Computers and Mathematics with Applications 62 (2011) 4547–4556

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

Importance analysis for models with correlated input variables by the state dependent parameters method

Luyi Li *, Zhenzhou Lu, Changcong Zhou

School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, Shaanxi Province, PR China

ARTICLE INFO

Article history:
Received 29 June 2011
Received in revised form 11 October 2011
Accepted 11 October 2011

Keywords:
Correlated input variable
Importance measure
Uncorrelated contribution
Correlated contribution
State dependent parameters method
Importance matrix

ABSTRACT

For clearly exploring the origin of the variance of the output response in case the correlated input variables are involved, a novel method on the state dependent parameters (SDP) approach is proposed to decompose the contribution by correlated input variables to the variance of output response into two parts: the uncorrelated contribution due to the unique variations of a variable and the correlated one due to the variations of a variable correlated with other variables. The correlated contribution is composed by the components of the individual input variable correlated with each of the other input variables. An effective and simple SDP method in concept is further proposed to decompose the correlated contribution into the components, on which a second order importance matrix can be solved for explicitly exposing the contribution components of the correlated input variable to the variance of the output response. Compared with the existing regression-based method for decomposing the contribution by correlated input variables to the variance of the output response, the proposed method is not only applicable for linear response functions, but is also suitable for nonlinear response functions. It has advantages both in efficiency and accuracy, which are demonstrated by several numerical and engineering examples.

1. Introduction

Since the 1960s, many researches have focused on the sensitivity analysis of the partial derivatives of structural responses, characters or indices with respect to input variables. However, those sensitivities are solved at nominal values, they cannot take account of the variation effect of the input variables, and thus those sensitivities are local. Compared with the local sensitivity, the uncertainty importance measure (UI) is defined as that uncertainty in the output can be apportioned to different sources of uncertainty in the model input [1], and the importance measure is also called global sensitivity. It is significant in engineering design and probability safety assessment, since it can comprehensively consider the average effect of the input variables on the output response in the value domain of the input variables. Thus, more and more studies nowadays are using importance analysis methods instead of local sensitivity analysis. Many importance analysis techniques are available, such as nonparametric techniques [2–4]; variance-based importance measure indices and their solutions [5–8]; and moment-independent importance measures [9–11], among which variance-based importance measures have a quite general applicability since they can reflect the effect of the input variables on the output response simply and effectively.

However, most of the existing importance analysis techniques assume input variables independence, and a few studies have focused on the importance analysis of the correlated input variables, which is usually the common case in engineering.

* Correspondence to: P.O. Box 120, School of Aeronautics, Northwestern Polytechnical University, Xian City, 710072, Shaanxi Province, PR China. Tel.: +86 29 88460480.

E-mail address: luyili@mail.nwpu.edu.cn (L. Li).

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.10.034

The correlation of the input variables may affect the importance rank of the variables dramatically, therefore, only the importance analysis methods which consider the correlation of the variables can reflect the effect of the input variables on the output response reasonably and correctly. By now, some studies have been conducted to obtain the importance measure of correlated input [12–15]. However, all of these methods only provide an overall importance measure of one input variable, which does not distinguish the correlated or uncorrelated contribution of one input variable. For response models with correlated input variables, to explore the origin of the uncertainty of the output response clearly, the contribution of uncertainty to output response by an individual input variable should be divided into two parts: the uncorrelated contribution (by the uncorrelated variations, i.e. the unique variations of a variable which cannot be explained by any other parameters) and correlated contribution (by the correlated variations, i.e. variations of a variable which are correlated with other input variables) [16]. As pointed out in [16], the distinction between uncorrelated and correlated contribution of uncertainty for an individual variable is very important, since it can help engineers decide if they need to focus on the correlated variations among specific variables (if the correlated contribution dominates) or the variable itself (if the uncorrelated contribution dominates). Based on this idea, a regression-based method is proposed in [16] to decompose the total variance of the output response into partial variances contributed by the correlated and uncorrelated variations of variables. However, this method is only suitable for the case where the relationship between output response and input variables is approximately linear, and it relies on the assumption that the estimation space is perpendicular to the error space and needs a double linear regression when estimating the uncorrelated contribution of the input variables. Therefore, it has limitations both in accuracy and efficiency. A more robust and similar treatment for correlated input variables was proposed in [17] where the total contribution of an input variable or a subset of input variables to the variance of the output response was decomposed into a structural contribution (reflecting the system structure) and a correlative (reflecting the correlated input probability distribution) one. This treatment can deal with both linear and nonlinear response functions. However, both the methods in [16,17] provide an overall correlated contribution of one input variable and do not decompose the correlated contribution into components, which is inconvenient for engineering decisions. To overcome the limitations in [16], a novel method based on the state dependent parameters (SDP) method is proposed in this paper to decompose the variance contribution of the correlated input variables, which is suitable for both linear and nonlinear response models. Additionally, to satisfy the engineering requirements, an SDP method is further proposed to estimate the correlated contribution of two specific variables, on which an importance matrix can be solved to provide more referential information for engineering decisions.

2. The SDP method for variance-based importance analysis with independent input variables

2.1. The variance-based importance measures of input variables

Consider the response model $y = f(\boldsymbol{x})$, in which $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ are n-dimensional independent input variables. In order to quantify the relative contribution of each input variable to the uncertainty of y, variance-based sensitivity indices of single variables or of groups of them are defined as [5]

$$S_{\boldsymbol{I}} = \frac{V_I}{V(y)} = \frac{V(E(y|\boldsymbol{x_I}))}{V(y)} \tag{1}$$

where $\boldsymbol{x_I}$ represents a random input variable x_i or a set of random input variables $(x_{i_1}, \ldots, x_{i_g})$, where $1 \le i_1 \le \cdots \le i_g \le n$, and they tell the portion of variance of y that is explained by $\boldsymbol{x_I}$.

