Statistica Sinica **28** (2018), 879-897
doi:https://doi.org/10.5705/ss.202016.0165

A SEQUENTIAL MAXIMUM PROJECTION DESIGN FRAMEWORK FOR COMPUTER EXPERIMENTS WITH INERT FACTORS

Shan Ba[1], William R. Myers[1] and Dianpeng Wang[2]

[1] *The Procter & Gamble Company and* [2] *Chinese Academy of Sciences*

Abstract: Many computer experiments involve a large number of input factors, but many of them are inert and only a subset are important. This paper develops a new sequential design framework that can accommodate multiple responses and quickly screen out inert factors so that the final design is space-filling with respect to the active factors. By folding over Latin hypercube designs with sliced structure, this sequential design can have flexible sample size in each stage and also ensure that each stage, as well as the whole combined design, are all approximately Latin hypercube designs. The sequential framework does not require prescribing the total sample size and, under the presence of inert factors, can lead to substantial savings in simulation resources. Even if all factors are important, the proposed sequential design can still achieve a similar overall space-filling property compared to a maximin Latin hypercube design optimized in a single stage.

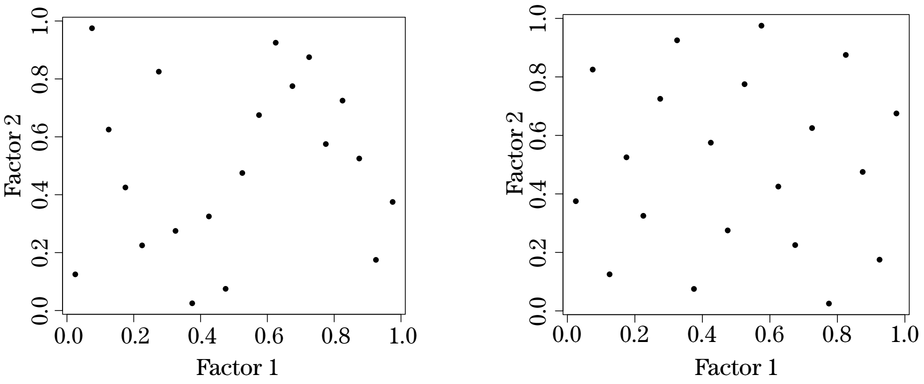
Key words and phrases: Effect sparsity, foldover design, sample size determination, sliced Latin hypercube design, space-filling criterion.

1. Introduction

Computer simulations, based on finite element analysis (FEA) and computational fluid dynamics (CFD), are commonly used to reduce both the need for physical experimentation and the building of a large number of prototypes. Because each simulation run can take hours or days to complete, a common strategy is to develop a surrogate model to approximate the time-consuming computer model with sufficient accuracy (Sacks et al. (1989)). Space-filling designs, which spread out design points evenly throughout the input space, are widely used in designing computer experiments, since computer models are often complex and highly nonlinear (Santner, Williams and Notz (2003)). In this paper, we propose a new sequential space-filling design framework to address challenges that are commonly encountered in practice.

