Computer Assisted Mechanics and Engineering Sciences, 14: 127–140, 2007.

Efficient sampling techniques for stochastic simulation of structural systems

Rafał Stocki, Piotr Tauzowski, Michał Kleiber
Institute of Fundamental Technological Research, Polish Academy of Sciences
Świętokrzyska 21, 00-049 Warsaw, Poland

(Received August 8, 2006)

One of the main obstacles in making stochastic simulation a standard design tool is its high computational cost. However, this problem can be significantly reduced by using efficient sampling techniques like optimal Latin hypercube (OLH) sampling. The paper advocates this kind of approach for scatter analysis of structural responses. After explaining the idea of the OLH sampling the principal component analysis method (PCA) is briefly described. Next, on numerical examples it is shown how this technique of statistical postprocessing of simulation results can be used in the design process. Important improvements of the estimation quality offered by OLH design of experiments are illustrated on two numerical examples, one simple truss problem and one involving finite element analysis of elastic plate. Based on numerical experiments an attempt is made to propose the sample size which for a given number of random variables provides an acceptable estimation accuracy of statistical moments of system responses and which enables more advanced statistical post-processing.

1. Introduction

It seems to be a commonly shared opinion nowadays that in the design process it is very important to properly account for parameter uncertainties in the model of underlying physical phenomenon. For many engineering problems deterministic approach to analysis can lead to non-robust and potentially unsafe designs. Natural goals of modern industry involving virtual prototyping, reduction of physical testing, faster design and shorter times to market cannot be achieved while resorting to forcedly perfect, deterministic models.

However, taking into account a scatter of model parameters usually leads to much more computationally expensive analysis which is still one of main obstacles in making the stochastic simulations a routine design practice. An accurate estimate of statistics of a system response usually requires hundreds or even thousands of runs of structural analysis software. Despite a constant progress in the performance of modern computers the cost of stochastic analysis is often too high to be accepted. Hence, the main motivation of the authors of this paper was to propose efficient sampling methods enabling sample size reduction.

Latin hypercube designs of experiments and the so-called optimal Latin hypercubes (OLH) provide good alternative to classical crude Monte Carlo sampling. They are known for their good estimation quality. However, OLHs are rather expensive to employ for high dimensional problems. It is important to investigate how many sample points are needed in order to provide sufficient estimation accuracy. Some suggestions concerning the number of crude Monte Carlo sample points as a function of the number of random parameters were given in [4]. Here, we would like to consider the problem of the choice of sample size for OLH sampling.

Usually, when performing stochastic simulations one is interested in computing mean values of system responses and/or correlations between input parameters and responses. However, very often a more advanced statistical postprocessing of simulation results can help in better understanding of structural behavior. The additional information gained by stochastic analysis allows for subsequent

design modifications rendering the structure more robust and reliable. For example, principal component analysis (PCA) allows to identify the random system parameters that contribute most to the scatter of results. It also gives a possibility to reduce the stochastic model and concentrate on a subset of responses which are representative for the overall system's variability.

Efficiency of the OLH sampling will be illustrated on the examples of some benchmark discussion as well as more complex finite element analysis problems. We will also show that efficient sampling techniques are particularly important for the quality of the PCA analysis.

2. Optimal Latin hypercube sampling

In stochastic simulations involving computationally expensive problems it is particularly important to select a sample generation technique which provides sufficiently accurate estimation of response statistics with the number of sample points which is not excessive. Unfortunately, the classical 'crude' Monte Carlo (MC) technique is far too costly for many problems of practical importance.

A sample design method known for its very good estimation properties is the so-called optimal Latin hypercube(OLH) design. However, contrary to MC, it requires a special optimization algorithm in order to find the "optimal" layout of points. In the current work, depending on the size of the problem (the sample size and the number of variables), two algorithms were used to generate OLH designs: the columnwise-pairwise (CP) algorithm put forward by Park [8] (with the modification described in [12]) and the genetic algorithm, inspired by the algorithm proposed in [10]. The algorithms are explained in detail in [5]. Below, only the most important information on the OLH design is given.

OLH is a viable sampling technique when one considers statistical optimality and projection properties. There are many criteria of the statistical optimality of a design of experiments. Most of them are based on fitting a (stochastic) model to experiments or computed data, see [6] and [8]. Another criterion, which is of interest in the current paper, measures how well the statistical properties of some model are predicted. By good projection properties we mean here that the sample points are well spread out when projected onto a subspace spanned by a number of coordinate axes. Once generated, an OLH for p variables and N points is independent of the considered application. It is stored in a matrix and does not need to be computed again.

A Latin hypercube is usually represented by a $N \times p$ (N rows and p columns) matrix $\mathbf{L}$ in which each column consists of a permutation of the integer numbers 1 to N. We will refer to each row of $\mathbf{L}$ as a sample point in p dimensions and use the notation

$$\mathbf{L} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}, \tag{1}$$

where $\mathbf{x}_i$, $1 \leq i \leq N$, is the i-th sample point.

The LH design obtained by simply generating p random permutations of the numbers 1 to N and placing them as columns in the matrix, without any subsequent changes, will be referred later in the text as the random Latin hypercube (RLH).

The matrix $\mathbf{L}$ can be subsequently used in generating samples of the random vector $\mathbf{X}$ taking into account the actual distribution of each variable. In the case of uncorrelated random variables, to find the realization $x_k(m)$ of the random variable X_k, $1 \leq k \leq p$, corresponding to the number m in the k-th column of the matrix $\mathbf{L}$, $1 \leq m \leq N$, the cumulative distribution function (CDF) of X_k is employed

$$x_k(m) = F^{-1}(\tilde{x}_m), \tag{2}$$

where

$$\tilde{x}_m = \frac{m}{N} - \frac{1}{2N}. \tag{3}$$

In other words, the range of variability of each random variable is divided into N intervals of equal probability and the values $x_k(i)$, $i = 1, \ldots, N$ correspond to probabilistic midpoints (medians) of X_k in these intervals.

In general the random variables can be arbitrarily distributed and correlated. However, to use the sample design generated with LH the variables must be first numerically transformed to a set of uncorrelated random variables. Depending on the available statistical information the Rosenblatt transformation [9] or the Nataf transformation [7] can be used. Both transform the original variables to the space of independent standardized Gaussian variables. Values of the random variables in the transformed space found from Eq. (2) are next transformed back to the original random variables $\mathbf{X}$.

The criterion that is used in the current paper to optimize LH design was proposed by Audze and Eglais in [1]. It is based on the function G which, in a physical analogy, is the sum of the norms of the repulsive forces if the samples are considered as electrically charged particles

$$G(\mathbf{L}) := \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{\|\mathbf{x}_i - \mathbf{x}_j\|^2}, \tag{4}$$

where $\|\cdot\|$ is the Euclidean norm. Using the function G the criterion that allows to compare two LH designs can be stated as

$$\mathbf{L}_1 \text{ is better than } \mathbf{L}_2 \text{ if } G(\mathbf{L}_1) < G(\mathbf{L}_2). \tag{5}$$

As it was shown in [5], this criterion is a reasonable compromise between good statistical properties and efficiency.

