Sequential Projection Maximin Distance Sampling Method

**J. Jang[1], W. Lim[1], S. Cho[1], M. Lee[2], J. Na[3]
and *T.H. Lee[1]**

[1]Department of automotive engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea
[2]Korea Institute of Ocean Science and Technology, 32-1312 beon-gil, Yuseong-daero, Yuseong-gu, Daejeon, Korea
[3]GMKorea, 199-1 Cheongcheon 2-dong, Bupyeong-gu, Incheon 403-714, Korea

*Corresponding author: thlee@hanyang.ac.kr

Abstract

Design optimization for engineering problems often requires severe computer simulations. Thus, to perform a design optimization efficiently, surrogate models replacing the time-consuming simulator by using the adequate number of computer experiments, i.e., design and analysis of computer experiments (DACE) have been developed. Our goal in this paper is to propose a sequential design of experiments to construct a global surrogate model. The proposed method employs the priority of variables defined from non-linearity, contribution ratio or global sensitivity. The priority gives a chance to have better projective property to more important variable, because relatively more important variable significantly influences on the accuracy of surrogate model. Consequently this causes a decrease in the error of surrogate model and a reduction of the total number of sample points. The proposed method is compared with sequential maximin distance design and optimal Latin hypercube design with two examples.

Keywords: Design of experiment (DOE), Sequential design, Maximin distance design, Space filling design, Projective property, Surrogate model

Introduction

In engineering problems, design often requires computer simulations to evaluate design objectives and constraints. If a single simulation is severe time-consuming, design optimization becomes impossible because it often requires the considerable number of simulations. One way of alleviating this burden is to employ surrogate models, for instance, response surface model (RSM), radial basis function (RBF) and kriging model. The basic concept of surrogate model is to approximate relation between input and output for predicting responses at untried input within the adequate number of computer experiments. This can reduce the computational cost by replacing the high-fidelity simulator. However, since inaccurate surrogate model can give incorrect responses, appropriate design of experiment is necessary to generate accurate surrogate model. Thus, many studies have been performed to suggest criteria of superior design of experiment and to implement the algorithm to enhance the efficiency. Among many criteria, we focus on two criteria such as space filling and projective property. And as a method for enhancing efficiency, sequential design method is adopted. In the following paragraphs, we briefly review earlier researches for two criteria and sequential design methods.

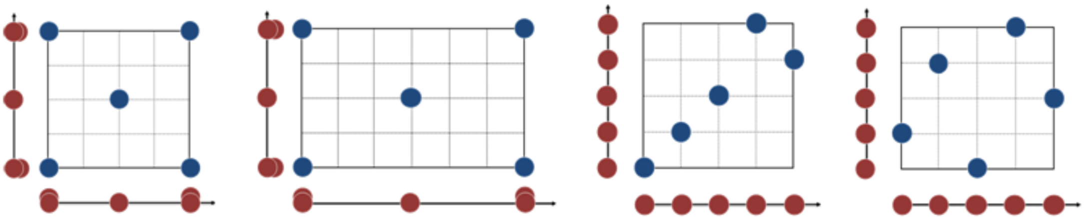
One of representative criteria for computational experiment is the space filling that sampling points fill design space uniformly. The space filling criterion has been developed to obtain information effectively on the overall design domain because computational experiment is deterministic. Many researchers have proposed different criteria to define the space filling design. Thus, the criteria provide optimal sample set accompanied with optimization algorithm and a sample set can be different according to the criteria in spite of the same number of sample points. One of these criteria is maximin distance that tries to maximize the smallest Euclidean distance between any two sets of points over design domain (Johnson, 1990). It is simple and easy to implement, so it is widely used in practice. Maximum entropy design was introduced (Shewry and Wynn, 1987). Entropy is defined as the matrix that consists of entropy function value of each sample point such as Gaussian form. By maximizing determinant of entropy, evenly distributed sample set can be achieved. As a further study, to reduce its computation cost and resolve singularity problem, maximim eigenvalue design was suggested (Lee and Jung, 2004). In general, however, above space filling criteria cannot simultaneously consider projective property that is a space filling in terms of each axis. Fig. 1-(a) shows the best space filling, but projective property is not considered at all, i.e., collapsing arose. In addition, the criteria cannot reflect the behavior of output but consider only relations of input. Thus, to obtain more information of an important variable, the scaled maximin distance design was proposed (Jin Chen, 2002) that gives dimensional weighting to more important variable but it still could not solve the problem of overlapping of sample points as shown in Fig. 1-(b).

