Orthogonal-Maximin Latin Hypercube Designs

V. Roshan Joseph and Ying Hung
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205, USA
roshan@isye.gatech.edu and yhung@isye.gatech.edu

Abstract

A randomly generated Latin hypercube design (LHD) can be quite structured: the variables may be highly correlated or the design may not have good space-filling properties. There are procedures to find good LHDs by minimizing the pairwise correlations or maximizing the inter-site distances. In this article we have shown that these two criteria need not agree with each other. In fact, maximization of inter-site distances can result in LHDs where the variables are highly correlated and vice versa. Therefore, we propose a multi-objective optimization approach to find good LHDs by combining correlation and distance performance measures. We also propose a new exchange algorithm for efficiently generating such designs. Several examples are presented to show that the new algorithm is fast and that the optimal designs are good in terms of both the correlation and distance criteria.

KEY WORDS: Computer experiments, Kriging, Multi-objective optimization, Simulated Annealing.

1. Introduction

Computer experiments are widely used for the design and development of products (for examples, see Fang, Li, and Sudjianto 2006). In computer experiments, instead of physically doing an experiment on the product, mathematical models describing the performance of the product are developed using engineering/physics laws and solved on computers through numerical methods such as the finite element method. Because deterministic models are used for experiments, 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 (see Sacks et al. 1989). For example, replication is not required. In fact, it is desirable to avoid replicates when projecting the design on to a subset of factors. This is because a few out of the numerous factors in the system usually dominate the performance of the product (known as effect sparsity principle). 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. This can be achieved by using a Latin Hypercube Design (LHD) (McKay, Beckman, and Conover 1979). An LHD has the property that by projecting an n-point design on to any factor, we will get n different levels for that factor. This property makes an LHD very suitable for computer experimentation.

Suppose the n levels of a factor are denoted by $1, 2, \cdots, n$. Figure 1a shows an LHD with two factors in six points. In general, an n-run LHD can be generated using a random permutation of $\{1, 2, \cdots, n\}$ for each factor. Each permutation leads to a different LHD. For k factors, we can thus obtain $(n!)^k$ LHDs. Figure 1b shows one such LHD. Clearly, this is not a good design. It is not good due to the following two reasons. First, the two factors are perfectly correlated. Therefore, we will not be able to distinguish between the effects of the two factors based on this experiment. Second, there is a large area in the experimental region that is not explored. Therefore, if we use such a design to develop a prediction model, then the prediction will be poor in those unexplored areas.

There has been some work in the literature to avoid the above problems and obtain a "good" LHD. The idea is to find the best design by optimizing a criterion that describes a

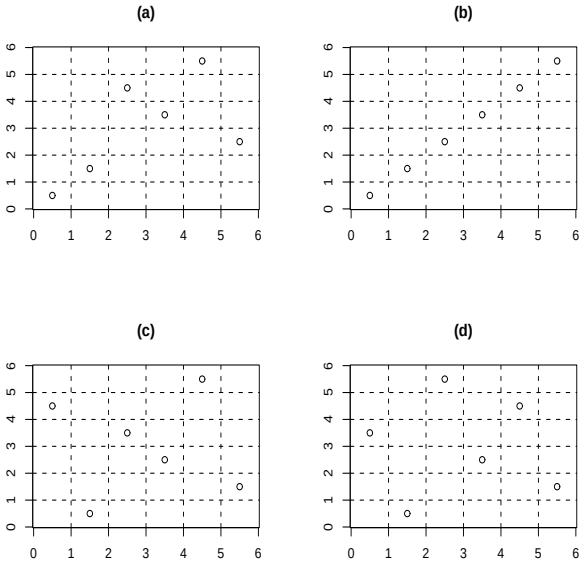

Figure 1: LHDs

desirable property of the design. Iman and Conover (1982), Owen (1994), and Tang (1998) proposed to find designs minimizing correlations among factors. Figure 1c shows the optimal LHD found by the procedure in Tang (1998), which is clearly much better than the one in Figure 1a and 1b. As discussed before, apart from the correlations we are also interested in spreading the points out across the experimental region. This is the idea behind space-filling designs. Morris and Mitchell (1995) proposed to find the best LHD by maximizing the minimum distance between the points. The optimal LHD under this criterion is shown in Figure 1d. Other approaches to find good LHDs are given by Owen (1992), Tang (1993), Park (1994), Ye (1998), Ye, Li, and Sudjianto (2000), and Jin, Chen, and Sudjianto (2005).

The minimum pairwise correlation between the factors and the maximum distance between the points are both good criteria for finding optimal LHDs. Intuitively, minimizing correlation should spread out the points and maximizing the distance between the points should reduce the correlation. But in reality, there is no one-to-one relationship between these two criteria. In fact, the designs obtained by these two criteria can be entirely different. To illustrate this, consider again an LHD with six points and two factors. There

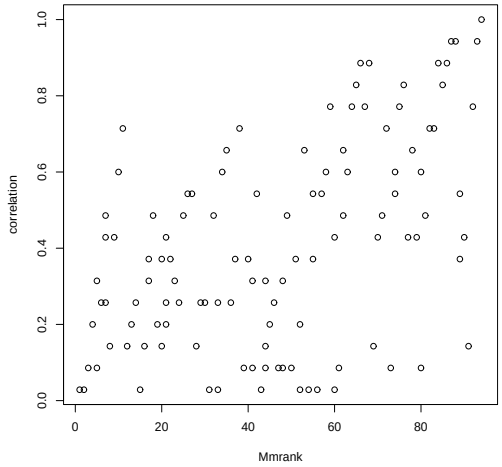

Figure 2: maximin rank vs correlation in $n = 6, k = 2$ case

are a total of $(6!)^2 = 518,400$ LHDs. The designs can be ranked based on the maximin distance criterion (Mitchell and Morris 1995), where the rank 1 is given to the best design. They are plotted in Figure 2 against absolute values of correlations (there are a total of 113 different combinations of correlations and maximin ranks in this example). We can see that the points are highly scattered showing that the minimization of one criterion may not lead to the minimization of the other criterion (see Figure 3 for an example.) The problem becomes more serious as the number of points or the number of factors is increased. This motivates us to develop a multi-objective criterion that minimizes the pairwise correlations as well as maximize the inter-site distances.

