11th World Congress on Computational Mechanics (WCCM XI)
5th European Conference on Computational Mechanics (ECCM V)
6th European Conference on Computational Fluid Dynamics (ECFD VI)
July 20 - 25, 2014, Barcelona, Spain

STATISTICALLY WEIGHTED MAXIMIN DISTANCE DESIGN WITH KERNEL DENSITY ESTIMATION

Junyong Jang*[1], Su-gil Cho[1], Minuk Lee[2] and Tae Hee Lee[1]

[1] Department of Automotive Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea, thlee@hanyang.ac.kr

[2] Ocean System Engineering Research Division, Korea Institute of Ocean Science & Technology, 32 1312beon-gil, Yuseong-daero, Yuseong-gu, Daejeon 305-343, Korea, mwlee97@kiost.ac

Key Words: *Maximin Distance Design, Kernel Density Estimation, Statistical Information, Weight Factor, Sequential design of experiments*

Recently in engineering fields, a large portion of real experiments is replaced by computational simulations to reduce cost, risk and time. But because a single simulation including a complex analysis can take a lot of time to complete the analysis, an engineer still has difficulties to handle a simulation for such tasks. Thus, in order to describe the response surface of simulation, surrogate models are proposed. And sampling strategies to construct an accurate surrogate model on overall design space are also suggested. Due to nature of a deterministic simulation, space filling that fills design space uniformly to obtain information effectively is the most important criterion of sampling strategies. However, space filling design has a drawback that it may extract points on unnecessary regions in terms of optimization because it does not consider output data. Thus, in order to minimize the sample size and to obtain more profitable points, various sequential design techniques are suggested. In this study, novel sequential design that can exploit the responses of simulation and allow exploration for a reduction of uncertainty on complex response surface is proposed. The proposed design is based on maximin distance design and a few parameters and kernel density estimation are suggested to consider responses.

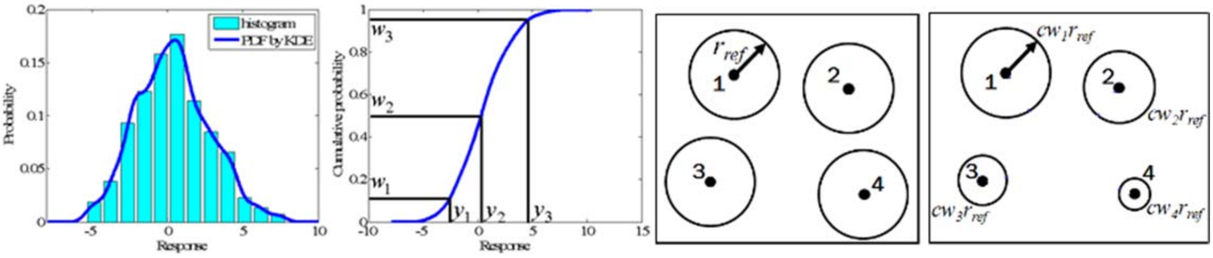
The formulation of proposed method is as follows:

$$\begin{aligned}
&\text{Maximize} \quad \min \mathbf{D} \\
&\quad \text{where} \quad \mathbf{D} = \left(d(\mathbf{x}_E, \mathbf{x}_{new}) - c\mathbf{w}r_{ref} \right) \\
&\qquad r_{ref} = \sqrt{n_d / (n_c + 1)} \\
&\qquad w = \hat{F}_h \left(y = y_i \,\middle|\, y \sim \hat{f}_h(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h} \right) \right) = \int_{-\infty}^{y_i} \hat{f}_h(u)\,du
\end{aligned} \tag{1}$$

Where $d(\cdot)$ is l-2 norm, $\mathbf{x}_E$ are pre-sampled points, $\mathbf{x}_{new}$ is candidate point as a new point, c is the convergence parameter that controls the effect of total weighting, n_d is the number of variables, n_c is the number of current samples, w_i (i=1,…,n_c) are weight factors of i-th response, $K(\cdot)$ is a kernel density function, h is a bandwidth, and y_i are responses. The Gaussian function is employed as a kernel density function.

The reference radius parameter (r_{ref}) is expressed as the n-sphere bounded by a certain radius and it is the basis domain at pre-sampled points. It prevents other points from penetrating the domain of a certain point. The reference radius parameter is monotone decreasing function as increasing number of sample points and design variables.

The weight factors are determined by cumulative distribution function (CDF) estimated by the kernel density estimation (KDE). The KDE uses responses of pre-sampled points and each point receive the weight factor by different response values. A point with lower response receive lower weight factor and a point with higher response is the opposite. It makes new generated point located close to a point with lower response, minimum.



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

Fig. 1 Application process of the weight factors: (a) estimated probility density function of response, (b) define weight factors by estimated cumulative distribution function, (c) non-weighted reference radius, (d) weighted reference radius

At last, 2-dimensional mathematical examples, branin_rocs function and haupt function are followed to show the performance of the proposed method. The number of initial sample point is 4 on vertex and samples are sequentially added to 40 points one by one. The convergence parameter *c* is fixed by 1. As in Fig.2, the proposed method generates the sample points close to the minimums without using surrogate model. In further study, the effect of convergence parameter and comparisons with existing designs will be showed.

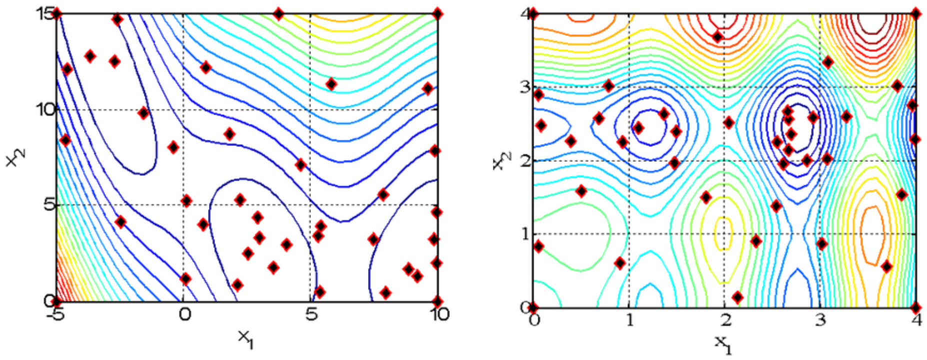


Fig. 2 Sequential sampling results from 4 to 40 points

REFERENCES-

[1] K. Crombecq, E. Laermans, T. Dhaene, Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. *European Journal of Operation Research*, Vol. **214**, No. 3, pp. 683–696, 2011.

[2] M.E. Johnson, L.M. Moore and D.Ylvisaker, Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, Vol. **26**, pp. 131–148, 1990.