Efficient global sensitivity analysis of computer simulation models using an adaptive Least Angle Regression scheme

Géraud Blatman, Bruno Sudret

Efficient global sensitivity analysis of computer simulation models using an adaptive Least Angle Regression scheme

Géraud Blatman[1,2] & Bruno Sudret[1,3]

[1] *IFMA-LaMI, Campus des Cézeaux, BP 265, 63175 Aubière cedex*
[2] *EDF R&D, Site des Renardières, 77250 Moret-sur-Loing cedex*
[3] *Phimeca Engineering S.A., Centre d'affaires du Zénith, 34 rue de Sarliève, 63800 Cournon*

Résumé On propose une méthode efficace d'analyse de sensibilité de modèles de simulation numérique dont les paramètres d'entrée sont aléatoires. Les *indices de sensibilité* du modèle sont couramment estimés par simulation de Monte Carlo, ce qui s'avère particulièrement coûteux dans le cas d'applications industrielles. Dans ce papier, on introduit les développements par chaos polynomial pour calculer les indices de sensibilité. Une méthode adaptative basée sur la procédure *LAR* (*Least Angle Regression*) est développée afin de détecter automatiquement les coefficients significatifs du chaos, qui peuvent donc être calculés au moyen d'un faible nombre d'évaluations du modèle. La méthode est appliquée à l'analyse de sensibilité de la *fonction de Sobol'*.

Abstract An efficient method is proposed to perform global sensitivity analysis of a the computer model of a physical system whose input parameters are random. The so-called *sensitivity indices* are usually computed by Monte Carlo simulation, which reveals inappropriate for industrial applications due to an unaffordable computational cost. In the present paper, *polynomial chaos (PC) expansions* are introduced in order to compute sensitivity indices. An adaptive Least Angle Regression (LAR) algorithm is devised for automatically detecting the significant coefficients of the PC expansion. The latter may thus be computed eventually by means of a relatively small number of possibly costly model evaluations. The method is applied to the sensitivity analysis of the so-called Sobol' function.

Keywords Sensitivity analysis, sparse polynomial chaos expansion, adaptive Least Angle Regression

1 Introduction

Global sensitivity analysis aims at quantifying the relative importance of random input variables onto the response of a mathematical model of a physical system. ANOVA-based indices such as the *Sobol' indices* [1] are well-known in this context. These indices are usually computed by Monte Carlo or quasi-Monte Carlo simulation, which reveals inappropriate for industrial applications due to an unaffordable computational cost. One may rather substitute the model under consideration for an *analytical* approximation (*i.e.* a *metamodel*), whose evaluations are inexpensive. In this respect, an approach based on *polynomial chaos (PC) expansions* has been proposed in [2]. The Sobol' indices are then obtained *analytically* from the PC coefficients. However, the required number of model evaluations (*i.e.* the computational cost) increases with the PC size, which itself dramatically increases with the number of input variables when the common truncation scheme of the PC expansion is applied.

To circumvent this problem, a truncation strategy based on the use of so-called *q-norms* with $0 < q < 1$ is proposed. It is motivated by the so-called *sparsity-of-effects principle* [3], which states that most models are principally governed by main effects and low-order interactions. The related truncated PC expansions contain a low number of likely important terms compared to the full representation. Using this truncation scheme, an adaptive method based on *Least Angle Regression* (LAR) [4] is proposed in order to retain progressively a small number of significant PC coefficients, leading to a *sparse* PC representation.

2 Global sensitivity analysis

Consider a physical system described by a numerical model $\mathcal{M}$ which can be analytical or more generally algorithmic (*e.g.* a finite element model). Suppose that this model has M uncertain input parameters which are represented by *independent* random variables $\{X_1, \ldots, X_M\}^\intercal$ gathered into a random vector $\boldsymbol{X}$ with prescribed joint probability density function $f_{\boldsymbol{X}}(\boldsymbol{x})$. Hence the model response denoted by $Y = \mathcal{M}(\boldsymbol{X})$ is also random. For the sake of simplicity, Y is assumed to be scalar throughout the paper. From now on it is assumed that Y has a finite variance.

