

# Statistical Analysis of Various Optimal Latin Hypercube Designs



A. R. M. Jalal Uddin Jamali, Md. Asadul Alam, and Abdul Aziz

**Abstract** Among several *Design of Experiments (DoEs)*, Latin Hypercube Design (LHD) is one of the most frequently used methods in the field of physical experiments and in the field of computer simulations to find out the behavior of response surface of the surrogate model with respect to design points. A good experimental design should have three important characteristics, namely (i) non-collapsing, (ii) space-filling and (iii) orthogonal properties. Though inherently LHD preserves non-collapsing property, but randomly generated LHDs have poor space-filling in terms of minimum pair-wise distance. In order to ensure the last two properties in LHD, researchers are frequently looking for finding optimal LHD in the sense of space-filling and orthogonal properties. Moreover, researchers are frequently encountered the question, which distance measure is the best in the case of optimal designs? In the literature, several types of optimal LHDs are available such as Maximin LHD, Orthogonal LHD, Uniform LHD, etc. On the other hand, two distance measures namely Euclidean and Manhattan distance measures are used frequently to find optimal DoEs. But which one of the two distance measures is better, is still unknown. In this article, intensive statistical analysis has been carried out on numerical instances to explore the deep scenario of each optimal LHD. The main goal of this research is to find out a scenario of the well-known optimal designs from statistical point of view. From this elementary experimental study, it seems to us that in the sense of space-filling, Euclidean distance measure-based Maximin LHD is the best. But if one needs space-filling along with better orthogonal property, then multi-objective (Maximin with approximate orthogonal)-based optimal LHD is relatively better than Maximin LHD.

---

A. R. M. Jalal Uddin Jamali (✉)

Department of Mathematics, Khulna University of Engineering & Technology, Khulna, Bangladesh

e-mail: [jamali@math.kuet.ac.bd](mailto:jamali@math.kuet.ac.bd)

Md. Asadul Alam

Department of Mathematics, Daud Public College, Jessore, Bangladesh

A. Aziz

Department of CSE, Khulna University of Engineering & Technology, Khulna, Bangladesh

**Keywords** Experimental design · Space-filling · Non-collapsing · Orthogonal · Euclidean distance

## 1 Introduction

Computer simulation-based DoEs are used in a wide range of applications to learn about the effect of input variables  $x$  on a response of interest  $y$ . Fang et al. (2006) said, as simulation programs are usually deterministic so the output of a computer experiment is not subject to random variations, which makes the design of computer experiments different from that of physical experiments. Many simulation models involve several hundred factors or even more. It is desirable to avoid replicates when projecting the design onto a subset of factors. This is because a few out of the numerous factors in the system usually dominate the performance of the product. Thus, a good model can be fitted using only these few important factors. Therefore, when we project the design on to these factors, replication is not required. As it is recognized by several authors, the choice of the design points for computer experiments should at least fulfill two requirements namely non-collapsing and space-filling (Johnson et al. 1990; Morris and Mitchell 1995). In addition, in order to find out the individual effect of factors on response surface, researchers hunt DoEs with orthogonal property among the factors. LHD inherently preserves non-collapsing property, i.e. factors are evenly spread whenever projected on any coordinate. McKay et al. (1979) first introduced LHD. Later, authors (Morris and Mitchell 1995; Grosso et al. 2009) defined LHD which are a bit different. In a LHD, there are  $N$  points (design points) and each point has  $k$  distinct coordinates (parameters/factors). The points are placed in such a way that they are uniformly distributed when projected on every single coordinate (factor) axis. In a LHD, the range of each parameter/factor is normalized to the interval  $[1, N]$  (actually authors Grosso et al. 2009) considered the interval  $[0, N-1]$  but for the symmetry of other designs considered, here we have considered the range  $[1, N]$ ). Mathematically, let us consider a set of  $N$  points in a uniform  $k$ -dimensional grid  $\{1, 2, \dots, N\}^k$ :

$$X = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \dots & \vdots \\ x_{N1} & \dots & x_{Nk} \end{pmatrix} \forall x_{ij} = \{1, 2, \dots, N\}.$$

Then  $X$  is a LHD iff  $x_{pj} \neq x_{qj} \forall j, p, q \in \{1, 2, \dots, N\} \exists p \neq q$ , i.e. each column has no duplicate entries. Note that the number of possible LHDs is huge: there are  $(N!)^k$  possible LHDs.

