-based sensitivity indices (SI)* are also used, compared to *variance-based SI*. Furthermore, we emphasize covariance decomposition can be generalized in a straightforward way to decompose high order moments. For academic problems, results show the method converges to exact solution regarding both skewness and kurtosis. Finally, the proposed method is applied on a realistic case, *i.e.* estimating chemical reactions uncertainties in a hypersonic flow around a space vehicle during an atmospheric reentry.

**Key-words:** stochastic problems; uncertainty quantification; (anchored) ANOVA; global sensitivity analysis; variance/covariance decomposition; high order moments decomposition

<sup>\*</sup> INRIA Bordeaux Sud-Ouest, Team BACCHUS, 200 avenue de la Vieille Tour, 33405 Talence, France

<sup>†</sup> Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

RESEARCH CENTRE  
BORDEAUX – SUD-OUEST

200 avenue de la Vieille Tour  
33405 Talence Cedex

# Sensitivity analysis using anchored ANOVA expansion and high order moments computation

**Résumé :** Pas de résumé

**Mots-clés :** Pas de motclef

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Some definitions: ANOVA, variance decomposition and global sensitivity analysis</b>	<b>4</b>
2.1	Variance-based global sensitivity estimates . . . . .	6
<b>3</b>	<b>Anchored ANOVA based method</b>	<b>6</b>
<b>4</b>	<b>Covariance decomposition of the variance <math>V(Y)</math> and covariance-based sensitivity indices</b>	<b>8</b>
4.1	Covariance-based sensitivity estimates using anchored ANOVA decomposition . .	9
<b>5</b>	<b>High order statistical moments evaluation using anchored decomposition</b>	<b>10</b>
5.1	Multi-indices $\alpha_i^{No,j}$ . . . . .	10
5.2	Skewness decomposition . . . . .	10
5.3	Kurtosis decomposition . . . . .	12
<b>6</b>	<b>Numerical results</b>	<b>13</b>
6.1	Sobol' function . . . . .	13
6.1.1	Covariance-based sensitivity analysis . . . . .	17
6.2	Ishigami function . . . . .	18
6.2.1	Numerical investigation: Covariance-based sensitivity indices and high order moments computation . . . . .	19
6.2.2	Analytical investigation: Sensitivity of component variance on anchor point	22
6.3	Application to the chemical reaction uncertainties during an atmospheric reentry	27
<b>7</b>	<b>Conclusions and perspectives</b>	<b>32</b>
<b>A</b>	<b>Sobol' function test with truncated expansion and adaptivity</b>	<b>32</b>
<b>B</b>	<b>Variance computation (Anchored ANOVA)</b>	<b>35</b>

## 1 Introduction

The Analysis of Variance (ANOVA) expansion is an elegant and very useful way to represent multivariate functions  $f(\mathbf{x})$  in high dimensions, for instance, when estimating the sensitivity indices via variance-based approaches. For independent random inputs, ANOVA from standard definition consists in a unique orthogonal decomposition of  $f(\mathbf{x})$ . Each component function provides its best approximation to  $f(\mathbf{x})$  in a least-square sense. From computational point of view, standard orthogonal ANOVA can be very expensive when encountering very high dimensional problems and complicated multivariate functions. Indeed, the drawback of standard ANOVA consists in the need to compute the high-dimensional integrals (often requiring Monte Carlo type sampling methods). Even the zeroth-order component function requires a full-dimensional integration in the stochastic space. Alternatively, anchored ANOVA decomposition [1, 2, 3, 4] gives a computationally efficient way for the numerical evaluation of component functions in ANOVA expansion, and therefore the estimation of mean and variance of multivariate functions often become computationally much cheaper. In particular, [4] presents some adaptive criteria as dimension reduction techniques, which can be applied to problems with a very high number of stochastic variables. One main drawback appears in such a decomposition: the accuracy of approximation is found to be very sensitive to the choice of the “anchor point”. [5, 6] show that a bad choice of the anchor point can lead to an unacceptable approximation error. This paper analyzes the reason of this sensitivity to anchor point, and proposes to use the covariance decomposition capable of evaluating very accurately the output variance of multivariate function in the framework of anchored ANOVA. We then extend this technique to the general case with the aim of evaluating statistical moments of arbitrary order. It is found that different (arbitrary) choices of “anchor point” lead to very close approximations. Numerical results show that the numerical solution converges very quickly to the exact solution.

The paper is organized as follows: Section 2 recalls some basic notions on the standard version of the ANOVA decomposition and on the global sensitivity analysis; we then introduce in Section 3 the anchored version of ANOVA, and present the numerical way of evaluating the component functions and the mean/variance of output function; Section 4 is devoted to the approach proposed in this paper: a covariance decomposition of the output variance. Moreover, some covariance-based sensitivity indices are introduced. In Section 5, we present the proposed approach to compute the high order statistics. Section 6 illustrates several numerical results. Finally, conclusions are drawn and some perspectives are outlined.

## 2 Some definitions: ANOVA, variance decomposition and global sensitivity analysis

Let us introduce some notations. The upper-case letters,  $\mathbf{X}$  and  $\mathbf{Y}$ , denote the independent random input variables and random output, respectively; the lower-case letters  $\mathbf{x}$  and  $\mathbf{y}$  represent the realizations. Let us suppose that the response of a given system can be represented by a  $N$ -dimensional integrable function  $f(\mathbf{x})$

$$y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_N). \quad (1)$$

We consider it in its functional expansion form as follows

$$y = f_0 + \sum_{1 \leq i \leq N} f_i(x_i) + \sum_{1 \leq i < j \leq N} f_{ij}(x_i, x_j) + \dots + f_{1,2,\dots,N}(x_1, x_2, \dots, x_N),$$

or in compact form using a multi index system:

$$y = f_{s_0} + \sum_{j=1}^{2^N-1} f_{s_j}(\mathbf{x}_{s_j}). \quad (2)$$

The multi indices  $s_j$  are defined such as

$$\begin{aligned} s_0 &= (0, 0, 0, \dots, 0) \\ s_1 &= (1, 0, 0, \dots, 0) \\ s_2 &= (0, 1, 0, \dots, 0) \\ &\vdots \\ s_N &= (0, 0, 0, \dots, 1) \\ s_{N+1} &= (1, 1, 0, \dots, 0) \\ s_{N+2} &= (1, 0, 1, \dots, 0) \\ &\vdots \\ s_{\mathcal{N}} &= (1, 1, 1, \dots, 1). \end{aligned} \quad (3)$$

where

$$\mathcal{N} = 2^N - 1.$$

The representation (2) is called ANOVA (Analysis Of Variance) decomposition [7] of  $f(\mathbf{x})$ , if

$$E_{X_i}(f_{s_j}(\mathbf{X}_{s_j})) = 0 \text{ for any element variable } X_i \text{ in } \mathbf{X}_{s_j}. \quad (4)$$

It follows from (4) the orthogonality of ANOVA component terms, namely

$$E(f_{s_j} f_{s_k}) = 0, \quad (5)$$

if  $j \neq k$ . Note that the terms in the ANOVA decomposition can be expressed as integrals of  $f(\mathbf{x})$ . Indeed,

$$E(Y) = f_0, \quad (6)$$

$$E(Y|x_i) = f_0 + f_i(x_i),$$

$$E(Y|x_i, x_j) = f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j),$$

and so on. We thus observe the ANOVA terms can be computed as follows:

$$f_{s_j}(\mathbf{x}_{s_j}) = E(Y|\mathbf{x}_{s_j}) - \sum_{s_k \subset s_j} f_{s_k}, \quad (7)$$

with  $s_k$  a subset multi-index of  $s_j$ .

By integrating  $f^2$  and exploiting the orthogonality property of component functions, the variance of  $f$  can be written as follows:

$$V(Y) = \sum_{j=1}^{\mathcal{N}} E(f_{s_j}^2(\mathbf{X}_{s_j})) = \sum_{j=1}^{\mathcal{N}} V_{s_j}, \quad (8)$$

which is in fact the sum of the variances of all the decomposition terms. Here, we have denoted  $V_{s_j} = V(f_{s_j}(\mathbf{X}_{s_j}))$ .



## 2.1 Variance-based global sensitivity estimates

The ANOVA decomposition is closely related to the global sensitivity indices [8, 9] which are defined by the ratios

$$\mathcal{S}_{s_j} = \frac{V_{s_j}}{V(Y)} = \frac{V(E(Y|\mathbf{x}_{s_j})) - \sum_{s_k \subset s_j} V_{s_k}}{V(Y)}. \quad (9)$$

From (8), all the  $\mathcal{S}_{s_j}$  are non-negative and their sum equals unity:

$$\sum_{j=1}^{\mathcal{N}} \mathcal{S}_{s_j} = 1. \quad (10)$$

Furthermore, the total effects of the variable  $X_i$  is estimated by

$$\mathcal{S}_i^T = 1 - \frac{V(E(Y|\mathbf{X}_{-i}))}{V(Y)}, \quad (11)$$

which is in fact the sum of all sensitivity indices containing  $X_i$ . Note here

$$\sum_{i=1}^N \mathcal{S}_i^T \neq 1. \quad (12)$$

## 3 Anchored ANOVA based method

It is noted the constant term from (6) and higher order terms from (7) can be very expensive to be computed for very high dimensional and complicated  $f(\mathbf{x})$ . Hence, the so-called ‘‘anchored-ANOVA’’ method [4, 3, 10, 1, 11] appears as a good candidate to reduce the computational cost. In this case, the Dirac measure is used instead of Lebesgue measure in integrations as follows:

$$d\mu(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{c}) d\mathbf{x}. \quad (13)$$

The point  $\mathbf{c}$  is called ‘‘anchor point’’. Hence, the constant term can be expressed in a very succinct way:

$$f_0 = f(\mathbf{c}). \quad (14)$$

We refer to [3, 10] for issues related to the choices of anchor points. Generally speaking, one can arbitrarily choose this reference point. We point out that a good approximation to  $f$  should be as less sensitive as possible to this anchor point  $\mathbf{c}$ .

Inserting (13) into (7) leads to the result of high order terms such that:

$$f_{s_i}(\mathbf{x}_{s_i}) = f(\mathbf{c}|\mathbf{x}_{s_i}) - \sum_{s_j \subset s_i} f_{s_j}(\mathbf{x}_{s_j}). \quad (15)$$

Here  $f(\mathbf{c}|\mathbf{x}_{s_i})$  represents the value of  $f(\mathbf{x})$  evaluated at anchor point  $\mathbf{c}$  except for the variables involved in  $\mathbf{x}_{s_i}$ .

Equation (15) means that only values of deterministic solver  $f$  evaluated on sampling points are needed in order to approximate the component function  $f_{s_i}$ , while with (7) one needs to evaluate multi-dimensional integrations.

As done in [4], let us take

$$\begin{aligned} f_{s_1} &= f_1(x_1), \\ f_{s_{N+1}} &= f_{12}(x_1, x_2), \end{aligned}$$

for instance, to explain how to compute the component functions in anchored decomposition. Let  $\mathbf{c} = (c_1, c_2, \dots, c_N)$  and  $q_j^1, q_j^2, \dots, q_j^\mu$  be the Gauss quadrature points for the dimension  $x_j$ . Using (15), we have the value of  $f_{s_1}$  at quadrature point  $q_1^i$ :

$$f_1(q_1^i) = f(q_1^i, c_2, c_3, \dots, c_N) - f_{s_0}. \quad (16)$$

After evaluating  $f_1$  at all quadrature points, since weight  $w_j^i$  at  $q_j^i$  can be generated by Gaussian quadrature method, one can readily compute the mean and variance of  $f_1$ , as follows

$$E(f_1) \approx \sum_{i=1}^{\mu} f_1(q_1^i) w_1^i, \quad (17)$$

$$\begin{aligned} V(f_1) &\approx \sum_{i=1}^{\mu} [f_1(q_1^i) - E(f_1)]^2 w_1^i \\ &= \sum_{i=1}^{\mu} [f_1(q_1^i)]^2 w_1^i - [E(f_1)]^2. \end{aligned} \quad (18)$$

The procedure above can be applied to all first-order component functions  $f_1, \dots, f_N$  for evaluating their mean and variance.

Concerning the second-order term  $f_{12} = f_{12}(x_1, x_2)$ , the computation is similar. Consider  $f_{12}$  at quadrature point  $(q_1^i, q_2^j)$  with  $i, j = 1, \dots, \mu$ . Using (15), we obtain

$$f_{12}(q_1^i, q_2^j) = f(q_1^i, q_2^j, c_3, \dots, c_N) - f_1(q_1^i) - f_2(q_2^j) - f_0, \quad (19)$$

where

$$f_1(q_1^i) = f(q_1^i, c_2, c_3, \dots, c_N) - f_0, \quad (20)$$

$$f_2(q_2^j) = f(c_1, q_2^j, c_3, \dots, c_N) - f_0. \quad (21)$$

Note that the tensor product rule or sparse grid method can be exploited in order to generate multi-dimensional quadrature points and their corresponding weights. Therefore, the mean and variance of second-order function  $f_{12}$  can be computed similarly as for first-order terms.

The process we have described can be recursively applied to all second-order terms and higher order ones, and will be used to compute mean, variance and high order statistics.

Integrating (2) leads to the formula for the mean computation of  $f(\mathbf{x})$ :

$$E(Y) = \sum_{i=0}^{\mathcal{N}} E(f_{s_i}(\mathbf{X}_{s_i})), \quad (22)$$

which is the sum of the means of all the component functions. Indeed,  $E(f_{s_i})$  can be easily evaluated by Gaussian quadrature method, as described previously for first and second-order components.