Many computer simulation studies involve a large number of input factors, where many of them are inert/inactive and only a few of them are important. This phenomenon is referred to as *effect sparsity* in the literature (Wu and Hamada (2009)). Because most simulation studies are also deterministic, a good space-filling design in computer experiments needs to be non-collapsing, so that the projection of design points onto the sub-dimension of important factors are non-overlapping. The most popular non-collapsing design for computer experiments is the *Latin hypercube design* (LHD) (McKay, Beckman and Conover (1979)), whose projections onto any single dimension are guaranteed to have distinct levels. Within the class of LHDs, Morris and Mitchell (1995) further proposed to construct the maximin-distance Latin hypercube design (Mm LHD), which ensures good space-filling properties in the full dimensional design space and uniform projections in each single dimension. For other possible sub-dimensions $2, 3, \ldots, p-1$ (where p is the total number of factors), the projections of a Mm LHD are only non-collapsing, but may not have good space-filling properties. For example, Figure 1a shows the projection of an 10-dimension 20-run Mm LHD onto two dimensions, which obviously is not space-filling. If we directly generate a 20-run LHD in two dimensions, its space-filling design points are shown in Figure 1b for comparison. In practice, the number of important factors in computer experiments is usually greater than one and less than the full dimension p, which makes the traditional Mm LHD less attractive. Recently, Joseph, Gul and Ba (2015) proposed the *Maximum Projection* (MaxPro) *design* that maximizes space-filling properties on projections to all possible subsets of factors. Although it largely improves the projection property, a single-stage MaxPro design is still not the most efficient in the sense that it also emphasizes good projection properties in the non-active factor space, as opposed to just the active factors. For some computationally expensive simulation studies, the design ideally should only focus on space-filling properties with respect to the active factors. This directly motivates our sequential design framework that can be considered an extension of the MaxPro idea.

Another challenge in designing computer experiments is to determine the number of runs that enables the surrogate model to achieve a sufficient level of accuracy. As a practical guide, Loeppky, Sacks and Welch (2009) introduced the popular $10d$ rule (10 times the input dimension) as a good rule of thumb for the number of runs in the computer experiment. Nevertheless, when there are many inert factors, it can be a waste of resources to perform all $10d$ simulation runs and we feel that it might be more appropriate to use 10 times the number of active factors d_0 ($d_0 \leq d$). Unfortunately, the value of d_0 cannot be known before we run the experiment. In addition, the number of necessary simulation



(a)

(b)

Figure 1. (a) 20-run Mm LHD in 10 factors (projected onto two factors); (b) 20-run Mm LHD in two factors.

runs should also depend on the complexity of the simulation response surface: for some simple surfaces $10d_0$ runs are more than adequate, but for emulating very complex response surface, we may need much more than $10d_0$ runs to achieve a sufficient level of accuracy.

This discussion motivates the need for a good sequential design for computer experiments, especially when the simulations are time intensive and the total number of input factors is large. Most existing sequential design strategies in the computer experiments literature focus on finding some specific features (such as global optimum) of the expensive black-box function and, as a result, their design points are centered around peaks or specific regions of design space instead of being spread out in the smooth regions. In this paper, we present a sequential design framework that can accommodate multiple responses and quickly screen out inert factors so that the final design is space-filling with respect to active factors. This results in a more efficient use of resources and allows the potential opportunity for fewer overall simulation runs. The remainder of this paper is organized as follows. Section 2 summaries a list of desired goals for the new sequential design to distinguish it from existing methods. Section 3 discusses the choice of sequential design criterion for different stages, and in Section 4 we develop the structure for the sequential design framework. Section 5 is devoted to simulation studies to demonstrate the performance of the proposed strategy. Some concluding remarks are given in Section 6.

2. Desired Properties for the New Sequential Design Framework

For the new sequential design framework, our goal is to achieve the properties: (1) it can quickly screen out inert factors in the computer experiments and construct a space-filing design with respect to only the important factors; (2) it is model independent which supports fitting various types of sophisticated surrogate models; (3) it runs the experiments sequentially in flexible batch size, which is more convenient and practical than the one-run-at-a-time approach; (4) the user does not need to prescribe the total number of stages or the total sample size beforehand, and the algorithm can stop at any stage when the current design is deemed adequate; (5) it can accommodate multiple responses generated by the same simulation; (6) it is not sensitive to the ratio of active/inert factors, and even if all factors turn out to be active, the overall design can still have a similar space-filling property as a single-stage design.

These desirable features intertwine with many existing works in the literature such as Lam and Notz (2008), Gramacy and Lee (2009), Loeppky, Moore and Williams (2010), Moon, Dean and Santner (2012), Duan et al. (2016), to name a few. Each of these existing methods has their own strength but, to the best of the authors' knowledge, there is not an existing approach that can easily fulfill all the above requirements. Technical details of our newly proposed approach are given in the subsequent sections.