The two most popular variance-based importance measures at present are the main effect

$$S_i = \frac{V_i}{V(y)} = \frac{V(E(y|x_i))}{V(y)} \tag{2}$$

and the total effect

$$S_{Ti} = \frac{V_{Ti}}{V(y)} = \frac{E(V(y|\boldsymbol{x}_{-i}))}{V(y)} \tag{3}$$

where $\boldsymbol{x}_{-i}$ indicates all input variables except x_i.

The main effect measures the unique contribution of the input variable x_i to the uncertainty (variance) of the output y, while the total effect measures the overall contribution of x_i on y, including all interaction terms of x_i with all other input variables.

2.2. The SDP method for variance-based importance measures

SDP modeling is a widely used method to represent nonlinear stochastic systems and time series. It is one class of non-parametric smoothing approach first suggested by Young [18,19], and has been applied successfully to the variance-based importance analysis by Ratto et al. in [20–22]. This application has improved the computational efficiency of the variance-based importance measures tremendously and made computationally intensive models accessible to variance-based importance analysis.

The SDP method for variance-based importance analysis actually falls within the context of metamodelling and emulation. It is based on a truncated high dimensional model representation (HDMR) [23], and calculates the variance-based importance measures of the input variables by approximating the input–output mapping of the response model. The process regarding the approximation of the first order HDMR of the response model $y = f(\boldsymbol{x})$ and the estimation of the main effect in Eq. (2) is summarized as follows [21,22]:

For N random samples $\boldsymbol{x}_t$ $(t = 1, 2, \dots, N)$ generated from the joint probability density function (PDF) of the input variables $\boldsymbol{x}$, the corresponding values of response y are $\boldsymbol{y}_t$ $(t = 1, 2, \dots, N)$. Considering the first order HDMR of the computation model,

$$y_t - f_0 = f_1(x_{1,t}) + f_2(x_{2,t}) + \cdots + f_n(x_{n,t}) + o(\boldsymbol{x}\boldsymbol{x}') \tag{4}$$

where $f_0 = E(y)$, $f_i\ (x_{i,t}) = E(y|x_{i,t}) - f_0$, $o(\boldsymbol{x}\boldsymbol{x}')$ is the high order truncated error. $t = (1, 2, \dots, N)$ is the index of the samplings.

Then the state-dependent model approximating the first order $f_i(x_{i,t})$ $(i = 1, 2, \dots, n)$ in Eq. (4) can be written as [21]

$$\begin{aligned} y_t - f_0 &= \boldsymbol{x}_t^T \boldsymbol{p}_t + e_t \\ &= p_{1,t}x_{1,t} + p_{2,t}x_{2,t} + \cdots + p_{n,t}x_{n,t} + e_t; \quad e_t \sim N(0, \sigma^2) \end{aligned} \tag{5}$$

where e_t is the observation noise, $p_{i,t}$ $(i = 1, 2, \dots, n)$ are state-dependent parameters depending on the corresponding state variables x_i $(i = 1, 2, \dots, n)$.

In order to estimate the $p_{i,t}$ $(i = 1, 2, \dots, n)$, it is necessary to characterize the variability of $p_{i,t}$ $(i = 1, 2, \dots, n)$ in some stochastic manner. In the SDP approach, this is achieved by modeling each state-dependent parameter $p_{i,t}$ by one member of the generalized random walk (GRW) class of non-stationary processes. Among the GRW processes, the integrated random walk (IRW) process turns out to provide good results, since it can ensure that the estimated SDP relationship has smooth properties of a cubic spline. Given the IRW characterization of $p_{i,t}$ $(i = 1, 2, \dots, n)$, model (5) can be put into state space form as [21]

$$\text{Observation equation: } y_t = \boldsymbol{x}_t^T \boldsymbol{p}_t + e_t \tag{6}$$

$$\text{State equations: } \quad \begin{aligned} p_{i,t} &= p_{i,t-1} + d_{i,t-1} \\ d_{i,t} &= d_{i,t-1} + \eta_{i,t} \end{aligned} \tag{7}$$

where $e_{i,t}$ (observation noise) and $\eta_{i,t}$ $(i = 1, 2, \dots, n)$ (system disturbances) are zero mean white noise inputs with variance σ_i^2 and $\sigma_{\eta_i}^2$, respectively. y_t represents $y_t - f_0$.

Given the SDP relationship in Eqs. (6) and (7), each state dependent parameter $p_{i,t}$ $(i = 1, 2, \dots, n)$ can be estimated in turn by exploiting backfitting procedure. At each iteration of the backfitting algorithm, different sorting strategies are used based on the state variable x_i associated with the current state dependent parameter $p_{i,t}$ being estimated, and then $p_{i,t}$ is estimated using the recursive Kalman Filter (KF) and associated recursive fixed interval smoothing (FIS) algorithm. Under the Gaussian assumption for the distributions of $e_{i,t}$ and $\eta_{i,t}$, the hyper-parameters associated with recursive process, in the form of the noise variance σ_i^2 and $\sigma_{\eta_i}^2$ in the case of IRW, are optimized by maximum likelihood (ML), using prediction error decomposition. Readers can refer to [19] for a more comprehensive discussion of SDP modeling and its algorithms and to [21] for details of the SDP approach to the approximation of the first order HDMR.

When the state dependent parameters $p_{i,t}$ $(i = 1, 2, \dots, n)$ in model (5) are estimated, the first order of the HDMR in Eq. (4) can be obtained by $f_i\ (x_{i,t}) = p_{i,t}x_{i,t}$, and then the calculation of the main effects $S_i = V_i/V(y) = V(E(y|x_i))/V(y)$ is straightforward.