Because of the huge combinatorial nature of the problem, finding the optimal LHD is a very difficult task. Several algorithms such as simulated annealing (Morris and Mitchell 1995), columnwise-pairwise algorithms (Ye, Li, and Sudjianto 2000), enhanced stochastic evolutionary algorithms (Jin, Chen, and Sudjianto 2005), etc. are proposed in the literature for finding the optimal LHD. Most of the algorithms use an exchange method for searching in the design space. For example, in the algorithm proposed by Morris and Mitchell, a column in the design is randomly selected and then two randomly chosen elements within that column are exchanged to find a new design. We observed that the columns in the design

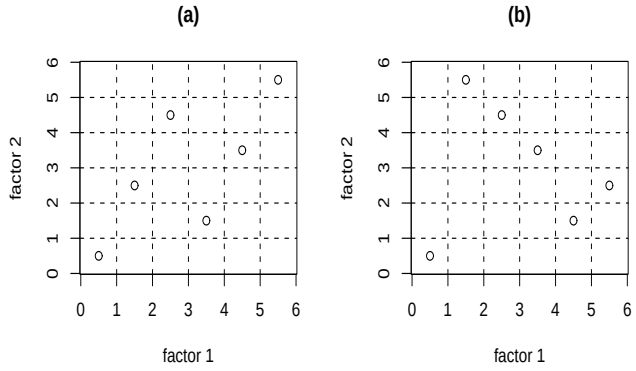

Figure 3: LHDs with $n = 6$ and $k = 2$. (a) correlation=0.714, maximin rank=11. (b) correlation=0.086, maximin rank=80.

matrix correspond to the experimental factors and thus we can choose them deterministically to reduce the pairwise correlations. Similarly, the rows in the design matrix correspond to the points in the experimental region and thus the elements can be chosen to maximize the inter-site distances. These observations lead to a new algorithm, which is highly suitable for finding the optimum based on our multi-objective criterion.

The article is organized as follows. In Section 2, performance measures for evaluating the goodness of an LHD with respect to pairwise correlations and inter-site distances are described. In Section 3, we propose a multi-objective criterion combining the two performance measures. In Section 4, we propose a new algorithm for generating optimal designs. Several examples are presented in Section 5 and a statistical justification for the new criterion is given in Section 6.

2. Performance Measures

Iman and Conover (1982), Owen (1994), and Tang (1998) proposed to choose designs by minimizing correlations among factors within the class of LHDs. We will use the following performance measure proposed by Owen, for evaluating the goodness of the LHD with respect to pairwise correlations. It is defined as

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

where ρ_{ij} is the linear correlation between columns i and j.

Now we will discuss a performance measure based on the inter-site distances. Let $\boldsymbol{X}$ be the design, which is an $n \times k$ matrix. Let $\boldsymbol{s}$ and $\boldsymbol{t}$ be any two design points (or sites). Consider the distance measure $d(\boldsymbol{s}, \boldsymbol{t}) = \{\sum_{j=1}^{k} |s_j - t_j|^p\}^{1/p}$, in which $p = 1$ and $p = 2$ correspond to the rectangular and Euclidean distances respectively. Johnson, Moore, and Ylvisaker (1990) proposed the maximin distance criterion, which maximizes the minimum inter-site distance. Morris and Mitchell (1995) applied this criterion to the class of LHDs to find the optimal LHD. Because there are many designs that maximize the minimum inter-site distance, they proposed an extended definition of the maximin criterion. For a given LHD, define a distance list $(D_1, D_2, ..., D_m)$ in which the elements are the distinct values of inter-site distances, sorted from the smallest to the largest. Hence $m \leq \binom{n}{2}$. Let J_i be the number of pairs of sites in the design separated by D_i. Then a design $\boldsymbol{X}$ is called a maximin design if it sequentially maximizes D_i's and minimizes J_i's in the following order: $D_1, J_1, D_2, J_2, \cdots, D_m, J_m$. Morris and Mitchell (1995) then proposed a scalar-valued function which can be used to rank competing designs in such a way that the maximin design received the highest ranking. The family of functions indexed by p is given by

$$\phi_p = (\sum_{i=1}^{m} J_i D_i^{-p})^{1/p}, \tag{2}$$

where p is a positive integer. Now for large enough p, the design that minimizes ϕ_p will be a maximin design. In the next section we propose a new criterion which combines the performance measures in (1) and (2).

3. Multi-Objective Criterion

Our objective is to find an LHD that minimizes both ρ^2 and ϕ_p. A common approach in multi-objective optimization is to optimize a weighted average of all the objective functions. Therefore consider the objective function

$$w_1 \rho^2 + w_2 \phi_p,$$

where w_1 and w_2 are some pre-specified positive weights. Because the two objectives are very different, it is not easy to choose appropriate weights. Moreover, the two objectives

have different scales. The objective function $\rho^2 \in [0,1]$, whereas the objective function ϕ_p can be more than 1. If we can scale ϕ_p also to $[0,1]$, then we might be able to assign some reasonable weights. In order to do this, we need to find an upper and lower bound for ϕ_p. This is what we try to do in the following.

Consider an LHD with n points and k factors, denoted by $LHD(n,k)$. Suppose each factor takes values in $\{1,2,\cdots,n\}$. Let $d_1, d_2, \cdots, d_{\binom{n}{2}}$ be the inter-site distances among the n points based on the rectangular distance measure $d(\boldsymbol{s},\boldsymbol{t}) = \sum_{j=1}^{k} |s_j - t_j|$. We will use the following two results for deriving bounds for ϕ_p. All the proofs are given in the Appendix.

