Computers and Structures 124 (2013) 47–60

Contents lists available at SciVerse ScienceDirect

Computers and Structures

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

Competitive comparison of optimal designs of experiments for sampling-based sensitivity analysis

Eliška Janouchová, Anna Kučerová *

Department of Mechanics, Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29 Prague 6, Czech Republic

ARTICLE INFO

Article history:
Available online 29 April 2013

Keywords:
Design of experiments
Space-filling
Orthogonality
Latin Hypercube Sampling
Sampling-based sensitivity analysis

ABSTRACT

A widely used strategy to explore the sensitivity of the model to its inputs is based on a finite set of simulations. These are usually performed for a chosen set of points in a parameter space. An estimate of the sensitivity can be then obtained by computing correlations between the model inputs and outputs. The accuracy of the sensitivity prediction depends on a quality of the points distribution in the parameter space, so-called the design of experiments. The aim of the presented paper is to review and compare available criteria determining an optimal design of experiments for sampling-based sensitivity analysis.

1. Introduction

Sensitivity analysis (SA) is an important tool for investigating properties of complex systems. It represents an essential part of inverse analysis procedures [1], response surface modelling [2] or uncertainty analysis [3]. To be more specific, SA provides some information about the contributions of individual system parameters/model inputs to the system response/model outputs. A number of approaches to SA has been developed, see e.g. [4] for an extensive review. The presented contribution is focused on widely used sampling-based approaches [2], particularly aimed at an evaluation of Spearman's rank correlation coefficient (SRCC), which is able to reveal a nonlinear monotonic relationship between the inputs and the corresponding outputs.

When computing the SA in a case of some real system using expensive experimental measurements or some computationally exhaustive numerical model, the number of samples to be performed within some reasonable time is rather limited. Randomly chosen sets of input parameters do not ensure appropriate estimation of related sensitivities. Therefore the sets must be chosen carefully. In this contribution we would like to present a review and comparison of several criteria, which can govern the stratified generation of input sets – the so called design of experiments (DoE).

Generation of optimal DoEs is a very broad topic and all pertinent aspects cannot be discussed within this paper. Hence, we focus especially on DoEs in discrete domains. The presented methods can be of course applied also to discretized continuous domain. Nevertheless, other possibilities for generation DoEs in continuous domains are, however, beyond the scope of this paper.

The following section reviews the criteria for optimisation of DoE, which are available in literature. Section 3 includes some comments on widely used methods for stratified generation of DoE and Section 4 presents the discussion on difficulties arising from optimisation of particular criteria. Section 5 is devoted to the comparison of mutual qualities of particular optimal DoEs and Section 6 compares their quality in terms of projective properties which are important in a screening phase of model analysis. Sequential improvement of the existing DoE is discussed in Section 7. Finally, Sections 8 and 9 present the assessment of the optimal designs quality for usage in sampling-based SA for theoretical analytical functions and structural models, respectively. Concluding remarks are summarised in Section 10.

2. Criteria for assessing optimal designs

A number of different criteria for assessing the quality of particular DoE can be found in literature. In general, they can be organised into groups w.r.t. the preferred DoE property. The most widely preferred features are

- *space-filling* property, which is needed to allow for the evaluation of sensitivities valid for the whole given domain of admissible input values, the so called design space;
- *orthogonality*, which is necessary to assess the impact of individual input parameters.

Other main objectives may be preferable in particular applications of DoE. In response surface methodology, reduction of noise and

* Corresponding author. Tel.: +420 2 2435 5326; fax: +420 2 2431 0775.
E-mail addresses: eliska.janouchova@fsv.cvut.cz (E. Janouchová), anicka@cml.fsv.cvut.cz (A. Kučerová).

0045-7949/$ - see front matter
http://dx.doi.org/10.1016/j.compstruc.2013.04.009

bias error can become more important than the orthogonality [5]. Nevertheless, no special objectives were formulated for the case of sampling-based SA, so we employ the common ones.

2.1. Space-filling criteria

Let us recall four widely used space-filling criteria.

Audze-Eglais objective function (**AE**) proposed by Audze and Eglais [6] is based on a potential energy among the design points. The points are distributed as uniformly as possible when the potential energy E^{AE} proportional to the inverse of the squared distances among points is minimised, i.e.

$$E^{AE} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{1}{L_{ij}^2}, \tag{1}$$

where n is the number of the design points and L_{ij} is the Euclidean distance between points i and j.