As has been proved in the literatures [21,22], the SDP approach for the approximation of the response model above is conceptually simple and very efficient. It normally estimates the first order terms of the HDMR and all the main effects of the input variables with only a single set and a few hundred model runs. And the computational cost associated with this method is almost independent of the dimensionality of the input variable, which can lead to a significant reduction in the computational effort of the approximation of the response model and the variance-based importance analysis. Additionally, the approach is flexible because, in principle, it can be applied with any available type of Monte Carlo sample. Especially when coupled with low-discrepancy samplings, such as Sobol sequence, this method is extremely efficient, allowing for a dramatic improvement in computational efficiency. The SDP method described above can not only be extended to the estimation of the interaction terms and even third order terms of the HDMR, but also can be extended to the case where the input variables are correlated [21]. In this paper, it is extended to the decomposition of the contribution of uncertainty by the correlated input variables to the model output.

3. The SDP method for variance decomposition with correlated input variables

For response models with independent input variables, the uncertainty contribution by an individual variable to the model output only results from the variations of the variables itself. In the importance analysis in Section 2.1, the samplings for different variables are mutually independent. Thus, the variance contribution V_i of an individual variable obtained by the SDP method above only contains the contribution associated with the variation of the variable itself. However, when correlation is present among the input variables, the variance contribution V_i of an individual variable consists of not only the uncorrelated contribution resulting from the variation of the variable itself, but also contains the correlated contribution, which is from the correlated variations of other variables [16]. i.e.

$$V_i = V_i^U + V_i^C \tag{8}$$

where V_i, V_i^U and V_i^C are the total contribution, uncorrelated contribution and correlated contribution to the variance of the response by the ith variable x_i of the correlated input variables, respectively. Obviously, if a specific variable x_k $(1 \le k \le n)$ is independent of the other $n-1$ variables $\boldsymbol{x}_{(-k)} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n)$, then $V_k = V_k^U$.

In the case of the response model with correlated input variables, the uncorrelated and correlated contribution by an individual variable to the variance of the model output can be derived by exploiting the SDP method to estimate the dependent relationship between y and a different part of the input variables. We will talk about this process in detail in the subsequent sections.

3.1. The total contribution by an individual variable x_i to the variance of the model output

Consider the response model $y = f(\boldsymbol{x})$ with n-dimensional correlated input variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. N samplings $\boldsymbol{x}_t$ $(t = 1, 2, \ldots, N)$ are generated from the joint PDF of the correlated input variables, and then the corresponding response values $\boldsymbol{y}_t$ $(t = 1, 2, \ldots, N)$ are obtained. To make full use of the high efficiency of the SDP method, Sobol's sequence and Copula transformation are used in this paper to generate correlated samplings.

The total contribution by x_i to the variance of the model output can be derived by exploiting the SDP method to estimate the dependent relationship between y and x_i in Eq. (9).

$$y_t = p_{i,t} x_{i,t} + e_{i,t} \tag{9}$$

where $t = 1, 2, \ldots, N$ denotes the index of the samplings, $p_{i,t}$ is a state dependent parameter relying on the corresponding state variables $x_{i,t}$, e_t is the observation error, and y_t represents $y_t - f_0$. Taking the dependent relationship in Eq. (9) as an observation equation, the corresponding state space model as shown in Eqs. (6) and (7) can be built, and then $p_{i,t}$ can be estimated by regressive KF and the corresponding regressive FIS.

The total variance contribution V_i of x_i can be estimated as follows:

$$\hat{V}_i = \frac{1}{N-1}\sum_{t=1}^{N}(\hat{y}_t^{(i)} - \bar{y}^{(i)})^2 \tag{10}$$

where $\hat{y}_t^{(i)} = \hat{p}_{i,t|N} x_{i,t}$, and $\hat{p}_{i,t|N}$ are the FIS estimates of $p_{i,t}$ for dependent relationship (9). $\bar{y}^{(i)}$ is the mean of all $\hat{y}_t^{(i)}$ $(t = 1, 2, \ldots, N)$.

3.2. The uncorrelated and correlated contributions by x_i to the variance of the model output

To estimate the variance contributed by the variation of x_i uncorrelated with all other variables, V_i^U, we need to estimate the dependent relationship between y and all input variables except x_i, denoted by $\boldsymbol{x}_{(-i)}$, as shown in Eq. (11) first by the SDP method.

$$y_t = \sum_{j \ne i}^{n} p_{j,t} x_{j,t} + e_t \tag{11}$$

where $p_{j,t}$ $(j = 1, 2, \ldots, n, j \ne i)$ are state dependent parameters relying on the corresponding state variables $x_{j,t}$ $(j = 1, 2, \ldots, n, j \ne i)$, e_t is the observation error, and y_t represents $y_t - f_0$. Analogously, $p_{j,t}$ $(j = 1, 2, \ldots, n, j \ne i)$ can be estimated by regressive KF and the corresponding regressive FIS.

Then the variance contributed by the input variables $\boldsymbol{x}_{(-i)}$ to the variance of the model output can be obtained as follows:

$$\hat{V}_{(-i)} = \frac{1}{N-1}\sum_{t=1}^{N}(\hat{y}_t^{(-i)} - \bar{y}^{(-i)})^2 \tag{12}$$

where $\hat{y}_t^{(-i)} = \sum_{j \ne i}^{n} \hat{p}_{j,t|N} x_{j,t}$, and $\hat{p}_{j,t|N}$ $(j = 1, 2, \ldots, n, j \ne i)$ are the final FIS estimates of $p_{j,t}$ $(j = 1, 2, \ldots, n, j \ne i)$ for dependent relationship (11). Here, $\bar{y}^{(-i)}$ is the average value of all $\hat{y}_t^{(-i)}$ $(t = 1, 2, \ldots, N)$.

