23rd International Conference
ENGINEERING MECHANICS 2017
Svratka, Czech Republic, 15 – 18 May 2017

SHIFTING OF LHS DESIGN FOR SURROGATE MODELING

E. Myšáková*, M. Lepš**

Abstract: *Surrogate modeling (Meta-modeling) is a commonly used approach for analysis of complex systems' behavior. Time and computing demands of analytical models describing such systems are usually very high and in cases of need of multiple evaluations (for example in Monte Carlo based reliability analysis) they cannot be used. Instead, a model of the original model called surrogate model can be used. The purpose of the surrogate model is to approximate an original model's response in an arbitrary point of the design domain while constructed on a very limited and thus computationally cheap training data. The training data consist of the Design of Experiments (DoE) and corresponding responses of the original model. The choice of the DoE is crucial for the quality of the surrogate's approximation and therefore the LHS design is often used for its convenient properties. The contribution proposes a procedure of shifting of a part of the design of experiments in cases where the area of interest is located after some original model's evaluations were performed. The goal is clear: to use the already computed training data while not deteriorate the quality of the DoE.*

Keywords: Surrogate modeling, Design of Experiments, Latin Hypercube Sampling, Space-filling.

1. Introduction

Analyses of complex systems constitute a common daily content of scientists and engineers worldwide. Systems' behavior is usually described by time-consuming and computationally demanding models, often Finite Element based. If the actual analysis is based on multiple evaluations of the model with different combinations of input parameters' values the original demanding model cannot be used. Such case is for example the Monte Carlo based reliability assessment (Metropolis, 1949), where hundreds of thousands evaluations can be needed, especially in cases of low failure probabilities (Bucher, 1988).

Possible solution is a utilization of the Surrogate model (or Meta-model) which can be described as a model of the original model (Forrester, 2008). Constructed on the computationally manageable training data it is able to provide an approximation of the original model's response anywhere within the design domain. The training data consist of **i)** the Design of Experiments (DoE) which gives positions of the support points and **ii)** responses of the original model in these support points. Note that the individual coordinates of the support points correspond to the values of individual input parameters (with some transformation and scaling).

The quality of the Design of Experiments is crucial for the meta-model's ability of good approximation. Therefore it should be chosen with consideration. An often used approach is to use an LHS design whose properties are discussed later in Chapter 2.

In our recent problem the area of interest (the failure region) is not known in advance. Therefore at first the initial surrogate model is constructed covering the area of $\mu \pm 3\sigma$, where μ is mean and σ is standard deviation of the original distribution. The optimized LHS design is used providing the positions of training points. Then the surrogate is used for search of the probable position of the Design Point (DP). Shortly, the Design Point marks the area of interest - the failure region. It is convenient to sample from this area where much more samples hit the failure region and then using Importance Sampling (IS)

* Ing. Eva Myšáková: Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29, Prague, CZ, eva.mysakova@fsv.cvut.cz

** Assoc. Prof. Ing. Matěj Lepš, PhD.: Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 166 29, Prague, CZ, leps@cml.fsv.cvut.cz

compute the probability of failure (Chapter 3). For new sampling the second surrogate model is needed constructed on the training points lying in the identified region as shown in Fig. 1.

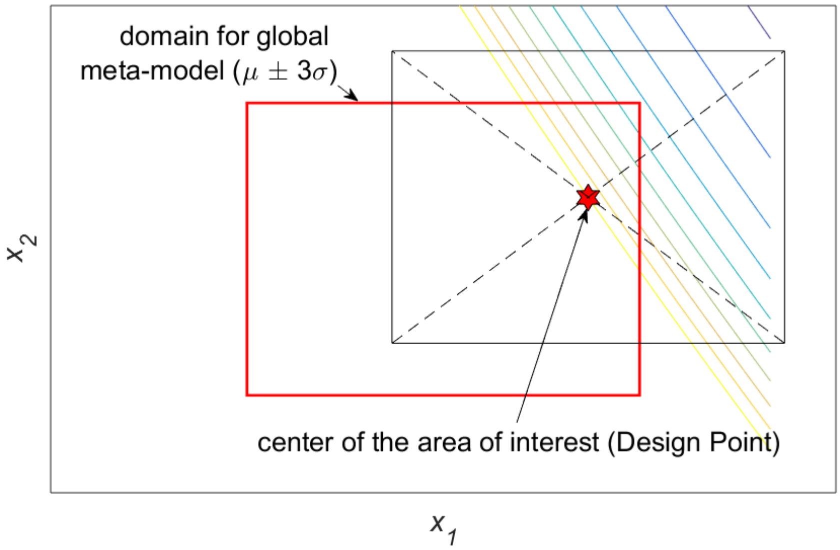

Fig. 1: Two domains for surrogates' construction.