It is known that randomly generated LHD has poor space-filling property. Moreover, it bears poor orthogonal property. Therefore, researchers search for optimal LHDs in the sense of space-filling, orthogonal, uniform etc., Jin et al. (2005), and Johnson et al. (1990) proposed Maximin (maximize the minimum inter-site pair-wise

distance) experimental design but it was not LHD. They measured the space-filling on the basis of minimum inter-site pair-wise distance “ $D_1$ ” among all pair-wise distances of design points. That is if “ $D_1$ ” value of any LHD is relatively smaller then we say that the LHD is not good enough in the sense of space-filling. Morris and Mitchell (1995) proposed  $\Phi_p$  optimal criterion and later Grosso et al. (2009) modified it for easy computation, which is given below:

$$\Phi_p(X) = \sum_{i=1}^N \sum_{j=i+1}^N \left[ \frac{1}{d_{ij}^p} \right]^{\frac{1}{p}} \tag{1}$$

where  $d_{ij} = d(x_i, x_j)$  be the distance between points  $x_i$  and  $x_j$  and  $p$  is a positive integer.

Audze and Eglais (1977) proposed a new optimal criterion that is similar to the potential energy function (Potential (U)) and which is defined as below:

$$\text{Min } U = \min \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{d_{ij}^2}. \tag{2}$$

It is known that *multicollinearity* (orthogonal property) measures the linear dependency among the factors of the design points. Multicollinearity may be measured by the partial pair-wise correlations among the factors. Several ways are available in the literature to measure the pair-wise correlations. Here, we consider the measure of average pair-wise correlations and maximum correlation, respectively, as below:

$$\rho^2 = \frac{\sum_{i=2}^k \sum_{j=1}^{i-1} \rho_{ij}^2}{k(k-1)/2} \tag{3}$$

and

$$\rho_{\max} = \max_{1 \leq i, j \leq k} \rho_{ij} \tag{4}$$

Here  $\rho_{ij}$  denotes the simple product–moment correlation between the factors  $i$  and  $j$ ;  $k$  denotes the number of factors in the design considered. Butler (2001) found optimal LHD on the basis of this criterion. On the other hand, Joseph and Hung (2008) proposed the combination of the above two criteria. Moreover, Fang et al. (2000) defined a uniform design that allocates experimental points uniformly spread over the domain. It is noted that uniform design does not require orthogonal (multicollinearity) property. It considers projective uniformity over all sub-dimensions. But, in Fang et al. (2000), authors classified uniform design as space-filling design. A good literature review is available in Viana et al. (2014). It is worthwhile to mention

here that Jamali et al. (2019) recently established some relations and bonds between Euclidian and Manhattan distance measures regarding LHD.

## 2 Experimental Result and Discussion

For the experimental study, we have considered a typical LHD of four factors ( $k = 4$ ) with nine ( $N = 9$ ) design points, which is optimized using several optimal criteria. The optimized LHDs are shown in Table 1. In Table 1, MLH-SA denotes Maximin LHD regarding Manhattan distance measure ( $L^1$ ) obtained by Simulated Annealing (SA) approach (Husslage et al. 2011). OMLH-MSA denotes Orthogonal Maximin LHD regarding  $L^1$  distance measure obtained by Modified Simulated Annealing (MSA) (Morris and Mitchell 1995), OLH-Y denotes Orthogonal LHD regarding  $L^1$  distance measure obtained by Ye (1998), ULH-F denotes uniform LHD regarding  $L^1$  distance measure obtained by Fang et al. (2000) and MLH-ILS indicates the Maximin LHD regarding Euclidean distance measure ( $L^2$ ) obtained by Iterated Local Search (ILS) approach (Grosso et al. 2009).