$\hat{V}_{(-i)}$ estimated in Eq. (12) includes the uncorrelated and correlated contribution to the variance of the model output by all the input variables except x_i, i.e. $\boldsymbol{x}_{(-i)}$, thus, it also contains the correlated contributions of each variable of $\boldsymbol{x}_{(-i)}$ correlated with x_i. Therefore, the uncorrelated contribution V_i^U to the variance of the model output by x_i can be estimated by subtracting $V_{(-i)}$ from the joint contribution of all the input variables, i.e. the total variance V of the model output, which will be estimated in the next subsection.

$$\hat{V}_i^U = \hat{V} - \hat{V}_{(-i)}. \tag{13}$$

Submitting the total variance contribution in Eq. (10) and the uncorrelated one in Eq. (13) of x_i into Eq. (8), the correlated contribution by the variations of x_i correlated with the other input variables $\boldsymbol{x}_{(-i)}$ can be estimated by the following equation:

$$\hat{V}_i^C = \hat{V}_i - \hat{V}_i^U. \tag{14}$$

By now, we can get the partial variances of the model output contributed by the uncorrelated and correlated variations of each variable by Eqs. (10), (13) and (14).

3.3. The estimation of the total variance of the model output and the importance measures of the partial contributions

A last remark concerns the computation of the importance measures of the partial contributions. This computation involves the normalization of the partial variances obtained above by the total variance of the model output, which can be estimated by the SDP method as follows.

With the input samplings $\boldsymbol{x}_t$ $(t = 1, 2, \ldots, N)$ and the corresponding output $\boldsymbol{y}_t$ $(t = 1, 2, \ldots, N)$, the conditional expectation $E(y|x_i)$ of the response can be easily obtained by the SDP method in Section 2.2. Similarly, we can get the conditional expectation $E(y^2|x_i)$ of the response squared only by replacing the output $\boldsymbol{y}_t$ with $\boldsymbol{y}_t^2$. According to the relationship between variance and expectation, the total variance V of the model output can be obtained by the following equation.

$$\hat{V} = V(y) = E(y^2) - E^2(y). \tag{15}$$

Normalizing the partial variances in Eqs. (10), (13) and (14) by the total variance in Eq. (15), respectively, we can get the total (S_i), uncorrelated (S_i^U), and correlated (S_i^C) contribution ratios of the variable x_i (namely, the first-order importance measures):

$$\begin{aligned} S_i &= \frac{\hat{V}_i}{\hat{V}} \\ S_i^U &= \frac{\hat{V}_i^U}{\hat{V}} \\ S_i^C &= \frac{\hat{V}_i^C}{\hat{V}}. \end{aligned} \tag{16}$$

The method for the decomposition of the contribution to the variance of the model output by the correlated input variables above is based on the SDP approach, which is used to estimate the nonlinear and non-stationary systems. Therefore, it is suitable for both the linear and nonlinear models. Additionally, the proposed method can estimate the uncorrelated contribution of an individual variable with only once calculation of the model parameters and makes no assumption regarding the estimation space and the error space. Therefore, it has advantages both in efficiency and accuracy, and has wide applicability.

4. The SDP method for the correlated contribution of two specific variables and importance matrix

4.1. The correlated contribution of two specific variables

It is common in engineering that the input variables are correlated. In this case, in addition to the necessity of decomposing the contribution of the uncertainty to the model output by an individual variable into uncorrelated and correlated contributions, it is also essential to further decompose this correlated contribution into components, i.e. the correlated contributions of the individual variable correlated with each of the other variables. However, there has been no research discussing this problem until now.

For the response model $y = f(\boldsymbol{x})$, the joint contribution $V_{l,q}$ to the variance of the model output by two specific correlated variables x_l and x_q $(1 \le l, q \le n, l \ne q)$ includes the uncorrelated contributions V_l^U and V_q^U of x_l and x_q respectively and the correlated contribution $V_{l,q}^C$ by the variation of x_l correlated with x_q. Their relationship can be expressed by the following equation:

$$V_{l,q} = V_l^U + V_q^U + V_{l,q}^C \tag{17}$$

where V_l^U and V_q^U can be estimated by the SDP method in 3.2.

To estimate the partial variance contributed by the joint variation of x_l and x_q, $V_{l,q}$, the dependent relationship between y and $\boldsymbol{x}_{(-l,-q)}$, i.e. all the input variables except x_l and x_q, as shown in Eq. (18) are estimated first by the SDP method:

$$y_t = \sum_{j \ne l, j \ne q}^{n} p_{j,t} x_{j,t} + e_t \tag{18}$$

where $p_{j,t}$ $(j = 1, 2, \ldots, n, j \neq l, j \neq q)$ are state dependent parameters depending on the corresponding state variables $x_{j,t}$ $(j = 1, 2, \ldots, n, j \neq l, j \neq q)$, and e_t is the observation error, and y_t represents $y_t - f_0$. Obviously, $p_{j,t}$ $(j = 1, 2, \ldots, n, j \neq l, j \neq q)$ can be estimated easily by the SDP method.

The partial variance contributed by all the input variables except x_l and x_q, i.e. $\boldsymbol{x}_{(-l,-q)}$ can be estimated by the following equation:

$$\hat{V}_{(-l,-q)} = \frac{1}{N-1}\sum_{t=1}^{N}(\hat{y}_t^{(-l,-q)} - \bar{y}^{(-l,-q)})^2 \tag{19}$$

where $\hat{y}_t^{(-l,-q)} = \sum_{j\neq l, j\neq q}^{n} \hat{p}_{j,t|N} x_{j,t}$. $\hat{p}_{j,t|N}$ is the final estimation of $p_{j,t}$ in the model (18) obtained by the SDP method. $\bar{y}^{(-l,-q)}$ is the mean of all $\hat{y}_t^{(-i,-q)}$ $(t = 1, 2, \ldots, N)$.