The procedure proposed in this contribution focuses on the LHS design used for the construction of the surrogate models. Once the original model's responses are computed it is not efficient to throw them away if the possibility of its utilization still exists which is our case as described in more detail in Chapter 4.

2. LHS design

Latin Hypercube Sampling (LHS) is a method for generation of samples using stratification of individual variables (McKay, 1979). Needed *np* sample points, in LHS each variable is divided into *np* intervals; the LHS rule than requires that in each interval of each variable lies exactly one point.

Two variants of the method are used differing in the position of the points within the interval. MATLAB uses marking **smooth off** (the points lie in the middle of interval) and **smooth on** (the points lie randomly within the interval).

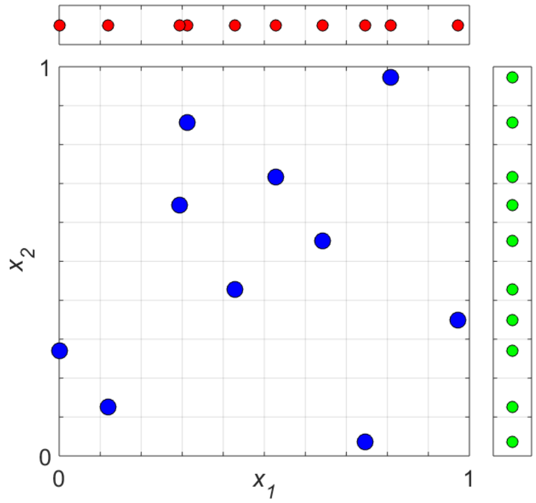

Fig. 2: LHS design in 2D with its reduction to individual axes.

The convenient and often desired attribute of the LHS design are the perfect projection properties. It is guaranteed by the LHS restriction. In case of dimension reduction induced by sensitivity analysis there is no problem to cut off the insignificant variables without getting duplicities, see Fig. 2.

3. Importance Sampling

Importance Sampling (IS) is a method for sampling in cases where the standard Monte Carlo integration cannot be used because it is not possible to sample from the target distribution of the variables (Rubinstein, 1981). It can be case of reliability assessment with very small expected probability of failure where the failure region is located in tail of variables' distribution. With standard Monte Carlo an enormous number of samples would be needed to at least hit the failure region. In IS the proposal distribution better covering the area of interest is used and the targeted integral is re-weighted using importance weights.

4. Design shifting

Because the surrogate model is constructed twice two sets of training data are demanded. But if the domains are overlapping, there is no need to construct two individual Design of Experiments. In fact it is desired to use as many of already evaluated training points as possible. But at the same time also the DoE for the second surrogate model should be as good as possible. Therefore the following procedure is proposed which ensures the utilization of the part of already evaluated training points while the properties of the LHS design remain preserved.