Another criterion is projective property. It is also called non-collapsing or non-overlapping property. It is important to consider distances that are projected to axis of each variable. If a certain variable 'x_i' has no influence on the output, two design points that are only different a coordinate of the variable 'x_i' are considered as the same point. Therefore, two design points should not share any coordinate values. In the early days, design method considering projective property were used in the field of safety diagnosis, reliability analysis or uncertainty propagation. Latin hypercube design (LHD) is representative method (Mckay, 1979). Even now, various design method based on LHD have been steadily proposed. However, there are some problems in LHD whose sample points are biased, distorted or clustered as shown in Fig. 1-(c). In order to resolve this problem, optimal Latin hypercube design (OLHD) was developed by employing optimization concept in previous work (Morris and Mitchell, 1995; Park, 1994). OLHD compromises between optimal criterion such as entropy and Latin hypercube with the good projection properties. However, depending on increase of the number of sample points and variables, it takes too much time to optimize and optimal sample set can be unstable. And design based on LHD is difficult to employ sequential design since area of one sample point is determined in advance according to the number of sample points and variables.

Meanwhile, most of the authors are concerned these criteria as one shot approach that sample points are selected over the design space in advance. However, since a simulator is nonlinear and complex, a designer is hard to predict how many sample points are necessary to achieve the sufficient accuracy of surrogate model. Thus, in order to solve this problem, sequential design methods have been proposed. It allows the sampling process to be stopped as soon as there is sufficient information as data accumulate. Also, it takes information such as predicted response, contribution, nonlinearity of variables and mean squared error (MSE) gathered from existing surrogate model updated sequentially with new sample points and the associated response evaluations. These are significantly advantageous compared to one shot approach.

Sequential design can be used for both global surrogate model and surrogate model-based design optimization. Sequential design for global surrogate model focuses on sequentially improving the accuracy of a surrogate model over the entire design space, but sequential design for surrogate model-based design optimization finds promising area where optimum point can be exist. The latter is also called infilling sampling method that gives up space filling. As one of the latter method, mean squared error gathered from kriging model based design of experiment (Sacks and Welch, 1989). And expected improvement (EI) was suggested in work by Mockus, Tiesis, and Zilinskas (1978), and has been popularized in work by Jones, Schonlau, and Welch (1998) as an efficient global optimization (EGO) algorithm. EI is the function whereby points that have either low objective function value or high uncertainty are preferred.

This paper focuses on a global surrogate model by using sequential design method. To enhance the efficiency, the priority of variables is defined, that derived from output information, i.e., nonlinearity of variable, contribution of variable, global sensitivity or even intuition of a designer. In addition, both space filling and projective property is simultaneously considered to improve the accuracy quickly. At last, the proposed method, sequential projection maximin distance design, overcome drawbacks of earlier space filling design and projective property based design. The proposed method is compared with sequential maximin distance design and optimal Latin hypercube design with two examples.



(a) space filling design, (b) scaled space filling design, (c) LHD, (d) OLHD
Figure 1. Examples of existing design methods: sample points on 2-D (blue) and those projected to axis (red)

Sequential projection maximin distance design

Formulation of the proposed method

Sequential projection maximin distance design is proposed based on maximin distance design. Original maximin distance design doesn't use output information gathered from existing surrogate model but use only input information, i.e., distance between pre-sampled points. Thus, to select a new sample point(s), we introduce sequential projection maximin distance design. The proposed method employs priority of variables. If one variable is more important than the other variables, priority should be assigned to that variable. The priority gives chance to have better projective property to more important variable. In other word, relatively less important variable's projective property does not significantly influences on accuracy of surrogate model. Thus, according to priority, each variable is sequentially optimized in iteration. And in order to satisfy space filling criterion, first optimized variables continually influence on an objective function, i.e. modified distance. The steps of new method are:

Step 1 Define priority measure and gather information of priority from existing surrogate model.
Step 2 Maximize the proposed criterion made up of Min. l_1 norm of 1st priority variable
Step 3 Maximize the proposed criterion made up of Min. l_2 norm and Min. l_1 norm of 2nd priority variable with optimized 1st priority variable.
Step 4 Repeat step 3 until the last variable.