In [4] (for instance in Section A.3. *A simple example*, pages 1612-1614), they evaluate the variance of  $f$  by the sum of the variances of all the component functions, similarly as for  $E(Y)$ :

$$V(Y) = \sum_{i=1}^{\mathcal{N}} V_{s_i} = \sum_{i=1}^{\mathcal{N}} E([f_{s_i}(\mathbf{X}_{s_i}) - E(f_{s_i})]^2), \quad (23)$$

with  $V_{s_i}$  calculated by Gaussian quadrature rule. (23) is in fact a direct generation of (8) from the standard version of ANOVA.

## 4 Covariance decomposition of the variance $V(Y)$ and covariance-based sensitivity indices

We emphasize that the way of evaluating variance  $V(f)$  in (23) can be efficient and accurate enough for a good choice of anchor point  $\mathbf{c}$  [4, 3, 10, 1, 11]. However, for an arbitrary choice of  $\mathbf{c}$ , as we will show later in the Section devoted to numerical experiments, (23) generally gives unacceptable errors.

In this paper, we propose a more flexible way to compute the variance  $V(f)$ . Let us observe that the formula (23) comes from (8) which is based on standard (non-anchored) ANOVA decomposition of  $f$ . Unfortunately, in anchored ANOVA decomposition, the orthogonality property of component functions is no longer valid. That is, in general:

$$E(f_{s_i}(\mathbf{X}_{s_i})f_{s_j}(\mathbf{X}_{s_j})) \neq 0, \text{ for } i \neq j. \quad (24)$$

In other words, for anchored decomposition, (8) should be modified in order to take (24) into account. Integrating  $(Y - E(Y))^2$ , variance  $V(Y)$  can be computed as the sum of the covariances:

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] \\ &= E\left[\left(\sum_{i=1}^{\mathcal{N}} f_{s_i} - \sum_{i=1}^{\mathcal{N}} E(f_{s_i})\right)^2\right] \\ &= \sum_{i=1}^{\mathcal{N}} [V(f_{s_i}) + \underbrace{\sum_{j=1, j \neq i}^{\mathcal{N}} \text{Cov}(f_{s_i}, f_{s_j})}_{(a)}] \\ &= \sum_{i=1}^{\mathcal{N}} [\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j})]. \end{aligned} \quad (25)$$

Comparing the covariance decomposition of  $V(Y)$  in (25) and the variance evaluation in (23), we see the term (a) in (25) is neglected in (23), while its contribution can be significant in general with an anchored decomposition. Indeed, the key difference of the two approaches lies in the fact that the covariance term  $\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j})$  in (25) can even be negative. Moreover, it is observed that the covariance decomposition of the unconditional variance  $V(Y)$  (25) is general. As a consequence, the result should not be sensitive to the choice of anchor point. In fact, different anchor points can lead to extremely different covariance decompositions. For this reason, methods become sensitive to anchor points, if taking account only the positive part  $V(f_{s_i})$  in  $\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j})$ . As a matter of fact, these methods do not converge to the exact solution.

Errors can be huge, as it will be illustrated in the Section 6.1. However, a good choice of the anchor point can reduce the difference  $\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j}) - V(f_{s_i})$ , thus providing acceptable results.

Note that in the case of standard orthogonal ANOVA, where

$$\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j}) = V(f_{s_j}),$$

(25) degenerates into (8) or (23).

Note that a similar covariance decomposition has been considered previously, e.g. in [12], where the correlated inputs are studied and covariance decomposition is used to account for correlations among inputs. We emphasize the use of (25) in this work in the framework of anchored ANOVA for independent variables is nevertheless new.

In the next Section, we generalize the approach of covariance decomposition (25) of unconditional variance  $V(Y)$  to the decompositions of higher order statistical moments, namely skewness and kurtosis. The anchored ANOVA functional decomposition of model output  $f(\mathbf{x})$  is reused.

#### 4.1 Covariance-based sensitivity estimates using anchored ANOVA decomposition

As already mentioned, the covariance decomposition given by (25) is general for any expansion of  $y = f(\mathbf{x})$  with the form (2). The variance decomposition given in (8) is in fact a special case when all component functions satisfy (4), i.e. when they are mutually orthogonal.

As similarly done in [12], three sensitivity indices are defined for a single or group of inputs  $\mathbf{X}_{s_i}$ :

$$\mathcal{S}_{s_i}^a = \frac{V_{s_i}}{V(Y)}, \quad (26)$$

$$\mathcal{S}_{s_i}^b = \frac{\sum_{j=1, j \neq i}^{\mathcal{N}} \text{Cov}(f_{s_i}, f_{s_j})}{V(Y)}, \quad (27)$$

$$\mathcal{S}_{s_i} = \frac{\text{Cov}(f_{s_i}, \sum_{j=1}^{\mathcal{N}} f_{s_j})}{V(Y)} = \mathcal{S}_{s_i}^a + \mathcal{S}_{s_i}^b. \quad (28)$$

They are related to *structural*, *correlative* and *whole* contributions of  $\mathbf{X}_{s_i}$ , respectively. According to (25), we have

$$\sum_{i=1}^{\mathcal{N}} \mathcal{S}_{s_i} = 1. \quad (29)$$

It is emphasized that only the structural sensitivity index  $\mathcal{S}_{s_i}^a$  is strictly non-negative, which means one can obtain a negative  $\mathcal{S}_{s_i}$ . Note also in the special case of orthogonal ANOVA expansion, we have  $\mathcal{S}_{s_i}^b = 0$ .

The total effects  $\mathcal{S}_i^{T,a}$ ,  $\mathcal{S}_i^{T,b}$  and  $\mathcal{S}_i^T$  of the variable  $X_i$  can be evaluated by adding all the sensitivity indices containing  $X_i$ .

As said before, we emphasize again that the covariance-based sensitivity analysis in [12] is performed for correlated input variables. When treating independent variables, their approach reduces to variance-based method, and only a single index  $\mathcal{S}_{s_i}$  is needed. Note however the covariance-based sensitivity analysis performed in this paper is only for independent variables, which makes our analysis quite different from [12]. Moreover, the sensitivity analysis in [12] is based on the meta-modeling approach, specifically on the random sampling - high dimensional model representation (RS-HDMR) expansion, while in this paper the so-called anchored ANOVA expansion is employed which is in fact an exact decomposition of the model output. Thus, the covariance-based sensitivity analysis in this work is a new way for the investigation of inputs importance.

## 5 High order statistical moments evaluation using anchored decomposition

In order to formulate innovative optimization methods for obtaining very robust designs by taking into account a complete description of the output statistics, [13] provides a general method aiming of computing the decomposition of high order statistics. In particular, an approach similar to standard ANOVA method but for skewness and kurtosis is formulated. As for variance computation by standard ANOVA, this approach of decomposition of high order statistics could be too prohibitive when evaluating the component functions and the products among them.

Note that we shall not repeat here the complete derivation of skewness and kurtosis decomposition for standard ANOVA functional decomposition; the final results in [13] will directly be used in this work to compute high order statistics within the anchored ANOVA framework, which leads to less expensive calculations.

### 5.1 Multi-indices $\alpha_i^{No,j}$

For notation convenience, we use the multi-indices  $\alpha_i^{No,j}$ , e.g. the one in [14] for multi-dimensional PC basis, to formulate the decomposition of skewness and kurtosis. The value  $No$  in  $\alpha_i^{No,j}$  denotes the order of statistics. That is,  $No = 3$  for skewness  $S(Y)$ ;  $No = 4$  for kurtosis  $K(Y)$ . The index  $i$  in  $\alpha_i^{No,j}$  is related to ANOVA component function  $f_{s_i}$ . The index  $j$  in  $\alpha_i^{No,j}$  is the numbering of the decomposition term for statistics. For instance, for a function  $f$  with two independent random variables in stochastic space, for which the number of component functions is three (excluding the zeroth order one), the third order multi-indices  $\alpha_i^{3,j}$  for skewness are listed in Table 1.

### 5.2 Skewness decomposition

The skewness  $S(Y)$ , in the framework of anchored decomposition, can be decomposed as follows:

$j$	$f_{s_1} = f_1$	$f_{s_2} = f_2$	$f_{s_3} = f_{12}$
1	3	0	0
2	2	1	0
3	2	0	1
4	1	2	0
5	1	1	1
6	1	0	2
7	0	3	0
8	0	2	1
9	0	1	2
10	0	0	3

Table 1: multi-indices  $\alpha_i^{3,j}$  for skewness decomposition of function  $y = f(x_1, x_2)$ .

$$\begin{aligned}
S(Y) &= E[(Y - E(Y))^3] \\
&= E \left[ \left( \sum_{i=1}^{\mathcal{N}} (f_{s_i} - E(f_{s_i})) \right)^3 \right] \\
&= \sum_{j=1}^{\mathcal{M}} E \left[ \beta_j^S \prod_{i=1}^{\mathcal{N}} (f_{s_i} - E(f_{s_i}))^{\alpha_i^{3,j}} \right] \\
&= \sum_{j=1}^{\mathcal{M}} E \left[ \beta_j^S \prod_{i=1}^{\mathcal{N}} (\tilde{f}_{s_i})^{\alpha_i^{3,j}} \right].
\end{aligned} \tag{30}$$

where we have defined

$$\tilde{f}_{s_i} = f_{s_i} - E(f_{s_i}).$$

The quantity  $\mathcal{M}$  is the number of the decomposition terms involved in skewness computation, and is given by:

$$\mathcal{M} = \binom{\mathcal{N}}{1} + 2 \binom{\mathcal{N}}{2} + \binom{\mathcal{N}}{3}.$$

$\beta_j^S$  is a constant coefficient and can be determined in the following way:

$$\beta_j^S = \begin{cases} 1 & \text{if } \{\alpha_i^{3,j}\} \supset \{3\}, \\ \binom{3}{1} = 3 & \text{if } \{\alpha_i^{3,j}\} \supset \{2, 1\}, \\ \binom{3}{1} \binom{2}{1} = 6 & \text{if } \{\alpha_i^{3,j}\} \supset \{1, 1, 1\}. \end{cases} \tag{31}$$

The multi-dimensional integrals involved in (30) can be evaluated by either product rule or sparse grid method.

Considering the simple example in Table 1 with 2 random variables, the skewness can be decomposed using (30) with  $\mathcal{M} = 10$ :

$$\begin{aligned}
S(Y) &= E_{X_1}[\tilde{f}_1^3] + E_{X_2}[\tilde{f}_2^3] + E[\tilde{f}_{12}^3] \\
&+ 3E[\tilde{f}_1^2 \tilde{f}_2] + 3E[\tilde{f}_1^2 \tilde{f}_{12}] \\
&+ 3E[\tilde{f}_1 \tilde{f}_2^2] + 3E[\tilde{f}_1 \tilde{f}_{12}^2] \\
&+ 3E[\tilde{f}_2^2 \tilde{f}_{12}] + 3E[\tilde{f}_2 \tilde{f}_{12}^2] \\
&+ 6E[\tilde{f}_1 \tilde{f}_2 \tilde{f}_{12}].
\end{aligned} \tag{32}$$

### 5.3 Kurtosis decomposition

The kurtosis  $K(Y)$  of multivariate function  $y = f(\mathbf{x})$  can be decomposed in the same way as for skewness:

$$\begin{aligned}
K(Y) &= E[(Y - E(Y))^4] \\
&= E \left[ \left( \sum_{i=1}^{\mathcal{N}} (f_{s_i} - E(f_{s_i})) \right)^4 \right] \\
&= \sum_{j=1}^{\mathcal{M}} E \left[ \beta_j^K \prod_{i=1}^{\mathcal{N}} (\tilde{f}_{s_i})^{\alpha_i^{4,j}} \right].
\end{aligned} \tag{33}$$

The number of the decomposition terms  $\mathcal{M}$  involved in kurtosis computation (33) is given by:

$$\mathcal{M} = \binom{\mathcal{N}}{1} + \binom{\mathcal{N}}{1} \binom{\mathcal{N}-1}{1} + \binom{\mathcal{N}}{2} + \binom{\mathcal{N}}{1} \binom{\mathcal{N}-1}{2} + \binom{\mathcal{N}}{4}.$$

The constant coefficient  $\beta_j^K$  is provided by:

$$\beta_j^K = \begin{cases} 1 & \text{if } \{\alpha_i^{4,j}\} \supset \{4\}, \\ \binom{4}{1} = 4 & \text{if } \{\alpha_i^{4,j}\} \supset \{3, 1\}, \\ \binom{4}{2} = 6 & \text{if } \{\alpha_i^{4,j}\} \supset \{2, 2\}, \\ \binom{4}{2} \binom{2}{1} = 12 & \text{if } \{\alpha_i^{4,j}\} \supset \{2, 1, 1\}, \\ \binom{4}{1} \binom{3}{1} \binom{2}{1} = 24 & \text{if } \{\alpha_i^{4,j}\} \supset \{1, 1, 1, 1\}. \end{cases} \tag{34}$$

Considering the example with 2 random variables and 4 ANOVA component functions, the

kurtosis can be computed using (33) with  $\mathcal{M} = 15$ :

$$\begin{aligned}
K(Y) &= E[\tilde{f}_1^4] + E[\tilde{f}_2^4] + E[\tilde{f}_{12}^4] \\
&+ 4E[\tilde{f}_1^3 \tilde{f}_2] + 4E[\tilde{f}_1^3 \tilde{f}_{12}] \\
&+ 4E[\tilde{f}_1 \tilde{f}_2^3] + 4E[\tilde{f}_1 \tilde{f}_{12}^3] \\
&+ 4E[\tilde{f}_2^3 \tilde{f}_{12}] + 4E[\tilde{f}_2 \tilde{f}_{12}^3] \\
&+ 6E[\tilde{f}_1^2 \tilde{f}_2^2] + 6E[\tilde{f}_1^2 \tilde{f}_{12}^2] \\
&+ 6E[\tilde{f}_2^2 \tilde{f}_{12}^2] + 12E[\tilde{f}_1^2 \tilde{f}_2 \tilde{f}_{12}] \\
&+ 12E[\tilde{f}_1 \tilde{f}_2^2 \tilde{f}_{12}] + 12E[\tilde{f}_1 \tilde{f}_2 \tilde{f}_{12}^2].
\end{aligned} \tag{35}$$

We emphasize that the implementation of formulae (30) for skewness and (33) for kurtosis becomes easy, once the multi-indices  $\alpha_i^{N_o, j}$  have been determined. In fact, one just needs to compute integrals of component functions' products.