The partial variance $\hat{V}_{(-l,-q)}$ estimated in Eq. (19) consists of the uncorrelated and correlated contributions to the variance of the model output by all the input variables except x_l and x_q, $\boldsymbol{x}_{(-l,-q)}$, thus, it also contains the correlated contributions by the variation of each variable of $\boldsymbol{x}_{(-l,-q)}$ correlated with x_l and x_q. Therefore, the respective uncorrelated contributions of x_l and x_q and their correlated contributions are not present in $V_{(-l,-q)}$. Subtracting $V_{(-l,-q)}$ from the total variance V of the model output, we can get the joint contribution $V_{l,q}$ by x_l and x_q to the variance of the model output, i.e.

$$\hat{V}_{l,q} = \hat{V} - \hat{V}_{(-l,-q)}. \tag{20}$$

Then, according to the relationship in Eq. (17), the estimation of the partial variance contributed by the variation of x_l correlated with x_q, $V_{l,q}^C$, is straightforward as follows:

$$\hat{V}_{l,q}^C = \hat{V}_{l,q} - \hat{V}_l^U - \hat{V}_q^U \tag{21}$$

and the corresponding importance measure $S_{l,q}^C$ can be estimated by Eq. (22).

$$S_{l,q}^C = \frac{\hat{V}_{l,q}^C}{\hat{V}}. \tag{22}$$

4.2. Importance matrix

With the uncorrelated and correlated contributions as well as the components of the correlated contribution by the input variables to the variance of the model output, a second order importance matrix $\mathbf{S}$ is established as follows:

$$\mathbf{S} = \begin{bmatrix} S_1^U & S_{1,2}^C & \cdots & S_{1,n}^C \\ S_{2,1}^C & S_2^U & \cdots & S_{2,n}^C \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1}^C & S_{n,2}^C & \cdots & S_n^U \end{bmatrix} \tag{23}$$

in which the element located at ith row and jth column $(1 \leq i, j \leq n, i \neq j)$ represents the importance measure $S_{i,j}^C$ of the correlated contribution of x_i and x_j. The elements at the diagonal are the importance measures S_i^U $(i = 1, \ldots, n)$ of the uncorrelated contributions of x_i $(i = 1, \ldots, n)$. The importance matrix $\mathbf{S}$ above can explicitly expose and compare the effects of the uncorrelated variations of the input variables and that of the correlated variations of an input variable correlated with another variable on the variability of the model output.

5. Examples

5.1. Numerical examples

Two numerical examples in this subsection are used to testify the accuracy of the SDP method. Because in most cases it is difficult to get the exact decomposition of contribution to the variance of the model output by an individual variable analytically, we give two simple examples so that we can compare the results of the SDP method with their exact results. In decomposing the contribution by an individual variable to the variance of the model output into an uncorrelated contribution and a correlated one, we take analytical results as the references of Examples 1 and 2, and the results of the method in literature [16] are also listed to testify the advantages of the presented method.

Example 1. In the first test example, we use the simple linear model $y = g(x_1, x_2) = 2x_1 + 3x_2$ in literature [16], where x_1 and x_2 are standard normally distributed with a Pearson correlation coefficient of 0.7. The results of the presented SDP method and those from the literature [16] as well as the analytical solutions are listed in Table 1.

Example 2. Considering the response model $y = 5 + 8x_1 + x_2^2$, where $x_i \sim N(2, 2^2)$ $(i = 1, 2)$ with a Pearson correlation coefficient of $\rho_{12} = 0.5$. Table 2 shows the results of the uncertainty decomposition by the SDP method and regression-based method in literature [16] as well as the analytical solutions.

Table 1
Uncertainty decomposition for model in Example 1.

Measures	SDP	[16]	Analytical solution	Measures	SDP	[16]	Analytical solution
V_1^T	16.57	17.73	16.81	V_2^T	19.19	20.36	19.36
$S_1^T(\%)$	77.76	79.48	78.55	$S_2^T(\%)$	90.05	91.26	90.47
V_1^U	2.12	1.95	2.04	V_2^U	4.74	4.58	4.59
$S_1^U(\%)$	9.95	8.74	9.53	$S_2^U(\%)$	22.24	20.52	21.45
V_1^C	14.45	15.78	14.77	V_2^C	14.45	15.78	14.77
$S_1^C(\%)$	67.81	70.75	69.02	S_2^C	67.81	70.75	69.02

Table 2
Uncertainty decomposition for model in Example 2.

Measures	SDP	[16]	Analytical solution	Measures	SDP	[16]	Analytical solution
V_1^T	403.82	400.54	402	V_2^T	287.85	255.65	288
$S_1^T(\%)$	83.64	83.52	83.75	$S_2^T(\%)$	59.62	53.31	60
V_1^U	194.96	192.74	192	V_2^U	78.99	47.85	78
$S_1^U(\%)$	40.38	40.19	40.00	$S_2^U(\%)$	16.36	9.98	16.25
V_1^C	208.86	207.80	210	V_2^C	208.86	207.80	210
$S_1^C(\%)$	43.26	43.33	43.75	S_2^C	43.26	43.33	43.75

Fig. 1. Diagram of the cantilever beam.

Tables 1 and 2 show that for both the linear response model in Example 1 and the quadratic response model in Example 2, there is a good agreement between the SDP contributions and the analytical ones, which testifies that the method presented in this paper for the decomposition of the variance contribution of the correlated input variables is feasible and accurate. However, the uncorrelated contribution of the input variable x_2 in Example 2 estimated by the method in literature [16] shows an obvious discrepancy from the exact solution. This is because there is a quadratic term of x_2 in the response model, which will lose its nonlinearity in the linear regression. Therefore, compared with the method in literature [16], the presented SDP method has wider applicability.

Furthermore, we apply the SDP method and the method in literature [16] to the two examples above using the same sample size of 1000. However, when estimating the uncorrelated contribution of an individual variable, the method in literature [16] needs to obtain the estimated residual first by regressing the variable over all other input variables, and then the uncorrelated contribution of the variable can be derived by regressing model output over the residual. That is to say that the method in literature [16] needs double linear regression. And the proposed SDP method needs only one estimation of the state dependent parameters in Eq. (11). Therefore, the presented method is concise in computation and can improve the computational efficiency.