LEMMA 1. *For an $LHD(n,k)$, the average inter-site distance (rectangular measure) is a constant given by*

$$\overline{d} = \frac{(n+1)k}{3}.$$

LEMMA 2. *Consider a set of positive values $\{d_{j1}, d_{j2}, \cdots, d_{jm}\}$ and denote its ordered sequence by $d_{j(1)} \le d_{j(2)} \le \cdots \le d_{j(m)}$ for $j = 1,2,\cdots,k$. Then*

$$\sum_{i=1}^{m} \frac{1}{\sum_{j=1}^{k} d_{ji}} \le \sum_{i=1}^{m} \frac{1}{\sum_{j=1}^{k} d_{j(i)}}.$$

Lemma 1 shows that for all LHDs, the average distance is a constant for a given n and k. As an interesting consequence, note that the last step in the definition of maximin criterion cannot be applied to an LHD, because D_m is determined by $D_1, \cdots, D_{m-1}$. Therefore, it is more appropriate to define the objective function for the distances as $(\sum_{i=1}^{m-1} J_i D_i^{-p})^{1/p}$. But we will continue to use (2), because it has a computationally simpler form (Jin, Chen, Sudjianto, 2005). It can be written as

$$\phi_p = \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{d_i^p} \right)^{1/p},$$

which can be easily calculated (no need to order the d_i's) .

Let

$$\phi_{p,L} = \left\{ \binom{n}{2} \left(\frac{\lceil \bar{d} \rceil - \bar{d}}{\lfloor \bar{d} \rfloor^p} + \frac{\bar{d} - \lfloor \bar{d} \rfloor}{\lceil \bar{d} \rceil^p} \right) \right\}^{1/p} \quad \text{and} \quad \phi_{p,U} = \left\{ \sum_{i=1}^{n-1} \frac{(n-i)}{(ik)^p} \right\}^{1/p},$$

where $\lfloor x \rfloor$ is the largest integer $\leq x$ and $\lceil x \rceil$ is the smallest integer $> x$. The following result states that the above two values can be used as a lower and upper bound for ϕ_p.

PROPOSITION 1. *For an* $LHD(n,k)$, $\phi_{p,L} \leq \phi_p \leq \phi_{p,U}$.

It is easy to see that the upper bound is achieved when all of the factors are equal. Thus the worst design in terms of ϕ_p is the same as the worst design in terms of ρ. However, there may not exist a design that achieves the lower bound.

Thus $(\phi_p - \phi_{p,L})/(\phi_{p,U} - \phi_{p,L}) \in [0,1]$ has the same range as ρ^2. Therefore, our new criterion is to minimize

$$\psi_p = w\rho^2 + (1-w)\frac{\phi_p - \phi_{p,L}}{\phi_{p,U} - \phi_{p,L}},$$

where $w \in (0,1)$. The case of $w = 0.5$ gives approximately equal importance to both the correlation and the distance measures. We will call a design that minimizes ψ_p as an orthogonal-maximin Latin hypercube design (OMLHD). In the next section we propose a new algorithm for finding an OMLHD.

4. A New Algorithm

Morris and Mitchell (1995) proposed a version of the simulated annealing algorithm for optimizing ϕ_p. We will call their algorithm as MMA. In MMA, the search begins with a randomly chosen LHD, and proceeds through the examination of a sequence of designs, each generated as a perturbation of the preceding one. A perturbation $\boldsymbol{X}_{try}$ of a design $\boldsymbol{X}$ is generated by interchanging two randomly chosen elements within a randomly chosen column in $\boldsymbol{X}$. The perturbation $\boldsymbol{X}_{try}$ replaces $\boldsymbol{X}$ if it leads to an improvement. Otherwise, it will replace $\boldsymbol{X}$ with probability $\pi = exp\{-[\phi_p(\boldsymbol{X}_{try}) - \phi_p(\boldsymbol{X})]/t\}$, where t is a preset parameter known as "temperature".

We propose a modification of the above algorithm. Instead of randomly choosing a column and two elements within that column, we will choose them judiciously in order to achieve improvement in our multi-objective function. Suppose at some stage of the iterations,

a column is almost orthogonal to the other columns. Then clearly, we will not gain much in perturbing this column. It is much better to choose a column that is highly correlated with the other columns, because through a perturbation of its elements we may be able to reduce the correlation, thereby improving our objective function. Similarly, if a point is far from the other points, then there is no need to perturb the elements in that row. Instead, we can choose a point that is close to the other points and perturb the elements in the chosen column. This may increase the distance of the point from the others, thereby improving our objective function. For doing this, at each step, compute

$$\rho_l^2 = \frac{1}{k-1}\sum_{j\neq l}\rho_{lj}^2, \tag{3}$$

for each column $l = 1, 2, \cdots, k$ and

$$\phi_{pi} = \left(\sum_{j\neq i} 1/d_{ij}^p\right)^{1/p}, \tag{4}$$

for each row $i = 1, 2, \cdots, n$, where ρ_{lj} is the correlation between columns l and j; and d_{ij} is the distance between the rows i and j. Thus ρ_l^2 and ϕ_{pi} can be used as measures for correlation and distance for each column and row respectively. For exchanging the elements, we want to choose a column with high probability that is highly correlated with the other columns. Similarly, we want to choose a row with high probability that is closest to the other rows. Therefore choose the column

$$l^* = l \text{ with probability } P(l) = \frac{\rho_l^\alpha}{\sum_{l=1}^k \rho_l^\alpha},$$

and the row

$$i^* = i \text{ with probability } P(i) = \frac{\phi_{pi}^\alpha}{\sum_{i=1}^n \phi_{pi}^\alpha},$$

with $\alpha \in [1, \infty)$. Note that if ρ_l (or ϕ_{pi}) is high for a column (or row), then it will be chosen with a higher probability than the other columns (or rows). This step makes our algorithm different from the existing algorithms. Now exchange $x_{i^*l^*}$ with a randomly chosen element $x_{i'l^*}$. This gives us the new design $\boldsymbol{X}_{try}$. If $\psi_p(\boldsymbol{X}_{try}) < \psi_p(\boldsymbol{X})$, then we will replace $\boldsymbol{X}$ by $\boldsymbol{X}_{try}$, otherwise we will replace it with probability $\pi = exp\{-[\psi_p(\boldsymbol{X}_{try}) - \psi_p(\boldsymbol{X})]/t\}$.