Although for small Latin hypercubes the computational cost of finding OLH is negligible compared to the expensive computer simulation of a given physical phenomenon, it grows very fast with the sample size and the number of variables. For large LH (hundreds of sample points and tens of variables) it may even take days of fast computers' CPU. The computational cost depends of course on the algorithms used for OLH optimization and the adopted optimality criterion. As already mentioned, for the purpose of this work, the CP algorithm and the genetic algorithm are used to generate OLHs.

3. Principal component analysis

When performing stochastic simulation one is usually interested in obtaining a statistical information concerning various performance measures of the system under study. Most often these are statistical moments and/or intervals of confidence of system responses, histograms and scatter plots. However, sometimes a more advanced statistical analysis is required to better identify the sources of scatter of results. Such investigation is offered by principal component analysis (PCA), see e.g. [4], and in the numerical examples to be presented in next sections we will also examine the usefulness of various sampling techniques with respect to the quality of PCA results.

After the stochastic simulation is performed the main problem to deal with is the large amount of available data. For investigations involving many descriptors (responses and/or variables), it is often useful to simplify the analysis by considering a smaller number of linear combinations of the original descriptors. By doing this we want to summarize, in a few dimensions, most of the variability of the covariance matrix of a large number of descriptors. Because of the diverse physical nature of the descriptors considered, it is convenient to use in PCA the correlation matrix $\mathbf{R}$ which is the covariance matrix of standardized descriptors. The principal component analysis is based on the solution of the following eigenproblem,

$$(\mathbf{R} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}, \qquad i = 1, \ldots, M, \tag{6}$$

where M is the number of descriptors while λ_i and $\mathbf{v}_i$, are the eigenvalues and normalized eigenvectors, respectively. We assume the eigenvalues (real and positive) are ordered according to their

values starting with the biggest. The line through the origin directed along $\mathbf{v}_1$ is called the first principal axis. The corresponding line directed along $\mathbf{v}_2$ is called the second principal axis, etc. Principal components give the positions of the M-dimensional data points (inputs + selected performances) with respect to this new system of principal axes. The matrix $\mathbf{V}$, the columns of which are the normalized eigenvectors, allows to compute the values of principal components for the data points by means of the following transformation:

$$\mathbf{F} = \begin{bmatrix} \frac{x_{11}-\bar{x}_1}{\sigma_{x_1}} & \frac{x_{12}-\bar{x}_2}{\sigma_{x_2}} & \cdots & \frac{x_{1M}-\bar{x}_M}{\sigma_{x_M}} \\ \frac{x_{21}-\bar{x}_1}{\sigma_{x_1}} & \frac{x_{22}-\bar{x}_2}{\sigma_{x_2}} & \cdots & \frac{x_{2M}-\bar{x}_M}{\sigma_{x_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{N1}-\bar{x}_1}{\sigma_{x_1}} & \frac{x_{N2}-\bar{x}_2}{\sigma_{x_2}} & \cdots & \frac{x_{NM}-\bar{x}_M}{\sigma_{x_M}} \end{bmatrix} \mathbf{V}, \tag{7}$$

where N is the number of data points, $\mathbf{F}$ is the $N \times M$ matrix containing the coordinates of data points in the space of principal components and $\bar{x}_i$ and σ_{x_i} are the mean values and standard deviations of the descriptors, respectively. It can be shown that the correlation between the i-th descriptor and the j-th principal component is given by the formula

$$\rho_{ij} = u_{ij}\sqrt{\lambda_j}, \tag{8}$$

where λ_j is the eigenvalue referring to the j-th principal component and u_{ij} is the i-th element of the corresponding eigenvector.

If there are few dominating eigenvalues λ_i, then a strong correlation between some input variables and the corresponding principal components is a good indicator of the influence a given variable has on the overall system's variability.

4. Example 1: Geometrically nonlinear truss

The model problem taken from [2] is illustrated in Fig. 1. The non-linear behaviour of the truss results from its kinematics, the constitutive relation for the bar and the spring are taken to be linear elastic. The unknown of the problem is w, the vertical displacement of the right node of the bar. The notation employed is: E – Young modulus of the bar material, A – cross-sectional area of the bar, l – length of the unloaded bar, $z > 0$ – vertical coordinate of the right node of the bar when it is unloaded, K_s – stiffness of the spring, S – vertical force applied to the right node of the bar.

The displacement w (deflection) can be determined from the following equation (cf. [2]):

$$S = \frac{EA}{l^3}\left(z^2 w + \frac{3}{2}zw^2 + \frac{1}{2}w^3\right) + K_s w. \tag{9}$$

In the derivation of this equation, it is assumed that the angle θ (see Fig. 1) is small which implies $z, w \ll l$. Equation (9) is a third degree polynomial equation for w. For some values of the parameters there is one (unique) real root, for other values of the parameters there are three roots. In the case of three roots we choose the one corresponding to the smallest magnitude of displacement. This situation occurs when the applied force is smaller than the critical force that cause the bar to snap-through to the other equilibrium position. The case with one root of Eq. (9) corresponds to the state of the bar after snap-through or when the stiffness of the spring is so big that it prevents this kind of instability.

Let us now consider the stochastic description of the problem. We assume that the input random variables are uncorrelated and their corresponding probability density functions with the respective parameters (mean and coefficient of variation (c.o.v.) for normal distribution and bounds for uniform distribution) are given in Table 1.

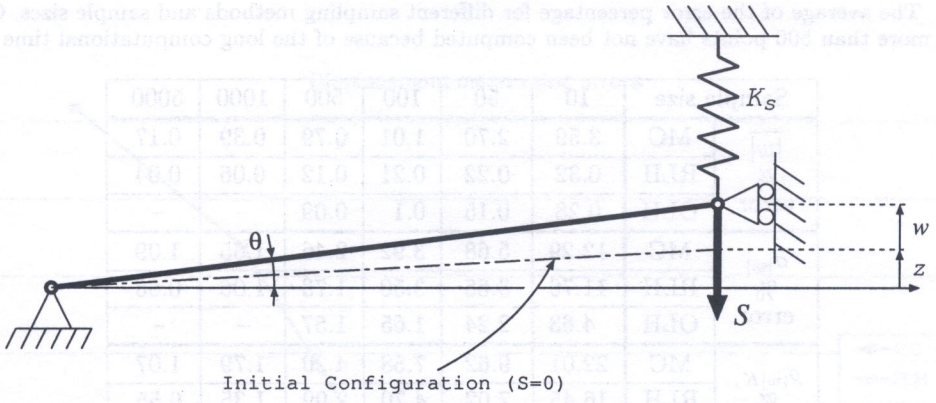

Fig. 1. The single bar structure prone to 'snap-through' type instability. The force S shown on the picture has negative sign