At first, we have extracted all important statistical characteristics of each optimal design given in Table 1. Then all the findings are shown in Table 2. The symbols SD, CoV,  $\beta_1$ ,  $\beta_2$ , Min Distance, Max Distance denote Standard Deviation, Coefficient of Variation (%), Moment Coefficient of Skewness, Moment Coefficient of Kurtosis ( $\alpha_1 + 3$ ), minimum pair-wise inter-site distance among all design points of the LHD and maximum pair-wise inter-site distance among all design points of the LHD respectively. Besides,  $D_1^{L^1}$  and  $D_1^{L^2}$  denote the minimum pair-wise inter-site distance of design points measured in  $L^1$  (Manhattan) distance measure and in  $L^2$  (Euclidean) distance measure respectively. It is mentioned earlier that space-filling is defined on only  $D_1$  value or  $\Phi$  value. Similarly, orthogonal LHD is defined by multi-collinearity and so on. But the insight views of several optimized LHDs are still unknown. Here, we have analyzed the optimal LHDs statistically and on distance measures. In addition, we have compared optimized LHDs in those issues.

**Table 1** Some optimal LHDs for  $(k, N) = (4, 9)$

Method	LH-SA	OMLH-MSA	OLH-Y	ULH-F	MLH-ILS
Optimal design matrix	1 3 3 4	1 5 3 3	1 2 6 3	4 1 7 5	1 5 8 4
	2 5 8 8	2 2 5 8	2 9 7 6	1 3 4 3	2 7 4 9
	3 8 6 2	3 9 7 5	3 4 2 9	9 9 5 4	3 2 1 6
	4 7 1 6	4 3 8 1	4 7 1 2	6 6 6 9	4 8 3 3
	5 2 9 3	5 7 1 7	5 5 5 5	5 7 2 1	5 1 5 1
	6 9 5 9	6 6 9 9	6 3 9 8	2 8 8 7	6 3 7 8
	7 1 4 7	7 1 2 4	7 6 8 1	3 5 1 6	7 6 9 2
	8 4 2 1	8 8 4 2	8 1 3 4	8 2 3 8	8 9 6 7
	9 6 7 5	9 4 6 6	9 8 4 7	7 4 9 2	9 4 2 5

Source Collected from the following papers: Morris and Mitchell (1995), Ye (1998), Fang et al. (2000) and Grosso et al. (2009)

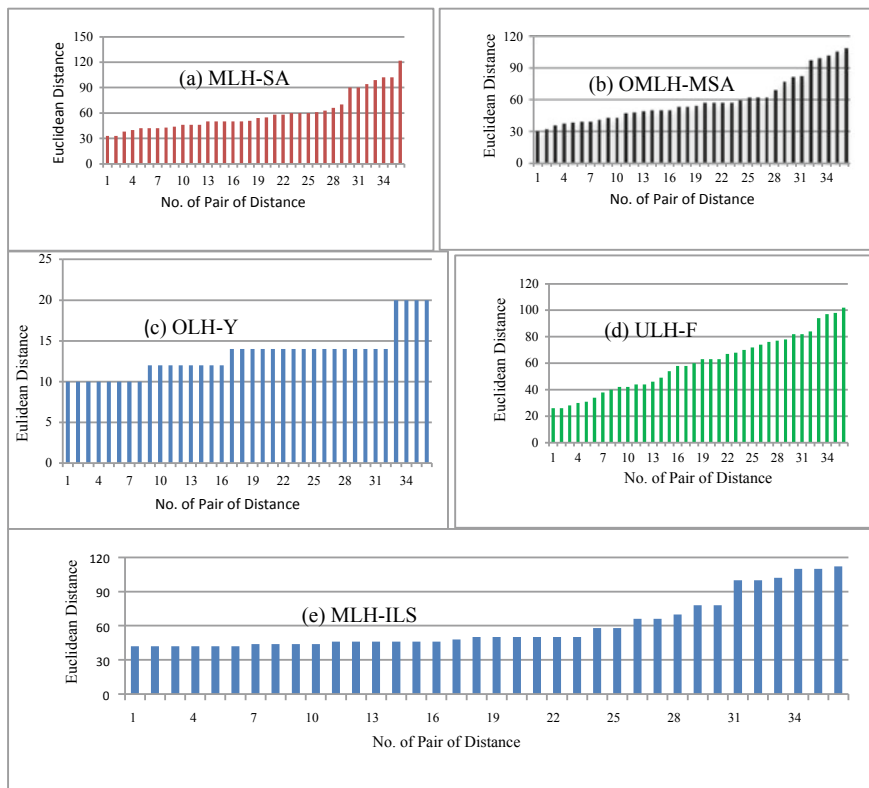
**Table 2** Statistical analysis among several optimal LHDs for  $(k, N) = (4, 9)$