Above steps can be formulated as Eq. (1)

$$
\begin{aligned}
&\text{Find } \mathbf{x}_{new} \\
&\quad \mathbf{i} \in \text{a set of priority order}, \mathbf{i} = \{i \mid 1,2,\ldots n_d\} \\
&\quad \text{while } i \le n_d \\
&\quad\; \text{if } i = 1, \text{Maximize} \min \left\| \mathbf{x}_E^i, \mathbf{x}_{new}^i \right\|_1, i = i+1 \\
&\quad\; \text{else, while } i = \text{Maximize} \frac{1}{\sqrt{n_d}} \min \left\| \mathbf{x}_E^{\mathbf{j}}, \mathbf{x}_{new}^i \right\|_2 + \min \left\| \mathbf{x}_E^i, \mathbf{x}_{new}^i \right\|_1, \mathbf{j} = [1,\ldots,i], i = i+1 \\
&\quad \text{end}
\end{aligned}
\tag{1}
$$

where $\mathbf{x}_E$ are existing sample points and n_d is the number of variables.
In this method, it is important to define the priority since it decisively determines accuracy of surrogate model. The priority of variable can be defined from nonlinearity of variable, contribution of variable, global sensitivity or even intuition of designers. Among them, we employ the nonlinearity of variable that can be alternated by correlation parameters, θ_k, in kriging model. The correlation parameters indicate smoothness of x_k coordinate. The smaller θ_k linear effect on the response of the variable, impact on the response is non-linear as the θ_k increases.

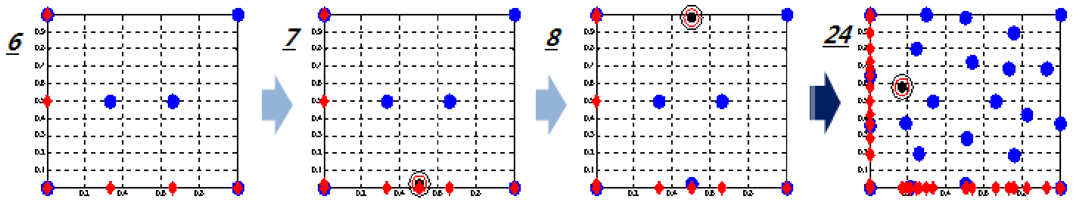
Proposed method comparison with sequential maximin distance design and OLHD

The proposed method is compared with sequential maximin distance design and OLHD in order to show or not it meets above criteria, space filling and projective property. Experiments are carried out sequentially one by one on 2-dimensional domain from initial 6 sample points, 4 on vertex and 2 on center. And since OLHD cannot provide sequential design, we perform experiments at the same number of sample points in order to compare the surface of distribution of sample points. We use genetic algorithm as an optimizer provided by matlab R2011 to select a new point, and OLHD is also designed by matlab toolbox.

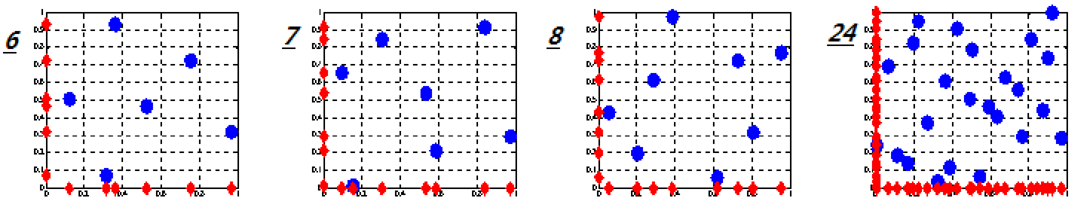
Fig. 2-(a) show results of sequential maximin distance design. We can check a first optimized sample point is located in bottom line. It is the best position as an aspect of space filling, but it can be the worst position in terms of vertical axis. A second optimized sample point also similar. Also after adding 18 points, the trend of result is same.