Table 1. Nonlinear truss example: random variables

Variable	Distribution type	Mean/Left bound	c.o.v./Right bound
E	normal	$5 \cdot 10^5\,\mathrm{N/mm^2}$	0.02
A	normal	$100\,\mathrm{mm^2}$	0.02
l	normal	2500 mm	0.03
z	uniform	15 mm	27 mm
K_s	normal	0.9 N/mm	0.2
S	uniform	−40 N	−30 N

In order to evaluate the estimation accuracy of various sampling techniques reference results must be provided first. For this purpose, the crude Monte Carlo simulation with 50 000 sample points was performed allowing to compute estimates of the following statistics:

- mean value of the deflection magnitude

$$\overline{|w|} = 40.49\,\mathrm{mm} \tag{10}$$

- standard deviation of the deflection magnitude

$$\sigma(|w|) = 5.89\,\mathrm{mm} \tag{11}$$

- correlation matrix

	E	A	l	z	K_s	$\lvert S\rvert$	$\lvert w\rvert$
E	1.0	0.0	0.0	0.0	0.0	0.0	0.0
A	0.0	1.0	0.0	0.0	0.0	0.0	0.0
l	0.0	0.0	1.0	0.0	0.0	0.0	−0.01
z	0.0	0.0	0.0	1.0	0.0	0.0	0.67
K_s	0.0	0.0	0.0	0.0	1.0	0.0	−0.65
$\lvert S\rvert$	0.0	0.0	0.0	0.0	0.0	1.0	0.26
$\lvert w\rvert$	0.0	0.0	−0.01	0.67	−0.65	0.26	1.0

(12)

By analyzing values of the correlation coefficients in the matrix (12) we find that due to large sample size there are no spurious correlations between independent input random variables. Moreover, coefficient values in the last row/column indicate that the dependent variable, i.e. deflection magnitude, is significantly positively correlated with the vertical coordinate z and negatively correlated

Table 2. The average of the error percentage for different sampling methods and sample sizes. OLH with more than 500 points have not been computed because of the long computational time

Sample size		10	50	100	500	1000	5000
$\overline{\lvert w\rvert}$ % error	MC	3.59	2.70	1.01	0.79	0.39	0.17
	RLH	0.32	0.22	0.21	0.12	0.06	0.04
	OLH	0.28	0.15	0.1	0.09	–	–
$\sigma_{\lvert w\rvert}$ % error	MC	12.29	5.68	3.92	2.46	1.65	1.09
	RLH	11.76	3.66	3.50	1.73	1.06	0.65
	OLH	4.63	2.24	1.65	1.57	–	–
$\rho_{\lvert w\rvert K_s}$ % error	MC	22.01	9.62	7.58	4.20	1.79	1.07
	RLH	16.45	7.02	4.70	2.99	1.35	0.55
	OLH	7.56	2.16	1.48	0.84	–	–
$\rho_{\lvert w\rvert z}$ % error	MC	33.45	14.75	8.09	4.35	3.61	1.32
	RLH	15.95	9.51	6.58	3.80	1.78	0.89
	OLH	8.07	5.10	4.09	3.70	–	–

with the spring stiffness K_s. On the other hand, the deflection is not influenced by the remaining variables.

To evaluate the different sampling methods we now set out to calculate values of $\overline{|w|}$, $\sigma_{|w|}$ and the two significant correlation coefficients $\rho_{|w|K_s}$ and $\rho_{|w|z}$ with the crude Monte Carlo method (MC), a random Latin hypercube (RLH), i.e. without subsequent optimization and the optimized Latin hypercube (OLH), respectively.

The results of the computations are shown in Table 2, where the average of the error percentage in the estimates of the statistics of interest are shown. To obtain these values K designs of experiments $\{\mathbf{X}^{(m)}\}_{m=1}^{K}$ with the method in question are determined first. The elements of these matrices are denoted by $x_{ij}^{(m)}$. Next, for each statistic s, K estimates $s^{(m)}$, $m = 1, \ldots, K$ are computed. A value given in Table 2 is then the average of the error of these estimates, given as a percentage of the respective reference value s_{ref} (see Eqs. (10)–(12)),

$$\frac{1}{s_{\text{ref}} K} \sum_{m=1}^{K} |s^{(m)} - s_{\text{ref}}|. \tag{13}$$

The same results are given as graphs in Figs. 2–5. By analyzing the obtained results the following observations can be made:

- OLH sampling leads to far more efficient estimations than the crude MC method. Depending on the statistics of interest and the sample size (especially for small samples) it produces many times smaller estimation error than MC. For the mean estimation, 10 points OLH gives almost 13 times smaller error and about 3 times smaller error for other statistics, for instance.

- Comparing OLH-based estimations with those obtained using RLH we find that, depending on the case, OLH results are up to 3 times better (see $\rho_{|w|K_s}$ estimation error) than the results for non-optimized Latin hypercubes. However, in this example, the gain of using OLH compared to RLH is rather minor for the mean value estimation while it is quite important for higher order moments estimates (at least for sample size less than 500).

- The algorithms for finding OLH designs usually have polynomial complexity (in [5] the computational time was estimated as $T \sim N^5$) so it is very expensive to obtain an OLH for hundreds of points. From the example it is evident that there is no substantial advantage of using OLH rather than RLH for $N > 500$ or, more generally, for $N > 10p$.

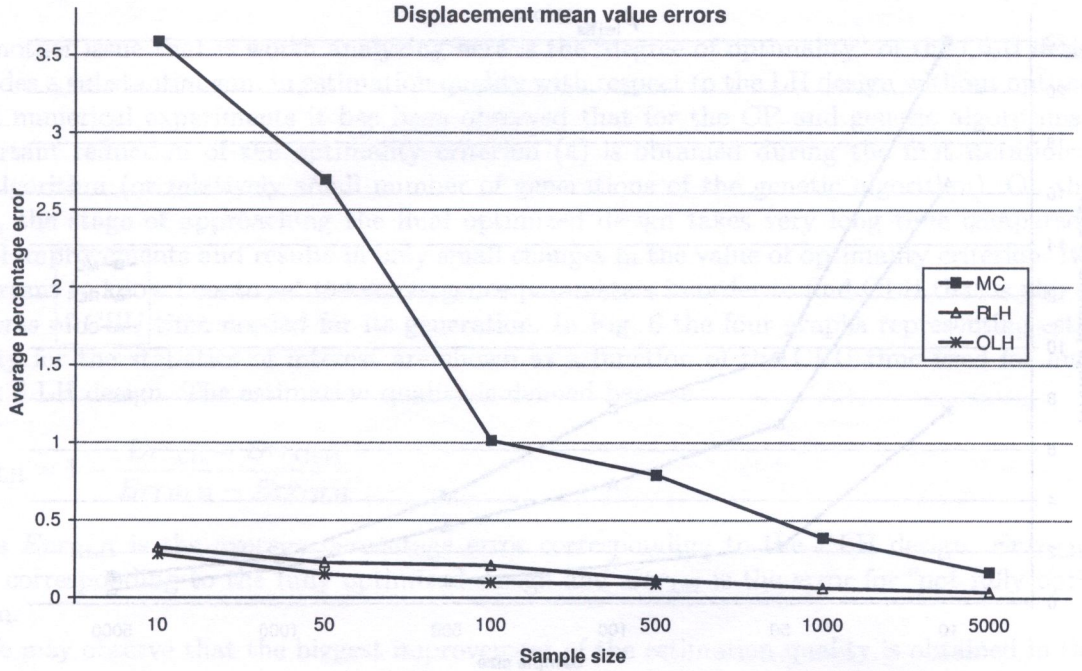

Fig. 2. Estimation of the deflection mean value. The average of the error percentage for different sampling methods and different sample sizes