Method	MLH-SA	OMLH-MSA	OLH-Y	ULH-F	MLH-ILS
Objective function	Maximin	Maximin + $\rho$	$\rho$	Uniform	Maximin
Dist. used in the method	$L^1$	$L^1$	$L^1$	$L^1$	$L^2$
No. of distances ( $n$ )	36	36	36	36	36
Sum of distances	2160	2160	2160	2160	2160
Average	60	60	60	60	60
SD	21.7894	<b>21.19355</b>	24.4949	21.65641	22.59794
CoV	36.31566	<b>35.32259</b>	40.82483	36.09401	37.66323
$\beta_1$	1.237749	<b>0.989075</b>	1.278651	0.148856	1.332428
$\beta_2(\alpha_1 + 3)$	0.760507	<b>0.093569</b>	1.918449	<b>-0.93998</b>	0.347584
Min Distance ( $L^2$ )	33	31	30	26	<b>42</b>
Max distance ( $L^2$ )	122	<b>109</b>	120	102	112
Range ( $L^2$ )	89	78	90	76	<b>70</b>
$\rho$	0.088889	0.041667	<b>0</b>	0.061111	0.133333
$\rho_{Max}$	0.216667	0.116667	<b>0</b>	0.15	0.233333
$\Phi_p^{L^2}$	0.183966	0.187146	0.202608	0.207282	<b>0.17497</b>
Potential (U)	0.667812	0.66931	0.7	0.699	<b>0.666661</b>
$D_1^{L^1}$	<b>11</b>	<b>11</b>	10	10	10
$D_1^{L^2}$	33	31	30	26	<b>42</b>

Source Authors’ own calculation by using programs

We have also detected all pair-wise inter-site Euclidean distances (actually square of Euclidean distances) “ $D_{ij}^2$ ” values for all pair of design points for each optimal LHDs. Now for each optimal LHD, the “ $D_{ij}^2$ ” are plotted corresponding to each pair of design points ( $I = \{1, 2, \dots, 36\}$ ), which are shown in Fig. 1. It is noted that since we have taken 9 design points, we have obtained  $[(N(N - 1))/2] = 36$  number of pair-wise distances. It is worthwhile to mention here that the sum of square of all pair-wise (Euclidian) distances,  $(d(x_i, x_j))^2$ , of a fixed LHD is identical (=2160) and so average of the square distances of all LHDs (for  $k = 4, N = 9$ ) are also identical. It is noted that there are  $(9!)^4$  such LHDs and for each such LHD the value of  $\sum (d(x_i, x_j))^2$  is equal to 2160.

Now, we will discuss the comparison between two Maximin LHDs namely MLH-SA and MLH-ILS in which both are optimized by using same single optimal criterion called Maximin optimal criterion. But distance measures namely  $L^1$  and  $L^2$  are used in MLH-SA and MLH-ILS, respectively, which are different. In addition, different optimal heuristic algorithms namely SA and ILS approaches are applied to optimize the MLS-SA and MLH-ILS designs, respectively. It is observed in both MLH-SA and MLH-ILS that the values of SD, CoV,  $\beta_1, \beta_2, \rho, \rho_{Max}, \Phi_p^{L^2}$  and Potential (U) are not significantly different. It is also observed that the difference of  $D_1^{L^1}$  of the



**Fig. 1** Graphical representation of square of Euclidean distances corresponding to the pair of design points for each optimal LHDs where  $(k, N) = (4, 9)$ . *Source* Created by the authors

two DoEs is not significant, though MLH-SA searched optimal  $D_1^{L1}$  value whereas MLH-ILS explored for  $D_1^{L2}$  value rather than  $D_1^{L1}$ .