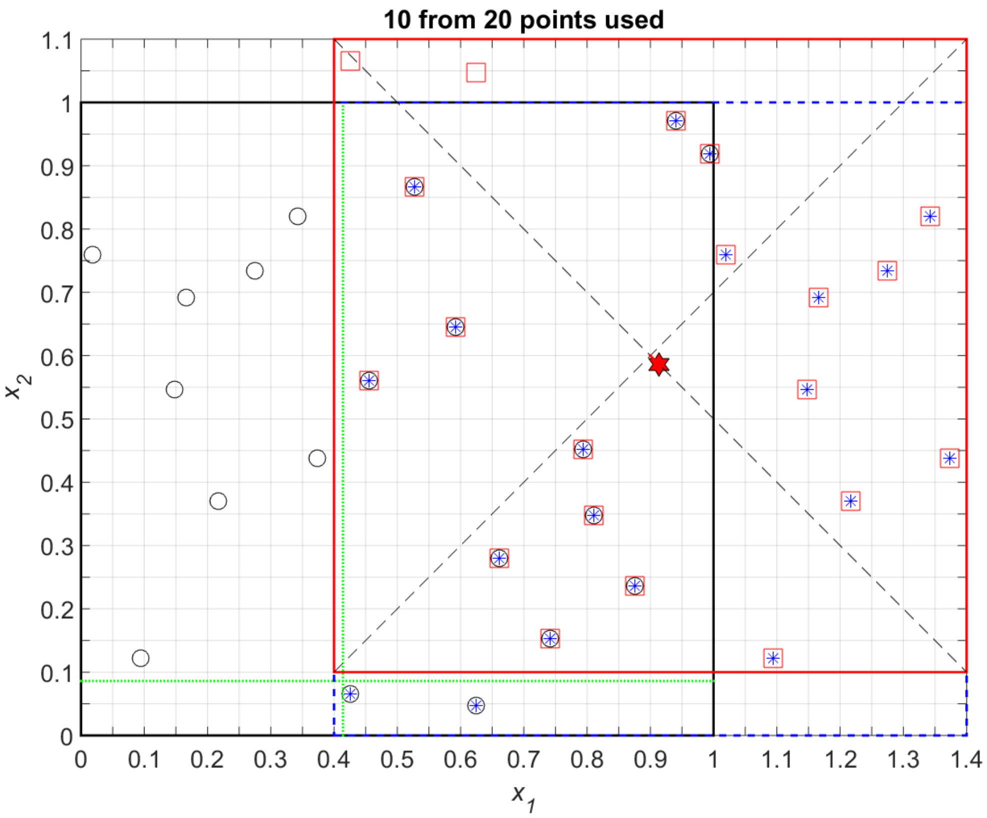

Fig. 3: Illustration of the proposed method.

An illustration of the proposed procedure is brought in Fig. 3. A legend for the figure is shown below:

○	**original LHS design**
✶	Design Point (DP)
·········	limits based on DP's position
✳	LHS design shifted according to x1
□	LHS design shifted according to x2 → **new LHS design**
— — —	new LHS design's diagonal
——	difference between DP's position and new LHS design's center

During the initial sampling position of the Design Point (DP) is found. It is the center of the identified area of interest. Then the lower limit for each axis is computed. Because it is intended to preserve the LHS properties of the design the closest interval bounds are found. The procedure then loops over each dimension and finds points lying outside the new bounds. These points are shifted which means that one is added or subtracted from its coordinate in actual dimension. Note that the procedure is run in normalized $[0, 1]^{dim}$-space, where *dim* is a number of problem's dimension, which is appropriate for the surrogates' construction.

5. Conclusions

The proposed procedure enables the utilization of already computed training data which is convenient because of the computational demands of every single evaluation. At the same time the properties of the design given by LHS restrictions remains in effect which ensures perfect projective properties and space-filling even for the second set of training data.

Of course the number of training points which can be used also for the second surrogate model is given by **i)** the position of the Design Point and **ii)** the actual LHS design to be shifted. If the Design Point lies the bounds of original training data none of them can be used again and the DoE for the second surrogate model has to be generated and evaluated by the original model. Also the actual appearance of the LHS design is important for the possibility of usage of the already evaluated points. In an extreme case none of the original training points can lie in the overlapping area of the domains.

Acknowledgement

Financial support from the Czech Science Foundation, Project No. 16-11473Y, and the Grant Agency of the Czech Technical University in Prague, Project No. SGS17/042/OHK1/1T/11, is gratefully acknowledged.

References

Bucher, C.G. (1988) Adaptive sampling - an iterative fast Monte Carlo procedure. Structural Safety, 5, 2, pp. 119-126, ISSN 0167-4730.

Forrester, A., Sóbester, A. and Keane, A. (2008) Engineering Design via Surrogate Modelling: A Practical Guide. John Wiley & Sons Ltd, ISBN 978-0-470-06068-1.

McKay, M.D., Beckman, R.J. and Conover, W.J. (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, 21, 2, pp. 239-245.

Metropolis, N. and Ulam, S. (1949) The Monte Carlo Method. Journal of the American Statistical Association, 44, 247, pp. 335-341.

Rubinstein, B.Y. (1981) Simulation and the Monte Carlo Method. New York: Wiley & Sons.