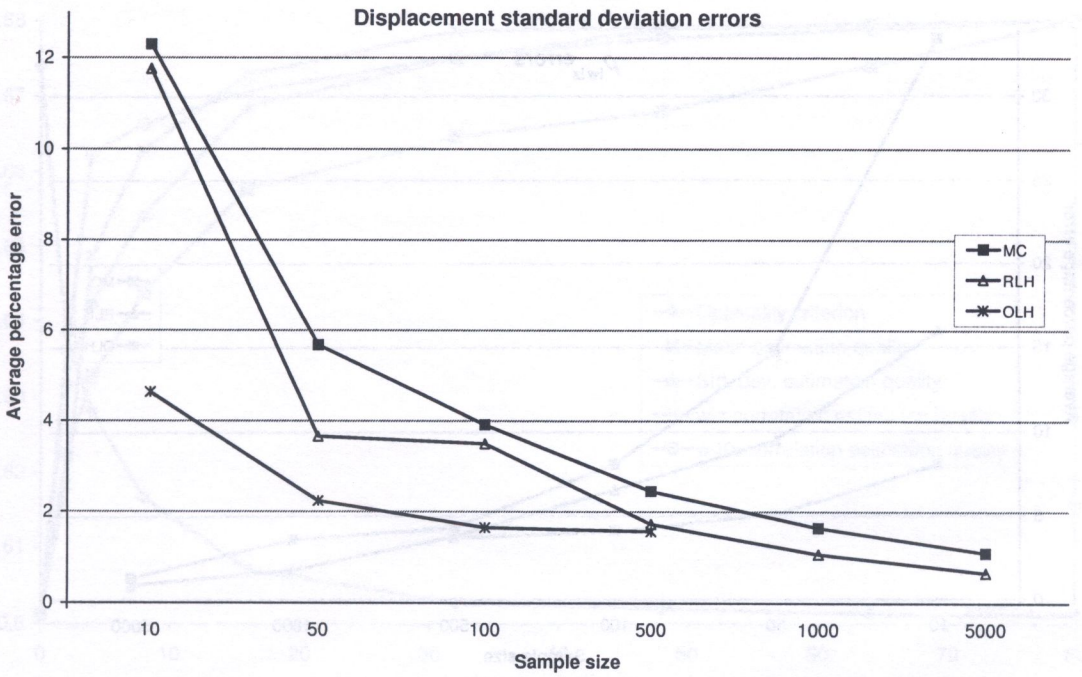

Fig. 3. Estimation of the deflection standard deviation. The average of the error percentage for different sampling methods and different sample sizes

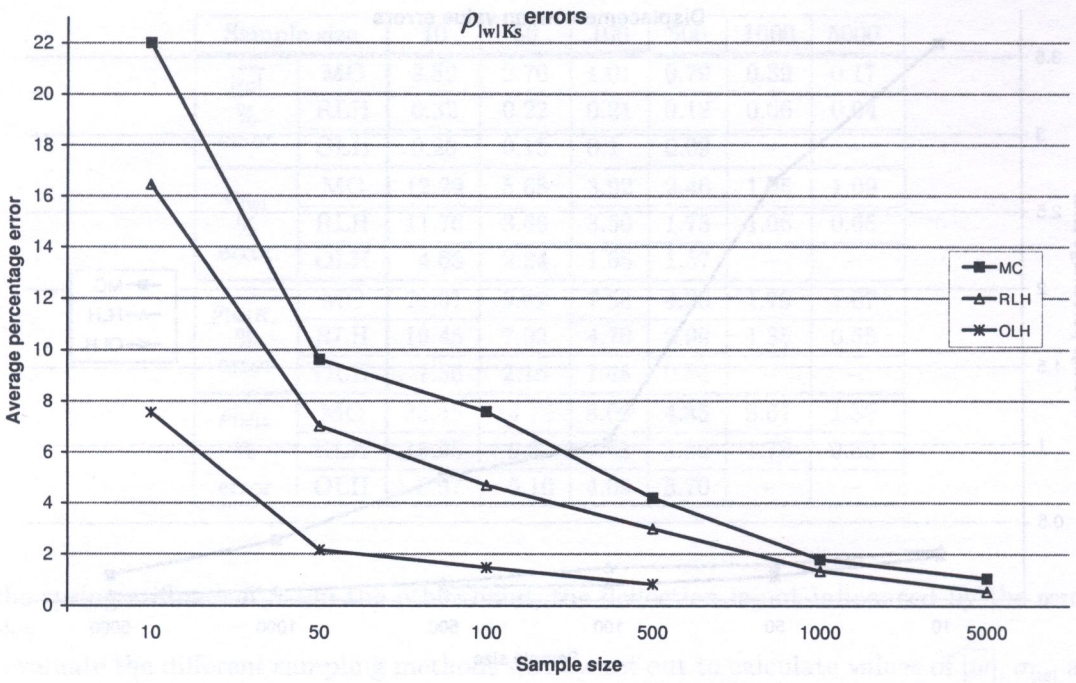

Fig. 4. Estimation of the correlation coefficient $\rho_{|w|K_s}$. The average of the error percentage for different sampling methods and different sample sizes

