Applied Mathematical Modelling 45 (2017) 748–767

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Analytical variance based global sensitivity analysis for models with correlated variables

Kaichao Zhang, Zhenzhou Lu*, Danqing Wu, Yongli Zhang

P.O. Box 120, School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 13 March 2015
Revised 19 April 2016
Accepted 14 December 2016
Available online 12 January 2017

Keywords:
Analytical method
Variance based GSA
Correlated variable
Subset decomposition
Orthogonal decorrelation
Kriging model

ABSTRACT

In order to quantitatively analyze the variance contributions by correlated input variables to the model output, variance based global sensitivity analysis (GSA) is analytically derived for models with correlated variables. The derivation is based on the input-output relationship of tensor product basis functions and the orthogonal decorrelation of the correlated variables. Since the tensor product basis function based simulator is widely used to approximate the input-output relationship of complicated structure, the analytical solution of the variance based global sensitivity is especially applicable to engineering practice problems. The polynomial regression model is employed as an example to derive the analytical GSA in detail. The accuracy and efficiency of the analytical solution of GSA are validated by three numerical examples, and engineering application of the derived solution is demonstrated by carrying out the GSA of the riveting and two dimension fracture problem.

1. Introduction

Sensitivity analysis aims at investigating the impact of input variations on the variation of a model output, which can be classified into two categories: local sensitivity analysis and global sensitivity analysis [1,2]. *Global sensitivity analysis* (GSA) is also called importance measure analysis [3]. The existing importance measures can be summarized to three categories: non-parameter techniques (correlation coefficient model) [4,5], variance based methods [1,6,7], and moment independent model [2,8]. Variance based methods can directly illustrate the variance contributions of the model output by inputs, and they have been widely used in engineering design. Variance based GSA was first employed by Cukier et al. in chemistry [9]. Then, Hora and Iman introduced the uncertainty importance, and Sacks et al. gave a visual inspection of sensitivity results by decomposition of the output [10]. Sobol was inspired by the formers' work and used *analysis of variance* (ANOVA) to define the variance based sensitivity indices [11,12].

There are abundant simulation-based methods for variance based GSA, such as Monte Carlo, SDP, FAST etc [12–16]. These simulation-based methods are easy to comprehend and program. Unfortunately, simulation-based method always needs a large number of samples, which results in huge computation burden in practice. For uncorrelated variables, an analytical variance based GSA method was proposed in Ref. [17] by the theory that multivariate integrals of *tensor product basis functions* can be translated to calculations of univariate integrals. For second order polynomial models with correlated variables, Refs. [18–20] derived the analytical variance based GSA.

* Corresponding author.
E-mail addresses: zhangkaichao123321@126.com (K. Zhang), zhenzhoulu@nwpu.edu.cn (Z. Lu), wudanqing86@163.com (D. Wu).

http://dx.doi.org/10.1016/j.apm.2016.12.036
0307-904X/

In this paper, we extend the analytical variance based GSA method for uncorrelated variables in Ref. [17] to models with correlated variables. Based on the input-output relationship constructed by tensor product basis functions, we derive a universal analytical solution of variance based GSA for models with correlated variables. Using orthogonal decorrelation of the correlated variables, we achieve a simplified and easy way to realize the analytical derivation for variance based GSA. The analytical method proposed in this paper is especially applicable in engineering practice because the tensor product basis functions are usually used to create an input-output relationship in engineering. Metamodels commonly used in engineering, such as polynomial regression model, Kriging model, Gaussian radial basis model, MARS model etc., can be translated into the form expressed by the tensor product basis functions [21]. If the metamodel has been constructed, we can conveniently and efficiently obtain the results using the analytical solution of variance based GSA, which needs little computation cost.

The definition of variance based global sensitivity indices, subset decomposition and the concept of subset sensitivity indices are shortly introduced in Section 2. In Section 3, based on the tensor product basis functions, the universal analytical solution of the variance based GSA is derived for both uncorrelated and correlated variables. In Section 4, after the orthogonal decorrelation of the correlated variables, the analytical variance based GSA for models with correlated variables is presented. In Section 5, some numerical examples are used to validate the method proposed in this paper and engineering practice problems are analyzed by the proposed method. The last section evaluates the methods proposed in this paper and draws some conclusions to GSA.