## 6 Numerical results

In this section, several results are presented for assessing the efficiency of the proposed approach. First, two classical functions, well-known in literature, are considered, *i.e.* the Sobol' and the Ishigami functions. Then, a real engineering problem featuring a high-dimension stochastic space, is taken into account: estimating the chemical reactions uncertainties, in a flow around a space vehicle during an atmospheric reentry.

### 6.1 Sobol' function

As similarly studied in [4], in this section, a simple 4-dimensional Sobol' function is taken into account:

$$y = f(\mathbf{x}) = \prod_{k=1}^N f^{(k)}(x_k), \quad x_k \in [0, 1], \tag{36}$$

where

$$\begin{aligned}
N &= 4, \\
f^{(k)}(x_k) &= \frac{|4x_k - 2| + a_k}{1 + a_k}, \\
a_k &= k^2.
\end{aligned} \tag{37}$$

This function is used in order to demonstrate the convergence of the proposed method using the anchored ANOVA decomposition. Both low and high order statistics of a function are considered. Considering uniformly distributed variables, the exact statistics of  $Y = f(\mathbf{X})$  can be easily computed as follows:

$$E(Y) = 1, \tag{38}$$

$$V(Y) = 1.0271 \times 10^{-1}, \tag{39}$$

$$S(Y) = 9.4209 \times 10^{-3}, \tag{40}$$

$$K(Y) = 2.4306 \times 10^{-2}. \tag{41}$$



	exact	computed	abs err	rel err
$E(Y)$	1.0000E+00	1.0000E+00	1.1102E-16	0.1110E-13 %
$V(Y)$	1.0271E-01	1.0271E-01	1.3878E-17	0.1351E-13 %
$V(Y)([4])$	1.0271E-01	1.0271E-01	1.3878E-17	0.1351E-13 %
$S(Y)$	9.4209E-03	9.4209E-03	1.8562E-16	0.1970E-11 %
$K(Y)$	2.4306E-02	2.4306E-02	4.7531E-16	0.1956E-11 %

Table 2: Computed statistics versus exact solution for anchor point 1.

These exact values are used to evaluate the absolute and relative error of numerical solutions.

The Gauss-Legendre quadrature rule is adopted for numerical integration. 6 quadrature points (2 elements) per dimension are chosen for demonstration. As it will be shown, 6 points are sufficient to achieve a very good accuracy. Tensor product rule is used to generate multi-dimensional points and their corresponding weights. In order to demonstrate the ability of our approach, three anchor reference points will be arbitrarily chosen for computing the statistical moments.

- **Anchor point 1.** First of all, let us set the anchor point  $\mathbf{c}$  such that

$$c_k = 0.75, \text{ for } k = 1, 2, 3, 4. \quad (42)$$

Table 2 presents the numerical results compared to the exact solution for the mean, variance, skewness and kurtosis. The absolute and relative error are also pointed out. The variance result of " $V([4])$ " in Table 2 is obtained using the state-of-the-art method in [4]. It gives the same accuracy as ours. In general, all computed results are almost exact.

The contributions of 120 terms in (25) for the covariance decomposition approach and 15 terms in (23) for the method in [4] are plotted in Fig. 1. It is observed the contribution of term ( $a$ ) in (25) is almost zero everywhere. In other words, with this specific anchor point, the orthogonality of component functions in anchored ANOVA decomposition is verified and the anchored approach performs as standard ANOVA method.

It is emphasized here, with this good choice of anchor point, [4] provides a more efficient method than the covariance approach, since a much smaller number of multi-dimensional integrations are required to be evaluated.

- **Anchor point 2.** We use a second anchor point  $\mathbf{c}$  such that

$$c_k = 0.5, \text{ for } k = 1, 2, 3, 4. \quad (43)$$

Table 3 shows the numerical results. We see that our approach is almost exact for all

	exact	computed	abs err	rel err
$E(Y)$	1.0000E+00	1.0000E+00	1.1102E-16	0.1110E-13 %
$V(Y)$	1.0271E-01	1.0271E-01	1.3878E-17	0.1351E-13 %
$V(Y)([4])$	1.0271E-01	4.8576E-02	5.4131E-02	0.5270E+02 %
$S(Y)$	9.4209E-03	9.4209E-03	2.1337E-16	0.2265E-11 %
$K(Y)$	2.4306E-02	2.4306E-02	5.4123E-16	0.2227E-11 %

Table 3: Computed statistics versus exact solution for anchor point 2.

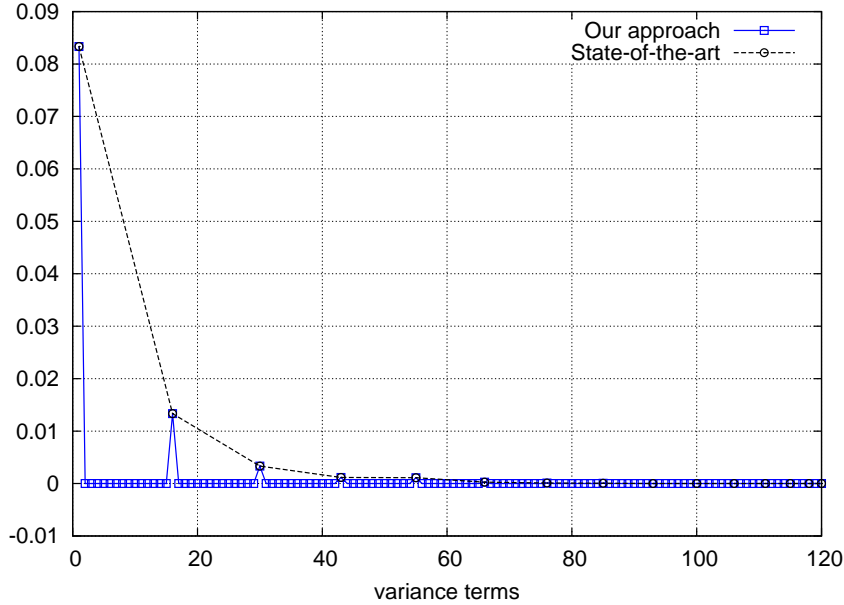


Figure 1: The contributions of 120 terms in (25) for our variance decomposition approach and 15 terms in (23) for the method in [4] using anchor point 1.

results with 6 quadrature points per dimension. The state-of-the-art approach in [4] gives a large error of 52.7% for variance. This confirms the sensitivity to anchor point of this approach.

As done previously, the contributions of 120 terms in (25) and 15 terms in (23) are plotted in Fig. 2. It is observed the 105 terms involved in term (a) in (25) gives non-negligible positive contributions for the variance. In order to obtain a method which is not sensitive to the anchor point, one would need to take into account the covariance decomposition.

- **Anchor point 3.** Finally, let us evaluate our approach using the anchor point  $\mathbf{c}$  such that

$$c_k = 0.000001, \text{ for } k = 1, 2, 3, 4. \quad (44)$$

Table 4 shows the corresponding numerical results. As in the previous cases, our approach gives almost exact results for all statistics. The state-of-the-art approach gives a very large error of 130% for variance, thus providing a very inaccurate result.

	exact	computed	abs err	rel err
$E(Y)$	1.0000E+00	1.0000E+00	1.1102E-16	0.1110E-13 %
$V(Y)$	1.0271E-01	1.0271E-01	5.5511E-17	0.5405E-13 %
$V(Y)([4])$	1.0271E-01	2.3637E-01	1.3366E-01	0.1301E+03 %
$S(Y)$	9.4209E-03	9.4209E-03	1.5613E-16	0.1657E-11 %
$K(Y)$	2.4306E-02	2.4306E-02	5.2389E-16	0.2155E-11 %

Table 4: Computed statistics versus exact solution for anchor point 3.

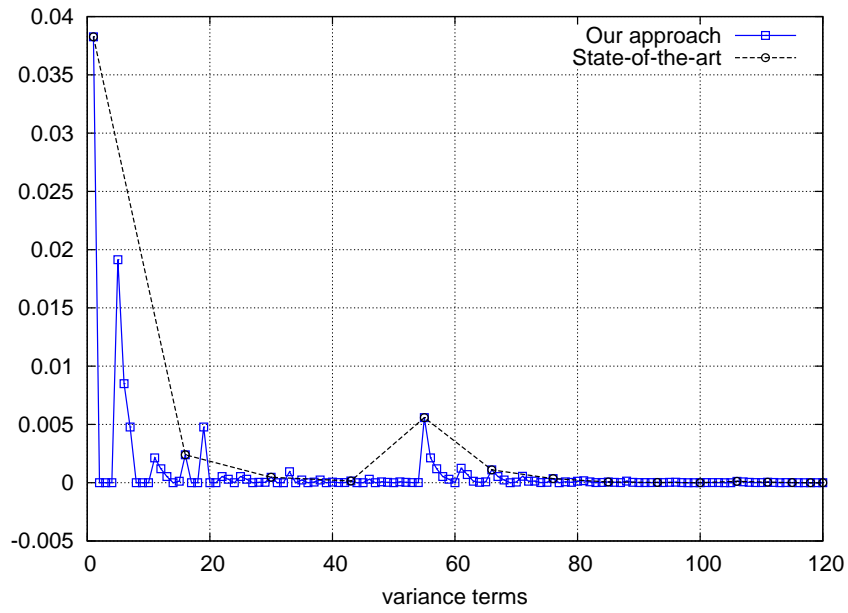


Figure 2: The contributions of 120 terms in (25) for our variance decomposition approach and 15 terms in (23) for the method in [4] using anchor point 2.

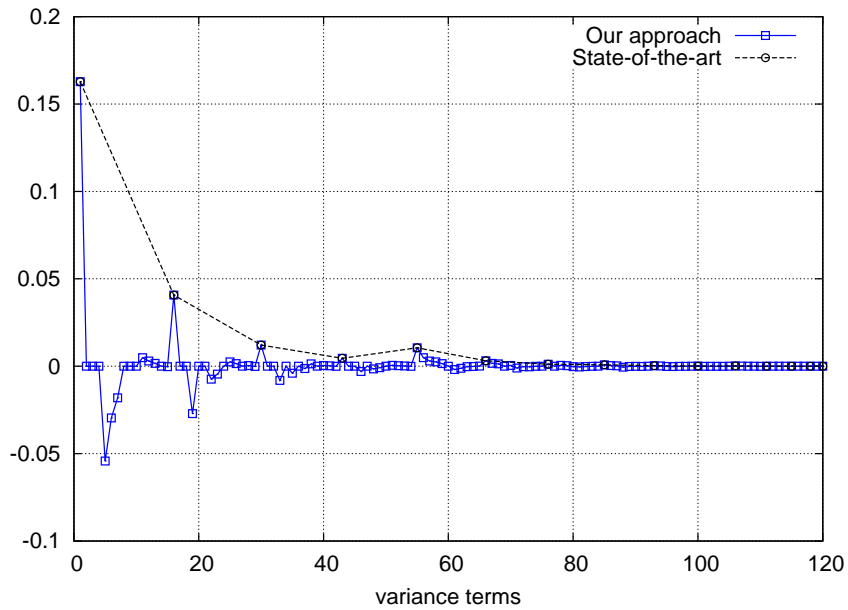


Figure 3: The contributions of 120 terms in (25) for our variance decomposition approach and 15 terms in (23) for the method in [4] using anchor point 3.

The contributions of 120 terms in (25) and 15 terms in (23) are plotted in Fig. 3. We

observe the 105 terms involved in term (a) in (25) gives significant (positive and negative) contributions to the total variance.

### 6.1.1 Covariance-based sensitivity analysis

The covariance-based sensitivity analysis presented in Section 4.1 is carried out for this Sobol' function. The three cases with different anchor points studied above are considered in this section. The results are shown in Table 5 - Table 7, respectively for each anchor point: (42), (43) and (44).

Results presented in Table 5 shows that all the correlative contributions of sensitivity indices are almost zero. We can conclude that, with the anchor point defined in (42), the anchored decomposition is very close to the unique orthogonal ANOVA expansion. The *structural* sensitivity index  $\mathcal{S}_{s_i}^a$  (or the *whole* sensitivity index  $\mathcal{S}_{s_i}$ ) represents the importance of component function  $f_{s_i}$ . It is thus pointed out that the importances are of descending order for functions from top to bottom. The total effects  $\mathcal{S}_{s_i}^{T,a}$ ,  $\mathcal{S}_{s_i}^{T,b}$  and  $\mathcal{S}_{s_i}^T$  are also found to be of descending order.

Table 6 shows the covariance-based sensitivity indices with the anchor point 2 (43). Since component functions are no longer mutually orthogonal, the *correlative* sensitivity indices  $\mathcal{S}_{s_i}^b$  can have significant contributions. The importances of component functions with the same number of variables remain of descending order from top to bottom. However, the higher order interaction functions can have more important contributions than the lower order ones. For instance, we have the following relations:

$$\begin{aligned}\mathcal{S}_{X_1, X_2} &> \mathcal{S}_{X_2}, \\ \mathcal{S}_{X_1, X_2, X_3} &> \mathcal{S}_{X_3}.\end{aligned}$$

Finally, we observe the total sensitivity indices  $\mathcal{S}_{s_i}^{T,a}$ ,  $\mathcal{S}_{s_i}^{T,b}$  and  $\mathcal{S}_{s_i}^T$  all have the same ordering as in the previous case presented in Table 5.