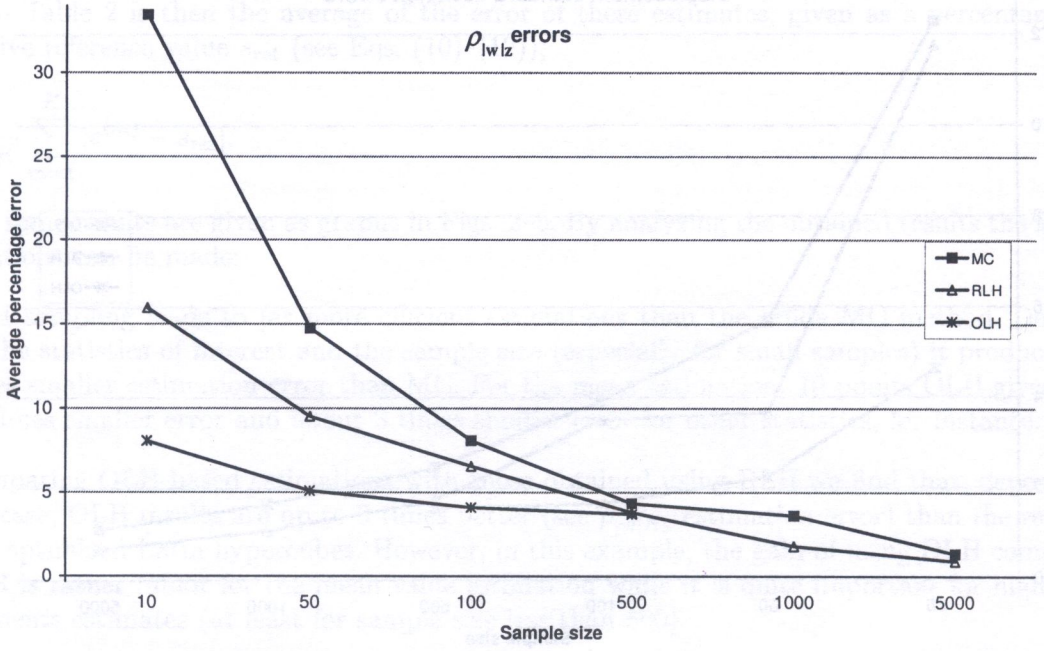

Fig. 5. Estimation of the correlation coefficient $\rho_{|w|z}$. The average of the error percentage for different sampling methods and different sample sizes

- Very good estimates obtained with OLH designs for small size samples make this sampling technique the best choice for stochastic simulation of computationally expensive systems.

Another issue that is worth analyzing here is the "degree of optimality" of the OLH design that provides a substantial gain in estimation quality with respect to the LH design without optimization. From numerical experiments it has been observed that for the CP and genetic algorithms a very important reduction of the optimality criterion (4) is obtained during the first iterations of the CP algorithm (or relatively small number of generations of the genetic algorithm). On the other hand, the stage of approaching the final optimized design takes very long time compared to the initial improvements and results in only small changes in the value of optimality criterion. It is then important to know how to set the convergence parameters in order to find OLH that is also optimal in terms of CPU time needed for its generation. In Fig. 6 the four graphs representing estimation quality for the statistics of interest are shown as a function of the CPU time used for improving 100×6 LH design. The estimation quality is defined here as

$$Q_{\mathrm{LH}} = 1 - \frac{Err_{\mathrm{LH}} - Err_{\mathrm{OLH}}}{Err_{\mathrm{RLH}} - Err_{\mathrm{OLH}}}, \tag{14}$$

where Err_{RLH} is the average percentage error corresponding to the RLH design, Err_{OLH} is the error corresponding to the fully optimized design and Err_{LH} is the error for "not fully optimized" design.

We may observe that the biggest improvement of the estimation quality is obtained in the early phase of the optimization process. Unfortunately, it is difficult to propose a general rule relating estimation quality and the time to find the OLH design. It depends on the estimated statistics and the size of the problem. However, it seems that even a limited number of iterations (say, of the CP type) can significantly ameliorate the quality of estimation.

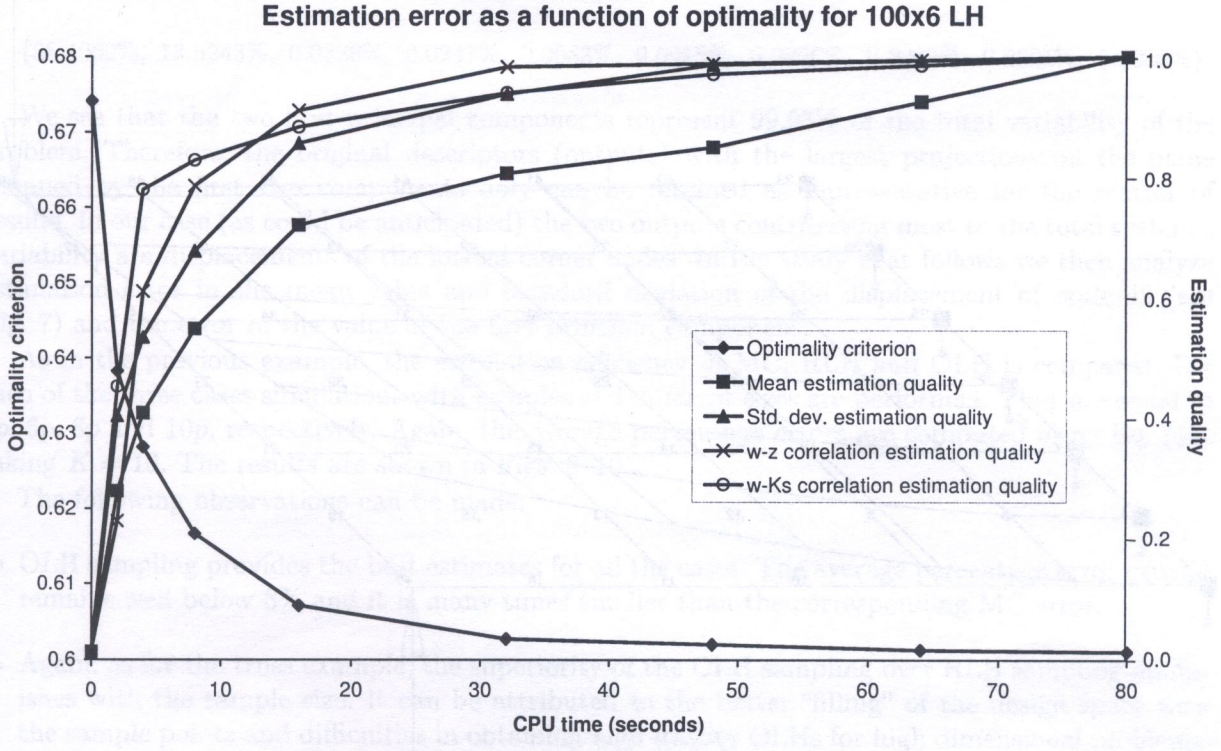

Fig. 6. Changes in LH estimation quality as function of CPU time used to improve the sample's spacial uniformity (to obtain OLH). The curve with solid diamond markers represents the CPU time history of the optimality criterion (4)

5. Example 2: Plate with random properties

The second numerical example was chosen in order to investigate the efficiency of OLH sampling for problems involving many correlated random variables. The problem is a modified version of the elastic plate example presented by Alonso and Collado in [4]. In addition to sampling efficiency issues it is an interesting example for illustrating usefulness of PCA in postprocessing of stochastic simulation results and for stochastic model reduction. The plate geometry and its deformed configuration are shown in Fig. 7. The square ($3\text{m} \times 3\text{m}$) plate is modelled by finite elements. It is clamped along one edge (nodes $1, 4, 7, 22, 25, 36, 39$ in Fig. 7) and loaded with two forces: F_1 applied downwards at the node 17 and F_2 applied upwards at the node 49.