2. Variance based GSA

2.1. The definition of global sensitivity indices

Suppose $y=f(\mathbf{x})$ is a square integrable function, in which $\mathbf{x}$ is a M-dimension input vector, i.e., $\mathbf{x}=(x_1, x_2,\ldots, x_M)$. The *probability density function* (PDF) of x_i is expressed by $p_i(x_i)$ and $p(\mathbf{x})$ is the joint PDF of $\mathbf{x}$. For models with uncorrelated variables, $p(\mathbf{x}) = \prod_{i=1}^{M} p_i(x_i)$. Using *high dimension model representation* (HDMR), $f(\mathbf{x})$ can be decomposited as [11]

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{M} f_i(x_i) + \sum_{i_1=1}^{M}\sum_{i_2=i_1+1}^{M} f_{i_1 i_2}(x_{i_1}, x_{i_2}) + \cdots + f_{12\ldots M}(x_1, x_2, \ldots, x_M), \tag{1}$$

where f_0 is the mean of $f(\mathbf{x})$, i.e.

$$f_0 = \int f(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} = E[f(\mathbf{x})] \tag{2}$$

in which $E[\cdot]$ is the expectation operator. $f_i(x_i)$ is called *main effect* which is only related to x_i and can be obtained by

$$f_i(x_i) = \int f(\mathbf{x})p(\mathbf{x}_{-i}|x_i)\mathrm{d}\mathbf{x}_{-i} - f_0 = E[f(\mathbf{x})|x_i] - f_0, \tag{3}$$

where $\mathbf{x}_{-i}=(x_1,\ldots, x_{i-1}, x_{i+1},\ldots, x_M)$, and $p(\mathbf{x}_{-i}|x_i)=p(\mathbf{x})/p(x_i)$ is the conditional PDF of $\mathbf{x}_{-i}$ on x_i. For uncorrelated variables, $p(\mathbf{x}_{-i}|x_i) = \prod_{j\neq i} p_j(x_j)$.

$f_{i_1 i_2}(x_{i_1}, x_{i_2})$ is called *second order interaction effect* which is related to two variables x_{i_1} and x_{i_2} and can be obtained by

$$\begin{aligned} f_{i_1 i_2}(x_{i_1}, x_{i_2}) &= \int f(\mathbf{x})p(\mathbf{x}_{-(i_1,i_2)}|x_{i_1}, x_{i_2})\mathrm{d}\mathbf{x}_{-(i_1,i_2)} - f_{i_1}(x_{i_1}) - f_{i_2}(x_{i_2}) - f_0 \\ &= E[f(\mathbf{x})|x_{i_1}, x_{i_2}] - f_{i_1}(x_{i_1}) - f_{i_2}(x_{i_2}) - f_0, \end{aligned} \tag{4}$$

where $p(\mathbf{x}_{-(i_1,i_2)}|x_{i_1}, x_{i_2}) = p(\mathbf{x})/p(x_{i_1}, x_{i_2})$ is the conditional PDF of $\mathbf{x}_{-(i_1,i_2)}$ on x_{i_1} and x_{i_2}. For uncorrelated variables $p(\mathbf{x}_{-(i_1,i_2)}|x_{i_1}, x_{i_2}) = \prod_{j\neq i_1,i_2} p_j(x_j)$.

In general, $f_{i_1\cdots i_s}(x_{i_1}, \ldots, x_{i_s})$ is called sth *order interaction effect* which is related to s variables $x_{i_1}, \ldots, x_{i_s}$ and can be obtained by

$$\begin{aligned} f_{i_1\cdots i_s}(x_{i_1}, \ldots, x_{i_s}) &= \int f(\mathbf{x})p(\mathbf{x}_{-(i_1,\ldots,i_s)}|x_{i_1}, \ldots, x_{i_s})\mathrm{d}\mathbf{x}_{-(i_1,\ldots,i_s)} - \sum_{k=1}^{s-1}\sum_{j_1,\ldots,j_k\in(i_1,\ldots,i_s)} f_{j_1\ldots j_k}(x_{j_1}, \ldots, x_{j_k}) - f_0 \\ &= E[f(\mathbf{x})|x_{i_1}, \ldots, x_{i_s}] - \sum_{k=1}^{s-1}\sum_{j_1,\ldots,j_k\in(i_1,\ldots,i_s)} f_{j_1\ldots j_k}(x_{j_1}, \ldots, x_{j_k}) - f_0, \end{aligned} \tag{5}$$

where $j_1 < j_2 < \ldots < j_k$ and $p(\mathbf{x}_{-(i_1,\ldots,i_s)}|x_{i_1}, \ldots, x_{i_s}) = p(\mathbf{x})/p(x_{i_1}, \ldots, x_{i_s})$ is the conditional PDF of $\mathbf{x}_{-(i_1,\ldots,i_s)}$ on $x_{i_1}, \ldots, x_{i_s}$. For uncorrelated variables $p(\mathbf{x}_{-(i_1,\ldots,i_s)}|x_{i_1}, \ldots, x_{i_s}) = \prod_{j\neq i_1,\ldots,i_s} p_j(x_j)$.

When all the variables are uncorrelated, the variance V of $f(\mathbf{x})$ can be expressed as the summation of variances $V_{i_1\ldots i_s}$, i.e.,

$$V = \sum_{i=1}^{M} V_i + \sum_{i_1=1}^{M}\sum_{i_2=i_1+1}^{M} V_{i_1 i_2} + \cdots + V_{1\ldots M}. \tag{6}$$