All the parameters in the new algorithm are set the same as that used in a standard simulated annealing algorithm for which the convergence is already established (Lundy and Mees 1986). Therefore the new algorithm will also converge to the global optimum. A limiting case of the algorithm is interesting. When $\alpha \to \infty$, the exchange rule becomes deterministic, given by

$$l^* = arg\max_l \rho_l^2 \text{ and } i^* = arg\max_i \phi_{pi}.$$

Under this rule, the transition probability matrix for moving from one design to another design can be reducible, violating one of the conditions required for convergence. But our simulations, given in the next section, show that the convergence is faster with the above modification. Therefore, we recommend it for use in practice.

Because the objective function is evaluated at each iteration of the algorithm, it is extremely important to have a computationally efficient implementation of the objective function (see Jin, Chen, and Sudjianto 2005). Instead of calculating ρ_l^2 and ϕ_{pi} using (3) and (4), we can use the following iterative formulas. Let $(\rho_l^2)^{(s)}$ and $\phi_p^{(s)}$ denote the values of ρ_l^2 and ϕ_p at the iteration step s. Then at step $(s+1)$

$$\phi_{pi}^{(s+1)} = \begin{cases} \left(\sum_{j\neq i} 1/(d_{ij}^{(s+1)})^p\right)^{1/p} & , i = i^{'}, i^* \\ \left((\phi_{pi}^{(s)})^p - (d_{i^*i}^{(s)})^{-p} - (d_{i^{'}i}^{(s)})^{-p} + (d_{i^*i}^{(s+1)})^{-p} + (d_{i^{'}i}^{(s+1)})^{-p}\right)^{1/p} & , i \neq i^{'}, i^* \end{cases}.$$

For all $j \neq i^*, i^{'}$ we have $d_{i^*j}^{(s+1)} = d_{i^*j}^{(s)} - t(i^*, i^{'}, j, l^*)$, and $d_{i^{'}j}^{(s+1)} = d_{i^{'}j}^{(s)} + t(i^*, i^{'}, j, l^*)$, where $t(i_1, i_2, u, v) = |x_{i_1v} - x_{uv}| - |x_{i_2v} - x_{uv}|$. Also note that the distance matrix (d_{ij}) is symmetric. For ρ_l^2 at step $(s+1)$, we obtain

$$(\rho_l^2)^{(s+1)} = \begin{cases} \frac{1}{k-1}\sum_{j\neq l}(\rho_{jl}^2)^{(s+1)} & , l = l^* \\ (\rho_l^2)^{(s)} + \frac{(\rho_{ll^*}^2)^{(s+1)} - (\rho_{ll^*}^2)^{(s)}}{k-1} & , l \neq l^* \end{cases}.$$

Thus

$$\phi_p^{(s+1)} = \left(\frac{1}{2}\sum_{i=1}^{n}(\phi_{pi}^{(s+1)})^p\right)^{1/p} \text{ and } (\rho^2)^{(s+1)} = (\rho^2)^{(s)} + \frac{2(\rho_{l^*}^2)^{(s+1)} - 2(\rho_{l^*}^2)^{(s)}}{k}.$$

We should point out that the proposed exchange procedure can also be implemented with any of the other stochastic optimization algorithms such as the columnwise-pairwise

Table 1: Example 1, MLHD vs OMLHD for $n = 5$ and $k = 3$

	MLHD	OMLHD
optimal design matrix	1 1 2 2 5 3 3 2 5 4 3 1 5 4 4	1 2 3 2 4 5 3 5 1 4 1 2 5 3 4
ϕ_p	0.2170	0.2201
$D_1(J_1)$	5(3)	5(4)
ρ	0.265	0.081
pairwise correlations	(0.4,0.2,0.1)	(-0.1,-0.1,0)

algorithm (Li and Wu 1997, Ye, Li, and Sudjianto 2000), the threshold accepting heuristic (Winker and Fang 1998), and the stochastic evolutionary algorithm (Jin, Chen, and Sudjianto 2005).

5. Examples

In this section, we compare our proposed method with some of the existing methods. For a fair comparison, we choose all the parameters in the simulated annealing algorithm equal to the recommended values in Morris and Mitchell (1995). In the following examples, we let $p = 15$ and $w = 0.5$. In all the examples, we started the iteration by using a randomly generated symmetric LHD (Ye, Li, and Sudjianto 2000).

Example 1 (*OMLHD vs MLHD*). Consider an $LHD(5, 3)$. In this case it is feasible to enumerate all the LHDs. We found that there are a total of 142 different designs according to the maximin criterion (Morris and Mitchell, 1995). The maximin Latin hypercube design (MLHD) and the proposed OMLHD are given in Table 1. We see that for OMLHD, the maximum pairwise correlation is only 0.1 compared to 0.4 of MLHD. The minimum inter-site distances of the two designs are the same ($D_1 = 5$), although the number of sites separated by this distance is one less in MLHD.