The *Sobol' sensitivity indices* [1] aim at quantifying the relative importance of each input parameter in the response variance:

$$S_i \;=\; \frac{\mathbb{V}\left[\mathbb{E}\left[\mathcal{M}(\boldsymbol{X})|X_i\right]\right]}{\mathbb{V}\left[\mathcal{M}(\boldsymbol{X})\right]} \tag{1}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ are the mathematical expectation and variance operators, respectively. S_i is referred to as the *first order sensitivity index* of the model response to the input random variable X_i. This measure may be extended to each subset $\{X_{i_1}, \ldots, X_{i_s}\}$ of input random variables to quantify the interaction effects.

In practice, the sensitivity indices are estimated using Monte Carlo simulation [1]. However, a very large number of model evaluations (say, more than 10^4) is usually required in order to obtain accurate results, which may lead to intractable calculations in case of a computationally demanding model $\mathcal{M}$. To overcome this problem, one may substitute the model response for its *polynomial chaos expansion*, as shown in the sequel.

3 Polynomial chaos expansion of the model response

As the random variable Y is assumed to have a finite variance, it may be recast as follows [5]:

$$Y \;=\; \mathcal{M}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \tag{2}$$

This expansion is referred to as the *polynomial chaos* (PC) *representation* of Y. The $a_{\boldsymbol{\alpha}}$'s are unknown deterministic coefficients and the $\psi_{\boldsymbol{\alpha}}$'s are multivariate polynomials which are orthonormal with respect to the joint PDF $f_{\boldsymbol{X}}$ of the input random vector $\boldsymbol{X}$, *i.e.* $\mathbb{E}[\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})\psi_{\boldsymbol{\beta}}(\boldsymbol{X})] = 1$ if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and 0 otherwise.

Let us consider a set of realizations of $\boldsymbol{X}$ denoted by $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$ and referred to as the *experimental design* (ED). Let us denote by $\mathcal{Y}$ the associated set of model response

quantities, say $\mathcal{Y} = \{\mathcal{M}(\boldsymbol{x}^{(1)}), \ldots, \mathcal{M}(\boldsymbol{x}^{(N)})\}$. The PC coefficients may be estimated by least square regression [6]. This method requires the choice of a truncation of the PC *ab initio*, *i.e.* a non empty finite set $\mathcal{A} = \{\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{P-1}\} \subset \mathbb{N}^M$ which contains the multi-indices of the retained basis polynomials $\{\psi_{\boldsymbol{\alpha}_0}, \ldots, \psi_{\boldsymbol{\alpha}_{P-1}}\}$. $\mathcal{A}$ is referred to as the *truncation set* in the sequel. The corresponding PC approximation is denoted by $Y_{\mathcal{A}} \equiv \mathcal{M}_{\mathcal{A}}(\boldsymbol{X})$. The PC expansion is commonly truncated by retaining those polynomials whose total degree is not greater than p, which leads to the truncation set $\mathcal{A}^{M,p} = \{\boldsymbol{\alpha} \in \mathbb{N}^M \, : \, \|\boldsymbol{\alpha}\|_1 \equiv \sum_{i=1}^M \alpha_i \leq p\}$. Accordingly, the number of PC terms is given by $P = \binom{M+p}{p}$. Hence it dramatically increases with both p and M. Consequently, the minimal size of the ED (*i.e.* the computational cost) that is required for an accurate solution of the regression problem blows up. This problem is tackled in Section 4.1.