5.2. Engineering examples

In this subsection, we apply the presented method to two engineering structures to show the engineering application of the decomposition of the variance contribution by the correlated input variables and to testify the engineering applicability of the presented method. The results obtained by the Monte Carlo (MC) numerical simulation are taken as the references of all the examples in this subsection.

Example 1. A simple cantilever beam with rectangular cross section is shown in Fig. 1. F and M are the axial force and moment imposed on the beam, respectively. Considering the maximal stress not exceeding the yield strength Q of the beam, the limit state response function can be constructed as $g(F, M, Q) = 1 - 4M/(bh^2Q) - F^2/(bhQ)^2$, where b and h are constant parameters of the section and $b = 8.5$, $h = 25$. The distributions of the random input variables F, M, Q and their Pearson correlation coefficients are listed in Table 3. The importance measures decomposition results of the SDP method and those of MC simulation for the cantilever beam are shown in Table 4.

Table 4 shows that the results of the presented method are in good agreement with the referential solutions for the nonlinear response model in this example, which testifies the accuracy and applicability of our method to the nonlinear response function. In addition, since there is a correlation only between F and M among the input variables, we have $S_F^C = S_M^C = S_{F,M}^C$ in Table 4. For F, the uncorrelated contribution and correlated one are basically consistent, which suggests that both the correlated and uncorrelated variations of F have important contributions to the uncertainty in model output. For M, the uncorrelated contribution is relatively small, only the contribution by the variation of M correlated with F has

Table 3
Distributions and the correlation coefficients of the random input variables in Example 1.

Variable (unit)	Distribution	Mean	Standard deviation	Correlation coefficients F	Correlation coefficients M	Correlation coefficients Q
F (MN)	Normal	500	100	1	0.5	0
M (MNm)	Normal	2000	400	0.5	1	0
Q (MPa)	Lognormal	5	0.5	0	0	1

Table 4
Importance measures decomposition of the correlated input variables of the cantilever beam.

Measures	SDP	MC	Measures	SDP	MC	Measures	SDP	MC
S_F^T	0.6174	0.6071	S_M^T	0.4664	0.4503	S_Q^T	0.2413	0.2429
S_F^U	0.2890	0.2689	S_M^U	0.1381	0.1121	S_Q^U	0.2684	0.2658
S_F^C	0.3283	0.3382	S_M^C	0.3283	0.3382	S_Q^C	−0.0271	−0.0229

Table 5
The distribution parameters of the random input variables of roof truss.

Random variable	q (N/m)	l (m)	A_S (m^2)	A_C (m^2)	E_S (N/m^2)	E_C (N/m^2)
Mean	20,000	12	9.82×10^{-4}	0.04	2×10^{11}	3×10^{10}
Coefficient of *variation*	0.07	0.01	0.06	0.12	0.06	0.06

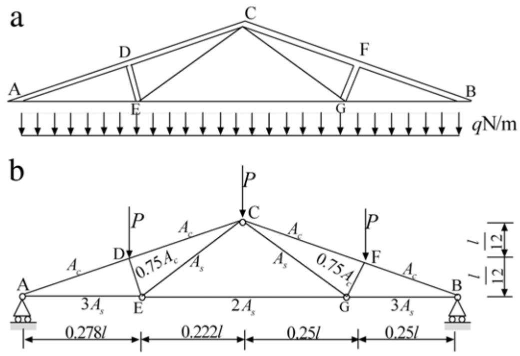

Fig. 2. Schematic diagram of a roof truss.

a notable effect on the output. For Q, there is a much smaller correlated contribution compared with the uncorrelated contribution, which is consistent with our parameter settings, since there is no correlation between Q and any other input variable. The non-zero correlated contribution may be due to the numerical error in the correlation calculation when using the Copula transformation. Therefore, to decrease the variability of the output response of the cantilever beam, we should start with decreasing the uncorrelated contributions of F and Q respectively as well as the correlated contribution of F and M.

Example 2. A roof truss is shown in Fig. 2, the top boom and the compression bars are reinforced by concrete, and the bottom boom and the tension bars are all made of steel. Assume the uniformly distributed load q is used on the roof truss, and the uniformly distributed load can be transformed into the nodal load $P = ql/4$. The perpendicular deflection Δ_C of node C can be obtained by mechanical analysis, and it is the function of the input random variables, i.e., $\Delta_C = \frac{ql^2}{2}\left(\frac{3.81}{A_C E_C} + \frac{1.13}{A_S E_S}\right)$, where A_C, A_S, E_C, E_S, l respectively represent sectional area, elastic modulus, length of the concrete and steel bars. Considering the safety and the applicability, Δ_C of the node C not exceeding 3 cm is taken as the constraint condition, the performance response function can be constructed by $g = 0.03 - \Delta_C$. Assume that all the input variables are normally distributed with the Pearson correlation coefficients $\rho_{lA_C} = \rho_{lA_S} = 0.3$, $\rho_{A_C A_S} = \rho_{E_C E_S} = 0.5$, respectively. Their distribution parameters are given in Table 5. The importance measures decomposition results of the SDP method and that of MC simulation are shown in Fig. 3.