Now we will talk about the comparison between Maximin LHD and Uniform LHD mainly MLH-ILS and ULH-F. It is shown in Table 2 that uniform optimality criterion-based LHD is flatter and un-skewed compared to all other LHDs considered here. Moreover, the coefficient of correlation of ULH-F is relatively smaller than that of MLH-ILS. It is also observed that SD, CoV, Max Distance and Range values are not significantly different from the two optimal LHDs namely MLH-ILS and ULH-F. It is also clear from Table 2 that the  $D_1^{L1}$  values of both DoEs are identical. On the other hand, for  $\Phi_p^{L2}$ , Potential (U) and  $D_1^{L2}$  values, MLH-ILS design is significantly better than those of ULH-F design. It is worthwhile to mention here that the  $D_1^{L2}$  value of ULH-F is worst compared to other optimal designs considered here.

We will also discuss orthogonal-based DoEs mainly of OLH- Y. In addition, comparison will be drawn with Maximin-based LHD mainly MLH-ILS. Since the initial criterion of OLH-Y is  $\rho = 0$ , so obviously the factors of the DoE are all uncorrelated. It is observed in Table 2 that the SD, CoV,  $\beta_1$  and  $D_1^{L1}$  values of OLH-Y

are not significantly different with that of MLH-ILS. But it is noticed that for  $\beta_2$ , Max Distance, Range,  $\Phi_p^{L2}$ , Potential (U) and  $D_1^{L2}$  values, OLH-Y is significantly worse compared with the values of MLH-ILS. Moreover, though in OLH-Y,  $L^1$  distance measure is used whereas, in MLH-ILS,  $L^2$  distance measure is used but both  $D_1^{L1}$  values are identical. On the other hand,  $D_1^{L2}$  value of MLH-ILS is significantly better compared with not only OLH-Y design but also all other designs considered here.

Now we will discuss the properties of OMLH-MSA in which multi-objective function namely Maximin and multi-collinearity was considered. It is observed that the SD, the CoV, the  $\beta_1$  and Max Distance values of OMLH-MSA are the best compared with all other optimal designs. But these values are not significantly different from those values of MLH-ILS.

It is observed in Table 2 that the average correlation as well as maximum correlation of OMLH-MSA is negligible and smaller than those values of MLH-ILS. But it is also noticed in Table 2 that the average correlation as well as maximum correlation of MLH-ILS is also not significantly larger. In addition, it is also observed in Table 2 that  $D_1^{L1}$  value of OMLH-MSA is not significantly smaller than that of other optimal LHDs. On the other hand,  $D_1^{L2}$ ,  $\Phi_p^{L2}$  and potential (U) values of MLH-ILS are the best compared with all other optimal LHDs. Moreover, these values are significantly better compared with OMLH-MSA. It is also observed in Fig. 1 that according to space-filling criterion MLH-ILS is relatively looking more homogeneous over the design space compared with other designs.

Now we will discuss the two distance measures considered here namely Euclidean and Manhattan. It is noticed in Table 2 that except MLH-ILS all other designs are optimized by using Manhattan distance measure whereas MLH-ILS is optimized by considering Euclidean distance measure. It is observed in Table 2 that  $D_1^{L1}$  value of anyone LHD compared with the remaining other LHDs is not significantly better. On the other hand, the  $D_1^{L2}$  value of MLH-ILS is the best and significantly larger than any other one.