Let us now define by $\mathcal{I}_{i_1,\ldots,i_s}$ the set of $\boldsymbol{\alpha}$-tuples in $\mathcal{A}$ such that only the indices $\{i_1, \ldots, i_s\}$ are nonzero:

$$\mathcal{I}^p_{i_1,\ldots,i_s} \, = \, \left\{ \, \boldsymbol{\alpha} \in \mathcal{A}^p \, : \, \alpha_k = 0 \iff k \notin (i_1, \ldots, i_s) \quad , \forall \, k = 1, \, \ldots \, , M \right\} \tag{3}$$

It is easy to derive sensitivity indices from the PC representation (2), see details in [2]. These indices, called *PC-based sensitivity indices* are denoted by $S^{\mathcal{A}^p}_{i_1,\ldots,i_s}$ and given by:

$$S^{\mathcal{A}^p}_{i_1,\ldots,i_s} \, = \, \frac{\sum_{\boldsymbol{\alpha} \in \mathcal{I}^p_{i_1,\ldots,i_s}} a^2_{\boldsymbol{\alpha}}}{\sum_{\boldsymbol{\alpha} \in \mathcal{A}^p} a^2_{\boldsymbol{\alpha}}} \tag{4}$$

4 Sparse polynomial chaos approximation

4.1 PC expansions based on the sparsity-of-effects principle

An alternative strategy is proposed in the present paper for truncating the PC expansion of the model response. It is motivated by the so-called *sparsity-of-effects principle* [3], which states that most models are principally governed by main effects and low order interactions. One proposes the use of the following truncation sets based on q-norms, $0 < q < 1$:

$$\mathcal{A}^{M,p}_q = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M \, : \, \|\boldsymbol{\alpha}\|_q \equiv \left(\sum_{i=1}^M \alpha_i^q \right)^{1/q} \leq p \right\} \tag{5}$$

Such norms penalize the higher order interaction terms all the more since q is low. Note that setting q equal to 1 corresponds to the usual truncation scheme mentioned in Section 3. The proposed truncation strategy thus leads to PC expansions with a reduced number of unknown coefficients, which may be computed using a moderate number N of model evaluations.

The computational cost may be further reduced by taking into account the fact that the PC expansion of the model response contains only a small number of significant terms (*sparse* PC expansion). This is the scope of the next section.

4.2 Sparse PC approximation using an adaptive LAR algorithm

4.2.1 Estimation of the PC coefficients using LAR

Regularization is a technique that allows one to perform a least-square regression when the number of model evaluations N is less than the number P of basis functions. It relies upon

a penalization of some norm (or more generally functional) of the regression coefficients. In particular, $\mathcal{L}^1$-regularized regression consists in fitting a metamodel $\mathcal{M}_{\mathcal{A}_q^{M,p}}$ by solving:

$$\text{Minimize} \quad \sum_{i=1}^{N} \left(y^{(i)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}} a_{\boldsymbol{\alpha}} \psi(\boldsymbol{x}^{(i)}) \right)^2 \qquad \text{subject to} \quad \sum_{\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}} |a_{\boldsymbol{\alpha}}| \leq s \tag{6}$$

where $s \geq 0$ is a tuning parameter. The $\mathcal{L}^1$-type constraint of this optimization problem yields a *sparse* solution (*i.e.* with many components equal to zero). In other words, a selection of a small number of significant terms in the basis $\mathcal{A}_q^{M,p}$ is performed. The solution is all the sparser since the value of s is low.

Least Angle Regression (LAR) [4] is an efficient algorithm for solving the problem in Eq.(6). It provides in one shot the entire paths of solution coefficients as s is increased from 0 up to a maximum value. The LAR procedure is described below:

1. Standardize the vectors $\{\boldsymbol{\Psi}_{\boldsymbol{\alpha}_i}, i = 1, \ldots, P-1\}$ to have empirical mean zero and empirical variance one, where $\boldsymbol{\Psi}_{\boldsymbol{\alpha}_i} \equiv \{\psi_{\alpha_i}(\boldsymbol{x}^{(1)}), \ldots, \psi_{\alpha_i}(\boldsymbol{x}^{(N)})\}^\mathsf{T}$. Initialize the approximated response vector $\hat{\mathcal{Y}} = \mathbf{0}$, which corresponds to the initial coefficients $a_{\boldsymbol{\alpha}_0}, \ldots, a_{\boldsymbol{\alpha}_{P-1}} = 0$. Define the residual $\boldsymbol{R} = \mathcal{Y} - \hat{\mathcal{Y}}$.