Euclidean maximin (**EMM**) **distance** is probably the best-known space-filling measure [7,8]. It states that the minimal distance $L_{\min,ij}$ between any two points i and j should be maximal. In order to apply the minimisation procedure to all presented criteria, we minimise the negative value of a minimal distance E^{EMM}, i.e.

$$E^{EMM} = -\min\{\ldots, L_{ij}, \ldots\}, \quad i = 1\ldots n,\ j = (i+1)\ldots\ n. \tag{2}$$

Modified L_2 discrepancy (ML$_2$) is a computationally cheaper variant of a discrepancy measure, which is widely used to assess precision for multivariate quadrature rules [9]. Here, the designs are normalised in each dimension to the interval [0,1] and then, the value of ML$_2$ is computed according to

$$E^{ML_2} = \left(\frac{4}{3}\right)^k - \frac{2^{(1-k)}}{n}\sum_{d=1}^{n}\prod_{i=1}^{k}(3 - x_{di}^2) + \frac{1}{n^2}\sum_{d=1}^{n}\sum_{j=1}^{n}\prod_{i=1}^{k}[2 - \max(x_{di}, x_{ji})], \tag{3}$$

where k is the number of input parameters, i.e. the dimension of the design space and x_{di} and x_{ji} are the ith coordinates of the dth and jth points, respectively. Since the evaluation of discrepancy for a large design can be time-consuming, some efficient algorithms are proposed e.g. in [10]. To achieve the best space-filling property of DoE, the value of ML$_2$ should be minimised.

D-optimality criterion (**Dopt**) was proposed by Kirsten Smith in 1918 [11] as a pioneering work in the field of DoE for regression analysis. This criterion minimises the variance associated with estimates of regression model coefficients by minimizing the determinant of the so called dispersion matrix $(\mathbf{Z}^T\mathbf{Z})^{-1}$ or equivalently, by maximising the determinant of the so called information matrix $(\mathbf{Z}^T\mathbf{Z})$ [12]. Again, in order to apply a minimisation procedure, but to avoid the inversion of the information matrix, we can minimise negative value of the determinant of the information matrix, i.e.

$$E^{Dopt} = -\det(\mathbf{Z}^T\mathbf{Z}), \tag{4}$$

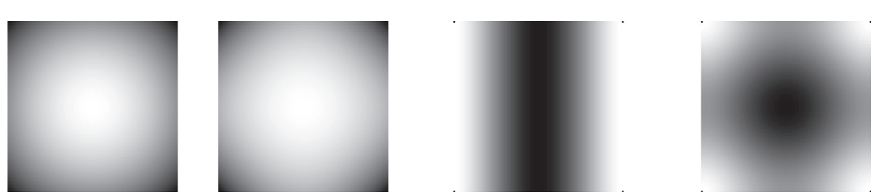
where $\mathbf{Z}$ is a matrix with evaluated regression terms in the design points. In the case of second order polynomial regression and two-dimensional design space, the matrix becomes

$$\mathbf{Z} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{bmatrix}. \tag{5}$$

It is known that under certain conditions, D-optimality criterion leads to the designs with duplicated points. For illustration, having five points in two-dimensional space, one can construct only linear regression having well defined information matrix, i.e. only three columns in matrix $\mathbf{Z}$. By fixing the position of four points into the corners of the squared domain, we can plot the value of E^{Dopt} as a function of the fifth point's coordinates. Fig. 1a shows that in this situation, the optimal position is located in one of the occupied corners and the optimisation will lead to the DoE with duplicates. To overcome this problem, different approaches have been developed. For instance, the authors in [13] start with a larger Latin Hypercube (LH) design optimised w.r.t. EMM criterion. Then, the final smaller DoE is selected as a combination of points from the previously obtained LH design. In this way, no duplicates are presented, however the optimisation w.r.t. D-optimal criterion is quite perfunctory. Another approach based on a Bayesian modification of an information matrix is proposed in [14]. The idea is to add higher order terms in the response surface approximation and subsequently to add corresponding columns into the matrix $\mathbf{Z}$. Then, the information matrix $(\mathbf{Z}^T\mathbf{Z})$ becomes singular, but according to [15], some constant $\tau\in]0,1]$ can be added to diagonal elements of $(\mathbf{Z}^T\mathbf{Z})$ corresponding to added columns to overcome this problem. The constant τ is a parameter of the proposed procedure defining the influence of the Bayesian modification, smaller values imply smaller influence. It is not obvious which terms should be added into the matrix $\mathbf{Z}$, for instance, Fig. 1b shows that one term needs not be sufficient, and thus this question should be the subject to additional research. One should keep in mind to add terms equally to all coordinates so as to preserve the isotropy of the resulting DoE, see Fig. 1c for an example of anisotropic criterion. In the case of five points in two-dimensional domain, two quadratic terms are sufficient to define DoE without the duplicates, see Fig. 1d. The D-optimal designs presented further in this paper are obtained by the described manual Bayesian extension of the information matrix.

2.2. Orthogonality-based criteria

There are two well-know approaches to evaluate the orthogonality of a DoE. The most popular one is based on correlation



$[\,1\ \ x_{i1}\ \ x_{i2}\,]$ $[\,1\ \ x_{i1}\ \ x_{i2}\ \ x_{i1}x_{i2}\,]$ $[\,1\ \ x_{i1}\ \ x_{i2}\ \ x_{i1}^2\,]$ $[\,1\ \ x_{i1}\ \ x_{i2}\ \ x_{i1}^2\ \ x_{i2}^2\,]$

(a) (b) (c) (d)

Fig. 1. Value of D-optimality criterion as a function of fifth point's coordinates for different structure of matrix $\mathbf{Z}$; optimal value is represented by black colour.