3. Sequential MaxPro Criterion

Suppose $D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ denotes an experimental design in n runs for p factors, where each $\boldsymbol{x}_i \in \mathscr{X} = [0,1]^p$. Let $d(\boldsymbol{u}, \boldsymbol{v}) = (\sum_{i=1}^{p} |u_i - v_i|^s)^{1/s}$ be the distance between points $\boldsymbol{u}$ and $\boldsymbol{v}$, and $s = 1$ and $s = 2$ correspond to the rectangular and Euclidean distances, respectively. The maximin distance criterion (Johnson, Moore and Ylvisaker (1990)) improves the property of a design by maximizing the minimum inter-point distance

$$\min_{\boldsymbol{x}_i, \boldsymbol{x}_j \in D_{Mm}} d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \max_{D} \min_{\boldsymbol{x}_i, \boldsymbol{x}_j \in D} d(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

Since a randomly generated LHD may not be space-filling, Morris and Mitchell (1995) proposed an average reciprocal distance criterion to select the Mm LHD which maximizes the minimum inter-point distance among all possible LHDs of the same size:

$$\min_{D} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(\boldsymbol{x}_i, \boldsymbol{x}_j)} \right\}^{1/k}. \tag{3.1}$$

As we have discussed in Section 1, the Mm LHD guarantees good space-filling property of design points in the full dimension p, and in all single dimensions. But if the number of active factors in the computer experiment is between 1 and p, the projections of its design points could be undesirable (Figure 1a), which would inevitably have a negative impact on the accuracy of the surrogate model. Here, we present a sequential design criterion that only focuses on improving the space-filling property of the subspace for those important factors. To achieve this, we consider the weighted distance measure

$$d(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \left(\sum_{i=1}^{p} w_i |u_i - v_i|^s \right)^{1/s},$$

where $\boldsymbol{w} = (w_1, \cdots, w_p)$, $\sum_{i=1}^{p} w_i = 1$, $w_1, \cdots, w_p \geq 0$ and the w_i can be interpreted as the importance of factor i. Using this definition, distance between design points in a sub-dimensional projection can be calculated by setting $w_i > 0$ for the relevant factors and $w_i = 0$ for the factors not involved in this sub-dimension. In our sequential design framework, we propose to set the w_i values proportional to the total sensitivity indices of input factors which can be estimated based on the available data from previous stages. Note that the weight w_i can take any value between 0 and 1, which is more general than just considering a factor to be active or inactive (binary w_i) since it can further distinguish active factors based on their relative importance. After this adjustment, the average reciprocal inter-point distance criterion in (3.1) can be extended to

$$\min_{D} \phi_k(D; \boldsymbol{w}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{w})}, \tag{3.2}$$

which emphasizes on the space-filling property of the sub-dimension spanned by all the active factors and which also assigns higher weights to the factors with higher importance. In our framework, we refer to (3.2) as the *sequential maximum projection* (*Sequential MaxPro*) *criterion.* Its relation with some other existing criteria will be discussed at the end of this section.

If we knew the true relative importance $\boldsymbol{w}$ for all input factors, the sequential MaxPro criterion in (3.2) would be a more accurate space-filling measure since the final surrogate model is only fitted based on the active factors. Many approaches can be used to estimate the total sensitivity indices (relative importance $\boldsymbol{w}$) of input factors using computer experiment outputs from the previous stages. Good reviews of the sensitivity analysis methods based on ANOVA-type decompositions can be found in Saltelli, Chan and Scott (2000) and Santner, Williams and Notz (2003). Although the choice is not unique, a good option is to use the analytical formulas provided in Chen, Jin and Sudjianto (2005) to obtain the sensitivity indices based on a Gaussian process model framework. Although estimates of the total sensitivity indices may not be very precise in earlier stages when fewer data points are available, their accuracy improves after completing each subsequent stage of experiments. If the computer experiment has more than one response variable, we can independently assess the input sensitivities under each response, then use the averages of those sensitivity indices to form the values of $\boldsymbol{w}$ for each factor in (3.2).