Table 7 illustrates results obtained with the anchor point (44). One key difference from the previous two cases is that the *correlative* sensitivity indices  $\mathcal{S}_{s_i}^b$  can have significant negative values. Consequently, the *whole* sensitivity indices  $\mathcal{S}_{s_i}$  can also be negative. Both  $\mathcal{S}_{s_i}$  and  $\mathcal{S}_{s_i}^a$  can be superior than unity (e.g.  $\mathcal{S}_{X_1}^a = 1.58$ ,  $\mathcal{S}_{X_1} = 1.13$ ). When looking at the magnitude of sensitivity indices ( $\mathcal{S}_{s_i}^a$ ,  $\mathcal{S}_{s_i}^b$  and  $\mathcal{S}_{s_i}$ ), as in the case shown in Table 6, the component functions with the same number of variables are found to have the same order of importances (i.e. descending from

	$\mathcal{S}_{s_i}^a$	$\mathcal{S}_{s_i}^b$	$\mathcal{S}_{s_i}$	$\mathcal{S}_{s_i}^{T,a}$	$\mathcal{S}_{s_i}^{T,b}$	$\mathcal{S}_{s_i}^T$
$X_1$	0.811E+00	-0.183E-16	0.811E+00	0.826E+00	-0.226E-16	0.826E+00
$X_2$	0.130E+00	0.164E-16	0.130E+00	0.141E+00	0.275E-16	0.141E+00
$X_3$	0.325E-01	0.147E-16	0.325E-01	0.357E-01	0.158E-16	0.357E-01
$X_4$	0.112E-01	-0.112E-16	0.112E-01	0.124E-01	-0.897E-17	0.124E-01
$X_1, X_2$	0.108E-01	0.334E-18	0.108E-01			
$X_1, X_3$	0.270E-02	0.154E-16	0.270E-02			
$X_1, X_4$	0.936E-03	-0.667E-17	0.936E-03			
$X_2, X_3$	0.433E-03	0.326E-17	0.433E-03			
$X_2, X_4$	0.150E-03	0.190E-16	0.150E-03			
$X_3, X_4$	0.374E-04	-0.481E-17	0.374E-04			
$X_1, X_2, X_3$	0.361E-04	-0.105E-16	0.361E-04			
$X_1, X_2, X_4$	0.125E-04	-0.314E-17	0.125E-04			
$X_1, X_3, X_4$	0.312E-05	-0.426E-17	0.312E-05			
$X_2, X_3, X_4$	0.499E-06	-0.241E-17	0.499E-06			
$X_1, X_2, X_3, X_4$	0.416E-07	0.444E-17	0.416E-07			

Table 5: Covariance-based sensitivity indices for anchor point 1.

	$\mathcal{S}_{s_i}^a$	$\mathcal{S}_{s_i}^b$	$\mathcal{S}_{s_i}$	$\mathcal{S}_i^{T,a}$	$\mathcal{S}_i^{T,b}$	$\mathcal{S}_i^T$
$X_1$	0.373E+00	0.177E+00	0.550E+00	0.443E+00	0.470E+00	0.913E+00
$X_2$	0.233E-01	0.317E-01	0.550E-01	0.801E-01	0.233E+00	0.313E+00
$X_3$	0.460E-02	0.762E-02	0.122E-01	0.173E-01	0.115E+00	0.132E+00
$X_4$	0.146E-02	0.259E-02	0.404E-02	0.557E-02	0.649E-01	0.705E-01
$X_1, X_2$	0.543E-01	0.147E+00	0.202E+00			
$X_1, X_3$	0.107E-01	0.646E-01	0.753E-01			
$X_1, X_4$	0.340E-02	0.357E-01	0.391E-01			
$X_2, X_3$	0.671E-03	0.870E-02	0.937E-02			
$X_2, X_4$	0.212E-03	0.430E-02	0.451E-02			
$X_3, X_4$	0.419E-04	0.119E-02	0.123E-02			
$X_1, X_2, X_3$	0.118E-02	0.250E-01	0.262E-01			
$X_1, X_2, X_4$	0.374E-03	0.135E-01	0.139E-01			
$X_1, X_3, X_4$	0.739E-04	0.518E-02	0.525E-02			
$X_2, X_3, X_4$	0.462E-05	0.705E-03	0.709E-03			
$X_1, X_2, X_3, X_4$	0.728E-05	0.177E-02	0.178E-02			

Table 6: Covariance-based sensitivity indices for anchor point 2.

	$\mathcal{S}_{s_i}^a$	$\mathcal{S}_{s_i}^b$	$\mathcal{S}_{s_i}$	$\mathcal{S}_i^{T,a}$	$\mathcal{S}_i^{T,b}$	$\mathcal{S}_i^T$
$X_1$	0.158E+01	-0.451E+00	0.113E+01	0.173E+01	-0.993E+00	0.739E+00
$X_2$	0.396E+00	-0.169E+00	0.227E+00	0.512E+00	-0.542E+00	-0.305E-01
$X_3$	0.118E+00	-0.560E-01	0.619E-01	0.159E+00	-0.219E+00	-0.608E-01
$X_4$	0.440E-01	-0.218E-01	0.222E-01	0.599E-01	-0.106E+00	-0.457E-01
$X_1, X_2$	0.103E+00	-0.355E+00	-0.252E+00			
$X_1, X_3$	0.306E-01	-0.151E+00	-0.120E+00			
$X_1, X_4$	0.114E-01	-0.806E-01	-0.692E-01			
$X_2, X_3$	0.764E-02	-0.379E-01	-0.302E-01			
$X_2, X_4$	0.285E-02	-0.189E-01	-0.161E-01			
$X_3, X_4$	0.849E-03	-0.624E-02	-0.539E-02			
$X_1, X_2, X_3$	0.150E-02	0.241E-01	0.256E-01			
$X_1, X_2, X_4$	0.559E-03	0.144E-01	0.150E-01			
$X_1, X_3, X_4$	0.166E-03	0.706E-02	0.722E-02			
$X_2, X_3, X_4$	0.416E-04	0.194E-02	0.198E-02			
$X_1, X_2, X_3, X_4$	0.728E-05	-0.151E-02	-0.151E-02			

Table 7: Covariance-based sensitivity indices for anchor point 3.

top to bottom). Moreover, similarly as for the anchor point 2, higher order interaction terms can be considered more important than some lower order functions. Eventually, both the total *structural* and *correlative* effects remain to be in the same order of importances as in Table 5 and Table 6, while the total effect  $\mathcal{S}_i^T$  does not.

## 6.2 Ishigami function

Let us consider the Ishigami function [15] widely used in the literature [12, 16] as a benchmark problem for uncertainty quantification and sensitivity analysis:

$$y = \sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1, \quad (45)$$

where the random input variables  $\mathbf{X} = (X_1, X_2, X_3)$  are uniformly distributed over  $[-\pi, \pi]$ . The constants are set to  $a = 7, b = 0.1$ , as done in [16, 12].

As already mentioned in Section 4, it is emphasized again that the covariance-based sensitivity analysis performed in this work for Ishigami function differs from [12]. Indeed, in [12], RS-HDMR is used to study the Ishigami function with correlated input variables, while we employ anchored ANOVA for investigation of independent input variables. Note also the covariance-based sensitivity analysis in [12] reduces to a single index  $\mathcal{S}_{s_i}$  when treating independent variables. That is, all correlative indices vanish.

	$\mathcal{S}_{s_i}$	$\mathcal{S}_i^T$
$X_1$	0.3138	0.5574
$X_2$	0.4424	0.4424
$X_3$	0	0.2436
$X_1, X_2$	0	
$X_1, X_3$	0.2436	
$X_2, X_3$	0	
$X_1, X_2, X_3$	0	
$V(Y)$	13.845	

Table 8: Analytical variance and variance-based sensitivity indices for Ishigami function.

Exploiting (7), the standard orthogonal ANOVA expansion can be derived analytically:

$$y = f_0 + f_1 + f_2 + f_3 + f_{12} + f_{13} + f_{23} + f_{123}, \quad (46)$$

where

$$\begin{aligned} f_0 &= \frac{1}{2}a, \\ f_1 &= \frac{1}{5} \sin x_1 (b\pi^4 + 5), \\ f_2 &= a \sin^2 x_2 - \frac{1}{2}a, \\ f_3 &= 0, \\ f_{12} &= 0, \\ f_{13} &= bx_3^4 \sin x_1 - \frac{1}{5} b\pi^4 \sin x_1, \\ f_{23} &= 0, \\ f_{123} &= 0. \end{aligned} \quad (47)$$

As presented in [15, 16], the variance and variance decomposition based on (46) can be obtained analytically:

$$\begin{aligned} V(Y) &= \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2}, \\ V_1 &= \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + \frac{1}{2}, \quad V_2 = \frac{a^2}{8}, \quad V_3 = 0, \\ V_{12} &= V_{23} = 0, \quad V_{13} = \frac{8b^2\pi^8}{225}, \\ V_{123} &= 0. \end{aligned} \quad (48)$$

Thus, the variance-based sensitivity indices can be gathered in Table 8.

### 6.2.1 Numerical investigation: Covariance-based sensitivity indices and high order moments computation

Let us then study the covariance-based sensitivities. We just randomly choose the anchor point as

$$\mathbf{c} = (c_1, c_2, c_3) = \left(-\frac{\pi}{2}, \frac{\pi}{4}, \frac{3\pi}{4}\right). \quad (49)$$

$\mu = 6$	$S_{s_j}^a$	$S_{s_j}^b$	$S_{s_j}$	$S_i^{T,a}$	$S_i^{T,b}$	$S_i^T$
$X_1$	0.777E+00	-0.216E+00	0.561E+00	0.178E+01	-0.106E+01	0.720E+00
$X_2$	0.280E+00	0.688E-17	0.280E+00	0.280E+00	0.485E-16	0.280E+00
$X_3$	0.628E+00	-0.628E+00	-0.666E-15	0.163E+01	-0.147E+01	0.159E+00
$X_1, X_2$	0.186E-31	-0.286E-16	-0.286E-16			
$X_1, X_3$	0.100E+01	-0.844E+00	0.159E+00			
$X_2, X_3$	0.225E-31	0.347E-16	0.347E-16			
$X_1, X_2, X_3$	0.585E-31	0.355E-16	0.355E-16			
$V(Y)$	10.741					
$\mu = 8$	$S_{s_j}^a$	$S_{s_j}^b$	$S_{s_j}$	$S_i^{T,a}$	$S_i^{T,b}$	$S_i^T$
$X_1$	0.613E+00	-0.170E+00	0.443E+00	0.141E+01	-0.837E+00	0.568E+00
$X_2$	0.432E+00	-0.205E-16	0.432E+00	0.432E+00	-0.342E-16	0.432E+00
$X_3$	0.497E+00	-0.497E+00	0.167E-15	0.129E+01	-0.116E+01	0.125E+00
$X_1, X_2$	0.252E-31	-0.830E-16	-0.830E-16			
$X_1, X_3$	0.792E+00	-0.667E+00	0.125E+00			
$X_2, X_3$	0.286E-32	-0.134E-16	-0.134E-16			
$X_1, X_2, X_3$	0.339E-31	0.827E-16	0.827E-16			
$V(Y)$	13.589					
$\mu = 10$	$S_{s_j}^a$	$S_{s_j}^b$	$S_{s_j}$	$S_i^{T,a}$	$S_i^{T,b}$	$S_i^T$
$X_1$	0.602E+00	-0.167E+00	0.435E+00	0.138E+01	-0.822E+00	0.558E+00
$X_2$	0.442E+00	0.181E-16	0.442E+00	0.442E+00	0.330E-16	0.442E+00
$X_3$	0.488E+00	-0.488E+00	-0.222E-15	0.127E+01	-0.114E+01	0.123E+00
$X_1, X_2$	0.218E-31	-0.319E-16	-0.319E-16			
$X_1, X_3$	0.778E+00	-0.655E+00	0.123E+00			
$X_2, X_3$	0.867E-32	0.147E-16	0.147E-16			
$X_1, X_2, X_3$	0.425E-31	0.321E-16	0.321E-16			
$V(Y)$	13.839					
$\mu = 12$	$S_{s_j}^a$	$S_{s_j}^b$	$S_{s_j}$	$S_i^{T,a}$	$S_i^{T,b}$	$S_i^T$
$X_1$	0.602E+00	-0.167E+00	0.435E+00	0.138E+01	-0.822E+00	0.558E+00
$X_2$	0.442E+00	0.112E-16	0.442E+00	0.442E+00	0.156E-16	0.442E+00
$X_3$	0.487E+00	-0.487E+00	0.722E-15	0.126E+01	-0.114E+01	0.123E+00
$X_1, X_2$	0.230E-31	-0.390E-16	-0.390E-16			
$X_1, X_3$	0.777E+00	-0.655E+00	0.123E+00			
$X_2, X_3$	0.101E-31	0.263E-16	0.263E-16			
$X_1, X_2, X_3$	0.398E-31	0.170E-16	0.170E-16			
$V(Y)$	13.845					

Table 9: Anchored ANOVA method. Computed variance and covariance-based sensitivity indices for Ishigami function.

In order to demonstrate the convergence of the method, 1D Gauss-Legendre quadrature points of increasing order ( $\mu = 6, 8, 10, 12$ ) are employed. The tensor product rule is used for generating multi-dimensional points. Thus, the number of required model output evaluations is given by

$$N_{eval} = 1 + N\mu + \binom{N}{2}\mu^2 + \dots + \binom{N}{N}\mu^N. \quad (50)$$

In the case of Ishigami function with three input variables, it holds that

$$N_{eval} = 1 + 3\mu + 3\mu^2 + \mu^3.$$

The resulting sensitivity indices and the variance of output  $V(Y)$  are shown in Table 9. We observe that it requires about  $\mu \geq 8$  points per dimension (That is  $\geq 729$  model evaluations) in order to obtain accurate result of output variance<sup>1</sup>. Note this is relatively high, since the Ishigami function is known to be strongly nonlinear and non-monotone; moreover, all input variables are important. Meanwhile, Table 9 clearly shows the computed variance converges to the analytical

<sup>1</sup>In [16], it requires about 7-th order polynomial chaos expansion to get accurate results.