The results of OLHD in Fig. 2-(b) are the opposite of the results of sequential maximin distance design. While projective property is sufficiently conscious, space filling is poor.

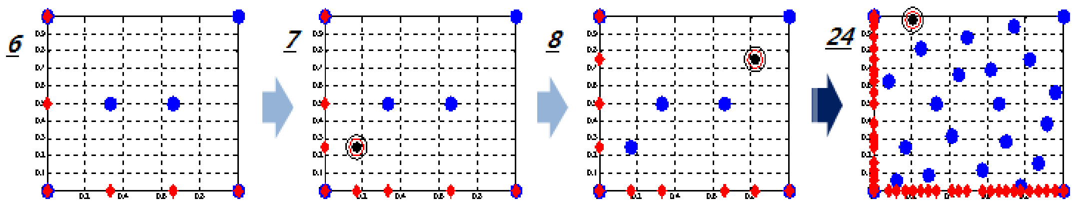
Even if one sample point is selected, the results of the proposed method in Fig 2-(c) show its characteristic well. The first optimized point satisfies projective property both horizontal and vertical axis. And after adding 18 points likewise above method, while not lose space filling, projective property is maintained very well.



(a) Sequential maximin distance design (sequentially sampled from 6 to 24)



(b) OLHD (one shot approach)



(c) Sequential projection maximin distance design (sequentially sampled)
Figure 2. Surface of distribution of samples using three methods: pre-sampled points (blue), projected points (black) and a new point (red)

Examples

The two examples are utilized in order to show the performance of proposed method. Since we focus on build up accurate global surrogate modeling, the accuracy, i.e. error is used as a performance measure.

Mathematical example

The first example is a mathematical example in 2-D that can easily obtain responses and know real response. The equation of example is as following;

$$f(\mathbf{x}) = 2\cos(8\pi x_1) + 0.25\cos(\pi x_2) + 12, \; 0 \le x_1, x_2 \le 1 \tag{2}$$

The experiment is carried out in the following procedure.
Step 1 Select initial sample points.
Step 2 Build up surrogate model, i.e. kriging model with initial sample points.
Step 3 Predict responses at validation points, and measure the error.
Step 4 Select a new sample point according to each method.

Step 5 Repeat steps 2~4 until pre-defined maximum iteration and skip step 4 in last iteration.

Validation points are 9^2 from full factorial design (FFD) and we employ a mean relative error (MRE) as follows;

$$MRE[\%] = E\left[\frac{\mathbf{Y}-\hat{\mathbf{Y}}}{\mathbf{Y}}\right]\times 100 \tag{3}$$

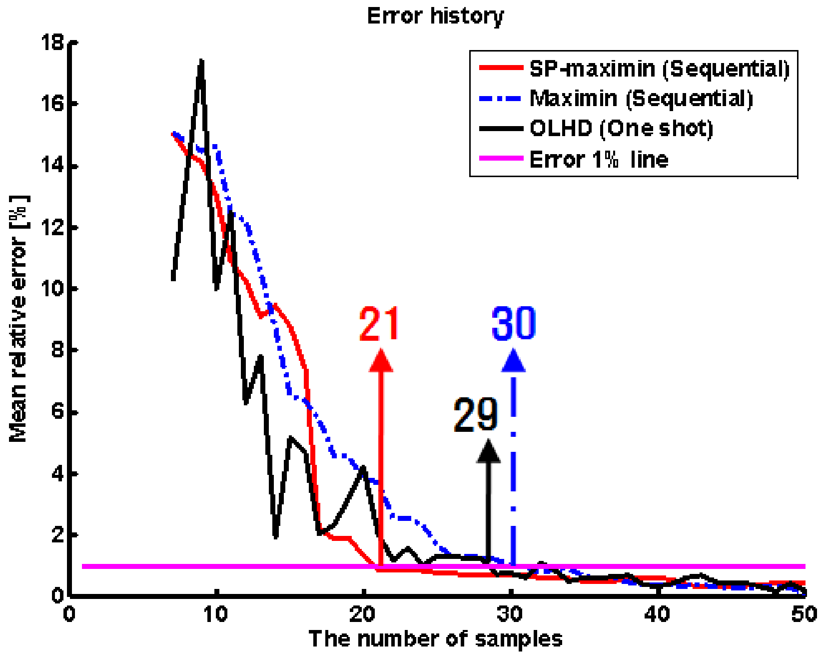


Figure 3. History of error for mathematical example as adding sample points; 21, 29 and 30 mean the first number of sample point satisfying 1% error