Example 2 (*OMLHD vs OLHD*). Ye (1998) proposed the orthogonal Latin hypercube designs (OLHD), in which all the columns are orthogonal (correlation = 0) to each other.

Table 2: Examples 2 and 4, OMLHD vs OLHD, MLHD, and ULHD for $n = 9$ and $k = 4$

	MLHD	OMLHD	OLHD	ULHD
optimal design matrix	1 3 3 4 2 5 8 8 3 8 6 2 4 7 1 6 5 2 9 3 6 9 5 9 7 1 4 7 8 4 2 1 9 6 7 5	1 5 3 3 2 2 5 8 3 9 7 5 4 3 8 1 5 7 1 7 6 6 9 9 7 1 2 4 8 8 4 2 9 4 6 6	1 2 6 3 2 9 7 6 3 4 2 9 4 7 1 2 5 5 5 5 6 3 9 8 7 6 8 1 8 1 3 4 9 8 4 7	4 1 7 5 1 3 4 3 9 9 5 4 6 6 6 9 5 7 2 1 2 8 8 7 3 5 1 6 8 2 3 8 7 4 9 2
ϕ_p	0.1049	0.1049	0.1154	0.1127
$D_1(J_1)$	11(3)	11(4)	10(8)	10(5)
ρ	0.108	0.063	0	0.076
maximum pairwise correlation	0.217	0.117	0	0.15
CL_2	0.1415	0.1386	0.1457	0.1374

Table 2 compares the OLHD with the proposed OMLHD for the case of $n = 9$ and $k = 4$. For comparison, the MLHD is also given in the Table. We can see that the OMLHD is a compromise between the MLHD and OLHD. OLHD exists only for certain n and k, whereas MLHD and OMLHD exist for all n and k. In this sense MLHD and OMLHD are more general.

Example 3 (*OMLHD vs OA-based LHD*). Owen (1992) and Tang (1993) proposed using orthogonal arrays for constructing good LHDs. Tang called such designs OA-based LHDs. Figure 4 shows an OA-based LHD and the OMLHD for the case of $n = 9$ and $k = 2$. Clearly the OMLHD is superior to this particular OA-based LHD. Interestingly, in this case, the OMLHD is also an OA-based LHD, but a good one in terms of both correlation and space-filling. However, in general an OMLHD need not be an OA-based LHD.

Example 4 (*OMLHD vs ULHD*). Another popular space-filling design is the unform design. It can be obtained by minimizing the centered L_2-discrepancy criterion (CL_2)(see Fang, Ma, and Winker 2000). Denote the optimal LHD under this criterion by ULHD. The ULHD for $n = 9$ and $k = 4$ is given in Table 2. We can see that the OMLHD is slightly worse

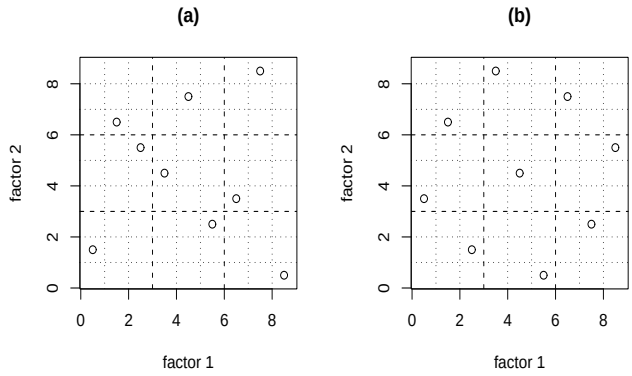

Figure 4: Example 3, (a) OA-based LHD ($\phi_p = 0.5380, D_1(J_1) = 2(3), \rho = -0.067$) (b)OMLHD ($\phi_p = 0.2879, D_1(J_1) = 4(8), \rho = 0$)

than the ULHD under this criterion, but is better in terms of both ϕ_p and ρ. Interestingly, the OMLHD performs much better than MLHD and OLHD in terms of CL_2.

We have also studied the performance of the proposed exchange algorithm. Figure 5 shows how ϕ_p and ρ^2 are reduced with each iteration for the case of $LHD(25, 4)$. The same starting design is used for both MMA and the new algorithm. We can see that the new algorithm converges more quickly than the MMA. We repeated this 200 times. The values of ψ_p at the 50^{th} iteration are plotted in Figure 6. We can see that they are much smaller for the new algorithm compared to the MMA. Thus for a fixed number of iterations, the new algorithm produces LHDs with smaller pairwise correlations and larger inter-site distances. The simulations are repeated for $LHD(50, 4)$, $LHD(10, 10)$, and $LHD(100, 10)$. The number of iterations for each of these cases was fixed at 100, 200, and 500 respectively. The results are shown in Figure 6. We can see that remarkable improvements are obtained by using the new algorithm compared to the MMA.

6. A Statistical Justification

Because of the absence of random errors, interpolating methods such as kriging are widely used for modeling and analysis in computer experiments. Consider a function $y(\boldsymbol{x})$, where

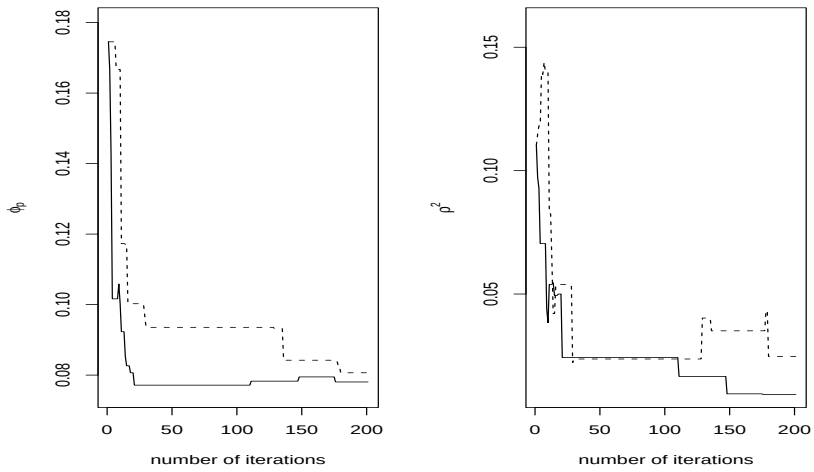

Figure 5: Performance of our new algorithm (solid) and MMA (dashed) against the number of iterations.