	Analytical	$\mu = 6$	$\mu = 10$	$\mu = 12$	$\mu = 14$	$\mu = 16$
$S(Y)$	0	-3.317	2.299	0.163	4.566E-3	6.306E-5
$K(Y)$	672.234	431.203	664.163	662.956	670.590	672.125
$N_{eval}$		343	1331	2197	3375	4913

Table 10: Anchored ANOVA method. Computed skewness and kurtosis for Ishigami function. Number of decomposition terms is  $\mathcal{M} = 84$  and  $\mathcal{M} = 210$  for skewness and kurtosis, respectively.  $N_{eval}$  is the number of model evaluations.

one when increasing the quadrature order. On the other hand, the converged covariance-based sensitivity indices  $\mathcal{S}_{s_i}$  and  $\mathcal{S}_{s_i}^T$  differ from the analytical variance-based ones. We emphasize this is expected, since anchored ANOVA provides a different expansion from the standard orthogonal ANOVA. However, the converged  $\mathcal{S}_{s_i}$  and  $\mathcal{S}_{s_i}^T$  indicate the same order of importances as by analytical variance-based indices in Table 8:

$$\mathcal{S}_{X_2} > \mathcal{S}_{X_1} > \mathcal{S}_{X_1, X_3}. \quad (51)$$

$$\mathcal{S}_{X_1}^T > \mathcal{S}_{X_2}^T > \mathcal{S}_{X_3}^T. \quad (52)$$

Note, for all the degrees of quadrature, the importance order given by the computed total sensitivity indices  $\mathcal{S}_i^T$  remains the same. Note also the zero value indices are computed exactly by  $\mathcal{S}_{s_i}$ , even at a low order of quadrature integration. Concerning the "structural" and "correlative" part of converged  $\mathcal{S}_{s_i}$ , namely  $\mathcal{S}_{s_i}^a$  and  $\mathcal{S}_{s_i}^b$ , the computed order of importances is as follows:

$$\mathcal{S}_{X_1, X_3}^a > \mathcal{S}_{X_1}^a > \mathcal{S}_{X_3}^a > \mathcal{S}_{X_2}^a, \quad (53)$$

$$|\mathcal{S}_{X_1, X_3}^b| > |\mathcal{S}_{X_3}^b| > |\mathcal{S}_{X_1}^b|. \quad (54)$$

Indices not shown in (53) or (54) are found to be of zero value. It is observed that (53) differs significantly from (51), even for first-order indices. This implies the "whole" indices  $\mathcal{S}_{s_i}$  provide more reliable importances analysis than  $\mathcal{S}_{s_i}^a$ . Note in particular the value of  $\mathcal{S}_{X_3}^a$  ( $\mathcal{S}_{X_3}^b$ ) is significant while  $\mathcal{S}_{X_3}$  vanishes.

As it has been analyzed in previous example, different choices of anchor point can make high order component functions more important than low order ones; however, the order of importances of components with the same order are proved, by numerical experiments, to remain the same for most cases. Note also the method in [4] is not investigated here for Ishigami function with the anchor point (49), because it provides unacceptable results.

The skewness and kurtosis of Ishigami function can be obtained analytically:

$$S(Y) = 0, \quad (55)$$

$$K(Y) = \frac{3}{136} b^4 \pi^{16} + \frac{3}{26} b^3 \pi^{12} + \frac{1}{24} \pi^8 b^2 a^2 + \frac{1}{4} b^2 \pi^8 + \frac{3}{20} \pi^4 b a^2 + \frac{3}{10} b \pi^4 + \frac{3}{8} + \frac{3}{8} a^2 + \frac{3}{128} a^4. \quad (56)$$

Still using the anchor point in (49), the results of computed skewness  $S(Y)$  and kurtosis  $K(Y)$  are reported in Table 10, compared to the analytical solution. We observe first of all that numerical solutions of the anchored ANOVA method converge to analytical solution when increasing the quadrature order. It is shown accurate skewness calculation can be obtained from about  $\mu \geq 12$ , and kurtosis from  $\mu \geq 14$ . As already mentioned before, this low convergence rate is due to the strong nonlinearity and non-monotonicity of the Ishigami function.



### 6.2.2 Analytical investigation: Sensitivity of component variance on anchor point

The anchored ANOVA expansion can be in fact exactly obtained for the Ishigami function. Indeed, using the relation (15), we have

$$y = f_0 + f_1 + f_2 + f_3 + f_{12} + f_{13} + f_{23} + f_{123}, \quad (57)$$

where

$$\begin{aligned} f_0 &= \sin c_1 + a \sin^2 c_2 + bc_3^4 \sin c_1, \\ f_1 &= (1 + bc_3^4)(\sin x_1 - \sin c_1), \\ f_2 &= a(\sin^2 x_2 - \sin^2 c_2), \\ f_3 &= (bx_3^4 - bc_3^4) \sin c_1, \\ f_{12} &= 0, \\ f_{13} &= (bx_3^4 - bc_3^4)(\sin x_1 - \sin c_1), \\ f_{23} &= 0, \\ f_{123} &= 0. \end{aligned} \quad (58)$$

In the case of the anchor point used in (49), the analytical component functions of anchored ANOVA are presented in Fig. 4 and Fig. 5, compared to the component functions derived from standard orthogonal ANOVA. We observe from Fig. 4, for anchored case, that all input variables give significant variation on their corresponding first-order function. Moreover, none of the component functions monotonically changes with respect to the input. The function  $f_1(x_1)$  has a pattern that is similar to the one of the standard case. However, it has a positive average value of 4.082, while component functions of standard ANOVA always have zero mean value. With the specific anchor point of (49), the component  $f_2(x_2)$  remains identical for both cases. The variation of  $f_3(x_3)$  with respect to  $x_3$  in anchored case has a profile of 4-th order polynomial, while  $f_3(x_3)$  in the framework of standard ANOVA approach remains constant and has zero value (see (48)). Fig. 5 shows, for both anchored and standard ANOVA, a very structured pattern for the interactions between  $x_1$  and  $x_3$ . Like in the first-order case in Fig. 4,  $f_{13}$  in anchored ANOVA expansion differs from the one in standard case. Note that the component functions evaluated by Gauss-Legendre rule in our numerical anchored approach should, by definition, give the exact values provided by the analytical functions. This is confirmed in Fig. 6 and Fig. 7: The component function values evaluated on sampling points match perfectly with the exact functions.

Comparing (47) and (58), we know that if the anchor point  $\mathbf{c}$  satisfies

$$\begin{aligned} \sin c_1 &= 0, \\ \sin^2 c_2 &= \frac{1}{2}, \\ c_3^4 &= \frac{1}{5}\pi^4, \end{aligned} \quad (59)$$

the anchored ANOVA expansion becomes identical to the standard orthogonal ANOVA. For real engineering problems, this best choice of anchor point is not available. The variance methods available in the literature [4] are reported to be very sensitive to the anchor point  $\mathbf{c}$ . Results can be unacceptable for some choices of  $\mathbf{c}$ , as shown in Sobol' function test. Let us study here this

sensitivity phenomenon as far as the Ishigami function is concerned.

The variances of component functions using the anchored expansion with an arbitrary anchor point  $\mathbf{c} = (c_1, c_2, c_3)$  can be obtained analytically:

$$\begin{aligned}
 V_1 &= \frac{1}{2} + bc_3^4 + \frac{1}{2} b^2 c_3^8, \\
 V_2 &= \frac{1}{8} a^2, \\
 V_3 &= \frac{16}{225} b^2 \sin^2 c_1 \pi^8, \\
 V_{12} &= 0, \\
 V_{13} &= -\frac{1}{450} b^2 (32 \pi^8 \cos^2 c_1 - 57 \pi^8 + 90 c_3^4 \pi^4 - 225 c_3^8), \\
 V_{23} &= 0, \\
 V_{123} &= 0,
 \end{aligned} \tag{60}$$

The sum of these above variances of components, denoted by  $V^a$ , is given by

$$V^a = \frac{1}{2} + bc_3^4 + b^2 c_3^8 + \frac{1}{8} a^2 + \frac{89}{450} b^2 \pi^8 - \frac{32}{225} b^2 \pi^8 \cos^2 c_1 - \frac{1}{5} b^2 c_3^4 \pi^4. \tag{61}$$

By definition, the "structural" sensitivity indices can be computed by dividing component variances in (60) by the total variance  $V(Y)$  in (48):

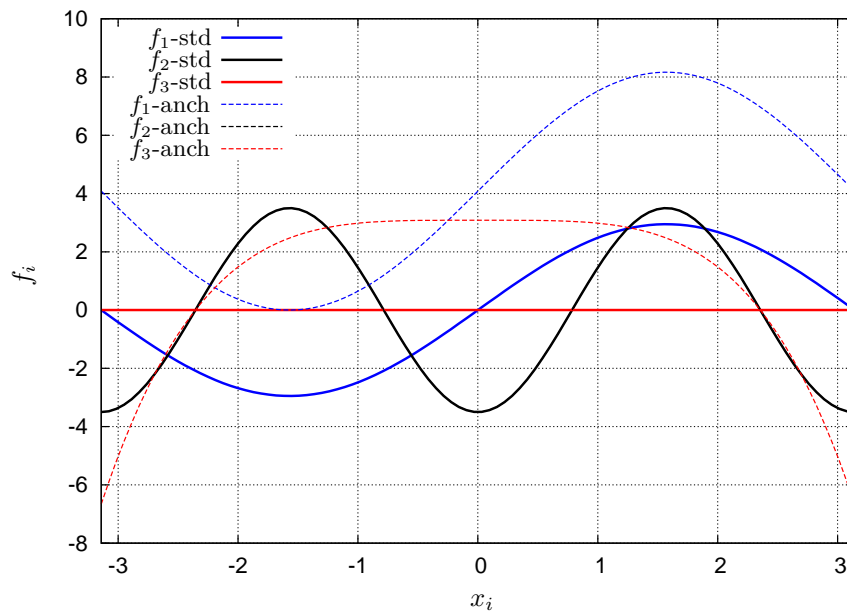


Figure 4: The first-order analytical component functions for Ishigami function. Standard v.s. anchored ANOVA.

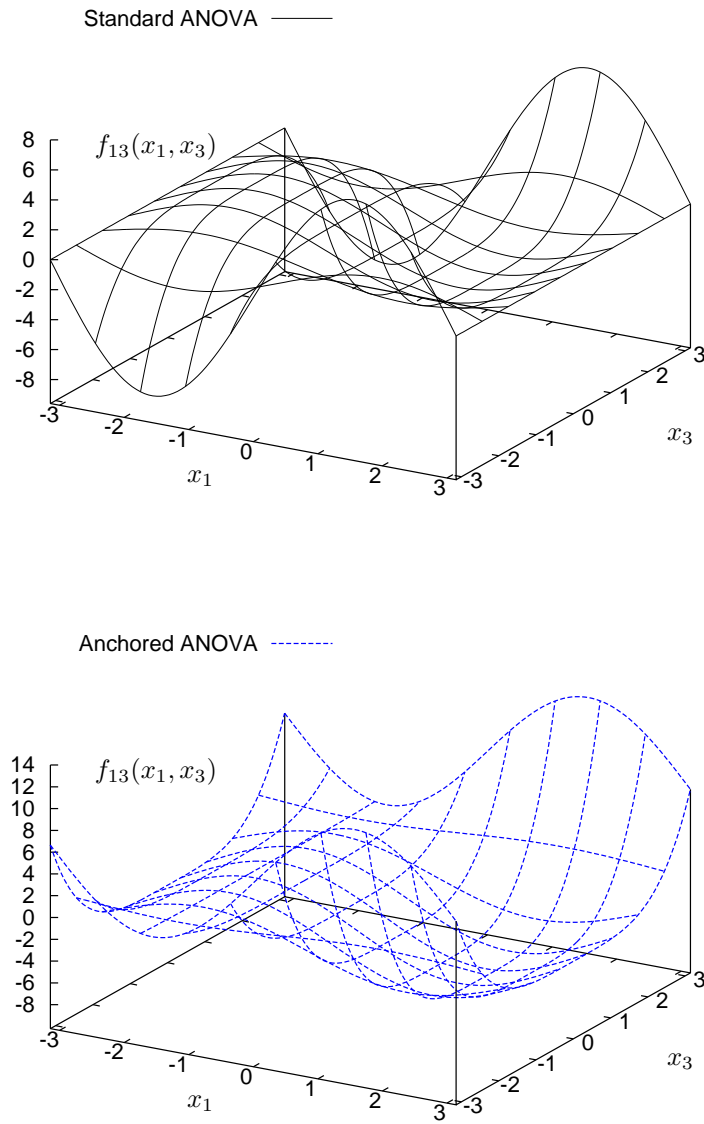


Figure 5: The significant analytical second-order component function  $f_{13}$  for Ishigami function. Standard v.s. anchored ANOVA.