2. Find the vector $\boldsymbol{\Psi}_{\boldsymbol{\alpha}_j}$ which is most correlated with $\boldsymbol{R}$.

3. Move $a_{\boldsymbol{\alpha}_j}$ from 0 towards the value $\boldsymbol{\Psi}_{\boldsymbol{\alpha}_j}^\mathsf{T} \boldsymbol{R}$, until some other vector $\boldsymbol{\psi}_{\boldsymbol{\alpha}_k}$ has as much correlation with the current residual as does $\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}$.

4. Move jointly $\{a_{\boldsymbol{\alpha}_j}, a_{\boldsymbol{\alpha}_k}\}^\mathsf{T}$ in the direction defined by their least-square coefficients of the current residual on $\{\boldsymbol{\psi}_{\boldsymbol{\alpha}_j}, \boldsymbol{\psi}_{\boldsymbol{\alpha}_k}\} \equiv \boldsymbol{\Phi}$, until some other vector $\boldsymbol{\psi}_{\boldsymbol{\alpha}_l}$ has as much correlation with the current residual. This direction is defined explicitly by the vector $\{\hat{a}_{\boldsymbol{\alpha}_j}, \hat{a}_{\boldsymbol{\alpha}_k}\}^\mathsf{T} = (\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{R}$.

5. Continue this way until all P predictors have been entered. After P steps, one gets the full least-square solution.

6. Compute accuracy estimates of the metamodels associated with all the solution coefficients, and select the truncation set $\mathcal{A}^{p,*}$ that corresponds to the largest accuracy estimate.

LAR may be thus regarded as a way of selecting an optimal sparse PC basis. One chooses to eventually apply ordinary least-square regression to recompute the corresponding coefficients, as it is considered to provide more accurate estimates. The accuracy estimates that are used in Step 6 are based on *leave-one-out cross validation* [7] and are denoted by $\hat{Q}^2$.

4.2.2 Sparse PC expansion using an adaptive LAR scheme

A limitation of LAR lies on the requirement of an *a priori* truncation set $\mathcal{A}_q^{M,p}$. To circumvent this difficulty, one proposes a procedure for progressively enriching the truncation set of the PC approximation, *i.e.* the set of active basis functions. One first selects the type q of the norm that is used to truncate the PC expansions. The proposed adaptive procedure is outlined below:

1. Select an ED $\mathcal{X}$ and collect the corresponding model evaluations in $\mathcal{Y}$ once and for all.
2. Initialize the PC degree $p = 0$.
3. Initialize the set $\mathcal{A}^{p,+} = \emptyset$.
4. Set $\mathcal{A}^p = \mathcal{A}_q^{M,p} \cup \mathcal{A}^{p,+}$. Apply the LAR algorithm to fit the metamodel $\mathcal{M}_{\mathcal{A}^p}$. Store the multi-indices of non zero coefficients in $\mathcal{A}^{p,+}$. Store the accuracy estimate $\hat{Q}^{2,p}$.
5. If $p \geq 2$: if $\hat{Q}^{2,p} \leq \hat{Q}^{2,p-1} \leq \hat{Q}^{2,p-2}$ (overfitting), then go directly to Step 7.
6. Set $p = p + 1$ and go back to Step 4.
7. Eventually select the metamodel $\mathcal{M}_{\mathcal{A}^k}$, $0 \leq k \leq p$ with largest $\hat{Q}^{2,k}$.

The condition in Step 5 is a heuristic criterion to avoid overfitting.

5 Application example

Let us consider the so-called Sobol' function [8]:

$$Y = \prod_{i=1}^{M} \frac{|4X_i - 2| + a_i}{1 + a_i} \tag{7}$$

where the input variables $X_i, i = 1, \ldots, M$ are uniformly distributed over $[0, 1]$ and a_i are non negative constants. The sensitivity indices of Y may be derived analytically. For numerical application, the number of input variables M is set equal to 8 and one selects $\boldsymbol{a} = \{1, 2, 5, 10, 20, 50, 100, 500\}^\mathsf{T}$. In order to validate our methodology, the sensitivity indices of the four greatest sensitivity indices (*i.e.* $S_1 - S_4$) are estimated by post-processing a PC approximation of Y as shown in Eq.(3). The PC representation is built up using the iterative LAR procedure outlined in Section 4.2.2. To this end, EDs based on a quasi-random Sobol' sequence with size $N = 100$ and $N = 500$ are used. Moreover, two norms are considered to truncate the PC expansion as in Eq.(5), namely by selecting $q = 1$ and $q = 0.6$. For the sake of comparison, the sensitivity indices are computed by crude Monte Carlo simulation using $N = 100,000$ samples. The results are reported in Table 1.