Errors of kriging model made by the three methods are decrease as adding sample points. It means that kriging model becomes more accurate as added pre-sampled points. However, error using the proposed method (SP-maximin) considering projective property according to priority of variables decreases faster than two other methods. Thus, the proposed method can reduce 8 or 9 experiments. Errors of OLHD as the one shot approach fluctuate since its distribution of sample points is change.

Engineering example

The second example is an engineering example with 10 design variables that takes more time and cost. Target model is a front cradle in a passenger vehicle released by GM Korea. 10 thickness values are considered as design variables. Analysis purpose is to make plastic deformation of front cradle under target value. Abaqus is used as analyzer. The experiment is carried out in the following procedure.

Step 1 Select initial sample points.
Step 2 Build up surrogate model, i.e. kriging model with initial sample points.
Step 3 Predict responses at validation points.
Step 4 Measure the error.
Step 5 Stop if error is under 1% five times in a row.
Step 6 Select a new sample point according to each method. And go step. 2

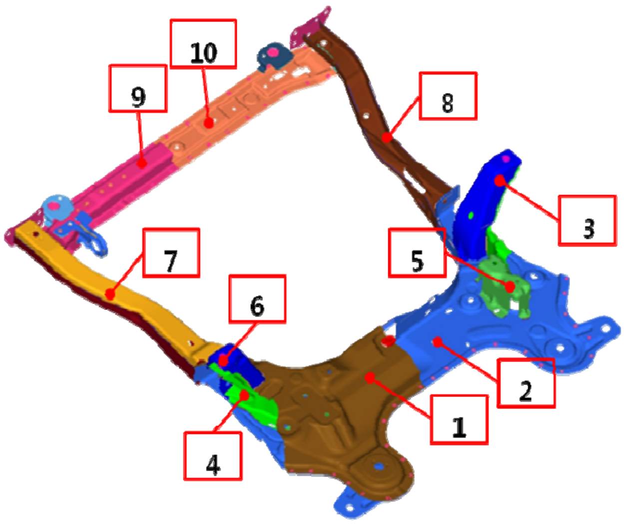


Figure 4. Finite element model of a front cradle

Cross validation (one-leave-out) is employed since adding points is also time-consuming. Initial sample points are obtained from OLHD and they are used for validation points. Lastly OLHD is not considered in this example.

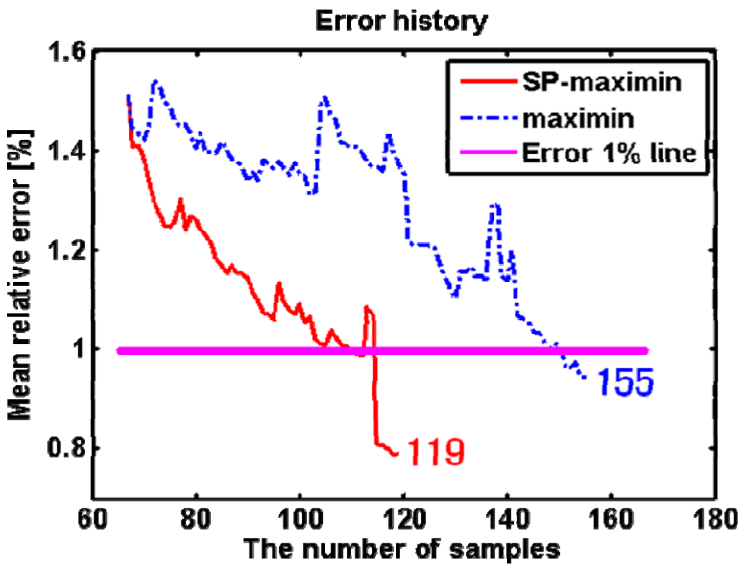


Figure 5. . History of error for an engineering example as adding sample points: 53 and 89 mean the last number of sample point satisfying 1% error 5 times in a row

Likewise a mathematical example, the error of the propose method decreases faster while the proposed method reduce the total number of sample points. It means the proposed method save about 6 hours since 1 simulation takes about 10 minute.