Fig. 3 shows that for the response function with major nonlinearity in this example, the decomposition results of the importance measures obtained by the presented SDP method also match those by the MC very well. This testifies the applicability of the proposed SDP method to the nonlinear response function once again. Using the SDP method for the estimation of the correlated contribution of two specific variables in Section 3.3, we can get the importance matrix $\mathbf{S}$(%) of the input variables as follows. However, we only show the contributions that are not zeros in theory. All the contributions that should be zeros theoretically are not illustrated, because the SDP results of these contributions are very small. For

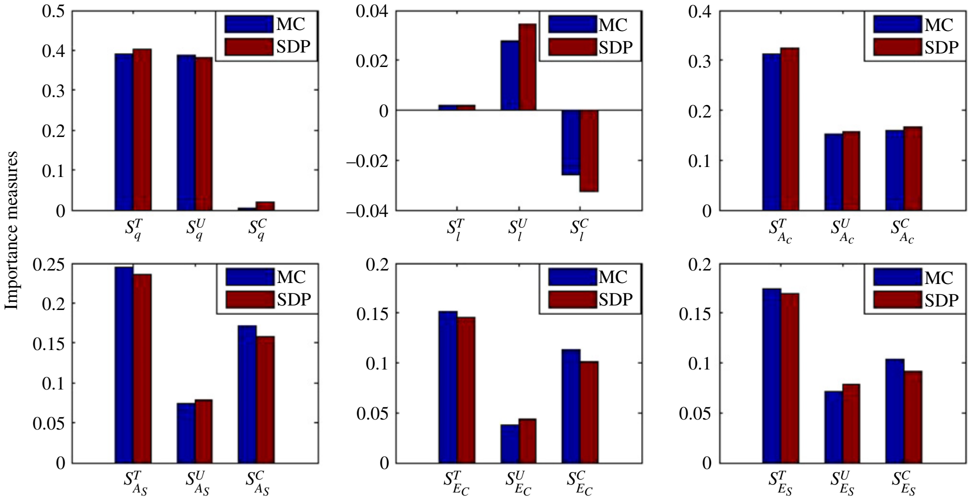

Fig. 3. Importance measures decomposition of the correlated input variables of the roof truss.

example, we only show the uncorrelated contribution of the input variable q, since there is no correlation between q and other input variables specified.

$$\mathbf{S}=\begin{bmatrix} S_q^U & S_{q,l}^C & S_{q,A_C}^C & S_{q,A_S}^C & S_{q,E_C}^C & S_{q,E_S}^C \\ S_{l,q}^C & S_l^U & S_{l,A_C}^C & S_{l,A_S}^C & S_{l,E_C}^C & S_{l,E_S}^C \\ S_{A_C,q}^C & S_{A_C,l}^C & S_{A_C}^U & S_{A_C,A_S}^C & S_{A_C,E_C}^C & S_{A_C,E_S}^C \\ S_{A_S,q}^C & S_{A_S,l}^C & S_{A_S,A_C}^C & S_{A_S}^U & S_{A_S,E_C}^C & S_{A_S,E_S}^C \\ S_{E_C,q}^C & S_{E_C,l}^C & S_{E_C,A_C}^C & S_{E_C,A_S}^C & S_{E_C}^U & S_{E_C,E_S}^C \\ S_{E_S,q}^C & S_{E_S,l}^C & S_{E_S,A_C}^C & S_{E_S,A_S}^C & S_{E_S,E_C}^C & S_{E_S}^U \end{bmatrix}=\begin{bmatrix} 38.28 & & & & & \\ & 3.43 & -2.51 & -1.97 & & \\ & -2.51 & 15.64 & 14.82 & & \\ & -1.97 & 14.82 & 7.90 & & \\ & & & & 4.42 & 9.38 \\ & & & & 9.38 & 7.80 \end{bmatrix}.$$

The results in Fig. 3 show that for the input variable q, the uncorrelated contribution is much bigger than the correlated one which is in agreement with the variable settings, since there is no correlation between q and other variables specified. The non-zero correlated contribution may be due to the numerical error. For the variables l, A_C and E_S, the uncorrelated contribution and correlated one are basically consistent, which suggests that both the uncorrelated and correlated variations of these variables have important contributions to the variance of the model output. However, the situation for the variables A_S and E_C is different, the results show that their correlated contributions dominate. The uncorrelated contributions are relatively small compared with the correlated ones. Additionally, we can get further information about the correlated contributions of the input variables from the importance matrix **S**. For the variable l, the contributions by the variation of l correlated with A_C and A_S respectively are all negative, which agrees with the inverse proportion relationship of l and A_C, A_S. Although the total contribution by l to the variance of the model output is near zero, we can also decrease the uncertainty of the output by increasing the variations of l correlated with other variables. For the variable A_C, the correlated contribution of A_C and A_S is much bigger than that of A_C and l. Therefore, to decrease the variability of the model output from the correlated contribution of A_C, we only need to care about the contribution of A_C correlated with A_S. This is also the situation of the variable A_S, since among its correlated contributions, the correlated contribution of A_S and A_C dominates. For the variables E_C and E_S, the correlated contributions of them are greater than their respective uncorrelated contributions. This suggests that the correlated variation of them has a more important contribution than their individual variations to the uncertainty in model output. Therefore, to decrease the variability of the output of the roof truss, we should decrease the uncorrelated contribution of all the input variables and the correlated contribution of A_C and A_S, E_C and E_S, as well as increasing the contribution by the variation of l correlated with A_C and A_S respectively.

Furthermore, it can be seen from the definition of the importance matrix **S** that the diagonal elements represent the uncorrelated contribution ratios of all the input variables, and the elements above the diagonal are the correlated contribution ratios of two specific variables. Therefore, the sum of the elements located at the diagonal and above the diagonal should equal 1, i.e. $\sum_{k=1}^{n}(S_k^U+\sum_{j=k+1}^{n}S_{k,j}^C)=1$. For the roof truss in this example, the sum actually is 97.19%. This is caused by the non-zero correlated contribution of two uncorrelated variables, which may be caused by the numerical error produced during the transformation of the independent samplings to the correlated ones. However, the total error does not exceed 3%, which shows that our method can be applied to the engineering structures with adequate accuracy. Besides, comparing the sum the correlated contributions of one variable correlated with each of the other variables, i.e. $\sum_{j=1,j\neq i}^{n}S_{i,j}^C$, in the importance matrix with the overall correlated contribution S_i^C of this input variable in Fig. 3, one may note that their SDP results are not exactly equal. For example, for the input variable l, $S_l^C=-3.24$ in Fig. 3, but $\sum_{j=1,j\neq l}^{n}S_{l,j}^C=-4.48$

in the importance matrix **S**. This error results from the 1st HDMR expansion, because the correlated contribution of two specific input variables may sometimes extensively be affected by their crossed term. However, for the response function with major crossed effect of the input variables in this case, the present method can also provide some useful information as discussed above. And if we can consider the correlated contribution caused by the crossed terms, the resultant $S_{i,j}^C$ may be more accurate. This also suggests an improving direction for applying our method to more general cases.