It appears that the most accurate estimates are obtained using the adaptive LAR method associated with $q = 0.6$, with relative errors of 3% and 1% on S_1 when using $N = 100$ and $N = 500$ samples, respectively. Those relative errors are equal to 8% and 4% when selecting $q = 1$. Note that the metamodel corresponding to $q = 0.6$ and $N = 100$ overperforms its counterpart corresponding to $q = 1$ and $N = 500$, hence a computational gain factor more than 5. Such results are well in agreement with the error estimates of the PC approximations. Note that as expected, the accuracy of the estimates increases with the size of the ED. Furthermore, the adaptive LAR scheme provides very sparse representations, with an *index of sparsity IS* varying from 1% to 23% with respect to their *full* counterparts.

On the other hand, all the LAR-based PC approximations globally yield more accurate estimates of the sensitivity indices than crude Monte Carlo simulation, with a computational cost divided by 200 or $1,000$.

Table 1: Estimates of the sensitivity indices of the Sobol' function

Sensitivity indices (%)	Reference (analytical)	LAR $N = 100$		LAR $N = 500$		MC
		$q = 1$	$q = 0.6$	$q = 1$	$q = 0.6$	($N = 10^5$)
S_1	60	52	63	64	61	57
S_2	27	33	27	26	27	29
S_3	7	6	6	5	6	6
S_4	2	1	2	1	2	3
PC accuracy estimate $\hat{Q}^2$		0.84	0.93	0.89	0.99	-
Reached degree p		4	4	8	10	-
Index of sparsity IS † (%)		4	23	1	8	-

† $IS \equiv$ Number of non zero terms/Total number of terms $= (\sum_{\boldsymbol{\alpha} \in \mathcal{A}_q^{M,p}} \mathbf{1}_{\{a_{\boldsymbol{\alpha}} \neq 0\}}) \; / \; (|\mathcal{A}_q^{M,p}|)$

6 Conclusion

An adaptive LAR procedure is proposed to build up a sparse PC representation of the random response of a model with random input parameters. In order to reduce the number of unknown PC coefficients to identify and hence the required number of computer experiments (*i.e.* the computational cost), an adaptive algorithm is proposed for automatically detecting the significant PC terms (sparse PC representation). The example of the Sobol' function shows that the algorithm may be used to efficiently estimate the sensitivity indices of the model response to its input parameters, leading to to a considerable reduction of the number of model evaluations compared to crude Monte Carlo simulation. A new version of the proposed procedure will automatically enrich the experimental design, in order to minimize the computational cost.

REFERENCES

[1] I.M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling & Comp. Exp.*, 1:407–414, 1993.

[2] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Sys. Safety*, 93:964–979, 2008.

[3] D.C. Montgomery. *Design and analysis of experiments.* 2004.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Statist.*, 32:407–499, 2004.

[5] C. Soize and R. Ghanem. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.*, 26(2):395–410, 2004.

[6] M. Berveiller, B. Sudret, and M. Lemaire. Stochastic finite elements: a non intrusive approach by regression. *Eur. J. Comput. Mech.*, 15(1-3):81–92, 2006.

[7] A. Elisseff and M. Pontil. Leave-one-out error and stability of learning algorithms with applications. In G. Horvath J. Suykens and S. Basu et. Al, editors, *Learning Theory and Practice*. NATO ASI Series, IOS Press, Amsterdam, 2002.

[8] I.M. Sobol'. Theorems and examples on high dimensional model representation. *Reliab. Eng. Sys. Safety*, 79:187–193, 2003.