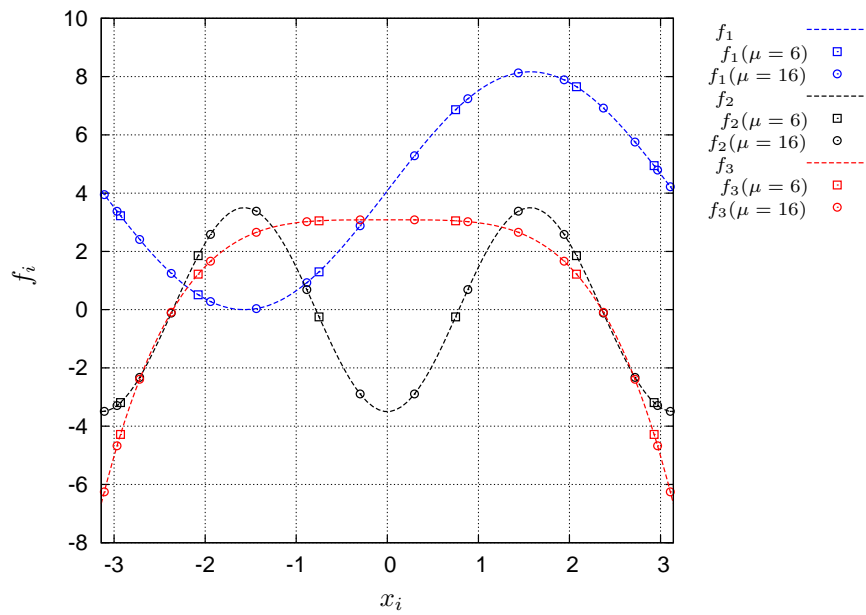


Figure 6: The analytical first-order component functions for Ishigami function using anchored ANOVA. Numerical approach using Gauss-Legendre sampling points provides sampling values of analytical  $f_i$ .

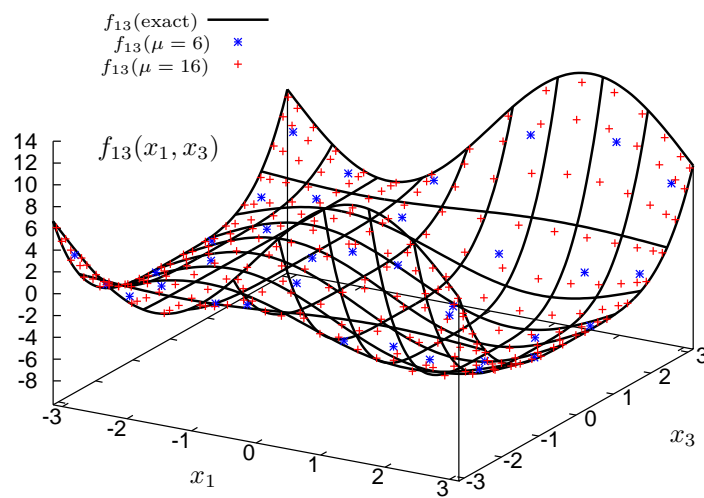


Figure 7: The analytical second-order component function  $f_{13}$  for Ishigami function using anchored ANOVA. Numerical approach using Gauss-Legendre sampling points provides sampling values of analytical  $f_{13}$ .

$$\begin{aligned}
\mathcal{S}_{X_1}^a &= 180 \frac{1 + 2bc_3^4 + b^2c_3^8}{20b^2\pi^8 + 72b\pi^4 + 180 + 45a^2}, \\
\mathcal{S}_{X_2}^a &= 45 \frac{a^2}{20b^2\pi^8 + 72b\pi^4 + 180 + 45a^2}, \\
\mathcal{S}_{X_3}^a &= \frac{128}{5} \frac{b^2 \sin^2 c_1 \pi^8}{20b^2\pi^8 + 72b\pi^4 + 180 + 45a^2}, \\
\mathcal{S}_{X_1, X_2}^a &= 0, \\
\mathcal{S}_{X_1, X_3}^a &= -\frac{4}{5} \frac{b^2 (32\pi^8 \cos^2 c_1 - 57\pi^8 + 90c_3^4\pi^4 - 225c_3^8)}{20b^2\pi^8 + 72b\pi^4 + 180 + 45a^2}, \\
\mathcal{S}_{X_2, X_3}^a &= 0, \\
\mathcal{S}_{X_1, X_2, X_3}^a &= 0.
\end{aligned} \tag{62}$$

We observe first of all that none of these indices depends on  $c_2$ . The index  $\mathcal{S}_{X_2}^a$  remains constant. The indices  $\mathcal{S}_{X_1, X_2}^a$ ,  $\mathcal{S}_{X_2, X_3}^a$  and  $\mathcal{S}_{X_1, X_2, X_3}^a$  are all zero regardless of the value of  $a$ ,  $b$  or  $\mathbf{c}$ . The variations of indices  $\mathcal{S}_{X_1}^a$  (only depending on  $c_3$ ),  $\mathcal{S}_{X_3}^a$  (only depending on  $c_1$ ), and  $\mathcal{S}_{X_1, X_3}^a$  (depending on  $c_1$  and  $c_3$ ) are plotted in Fig. 8. The first-order indices  $\mathcal{S}_{X_1}^a$  and  $\mathcal{S}_{X_3}^a$  are shown to be strongly nonlinear and non-monotone with respect to  $c_3$  and  $c_1$ . In particular,  $\mathcal{S}_{X_1}^a$  can attain around the value of "4" when  $|c_3|$  is set close to  $\pi$ , while it is known the variance-based index, with standard ANOVA, can only vary in  $[0, 1]$ . The second-order index  $\mathcal{S}_{X_1, X_3}^a$  is also shown to be very nonlinear and non-monotone with respect to the variations of  $c_1$  and  $c_3$ . Similar to  $\mathcal{S}_{X_1}^a$ , the variation of  $\mathcal{S}_{X_1, X_3}^a$  is particularly important when  $c_3$  approaches to its domain extremity.

We remind that the variation of "structural" indices  $\mathcal{S}_{s_i}^a$  represents the sensitivity of component variance  $V_i$  on the anchor point  $\mathbf{c}$ . Thus, it can be confirmed that this sensitivity can be tremendous.

The sum of "structural" sensitivity indices, denoted by  $\mathcal{S}_{sum}^a = \frac{V^a}{V(Y)}$ , is given by

$$\mathcal{S}_{sum}^a = -\frac{1}{5} \frac{-900 - 1800bc_3^4 - 1800b^2c_3^8 - 225a^2 - 356b^2\pi^8 + 256b^2\pi^8 \cos^2 c_1 + 360b^2c_3^4\pi^4}{20b^2\pi^8 + 72b\pi^4 + 180 + 45a^2}. \tag{63}$$

Note that the index  $\mathcal{S}_{sum}^a$  can be regarded as a factor evaluating the error of existing approach (for instance [4]), when computing the output variance. In fact, as already mentioned,  $V^a$  contains component variances, but excludes all covariances between component functions (the term  $(a)$  in (25)). If  $\mathcal{S}_{sum}^a = 1$ , one obtains the exact result. This can happen when, for instance, the condition (59) is satisfied. Fig. 9 (top) shows the variation of  $\mathcal{S}_{sum}^a$  over  $c_1$  and  $c_3$ . Fig. 9 (bottom) presents the corresponding variation by fixing  $c_1$  or  $c_3$  according to (59). It is clearly shown the value of  $\mathcal{S}_{sum}^a$  is in general very sensitive to both  $c_1$  and  $c_3$ . In particular,  $\mathcal{S}_{sum}^a$  can have an approximated value of "7", when fixing  $c_1 = 0$  and  $c_3 = \pi$ . That is, one obtains an output variance which is 700% bigger than the exact solution!

Similar analysis can be realized for other sensitivity indices (e.g. the "correlative" indices  $\mathcal{S}_{s_i}^b$  and the "whole" indices  $\mathcal{S}_{s_i}$ ). For the sake of conciseness, it will not be provided in this paper.

We conclude, for this section, that existing anchored approach with a bad choice of anchor point can give unacceptable results when considering strongly nonlinear and non-monotone functions.

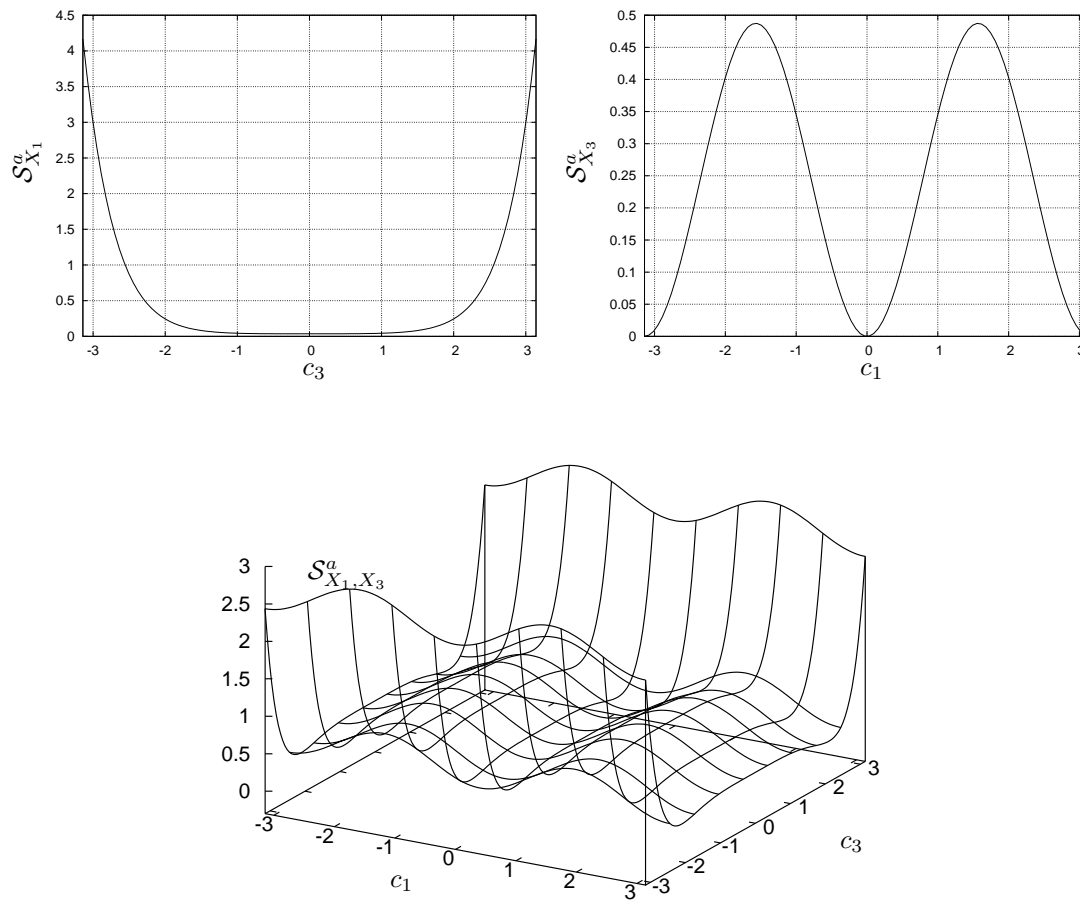


Figure 8: The variation (sensitivity) of "structural" sensitivity indices  $\mathcal{S}_{s_i}^a$  over the choice of anchor point  $\mathbf{c} = (c_1, c_2, c_3)$ .

This problem can be resolved by using the covariance decomposition, i.e. by adding the covariances between all component functions. It is however emphasized here that, when treating real engineering problems without such nonlinearity, one can still have an accurate approximation, provided a good anchor point is given.

### 6.3 Application to the chemical reaction uncertainties during an atmospheric reentry

When considering the design of a reentry vehicle, estimating kinetic and radiative processes variability in the flow is of fundamental importance for yielding an efficient design. During the reentry phase, the spacecraft is decelerated by converting a large amount of kinetic energy into thermal energy and by inducing a strong bow shock in front of the vehicle nose. The significant increase in the gas temperature promotes strong collisions among the gas particles, changes in the chemical composition of the gas, excitation of its internal energy modes, and emission of

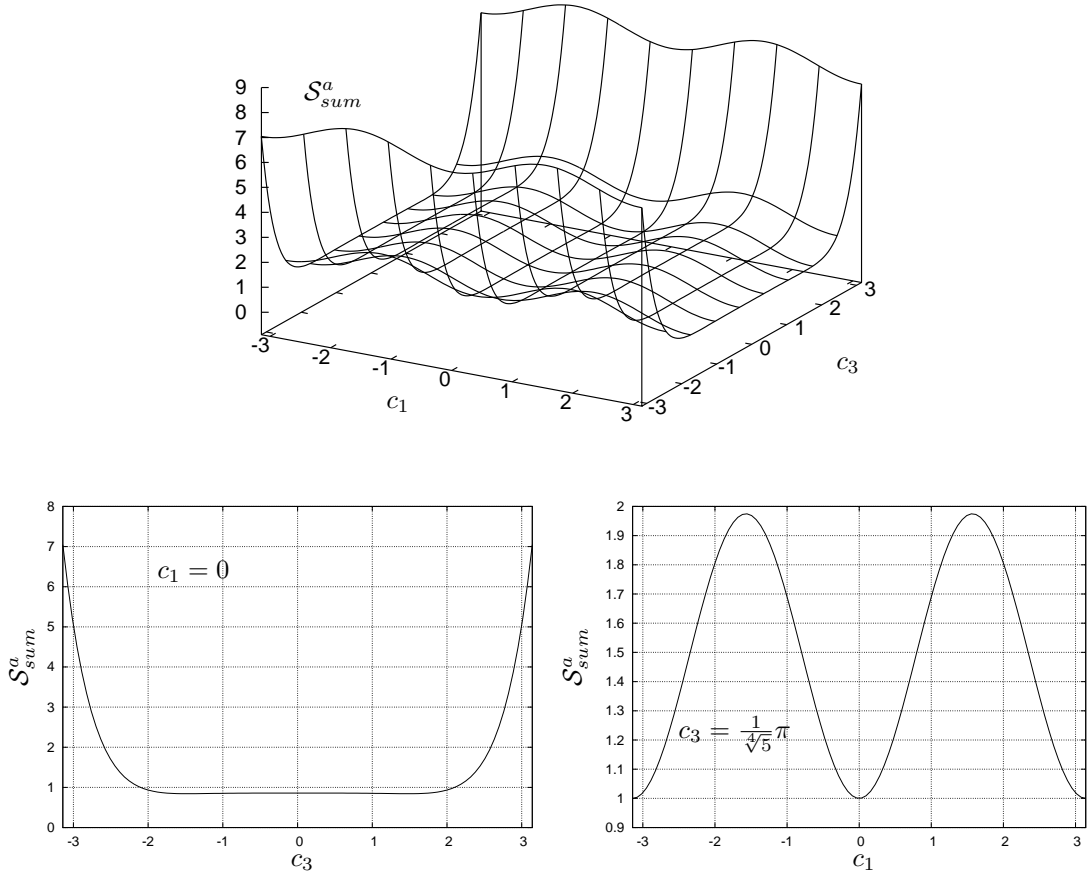


Figure 9: The variation (sensitivity) of "structural" index  $S_{sum}^a$  over the choice of anchor point  $\mathbf{c} = (c_1, c_2, c_3)$ .

radiation.