The stochastic description of the problem consists of the following random variables: nodal force F_1 is normally distributed with the mean value $\mu_{F_1} = 80\,\text{kN}$ and the standard deviation $\sigma_{F_1} = 15\,\text{kN}$, nodal force F_2 – normally distributed with $\mu_{F_2} = 60\,\text{kN}$ and $\sigma_{F_2} = 10\,\text{kN}$, Young modulus E – lognormally distributed with $\mu_E = 210000\,\text{MPa}$ and $\sigma_E = 21000\,\text{MPa}$, Poisson ratio ν – lognormally distributed with $\mu_\nu = 0.3$ and $\sigma_\nu = 0.03$. Plate thickness t is modelled by a homogeneous Gaussian random field with mean $\mu_t = 0.06\,\text{m}$, standard deviation $\sigma_t = 0.003\,\text{m}$ and autocorrelation coefficient function $\rho(\mathbf{x}, \mathbf{x}')$ given by

$$\rho(\mathbf{x}, \mathbf{x}') = \exp\left(\frac{\|\mathbf{x} - \mathbf{x}'\|}{a}\right), \tag{15}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are vectors of coordinates of two points in the two-dimensional random field, a is a measure of the rate of fluctuation of the random field, commonly known as the correlation length and $\|\cdot\|$ is the Euclidean norm. In this example it is assumed that $a = 4\,\text{m}$.

For the purpose of the finite element computations the random field has to be discretized. There exist many methods to address this task. A comprehensive review of random field discretization

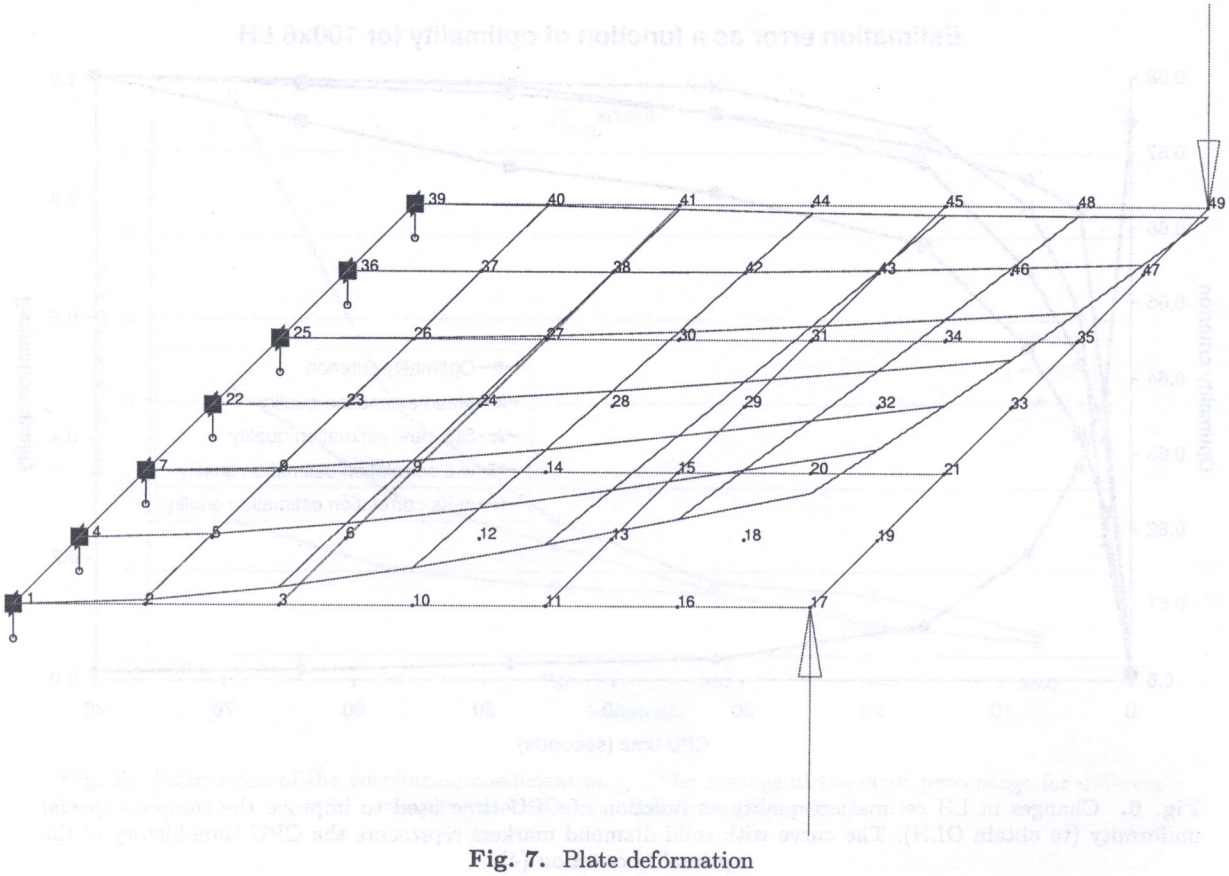

Fig. 7. Plate deformation

techniques can be found in [11], for instance. Among the most popular methods one can mention the midpoint method, the shape function method, the optimal linear estimation method (Kriging) or series expansion methods like, the Karhunen–Loève and orthogonal series expansion. In general, a discretization procedure is the approximation of a continuous random field by means of a finite set of random variables. The main goal in the random field discretization is to define a method providing the "best" approximation with respect to some error estimator based on the minimal number of random variables. However, for the purpose of our example the random field discretization method serves primarily as a "source of random variables". For this reason we decided to adopt the midpoint discretization method (see [3]). The method consists in approximating the random field in each finite element by a single variable defined as the value of the field at the centroid of this element. This is definitely not the most sophisticated technique but it gives us flexibility in generating multidimensional problems by simply changing the mesh density.

We consider three cases of the problem:

a) mesh 3×3, 9 finite elements, 49 nodes, $p = 13$ random variables,

b) mesh 6×6, 36 finite elements, 169 nodes, $p = 40$ random variables,

c) mesh 9×9, 81 finite elements, 361 nodes, $p = 85$ random variables.

9-node thick plate finite elements (Reisner–Mindlin type) are used to model the structure. As structural responses we consider displacements of the free nodes.

First, we shall discuss how to select the outputs that explain most of the variability of our stochastic simulation results. In order to do this the principal component analysis of the reference Monte Carlo simulations ($N = 200\,000$ sample points) for the case a) is performed. The fraction of the total variability of the system explained by each principal component has been computed using the eigenvalues of the covariance matrix. Contribution of the values of first ten principal components to the sum of all components is as follows:

{86.4062%, 13.5243%, 0.0338%, 0.0247%, 0.0083%, 0.0015%, 0.0006%, 0.0003%, 0.0001%, 0.0000%}

We see that the two first principal components represent 99.93% of the total variability of the problem. Therefore, the original descriptors (outputs) with the largest projections on the plane spanned by the first two components only can be retained as representative for the scatter of results. In our case (as could be anticipated) the two outputs contributing most to the total system's variability are displacements of the loaded corner nodes. In the study that follows we then analyze estimation error in the mean value and standard deviation of the displacement of node 49 (see Fig. 7) and the error in the value of the first principal component.