6. Conclusions

Variance-based importance analysis of the correlated input variables divides the contribution by the correlated input variables to the variance of the model output into the uncorrelated contribution and correlated one, which provides an effective way to improve the structure model. To overcome the problems in the existing method, a new approach is presented to analyze the importance of the correlated variables. Compared with the existing method, the presented method relies on no assumption regarding the estimation space and the error space and is suitable for the nonlinear response model due to the employment of the SDP method which is to estimate the nonlinear and non-stationary system. Therefore, it has wider applicability. Additionally, the presented method avoids the double regression employed by the existing method when estimating the uncorrelated contributions of the input variables, which can improve the computational efficiency as well as decrease the error. Thus, it has advantages both in efficiency and accuracy. Based on the proposed SDP method for the estimation of the correlated contribution of two specific variables, an importance matrix is defined to explicitly reflect the contribution by the correlated input variables to the model output, which can meet different requirements in engineering. The importance measures of the uncorrelated contribution and correlated one of the correlated input variables in the numerical and engineering examples are calculated. The results show that the presented method can accurately decompose the contribution of the correlated input variables for both linear and nonlinear response models with minor crossed effect of the input variables, and decrease the computational effort compared with the existing method. For the response models with major crossed effect of the input variables, the present method also can provide some guiding information for improving the structure model, and the results may be more accurate by considering the correlated contribution caused by the crossed terms. At last, the application of the presented method to the engineering examples provides a direction for the engineers to decrease the variability of the output of the beam and roof truss.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. NSFC 50875213, 51175425).

References

[1] A. Satelli, Sensitivity analysis for importance assessment, Risk. Anal. 22 (2002) 579–590.
[2] J.C. Helton, F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliab. Eng. Syst. Saf. 81 (2003) 23–69.
[3] J.C. Helton, F.J. Davis, Sampling-based methods, in: A. Saltelli, K. Chan, E.M. Scott (Eds.), Sensitivity Analysis, Wiley, New York, 2000, pp. 101–153.
[4] A. Saltelli, J. Marivoet, Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques, Reliab. Eng. Syst. Saf. 28 (1990) 229–253.
[5] I.M. Sobol, Sensitivity analysis for non-linear mathematical models, Math. Model. Comput. Exp. 1 (1993) 407–414. English translation of Russian original paper.
[6] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mat. Comput. Simulation 55 (2001) 221–280.
[7] R.L. Iman, S.C. Hora, A robust measure of uncertainty importance for use in fault tree system analysis, Risk. Anal. 10 (1990) 401–406.
[8] E. Castillo, R. Minguez, C. Castillo, Sensitivity analysis in optimization and reliability problems, Reliab. Eng. Syst. Saf. 93 (2008) 1788–1800.
[9] M.H. Chun, S.J. Han, N.I. Tak, An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, Reliab. Eng. Syst. Saf. 70 (2000) 313–321.
[10] H.B. Liu, W. Chen, A. Sudjianto, Relative entropy based method for probabilistic sensitivity analysis in engineering design, J. Mech. Design. 128 (2006) 326–333.
[11] E. Borgonovo, A new uncertainty importance measure, Reliab. Eng. Syst. Saf. 92 (2007) 771–784.
[12] A. Saltelli, M. Ratto, S. Tarantola, Model-free importance indicators for dependent input, in: Proceedings of SAMO 2001, Third International Symposium on Sensitivity Analysis of Model Output, Madrid, 2001.
[13] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Wiley, West Sussex, 2004, pp. 132–136.
[14] T. Bedford, Sensitivity indices for (tree)-dependent variables, in: Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output, Venice, Italy, 1998.
[15] S. Fang, G.Z. Gertner, A. Anderson, Estimation of sensitivity coefficients of nonlinear model input parameters which have a multinormal distribution, Comput. Phys. Commun. 157 (2004) 9–16.
[16] C. Xu, G.Z. Gertner, Uncertainty and sensitivity analysis for models with correlated parameters, Reliab. Eng. Syst. Safe. 93 (2008) 1563–1573.
[17] G. Li, H. Rabitz, et al., Global sensitivity analysis for systems with independent and/or correlated inputs, J. Phys. Chem. 114 (2010) 6022–6032.
[18] P.C. Young, Time variable and state dependent modelling of nonstationary and nonlinear time series, in: T.S. Rao (Ed.), Developments in Time Series Analysis, Chapman and Hall, London, 1993, pp. 374–413.
[19] P.C. Young, Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation, in: W.J. Fitzgerald, R.L. Smith, A.T. Walden, P.C. Young (Eds.), Nonlinear and Nonstationary Signal Processing, Cambridge University Press, Cambridge, 2000, pp. 74–114.
[20] M. Ratto, S. Tarantola, A. Saltelli, P.C. Young, Accelerated estimation of sensitivity indices using state dependent parametermodels, in: K.M. Hanson, F.M. Hemez (Eds.), Sensitivity Analysis of Model Output, Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output SAMO 2004 Santa Fe, New Mexico, March 8–11, 2004, pp. 61–70.
[21] M. Ratto, A. Pagano, P.C. Young, State dependent parameter meta-modelling and sensitivity analysis, Comput. Phys. Commun. 177 (2007) 863–876.
[22] M. Ratto, A. Pagano, P.C. Young, Non-parametric estimation of conditional moments for sensitivity analysis, Reliab. Eng. Syst. Saf. 94 (2009) 237–243.
[23] H. Rabitz, O.F. Alis, J. Shorter, K. Shim, Efficient input–output model representations, Comput. Phys. Commun. 117 (1999) 11–20.