Conclusions

The sequential design method to create global surrogate model accurately is proposed in this paper. The proposed method employs the priority of variables defined from non-linearity, contribution ratio or global sensitivity. The priority gives a chance to have better projective property to more important variable. Consequently more information of variables in the priority can be obtained by using the proposed method. This decreases the error of surrogate model and reduces the total number of sample points. In order to show the performance of the proposed method, kriging model is introduced and correlation coefficients of kriging model are considered as a criterion defining the priority. And sequential maximin distance design and optimal Latin hypercube design are used for comparison. The mathematical example that consists of a highly nonlinear variable and a moderately linear variable shows an advantage of the proposed method well. And there is a remarkable difference between the convergence histories. This is because a curve of the highly nonlinear variable is well fitted when projective property of it is fully represented. Even if the response is unpredictable, this merit still exists. In engineering example, the finite element model of front cradle in the vehicle, a sampling with the proposed method is stopped after 53th iteration under defined stop criterion. In other word, the proposed method uses 119 sample points to create a sufficiently accurate surrogate model and it uses less 36 sample points than using sequential maximin distance design. In terms of time, an engineer can save 6 hours since 1 simulation takes about 10 minute. As a result, the sequential projection maximin distance design helps engineers to solve the problems in that only a part of variables are important.

References

Johnson, M., Moore, L. and Ylvisaker, D., (1990), Minimax and maximin distance designs. Journal of Statistical Planning and Inference. 26, pp. 131–148.

Shewry, M.C, and Wynn, H.P (1987), Maximum entropy sampling. Journal of Applied Statistics. 40(4), pp. 165–170.

Lee, T.H., and Jung, J.J. (2004), Maximum eigenvalue sampling of kriging model. 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. Albany, New York, August, AIAA 2004-4488.

Morris, M. D. and Mitchell, T. J. (1995), Exploratory designs for computational experiments. Journal of Statistical Planning and Inference. 43(3), pp.381-402.

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.

Park, J.S. (1994), Optimal Latin-hypercube designs for computer experiments. Journal of statistical planning and Inference. 39(1), pp.95-111.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989), Design and analysis of computer experiments. Statistical Science. 4, pp.409–435.

Mockus, J., Tiesis, V. and Zilinskas, A. (1978), The application of Bayesian methods for seeking the extremum. Towards Global Optimisation . North Holland, Amsterdam, 2, pp.117–129.

Jones, D.R., Schonlau. M., and Welch, W.J (1998), Efficient global optimization of expensive black-box functions. Journal of Global Optimization. 13, pp455-492.

Jin, R., Chen, W., and Sudjianto, A. (2002), On sequential sampling for global metamodeling for in engineering design. Proceedings ASME 2002 Design Engineering Technical Conferences and Computer and Information in Engineering Conference, Montreal, Canada, September, DETC2002/DAC-34092.

Crombecq , K., Laermans, E. and Dhaene, T. (2011), Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. European Journal of Operation Research. 214(3), pp683-696.

Xiong, F., Xiong, Y., Chen, W. and Yang, S. (2009), Optimizing Latin Hypercube design for sequential sampling of computer experiments. Engineering Optimization. 41(8), pp793-810.

Timothy W. S., Dennis, K. J. Lin, Chen, W. (2001), Sampling strategies for computer experiments: design and analysis. International Journal of Reliability and Applications. 2(3), pp209-240.

Jin, R., Chen, W., Sudjianto, A. (2005). An effcient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inference.134, pp268–287.

van Dam, E.R., Husslage, B., den Hertog, D., Melissen, H. (2007). Maximin latin hypercube design in two dimensions. Operations Research. 55, pp158–169.