As in the previous example, the estimation efficiency of MC, RLH and OLH is compared. For each of the three cases simulations with samples of 4 different sizes are performed. They are equal to $2p$, $6p$, $8p$ and $10p$, respectively. Again, the average percentage errors are computed using Eq. (13) taking $K = 10$. The results are shown in Figs. 8–10.

The following observations can be made:

- OLH sampling provides the best estimates for all the cases. The average percentage error usually remains well below 3% and it is many times smaller than the corresponding MC error.

- Again, as for the truss example, the superiority of the OLH sampling over RLH sampling diminishes with the sample size. It can be attributed to the better "filling" of the design space with the sample points and difficulties in obtaining high quality OLHs for high dimensional problems. However, contrary to the first example it is difficult to give here a limit of the type $N = 10p$ for the OLH-based sampling to be superior, see large differences in estimation error between OLH and RLH for the case of 13 random variables, for instance.

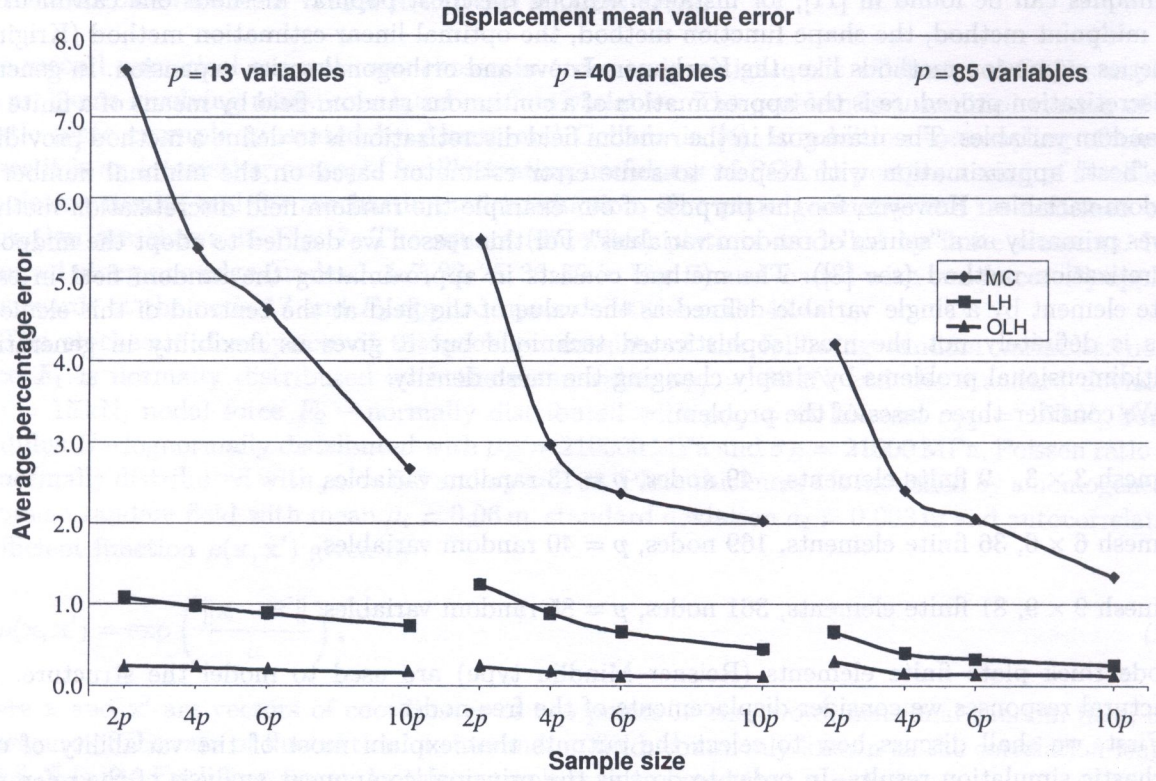

Fig. 8. Estimation of the mean value of the node 49 displacement. The average of the error percentage for different sampling methods, sample sizes and for three different finite element discretizations (implying different numbers of random variables)