During the initial stage when we have no prior information about the input sensitivities, we can assign a uniform prior distribution $p(\boldsymbol{w})$ to the values of $\boldsymbol{w}$ and (3.2) becomes

$$\min_D \int \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{w})} p(\boldsymbol{w}) d\boldsymbol{w}. \tag{3.3}$$

Joseph, Gul and Ba (2015) showed that if we choose $k = sp$, then this criterion can be simplified to

$$\min_D \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{\prod_{l=1}^{p} |\boldsymbol{x}_{il} - \boldsymbol{x}_{jl}|^s}, \tag{3.4}$$

which is the single-stage MaxPro criterion. Based on (3.3), the single-stage MaxPro criterion simultaneously maximizes the space-filling properties in all possible sub-dimensional spaces, and thus it is an ideal criterion for optimizing the first stage of experiments. Starting from the second stage, the factor importance indices $\boldsymbol{w}$ can be estimated and sequentially updated after each stage, and thus the sequential MaxPro criterion in (3.2) is used to optimize the new design points.

It is worth noting that some well-known existing criteria such as the maximum entropy criterion (Shewry and Wynn (1987)) or the integrated mean squared error criterion (Sacks et al. (1989)) can also be used in a sequential design framework for screening out inert factors. We prefer the sequential MaxPro criterion because it works with multiple simulation outputs and, once given the sensitivity index estimates, it is a model-independent criterion that enables the design to support fitting different types of surrogate models. Loeppky, Moore and Williams (2010) discussed using a version of the maximin weighted distance criterion whose weights were chosen as the estimated Gaussian process correlation parameters. Except for the first stage, it is similar to the sequential MaxPro criterion. However, a correlation parameter mainly determines the smoothness but may not accurately reflect the sensitivity or importance of that factor in influencing the response. The sensitivity index, on the other hand, depends not only on the estimated correlation, but also on how the observed responses vary in the corresponding dimension, which is a more accurate measure. In fact, when a Gaussian correlation function is used, it is possible that a factor with smaller correlation parameter is more important than a factor with larger correlation parameter. For example, for certain factors having linear effects on the response variable, their estimated Gaussian process correlation parameters would be as small as zero, which obviously cannot accurately reflect their importance in the sequential MaxPro criterion.

4. Sequential LHD Structure

Finding an optimal design by directly maximizing/minimizing the corresponding space-filling criterion in continuous design space is challenging, and the optimization algorithm can easily get stuck in a low-quality local optimum for a number of reasons. (i) The number of variables in the optimization (np) is extremely high even for moderate size problems; (ii) the design criterion has many local optimums and (iii) the designs are isomorphism under the reordering of rows and columns. A practical solution to generating good space-filling designs is to search for the optimal design only among all possible LHDs, whose structure reduces the complexity of optimization. By discretizing the design space and using exchange algorithms to restrict the class of candidate designs to only LHDs, a simulated annealing algorithm proposed by Morris and Mitchell (1995) can efficiently move away from local optimal results and search for the global optimum. Other similar exchange algorithms have been discussed by Jin, Chen and Sudjianto (2005) and Joseph and Hung (2008). After obtaining the optimal LHD, it can also be used as the starting design in a continuous optimization algorithm to find the (unrestricted) optimal design in its neighborhood (Joseph, Gul and Ba (2015)). Our proposed sequential design framework takes advantage of such LHD structure in the design construction.

4.1. Sequential fold-over LHDs

As discussed at the end of previous section, the first stage of our sequential design is a MaxPro LHD which maximizes the projection properties in all possible sub-dimensions. In each of the subsequent stages, we fix the existing design points and optimize new design points based on the sequential MaxPro criterion in (3.2) where the $\boldsymbol{w}$ values for factor importance can be estimated based on the sensitivity analysis results from previous stages. Because directly optimizing criterion (3.2) by a continuous optimization algorithm is not easy, we propose to search for the new design points using a *sequential fold-over* LHD structure which is described below.