$\boldsymbol{x} = (x_1, \cdots, x_k)'$. The ordinary kriging model is given by,

$$Y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}), \tag{5}$$

where $Z(\boldsymbol{x})$ is a weak stationary stochastic process with mean 0 and covariance function $\sigma^2 R$. A popular choice for the correlation function is the exponential correlation function:

$$R(\boldsymbol{h}) = e^{-\theta \sum_{i=1}^{k} |h_i|^\gamma}, \tag{6}$$

with $\theta \in (0, \infty)$ and $\gamma \in (0, 2]$. Suppose we evaluated the function at n points $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ and let $\boldsymbol{y} = (y_1, \cdots, y_n)'$ be the corresponding function values. Then, the best linear unbiased predictor (BLUP) is given by $\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1})$, where $\mathbf{1}$ is a column of 1's having length n, $\boldsymbol{r}(\boldsymbol{x})' = (R(\boldsymbol{x} - \boldsymbol{x}_1), \cdots, R(\boldsymbol{x} - \boldsymbol{x}_n))$, $\boldsymbol{R}$ is an $n \times n$ matrix with elements $R(\boldsymbol{x}_i - \boldsymbol{x}_j)$, and $\hat{\mu} = \mathbf{1}'\boldsymbol{R}^{-1}\boldsymbol{y}/\mathbf{1}'\boldsymbol{R}^{-1}\mathbf{1}$. Note that the model in (5) assumes a constant mean and therefore, the predictor does not perform well when there are some trends (see Joseph 2006). If the trends are known, then universal kriging can be used instead of ordinary kriging. The universal kriging model with linear trends is given by

$$Y(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + Z(\boldsymbol{x}), \tag{7}$$

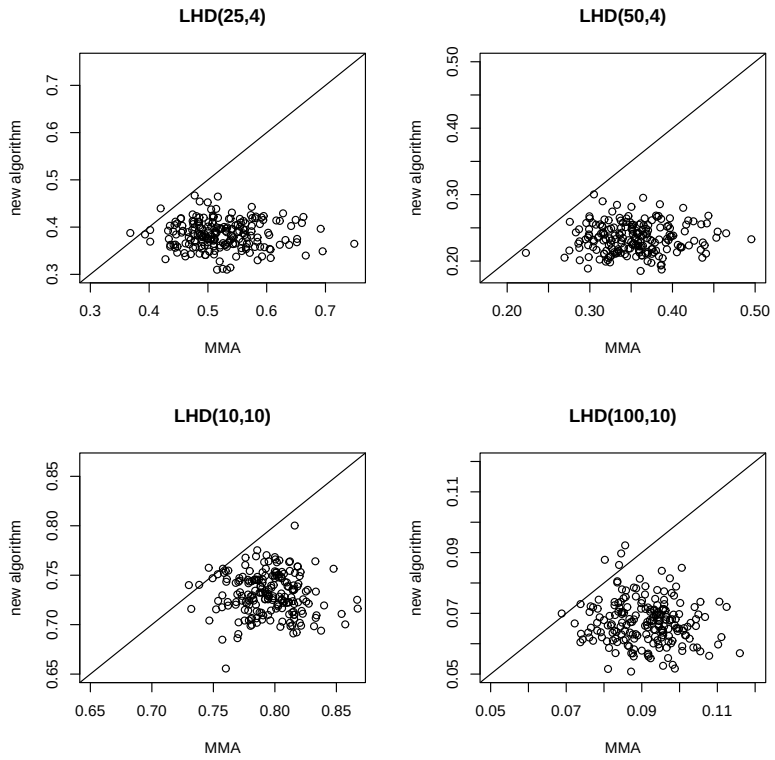

Figure 6: Plot of ψ_p values from the new algorithm against that of the MMA.

where $\beta_0, \beta_1, \cdots, \beta_k$ are some unknown constants. Simulations carried out by Martin and Simpson (2005) show that universal kriging can improve the prediction over ordinary kriging. See Qian et al. (2006) for a real application of universal kriging with linear trends.

Johnson, Moore, and Ylvisaker (1990) have shown that the maximin design with minimum J_1 is asymptotically D-optimum under the ordinary kriging model (as correlation becomes weak). Thus the objective of a maximin design can be thought of as finding a design to improve prediction through the stochastic part $Z(\boldsymbol{x})$. Whereas minimizing the correlation among the variables will help in estimating the deterministic mean part $\beta_0 + \sum_{i=1}^{k} \beta_i x_i$ efficiently. For the universal kriging predictor to perform well, both parts need to be estimated precisely. Thus the orthogonal-maximin LHD can be considered suitable for the efficient estimation of the universal kriging model with linear trends.