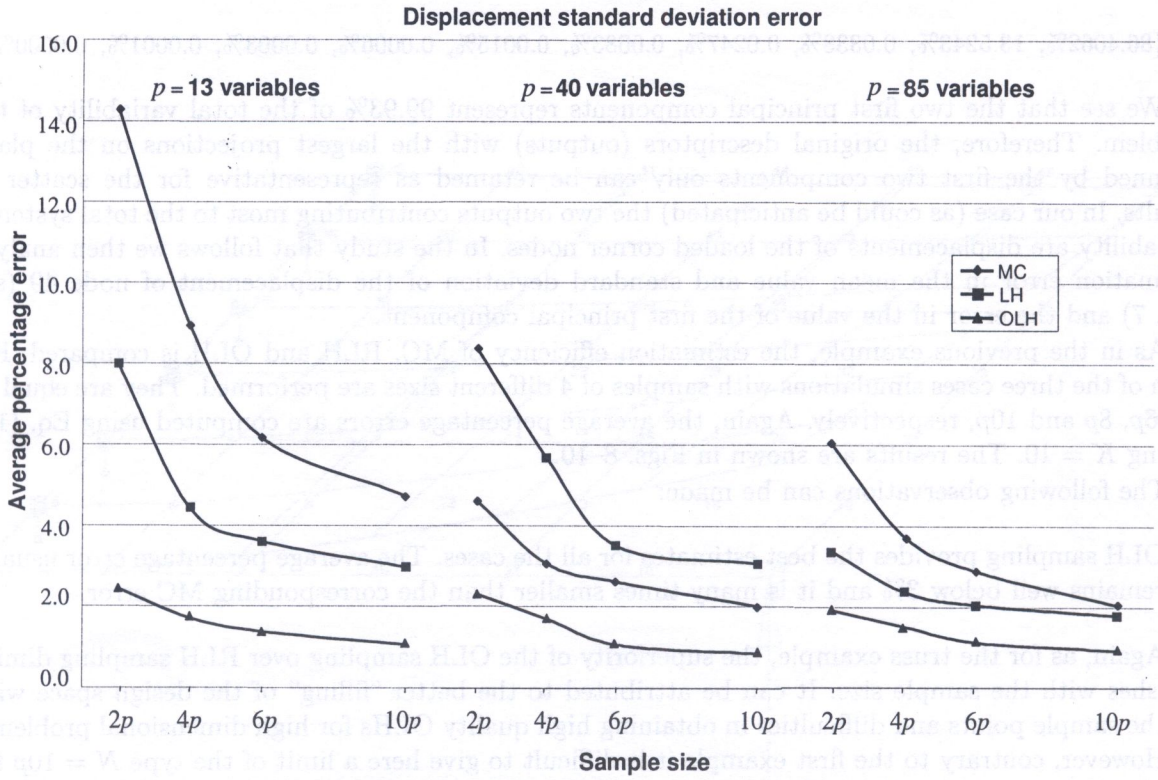

Fig. 9. Estimation of the standard deviation of the node 49 displacement

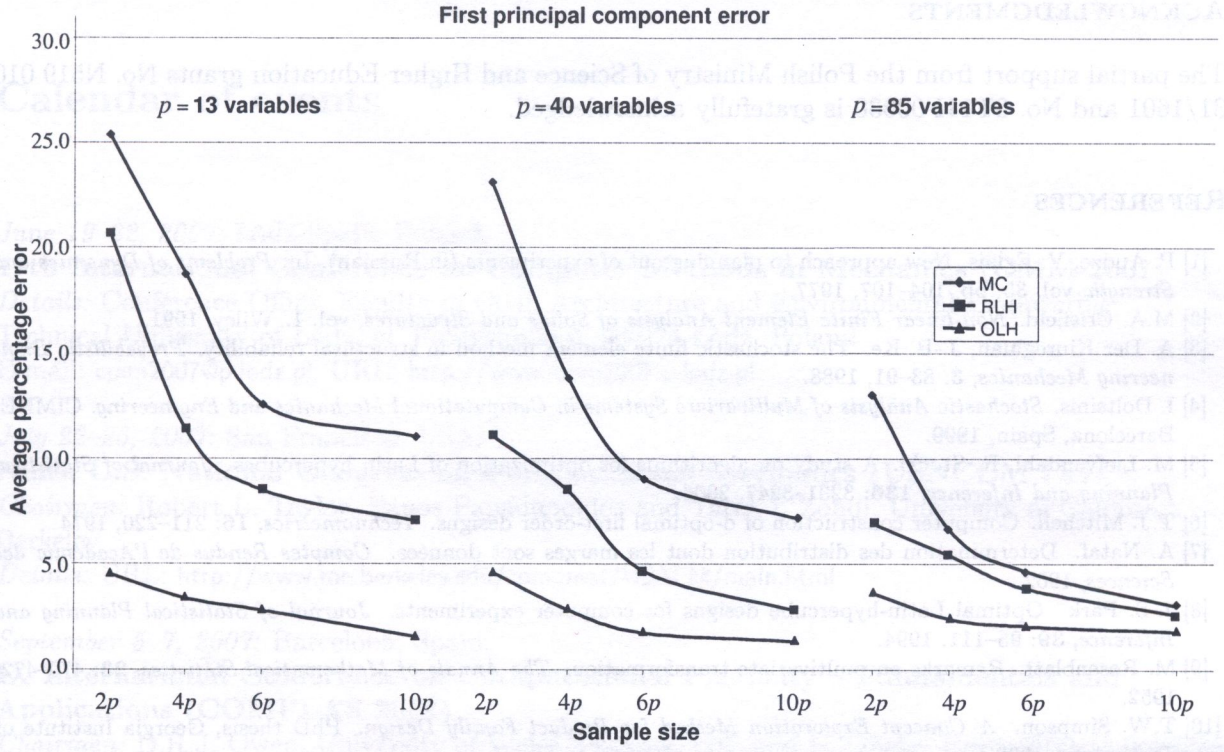

Fig. 10. Estimation of the value of first principal component

- For MC and RLH sampling and for the sample size N being a constant multiple of the number of random variables the estimation error reduces with the dimension of the problem. In other words, for MC and RLH, the error for $N = kp$, where $k \in \{2, 4, 6, 10\}$, is the greatest for $p = 13$ and the smallest for $p = 85$. This is, however, not a general tendency but an effect of adopting the midpoint method for random field discretization. A refined mesh leads to less abrupt changes of the plate thickness between elements and results in smaller displacements variation. Taking this into account, we may now make an interesting observation concerning OLH sampling results: for a given k the estimation errors are almost equal, irrespective of p, i.e. irrespective of the scatter of results.

6. Conclusions

A proper accounting for the scatter of model parameters usually leads to computationally expensive structural analysis which is still one of main obstacles in making the stochastic simulations a routine design practice. Therefore, it is important to use efficient sampling methods that enable reduction of the computational burden of stochastic analysis, especially for complex structures.

Latin hypercube designs of experiments and particularly optimal Latin hypercubes provide good alternative to classical crude Monte Carlo sampling. It has been shown in the two numerical examples involving up to 85 random variables, with or without a correlation structure, that OLH sampling leads to substantial reduction of the estimation error when compared to MC or LH results. It guarantees very good estimation accuracy even for small size samples. From numerical experiments it was observed that OLH sampling with $2p$ sample points, where p is the number of random variables, assures accurate estimates of statistical moments of structural responses as well as enables reliable advanced statistical postprocessing, like, principal component analysis, for instance.

ACKNOWLEDGMENTS

The partial support from the Polish Ministry of Science and Higher Education grants No. N519 010 31/1601 and No. 3T11F00930 is gratefully acknowledged.

REFERENCES

[1] P. Audze, V. Eglais. New approach to planning out of experiments (in Russian). In: *Problems of Dynamics and Strength*, vol. 35, pp. 104–107, 1977.

[2] M.A. Crisfield. *Non-linear Finite Element Analysis of Solids and Structures*, vol. 1. Wiley, 1991.

[3] A. Der Kiureghian, J.-B. Ke. The stochastic finite element method in structural reliability. *Probabilistic Engineering Mechanics*, **3**: 83–91, 1988.

[4] I. Doltsinis. *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*. CIMNE, Barcelona, Spain, 1999.

[5] M. Liefvendahl, R. Stocki. A study on algorithms for optimization of Latin hypercubes. *Journal of Statistical Planning and Inference*, **136**: 3231–3247, 2006.

[6] T.J. Mitchell. Computer construction of d-optimal first-order designs. *Technometrics*, **16**: 211–220, 1974.

[7] A. Nataf. Determination des distribution dont les marges sont données. *Comptes Rendus de l'Académie des Sciences*, 1962.

[8] J.-S. Park. Optimal Latin-hypercube designs for computer experiments. *Journal of Statistical Planning and Inference*, **39**: 95–111, 1994.

[9] M. Rosenblatt. Remarks on multivariate transformation. *The Annals of Mathematical Statistics*, **23**: 470–472, 1952.

[10] T.W. Simpson. *A Concept Exploration Method for Product Family Design*. PhD thesis, Georgia Institute of Technology, 1998.

[11] B. Sudret, A. Der Kiureghian. Stochastic finite element and reliability. A state-of-the-art report. Technical report, Department of Civil and Environmental Engineering, University of California, Berkeley, 2000.

[12] K.Q. Ye, W. Li, A. Sudjianto. Algorithmic construction of optimal symmetric Latin hypercubes. *Journal of Statistical Planning and Inference*, **90**: 145–159, 2000.