### 3 Conclusion

When LHD is considered as an experimental design, the design inherently preserves good non-collapsing property. On the other hand, when Maximin optimal criterion is considered, then the DoE should have good space-filling property. Again, when we use multicollinearity as an optimal criterion, then the factors of the DoE are uncorrelated or approximately uncorrelated. Similarly, whenever the uniform optimal design is considered then design becomes flatter. Besides optimal criteria, another question is frequently raised; which distance measure is relatively better in case of an optimal design? In this elementary study, the main objective is to find out the schematic view of several well-known optimal designs from the statistical point of view. A rigorous statistical analysis has been carried out over some well-known optimal LHDs, which are optimized with different optimal criteria. From this

elementary analysis, it seems to us that, if we prefer a DoE containing all three properties (good) then multi-objective-based LHD namely OMLH-MSA is relatively better. But if one considers only space-filling LHD, then MLH-ILS is the best in which Euclidean distance measure is considered. Similarly, if one needs the design with uncorrelated factors along with non-collapsing property then it is obvious that OLH-Y is the best. In addition, it seems to us that  $L^2$  is relatively better to find out Maximin LHD. Though multi-objective-based optimal LHD is relatively better, it is obviously time-consuming. Moreover, in that optimal DoE, none of the individual optimal criterion (of multi-objective function) is found good enough. It is worthwhile to mention here that ILS is preferable to find Maximin LHD with cheaper computational time. Therefore, according to us, one may first find out an optimal LHD by considering Maximin optimal criterion with ILS approach to detect the important factors on response surface by using  $L^2$  distance measure. Then orthogonal optimal criteria may be applied on the reduced DoE (eliminated unimportant factors) to get an uncorrelated-based DoE for further experiments. But it is worthwhile to mention here that, in this experimental research, only some optimal LHDs of  $(k, N) = (4, 9)$  are considered. At the same time, we would like to note here that in the existing literature only four optimal LHDs of  $(k, N) = (4, 9)$  viz MLH-SA, OMLH-MSA, OLH-Y, ULH-F, MLH-ILS are available, which means, some other instances that are optimized with all these five approaches are not yet available in the literature. So, further experiments (with different instances) should be carried out to draw a more concrete conclusion.

**Acknowledgements** This research work is supported by the Special Grant of Ministry of Science and Technology and Khulna University of Engineering and Technology, Khulna, Bangladesh.

## References

- Audze, P., & Eglais, V. (1977). New approach for planning out of experiments. *Problems of Dynamics and Strength*, 35, 104–107.
- Butler, N. A. (2001). Optimal and orthogonal Latin hypercube designs for computer experiments. *Biometrika*, 88(3), 847–857.
- Fang, K. T., Lin, D. K. J., Winker, P., & Zhang, Y. (2000). Uniform design: Theory and application. *Technometrics*, 42(3), 237–248.
- Fang, K. T., Li, R., & Sudjianto, A. (2006). *Design and modeling for computer experiments*. CRC Press.
- Felipe, A. C. V., Simpson, T. W., Balabanov, V., & Toropov, V. (2014). Metamodeling in multi-disciplinary design optimization: How far have we really come? *AIAA (American Institute of Aeronautics and Astronautics) Journal*, 52(4), 670–690.
- Grosso, A., Jamali, A. R. J. U., & Locatelli, M. (2009). Finding maximin latin hypercube designs by I. L. S. Heuristics. *European Journal of Operation Research*, 197, 541–547.
- Husslage, B. G. M., Rennen, G., van Dam, E. R., & den Hertog, D. (2011). Space-filling Latin hypercube designs for computer experiments. *Journal of Optimization and Engineering*, 12, 611–630. <https://doi.org/10.1007/s11081-010-9129-8>



- Jamali, A. R. M. J. U., Alam, Md., & A. . (2019). Approximate relations between Manhattan and Euclidean distance regarding Latin hypercube experimental design . *Journal of Physics: Conference Series IOP*, 1366, 012–030. <https://doi.org/10.1088/1742-6596/1366/1/012030>
- Jin, R., Chen, W., & Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134(1), 268–287.
- Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26, 131–148.
- Joseph, V. R., & Hung, Y. (2008). Orthogonal-maximin latin hypercube designs. *Statistica Sinica*, 18, 171–186.
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Techno Metrics*, 21, 239–245.
- Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computer experiments. *Journal of Statistical Planning and Inference*, 43, 381–402.
- Ye, K. Q. (1998). Orthogonal column Latin hypercube and their application in computer experiments. *Journal of the American Statistical Association*, 3, 1430–1439.