We focused here on the reaction rate coefficients, that are usually very uncertain, since they are difficult to measure experimentally or to estimate accurately from ab initio calculations. Detailed chemical mechanisms are necessary for an accurate heat-flux prediction but these models increase a lot the dimensionality of the stochastic space (number of uncertain parameters) with respect to conventional multi-temperature models. For this reason, efficient and low-cost uncertainty quantification methods are necessary in order to compute the most important uncertainties and to reduce the model.

One point of the heat flux trajectory is taken into account (flight path angle  $= -12.5^\circ$ ). The stand-off distance of the shock wave is taken equal to 0.022 m. A low pressure point is considered (typically at a free stream pressure of 0.1 Torr) at high altitude and high velocity, where the ionization rate is important. Flow is taken in radiative non equilibrium (for more details see [17]).

Reaction	ID	$\min_i$	nominal <sub><i>i</i></sub>	$\max_i$
$N(1) + e^- \rightleftharpoons N(2) + e^-$	1	1.4476e+15	1.4476e+16	1.4476e+17
$N(1) + e^- \rightleftharpoons N(3) + e^-$	2	3.8673e+14	3.8673e+15	3.8673e+16
$N(2) + e^- \rightleftharpoons N(3) + e^-$	3	3.6169e+14	3.6169e+15	3.6169e+16
$O(1) + e^- \rightleftharpoons O(2) + e^-$	4	8.2744e+11	8.2744e+12	8.2744e+13
$O(1) + e^- \rightleftharpoons O(3) + e^-$	5	5.6319e+10	5.6319e+11	5.6319e+12
$O(2) + e^- \rightleftharpoons O(3) + e^-$	6	2.704881e+13	2.704881e+14	2.704881e+15
$N(1) + e^- \rightleftharpoons N^+ + e^- + e^-$	7	5.1688e+13	5.1688e+14	5.1688e+15
$N(2) + e^- \rightleftharpoons N^+ + e^- + e^-$	8	4.6563e+11	4.6563e+12	4.6563e+13
$O(1) + e^- \rightleftharpoons O^+ + e^- + e^-$	9	3.2477e+11	3.2477e+12	3.2477e+13
$O(2) + e^- \rightleftharpoons O^+ + e^- + e^-$	10	5.1616e+11	5.1616e+12	5.1616e+13

Table 11: Ten reactions are considered uncertain. The reaction rates are log-uniformly distributed between the minimum and maximum values.

Using this condition of the trajectory, the uncertainties on the radiative heat-flux is considered. Then, we study the interactions between the chemical reactions and their influence on the error of the radiative heat flux. The quantity of interest is the radiative heat flux at a distance corresponding to the stand-off distance for the ERC capsule.

The reaction rate coefficients of ten reactions were considered as unknown, the first six were excitation reactions and the next four were ionization. Reactions for which the rate coefficients ( $X_i$ ) are considered uncertain ( $\log_{10} X_i \sim \mathcal{U}(\log_{10} \min_i, \log_{10} \max_i)$ ) are reported in Table 11.

For simplicity, we show our results for ANOVA expansion truncated at order 2. Let us simply choose the nominal values as the anchor point. First of all, the first-order component variances can be easily computed, once the necessary deterministic sampling outputs are obtained. The results are presented in Fig. 10. Obviously, since the ten reaction rates are considered as independent, covariance product between any two first-order component functions vanishes. Next, we can employ the variance-based adaptive criterion (see [4] for more details) in order to retain the active dimensions. Fig. 11 shows the component variances in the descending order by comparing their contributions. We observe that the first six dimensions

$$\{X_1, X_4, X_2, X_5, X_8, X_7\} \quad (64)$$

represent about 99.9% of the total first-order variance. According to [4], these six dimensions are selected as active ones, and they are further used to calculate second-order interaction terms.

After computing the 10 first-order component functions and 15 second-order component functions, one can obtain the following statistics using the covariance decomposition:

$$\begin{aligned} E(Y) &= 1965.42 \text{ kW/m}^2, \\ V_{\text{cov}}(Y) &= 6.911 \text{ (kW/m}^2\text{)}^2. \end{aligned} \quad (65)$$

On the other hand, with the classical variance decomposition approach, the output variance is



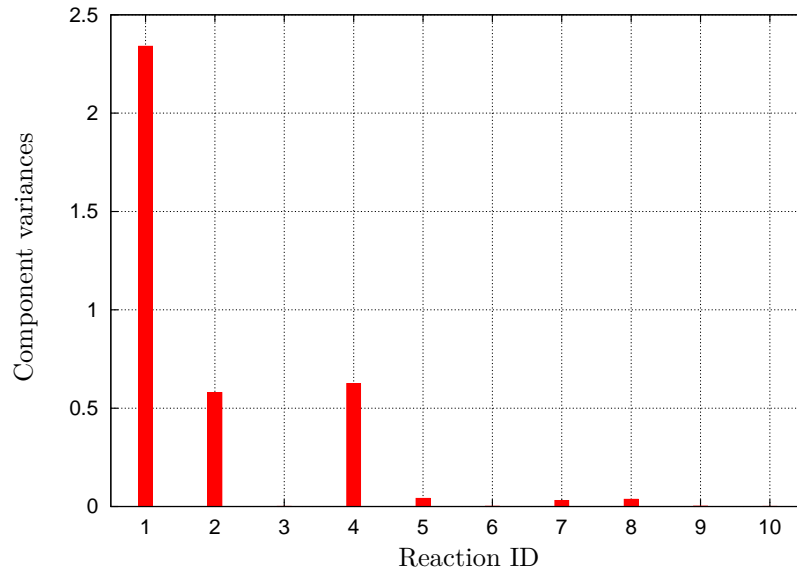


Figure 10: First-order component variances in the chemical reaction problem.

found to be

$$V_{\text{var}}(Y) = 8.090 \text{ (kW/m}^2\text{)}^2. \quad (66)$$

Note that in Appendix A, we show that, for the academic Sobol' function, the covariance decom-

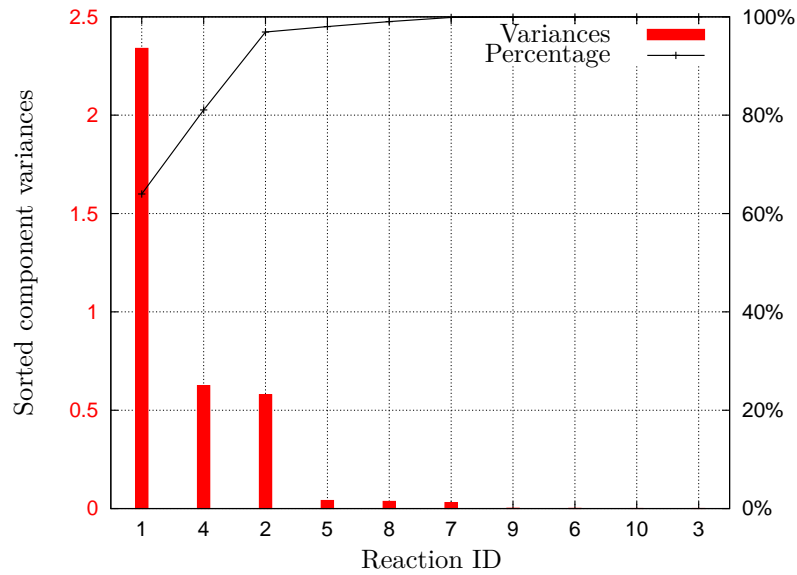


Figure 11: First-order component variances in the chemical reaction problem. The reaction ID is sorted in the descending order regarding the significance of the corresponding variance. The black line illustrates the percentage of the cumulative variance over the total first-order variance.

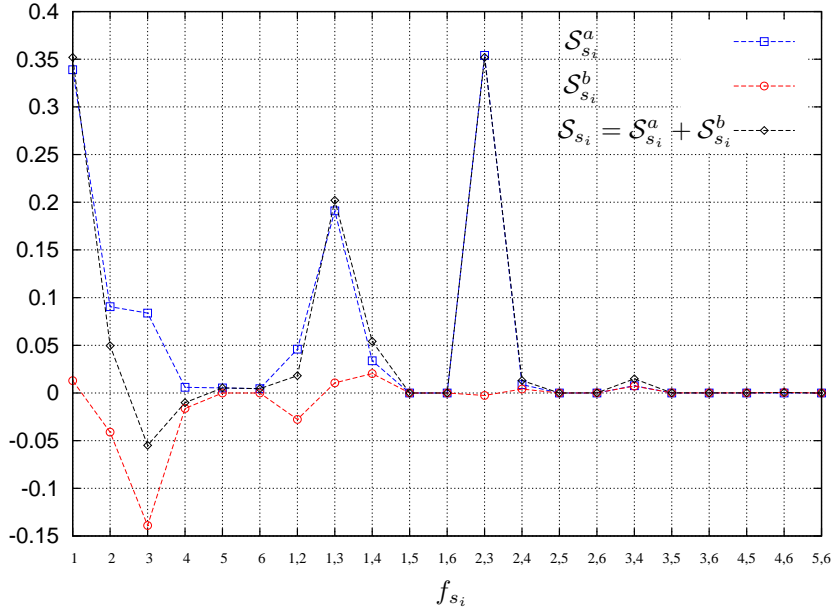


Figure 12: Covariance-based sensitivity indices for the chemical reaction problem. Only first- and second-order component functions are considered. We emphasize, in this figure, only the six active dimensions are taken into account, and there is  $\sum \mathcal{S}_{s_i} = 1$ . The numbering in the axis of abscissas follows (67).

position approach combined with *truncated* anchored ANOVA expansion gives more accurate results than the classical variance decomposition. As also shown in Appendix A, this still remains true when an *adaptivity* strategy is applied in order to reduce the number of second-order component functions.

In order to better understand the difference between (65) and (66), let us focus only on six active dimensions (64). Using the new ordering of reaction rates as expressed in (67),

$$\{X'_1, X'_2, X'_3, X'_4, X'_5, X'_6\} = \{X_1, X_4, X_2, X_5, X_8, X_7\}, \quad (67)$$

Fig. 12 presents the covariance-based sensitivity indices of six active dimensions (including 6 first-order component-based indices and 15 second-order indices). We observe the second-order interaction terms

$$\begin{aligned} (X'_1, X'_3) &= (X_1, X_2), \\ (X'_2, X'_3) &= (X_4, X_2), \end{aligned} \quad (68)$$

are shown to be very important compared to first-order indices. The blue line illustrates the structural indices  $\mathcal{S}_{s_i}^a$  leading to the result of (66), while the black line illustrates the whole indices  $\mathcal{S}_{s_i}$  linked to the result of (65). Note finally the correlative indices  $\mathcal{S}_{s_i}^b$  represented by the red line is in fact related to the difference between (65) and (66). It is observed the difference is significant for the indices corresponding to the following variables and group:

$$\begin{aligned} X'_2 &= X_4, \\ X'_3 &= X_2, \\ (X'_1, X'_2) &= (X_1, X_4). \end{aligned} \quad (69)$$

For the specific case of anchor point chosen as nominal values, classical variance decomposition approach that neglects correlative indices for (69) provides output variance quite different from covariance approach.

Finally, we have checked that the computed output variance (65) is less sensitive to anchor point than (66). For the sake of conciseness, results with other anchor points are not shown in this paper.

## 7 Conclusions and perspectives

In this paper, the standard and anchored versions of ANOVA decomposition have been reviewed. Both of them are an exact expansion of model output. Anchored ANOVA is computationally more feasible, since only sampling outputs are needed to evaluate the component functions, while numerical multi-dimensional integrations must be computed when using standard ANOVA method.

Concerning the computation of the output variance, the covariance decomposition, though requiring to resolve more integrals, provides a general exact approach which is not sensitive to the choice of anchor point, compared to the one in [4]. The covariance-based sensitivity indices are then introduced to estimate the importances of variables. A 4-dimensional Sobol' function and the Ishigami function have been exhaustively studied using the covariance decomposition and the covariance-based sensitivity indices. In particular, the sensitivity of existing approach over the anchor point for computing output variance has been analyzed via the Ishigami function test.

Also, we have proposed a unique algorithm for an accurate computation of high order statistical moments. In fact, the formula of covariance decomposition (25) of output variance can also be rewritten using the multi-indices as for skewness and kurtosis (see Appendix B). Numerical experiments confirm the numerical solution using the proposed method for decomposition of statistics converges to the exact solution.

The decomposition of high order statistics is generally very prohibitive. In fact, we need to compute 680 terms for skewness and 3060 terms for kurtosis in the example studied in Section 6.1. The number of component functions  $\mathcal{N}$  in ANOVA decomposition (thus also the number of statistics decomposition terms  $\mathcal{M}$ ) increases exponentially with respect to the dimension  $N$  of deterministic solver. In order to reduce the computational cost, future work will be directed towards new adaptive criteria aiming to retain the active dimensions and the effective terms that give the most significant contributions.

Another perspective consists in proposing moment-independent sensitivity indices (see for instance [18]) within the anchored ANOVA framework.

