Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 301 (2016) 80–115
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

High-order statistics in global sensitivity analysis: Decomposition and model reduction

G. Geraci[a], P.M. Congedo[b,∗], R. Abgrall[c], G. Iaccarino[a]

[a] Flow Physics and Computational Engineering, Stanford University, 488 Escondido Mall, CA 94305-3035, USA
[b] INRIA Bordeaux–Sud-Ouest, CARDAMOM Team, 200, Avenue de la Vieille Tour, 33400 Talence, France
[c] Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 30 April 2015; received in revised form 15 November 2015; accepted 22 December 2015
Available online 30 December 2015

Highlights

- We present a decomposition of high-order statistics for model reduction.
- Sensitivity indices based on skewness and kurtosis decomposition are introduced.
- The importance of ranking the contributions w.r.t. high-order moments is assessed.

Abstract

ANalysis Of VAriance (ANOVA) is a common technique for computing a ranking of the input parameters in terms of their contribution to the output variance. Nevertheless, the variance is not a universal criterion for ranking variables, since non symmetric outputs could require higher order statistics for their description and analysis. In this work, we illustrate how third and fourth-order moments, *i.e.* skewness and kurtosis, respectively, can be decomposed mimicking the ANOVA approach. It is also shown how this decomposition is correlated to a Polynomial Chaos (PC) expansion leading to a simple strategy to compute each term. New sensitivity indices, based on the contribution to the skewness and kurtosis, are proposed. The outcome of the proposed analysis is depicted by considering several test functions. Moreover, the ranking of the sensitivity indices is shown to vary according to their statistics order. Furthermore, the problem of formulating a truncated polynomial representation of the original function is treated. Both the reduction of the number of dimensions and the reduction of the order of interaction between parameters are considered. In both cases, the impact on the reduction is assessed in terms of statistics, namely the probability density function. Feasibility of the proposed analysis in a real-case is then demonstrated by presenting the sensitivity analysis of the performances of a turbine cascade in an Organic Rankine Cycles (ORCs), in the presence of complex thermodynamic models and multiple sources of uncertainty.

Keywords: Sensitivity analysis; High-order statistics; Polynomial chaos; Model reduction; ORCs turbines

∗ Corresponding author.
E-mail addresses: ggeraci@stanford.edu (G. Geraci), pietro.congedo@inria.fr (P.M. Congedo).

http://dx.doi.org/10.1016/j.cma.2015.12.022

1. Introduction

Optimization and design in the presence of uncertain operating conditions, material properties and manufacturing tolerances poses a tremendous challenge to the scientific computing community. Uncertainty quantification (UQ) approaches represent the inputs as random variables and seek to construct a statistical characterization of the quantities of interest. Several methodologies are proposed to tackle this issue, among those stochastic spectral methods [1–5], that can provide considerable speed-up in computational time when compared to Monte Carlo (MC) simulation. The presence of a large number of uncertain inputs leads to an exponential increase of the cost thus making these methodologies unfeasible [6]. This situation becomes even more challenging when robust design optimization is tackled [7,8].

Several UQ methods have been developed with the objective of reducing the number of solutions required to obtain a statistical characterization of the quantity of interest, such as Sparse Grid techniques [9] or adaptive mesh generation [10–14]. These techniques can lead to a dramatical reduction of the quadrature points for moderate dimensional problem, provided that the function has some regularity properties. Classical sparse grids [9] are constructed from tensor products of one-dimensional quadrature formulas. Some Galerkin-based methods deal with multi-resolution wavelet expansions [15,16], domain decomposition in the random space [17], adaptive h-refinement [3] for dealing with arbitrary probability distributions.

Among the collocation-based stochastic spectral methods, in [18] the authors proposed the use of sparse grid quadrature for stochastic collocation. Older studies show the errors and efficiency of sparse grid integration and interpolation [19,20], Smolyak constructions based on one-dimensional nested Clenshaw–Curtis rules [19,21] and the integration error of sparse grids based on one-dimensional Kronrod–Patterson rules [22].

An alternative solution is based on approaches attempting to identify the relative importance of the input uncertainties on the output. If some dimensions could be identified as negligible, they could be discarded in a reduced stochastic problem and better statistics estimations could be achieved with a lower computational cost.

Identifying the most influent parameters requires to determine the uncertain inputs which have the largest impact on the variability of the model output. In literature, Global sensitivity analysis (GSA) aims at quantifying how uncertainty in the input parameters of a model contributes to the uncertainty in its output (see for example [23]) where global sensitivity analysis techniques are applied to probabilistic safety assessment models. GSA classifies the inputs according to their importance on the output variations and it gives a hierarchy of the most important ones.

Traditionally, GSA is performed using methods based on the decomposition of the output variance [24], *i.e.* ANOVA. The ANOVA approach involves splitting a multi-dimensional function into its contributions from different groups of subdimensions. In 2001, Sobol used this formulation to define global sensitivity indices [24], displaying the relative variance contributions of different ANOVA terms. In [25], the authors introduced two High-Dimensional Model Reduction (HDMR) techniques to capture input–output relationships of physical systems with many input variables. These techniques are also based on ANOVA decompositions.

Several techniques have been developed to compute sensitivity indices at low computational cost [26]. In [27–29], generalized Polynomial Chaos Expansions (gPC) are used to build surrogate models for computing the Sobol's indices analytically as a post-processing of the PC coefficients. In [6], the authors combine multi-element polynomial chaos with an ANOVA functional decomposition to enhance the convergence rate of polynomial chaos in high dimensions and in problems with low stochastic regularity. In [30], the use of adaptive ANOVA decomposition is investigated as an effective dimension-reduction technique in modeling incompressible and compressible flows with high-dimension of random space. In [31], sparse Polynomial Chaos (PC) expansions are introduced in order to compute sensitivity indices: a PC-based metamodel (a computationally inexpensive polynomial approximation of the relation between inputs and output) which contains the significant terms whereas the PC coefficients are computed by least-square regression.

Other approaches are developed if the assumption of independence of the input parameters is not valid. For instance, new indices have been proposed in [32,33] when a linear correlation exists. In [34], the authors introduce a global sensitivity indicator which looks at the influence of input uncertainty on the entire probability distribution without reference to a specific moment of the output (moment independence) and which can be defined also in the presence of correlations among the parameters. In [35], a gPC methodology to address global sensitivity analysis for this kind of problems is introduced, while in [36], a numerical procedure is proposed for moment-independent sensitivity methods.

The ANOVA-based analysis create a hierarchy of most important input parameters for a given output when variance is chosen as metrics. A strong limitation of this approach is the fact that the variance cannot be considered a general