More specifically, consider the following hierarchical Bayesian model:

$$\boldsymbol{y}|\boldsymbol{\beta} \sim N(\boldsymbol{F}\boldsymbol{\beta}, \sigma^2\boldsymbol{R}), \;\; \boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \tau^2\boldsymbol{I}),$$

where $\boldsymbol{F} = [\boldsymbol{1}, \boldsymbol{X}]$ is the model matrix corresponding to $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_k)'$ and $\boldsymbol{I}$ is an identity matrix. The maximum entropy design (Shewry and Wynn 1987) is obtained by maximizing the determinant of the variance-covariance matrix of $\boldsymbol{y}$. Thus we need to maximize $det(\sigma^2\boldsymbol{R} + \tau^2\boldsymbol{F}\boldsymbol{F}')$, which is equal to (see Santner, Williams, and Notz 2003, page 167)

$$det(\sigma^2\boldsymbol{R})\; det(\tau^2/\sigma^2\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F} + \boldsymbol{I}).$$

Johnson, Moore, and Ylvisaker (1990) have shown that as $\theta \to \infty$ in (6), a maximin design maximizes the first term $det(\sigma^2\boldsymbol{R})$. As $\theta \to \infty$, $\tau^2/\sigma^2\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F} + \boldsymbol{I} \to \tau^2/\sigma^2\boldsymbol{F}'\boldsymbol{F} + \boldsymbol{I}$, whose determinant is maximized when $\boldsymbol{F}$ is orthogonal. Thus an orthogonal design maximizes the second term. A design will be asymptotically ($\theta \to \infty$) optimum with respect to the maximum entropy criterion if both the terms are maximized. Therefore, an OMLHD, which possesses good maximin and orthogonality properties, can be expected to perform well in terms of the maximum entropy criterion for the model in (7) among all LHDs.

Acknowledgements

We would like to thank Zhiguang Qian and two referees for their valuable comments and suggestions. This research was supported by US National Science Foundation CAREER grant DMI-0448774.

Appendix: Proofs

Proof of Lemma 1

Since each column in the $LHD(n, k)$ is a permutation of $\{1, 2, \cdots, n\}$, we have

$$\sum_{i=1}^{\binom{n}{2}} d_i = \sum_{i=2}^{n}\sum_{j=1}^{i-1} d(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sum_{l=1}^{k}\sum_{i=2}^{n}\sum_{j=1}^{i-1} |s_{il} - s_{jl}| = k\sum_{i=2}^{n}\sum_{j=1}^{i-1} |s_{i1} - s_{j1}|.$$

Without loss of generality, we can take the first column as $(1, 2, \cdots, n)'$. Therefore,

$$\sum_{i=1}^{\binom{n}{2}} d_i = k \sum_{i=2}^{n} \sum_{j=1}^{i-1} |i-j| = k \sum_{i=2}^{n} \frac{i(i-1)}{2} = \frac{kn(n^2-1)}{6}.$$

Thus,

$$\overline{d} = \frac{kn(n^2-1)/6}{\binom{n}{2}} = \frac{(n+1)k}{3}\ .$$

Proof of Lemma 2

For $m = 2$,

$$\frac{1}{\sum_{j=1}^{k} d_{j1}} + \frac{1}{\sum_{j=1}^{k}(c_j - d_{j1})} = \frac{\sum_{j=1}^{k} c_j}{\sum_{j=1}^{k} d_{j1} \times \sum_{j=1}^{k}(c_j - d_{j1})},$$

where $c_j = d_{j1} + d_{j2}$, for all $j = 1, 2, \cdots, k$. Since $\sum_{j=1}^{k} c_j$ is a constant, it is easy to see that the right side is a maximum when $\sum_{j=1}^{k} d_{j1} = \sum_{j=1}^{k} d_{j(1)}$. Therefore,

$$\sum_{i=1}^{2} \frac{1}{\sum_{j=1}^{k} d_{ji}} \leq \sum_{i=1}^{2} \frac{1}{\sum_{j=1}^{k} d_{j(i)}}.$$

Thus, the result holds for $m = 2$. Assume the upper bound is achieved by the ordered sequence for $m = M$. When $m = M + 1$, suppose the upper bound is achieved by some unordered sequence $\{d_{j1^*}, \cdots, d_{jM+1^*}\}$. So the upper bound is $\sum_{i=1}^{M+1} \frac{1}{\sum_{j=1}^{k} d_{ji^*}}$. Without loss of generality, assume that at least the first sequence does not follow the order. Because of this, there always exists an M-element subset $\{d_{11^*}, \cdots, \widehat{d_{1t^*}} \cdots, d_{1M+1^*}\}$ that does not follow the order, where the notation $\widehat{d_{1t^*}}$ means that the sequence is without d_{1t^*}. But since the upper bound holds for $m = M$, we have

$$\frac{1}{\sum_{j=1}^{k} d_{j1^*}} + \cdots + \widehat{\frac{1}{\sum_{j=1}^{k} d_{jt^*}}} \cdots + \frac{1}{\sum_{j=1}^{k} d_{jM+1^*}} \leq \frac{1}{\sum_{j=1}^{k} d_{j(1)}} + \cdots + \widehat{\frac{1}{\sum_{j=1}^{k} d_{jt^*}}} \cdots + \frac{1}{\sum_{j=1}^{k} d_{j(M+1)}}.$$

This is a contradiction, because by adding $1/\sum_{j=1}^{k} d_{jt^*}$ to both sides we obtain

$$\sum_{i=1}^{M+1} \frac{1}{\sum_{j=1}^{k} d_{ji^*}} \leq \frac{1}{\sum_{j=1}^{k} d_{j(1)}} + \cdots + \frac{1}{\sum_{j=1}^{k} d_{jt^*}} \cdots + \frac{1}{\sum_{j=1}^{k} d_{j(M+1)}},$$

which is a better upper bound. By mathematical induction, we can prove that the function achieves the upper bound when all k sequences are in increasing order.