## A Sobol' function test with truncated expansion and adaptivity

The purpose of this section is to show, for academic Sobol' function, that our covariance decomposition approach converges more quickly, with respect to the truncation order of anchored ANOVA expansion, than the classical variance decomposition. Furthermore, we show, when

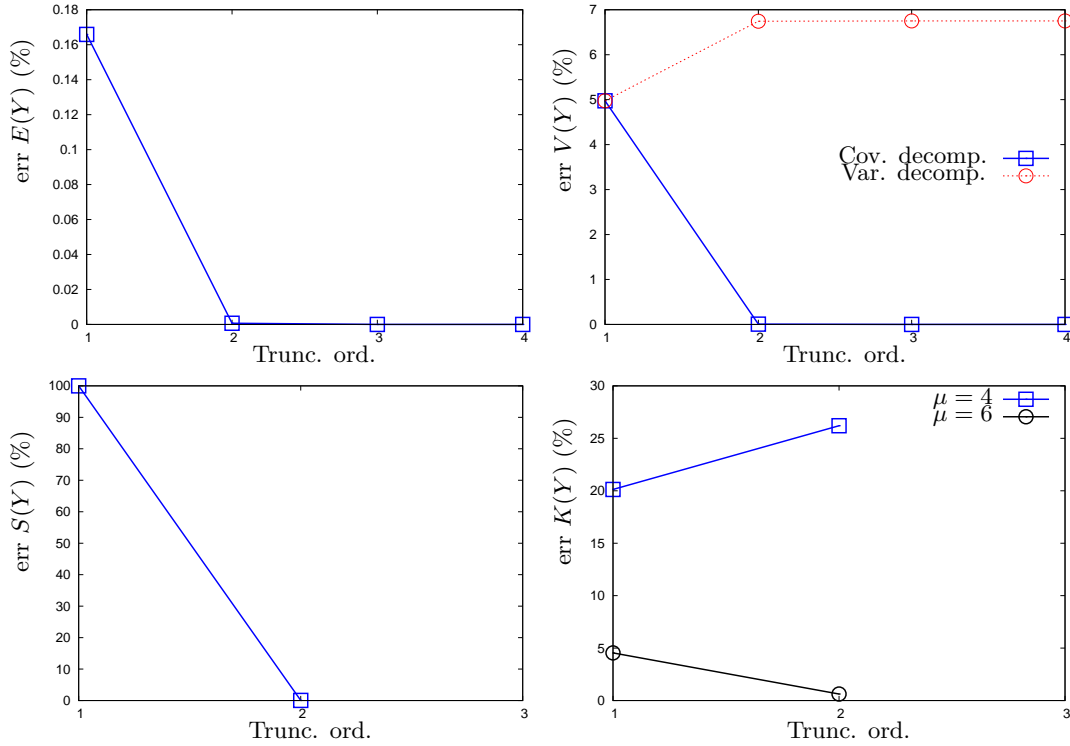


Figure 13: 8-dimensional Sobol' function test. Anchor point  $\mathbf{c}_1$ . Relative error of four statistics in function of truncation order of anchored ANOVA expansion.

using the adaptivity strategy for instance the variance-based one in [4], that our approach can still provide more accurate results.

Note that we consider the same case as in Section 6.1, except that the number of stochastic dimensions is now  $N = 8$ . If not otherwise mentioned, 4 quadrature points for 2 elements per dimension are used. Let us choose the anchor point  $\mathbf{c}_1$  as follows

$$\left\{ \mathbf{c}_1 | f^{(k)}(c_k) = (V^{(k)} + (E^{(k)})^2)/E^{(k)} \right\}. \quad (70)$$

Note that (70) has been used in [4]. With blue line, Fig. 13 illustrates the relative error of mean, variance, skewness and kurtosis, obtained with our approach, as a function of the truncation order. We observe that the error is very small from order 2 except for kurtosis. However, when increasing the number of quadrature points per dimension, we see the error for kurtosis is also decreasing very fast. The red line is the variance result obtained using classical variance decomposition approach. It shows that the error does not decrease, and from order 2 it remains quasi-constant with a value of 6.75%.

Let us now evaluate our numerical approach by using a variance-based adaptivity strategy [4]. It can be found that the contributions of the first 5 components represent nearly 99% of the total first-order variance. Thus, we retain them as active dimensions from order 2. Fig. 14 presents the results for mean and variance. For the covariance decomposition of the output variance, we

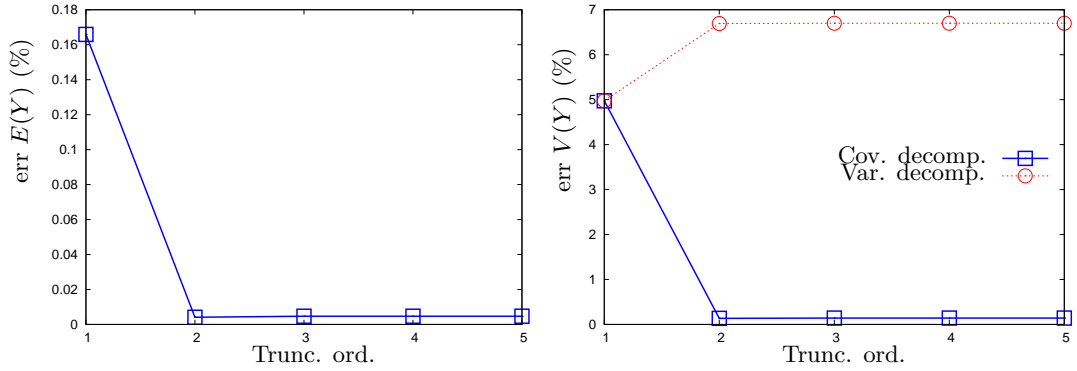


Figure 14: 8-dimensional Sobol' function test. Anchor point  $\mathbf{c}_1$ . Variance-based adaptive strategy is used [4]. 5 active dimensions are retained. Relative error of mean and variance in function of truncation order of anchored ANOVA expansion.

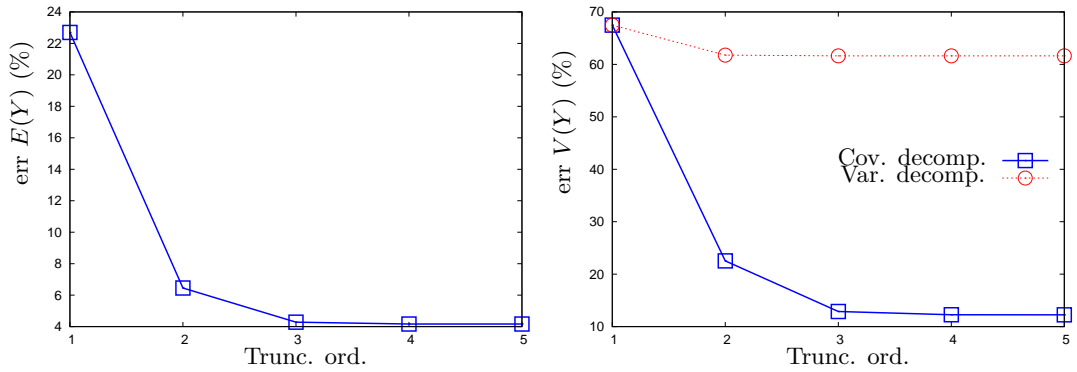


Figure 15: 8-dimensional Sobol' function test. Anchor point  $\mathbf{c}_2$ . Variance-based adaptive strategy is used [4]. 5 active dimensions are retained. Relative error of mean and variance in function of truncation order of anchored ANOVA expansion.

observe that the error is slightly bigger (constant value of 0.14% from order 2) than in Fig. 13, while the error of variance decomposition approach is more or less the same as before.

When a second anchor point  $\mathbf{c}_2$  is chosen as follows

$$\{\mathbf{c}_2 | c_k = 0.5\}, \quad (71)$$

results are shown in Fig. 15. In general, mean and variance error of both approaches increases. The classical variance decomposition gives a constant error of 61.6% from order 2, while our approach for variance provides a decreasing error from order 2, with a final value of 12.2% at the 5<sup>th</sup> order.

## B Variance computation (Anchored ANOVA)

Utilizing the multi-indices introduced in Section 5.1, the variance computation formula (25) in the anchored ANOVA case can be rewritten as follows:

$$\begin{aligned}
 V(Y) &= E[(Y - E(Y))^2] \\
 &= E \left[ \left( \sum_{i=1}^{\mathcal{N}} (f_{s_i} - E(f_{s_i})) \right)^2 \right] \\
 &= \sum_{j=1}^{\mathcal{M}} E \left[ \beta_j^V \prod_{i=1}^{\mathcal{N}} (\tilde{f}_{s_i})^{\alpha_i^{2,j}} \right].
 \end{aligned} \tag{72}$$

Here we have defined

$$\tilde{f}_{s_i} = f_{s_i} - E(f_{s_i}).$$

The total number  $\mathcal{M}$  is given by

$$\mathcal{M} = \binom{\mathcal{N}}{1} + \binom{\mathcal{N}}{2}.$$

The coefficient  $\beta_j^V$  is provided by

$$\beta_j^V = \begin{cases} 1 & \text{if } \{\alpha_i^{2,j}\} \supset \{2\}, \\ \binom{2}{1} = 2 & \text{if } \{\alpha_i^{2,j}\} \supset \{1, 1\}. \end{cases}$$

## References

- [1] Ma X, Zabaras N. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *Journal of Computational Physics* May 2010; **229**(10):3884–3915, doi:10.1016/j.jcp.2010.01.033. URL <http://linkinghub.elsevier.com/retrieve/pii/S0021999110000562>.
- [2] Gao Z, Hesthaven JS. On ANOVA expansions and strategies for choosing the anchor point. *Applied Mathematics and Computation* Dec 2010; **217**(7):3274–3285, doi:10.1016/j.amc.2010.08.061. URL <http://linkinghub.elsevier.com/retrieve/pii/S0096300310009252>.
- [3] Zhang Z, Choi M, Karniadakis GE. Anchor Points Matter in ANOVA Decomposition. *Spectral and High Order Methods for Partial Differential Equations, Lecture Notes in Computational Science and Engineering* 2011; **76**:347–355, doi:10.1007/978-3-642-15337-2. URL <http://www.springerlink.com/index/10.1007/978-3-642-15337-2>.
- [4] Yang X, Choi M, Lin G, Karniadakis GE. Adaptive ANOVA decomposition of stochastic incompressible and compressible flows. *Journal of Computational Physics* 2012; **231**(4):1587–1614, doi:10.1016/j.jcp.2011.10.028. URL <http://dx.doi.org/10.1016/j.jcp.2011.10.028>.

- [5] Wang X. On the approximation error in high dimensional model representation. *Proceedings of the 2008 Winter Simulation Conference*, x, 2008; 453–462.
- [6] Sobol' I. Theorems and examples on high dimensional model representation. *Reliability Engineering & System Safety* Feb 2003; **79**(2):187–193, doi:10.1016/S0951-8320(02)00229-6. URL <http://linkinghub.elsevier.com/retrieve/pii/S0951832002002296>.
- [7] Archer GEB, Saltelli A, Sobol' IM. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation* 1997; **58**:99–120.
- [8] Sobol' IM. On sensitivity estimation for nonlinear mathematical models. *Matem. Mod.* 1990; **2**(1):112–118.
- [9] Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* Feb 2001; **55**(1-3):271–280, doi:10.1016/S0378-4754(00)00270-6. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378475400002706>.
- [10] Zhang Z, Choi M, Karniadakis GE. Error estimates for the ANOVA method with Polynomial Chaos interpolation: Tensor product functions. *SIAM J. Sci. Comput.* 2012; **34**(2):A1165–A1186.
- [11] Xu H, Rahman S. A generalized dimension-reduction method for multidimensional integration in stochastic mechanics. *International Journal for Numerical Methods in Engineering* Nov 2004; **61**(12):1992–2019, doi:10.1002/nme.1135. URL <http://doi.wiley.com/10.1002/nme.1135>.
- [12] Li G, Rabitz H, Yelvington PE, Oluwole OO, Bacon F, Kolb CE, Schoendorf J. Global sensitivity analysis for systems with independent and/or correlated inputs. *The journal of physical chemistry. A* May 2010; **114**(19):6022–32, doi:10.1021/jp9096919. URL <http://www.ncbi.nlm.nih.gov/pubmed/20420436>.
- [13] Abgrall R, Congedo PM, Geraci G, Iaccarino G. Decomposition and Computation of high-order statistics. *INRIA Research Report, RR-8193* 2012; .
- [14] Le Maître O, Knio OM. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.
- [15] Ishigami T, Homma T. An importance quantification technique in uncertainty analysis for computer models. *ISUMA'90, First International Symposium on Uncertainty Modeling and Analysis* 1990; :398–403 URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=151285](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=151285).
- [16] Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety* 2008; **93**:964–979, doi:10.1016/j.res.2007.04.002.
- [17] Bellas-Chatzigeorgis G, Villedieu N, Panesi M, Congedo P, Magin T. Propagation of uncertainties related to a complex detailed chemical mechanism. *UQ4AERO: Uncertainty Quantification for Aerospace Applications* 2013; :1–90.
- [18] Borgonovo E. A new uncertainty importance measure. *Reliability Engineering & System Safety* Jun 2007; **92**(6):771–784, doi:10.1016/j.res.2006.04.015. URL <http://linkinghub.elsevier.com/retrieve/pii/S0951832006000883>.



**RESEARCH CENTRE  
BORDEAUX – SUD-OUEST**

200 avenue de la Vieille Tour  
33405 Talence Cedex

Publisher  
Inria  
Domaine de Voluceau - Rocquencourt  
BP 105 - 78153 Le Chesnay Cedex  
[inria.fr](http://inria.fr)

ISSN 0249-6399