Proof of Proposition 1

To find a lower bound for ϕ_p, consider the following minimization problem with respect to $d_1, \cdots, d_{\binom{n}{2}}$.

$$\min \phi_p = \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{d_i^p} \right)^{1/p} \quad \text{subject to} \quad \sum_{i=1}^{\binom{n}{2}} d_i = \binom{n}{2}\bar{d},$$

where $\bar{d} = (n+1)k/3$. Using the Lagrange multiplier method, it is easy to show that the optimal solution is given by $d_1 = d_2 = \cdots = d_{\binom{n}{2}} = \bar{d}$. Therefore, $\binom{n}{2}^{1/p}/\bar{d}$ is a lower bound for ϕ_p. But since we know the d_i's in an LHD are integers, a better lower bound can be obtained by adding this constraint to the above optimization problem. To find the optimal solution under the integer restriction, consider the following two groups: $I = \{i : d_i \leq \bar{d}\}$ and $II = \{i : d_i > \bar{d}\}$. Since the sum of the d_i's is a constant, if we increase a d_i for an $i \in I$, then we should decrease a d_i, $i \in II$, by the same amount . It is easy to show that such a change will decrease ϕ_p. Therefore, the minimum of ϕ_p can be achieved by

$$d_1 = \cdots = d_N = \lfloor \bar{d} \rfloor \ ; \ d_{N+1} = \cdots = d_{\binom{n}{2}} = \lceil \bar{d} \rceil,$$

provided such an N exists. We must have $N\lfloor \bar{d} \rfloor + \{\binom{n}{2} - N\}\lceil \bar{d} \rceil = \binom{n}{2}\bar{d}$, which gives $N = \binom{n}{2}(\lceil \bar{d} \rceil - \bar{d})$. This is a feasible solution, because $\binom{n}{2}\bar{d} = (n+1)k$ is an integer. Thus

$$\phi_p \geq \left(\frac{N}{\lfloor \bar{d} \rfloor^p} + \frac{\binom{n}{2} - N}{\lceil \bar{d} \rceil^p} \right)^{1/p} = \phi_{p,L}.$$

Now consider the upper bound. All the k factors have the same inter-site distances $\{d_{j,1}, \cdots, d_{j,\binom{n}{2}}\}$, where $j = 1, \cdots, k$. For example, if $n = 5$, the inter-site distances for each factor is $\{1, 1, 1, 1, 2, 2, 2, 3, 3, 4\}$. In general, $(n-1)$ of the $d_{j,i}$'s are 1, $(n-2)$ of the $d_{j,i}$'s are 2, $\cdots$, and one is $(n-1)$. Different LHDs have different combinations of the inter-site distances of each factor. Therefore $d_i = \sum_{j=1}^{k} d_{j,i}$, where $i = 1, \cdots, \binom{n}{2}$. By Lemma 2

$$\phi_p = \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{d_i^p} \right)^{1/p} = \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{\sum_{j=1}^{k} d_{j,i}^p} \right)^{1/p} \leq \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{\sum_{j=1}^{k} d_{j,(i)}^p} \right)^{1/p}$$

Note that the inter-site distances of each of the k factors is ordered in the same way. Therefore, $(n-1)$ of the d_i's are k, $(n-2)$ of the d_i's are $2k$, $\cdots$, and one is $(n-1)k$. Thus

$$\phi_p \leq \left(\sum_{i=1}^{\binom{n}{2}} \frac{1}{\sum_{j=1}^{k} d_{j,i}^p} \right)^{1/p} = \left\{ \sum_{i=1}^{n-1} \frac{(n-i)}{(ik)^p} \right\}^{1/p} = \phi_{p,U}.$$

References

Fang, K. T., Ma, C. X. and Winker, P. (2002). Centered L_2-discrepancy of random sampling and Latin hypercube design, and construction of uniform designs. *Math. Comp.* **71**, 275-296.

Fang, K. T., Li, R. and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments.* CRC Press, New York.

Iman, R. L. and Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Comm. Statist. Part B–Simulation Comput.* **11**, 311-334.

Jin, R., Chen, W. and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *J. Statist. Plann. Inference* **134**, 268-287.

Johnson, M., Moore, L. and Ylvisaker, D. (1990). Minimax and maximin distance design. *J. Statist. Plann. Inference* **26**, 131-148.

Joseph, V. R. (2006). Limit kriging. *Technometrics*, to appear.

Li, W. W. and Wu, C. F. J. (1997). Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. *Technometrics* **39**, 171-179.

Lundy, M. and Mees, A. (1986). Convergence of an annealing algorithm. *Mathematical Programming* **34**, 111-124.

Martin, J. D. and Simpson, T. W. (2005). On the Use of Kriging Models to Approximate Deterministic Computer Models. *AIAA Journal* **43**, 853-863.

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**, 239-245.

Morris, M. D. and Mitchell, T. J. (1995). Exploratory Designs for Computer Experiments. *J. Statist. Plann. Inference* **43**, 381-402.

Owen, A. (1994). Controlling correlations in Latin hypercube samples. *J. Amer. Statist. Assoc.* **89**, 1517-1522.

Park, J. S. (1994). Optimal Latin-hypercube designs for computer experiments. *J. Statist. Plann. Inference* **39**, 95-111.

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K. and Wu, C. F. J (2006). Building surrogate models based on detailed and approximate simulations. *ASME Journal of Mechanical Design*, to appear.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4**, 409-423.

Santner, T., Williams, B. and Notz, W. (2003). *The Design and Analysis of Computer Experiments.* Springer, New York.

Shewry, M. C. and Wynn, H. P. (1987). Maximum entropy sampling. *Journal of Applied Statistics* **14**, 165-170.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88**, 1392-1397.

Tang, B. (1998). Selecting Latin hypercubes using correlation criteria. *Statistica Sinica* **8**, 965-978.

Winker, P. and Fang, K. T. (1998). Optimal U-type design, in *Monte Carlo and Quasi-Monte Carlo Methods 1996*, eds. H. Niederreiter, P. Zinterhof and P. Hellekalek, Springer, 436-448.

Ye, K.Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *J. Amer. Statist. Assoc.* **93**, 1430-1439.

Ye, K. Q., Li, W. and Sudjianto, A. (2000). Algorithmic construction of optimal symmetric Latin hypercube designs. *J. Statist. Plann. Inference* **90**, 145-159.