Suppose the design space has been standardized into the unit region $[0,1]^p$. For a given $(n_1 \times p)$ LHD with $n_1 = m+1$ equally spaced levels $\{0, 1/m, 2/m, \cdots, (m-1)/m, 1\}$, it can always be combined with any $(n_2 \times p)$ LHD with $n_2 = m$ equally spaced levels $\{1/2m, 3/2m, \cdots, (2m-1)/2m\}$ to form a $((n_1+n_2) \times p)$ LHD with $2m+1$ equally spaced levels $\{0, 1/2m, 2/2m, \cdots, (2m-1)/2m, 1\}$. Here we call the second design the *fold-over* LHD to the original LHD, analogous to the fold-over fractional factorial designs in the physical design of experiments literatures (Wu and Hamada (2009)). Consider the example in Figure 2 for illustration. Figure 2a contains a (7×2) LHD with levels $\{0, 1/6, 2/6, \cdots, 5/6, 1\}$ that can be paired with the (6×2) LHD in Figure 2b with levels $\{1/12, 3/12, \cdots, 9/12, 11/12\}$. Their combined design is a (13×2) LHD with levels $\{0, 1/12, 2/12, \cdots, 11/12, 1\}$ shown in Figure 2c. By combining the original LHD with one of its fold-over LHDs, the number of equally-spaced levels for each factor and also the total sample size get almost doubled, which enables the combined design to capture more complex nonlinear effects in the computer experiment. It can also be seen from Figure 2 that the new levels in the fold-over LHD are the *midpoints* between two existing factor levels in the original LHD, which is similar in spirit to the sequential design strategy proposed in Ba et al. (2013), and can be interpreted as “filling in” the vacant spaces in the original LHD to improve its space-filling property.

Given an original LHD, there are many possible choices of fold-over LHDs of the same size and levels. Similar to the concept of optimal folder-over designs (Li and Lin (2003)) in the physical experiment literature, in practice we also need to select the best fold-over LHD so that the combined design is optimal with respect to the sequential MaxPro criterion (3.2). Fortunately, because the fold-over LHD itself is also a LHD, this structure enables us to apply exchange algorithms as discussed in the beginning of this section to efficiently optimize the criterion in (3.2) for new design points without being trapped in a low-quality local optimum. The optmial folding-over process of LHD can be repeated indefinitely which constitutes the basic structure of our proposed sequential design framework: the designs at each stage, as well as the whole combined design, are all LHDs. When the combined design at a certain stage turns out to be adequate,

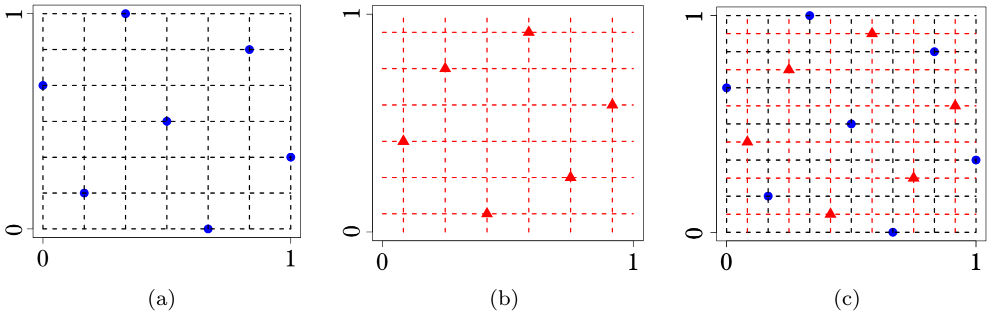


Figure 2. (a) A (7×2) original LHD with levels $\{0, 1/6, 2/6, \cdots, 5/6, 1\}$; (b) its (6×2) folder-over LHD with levels $\{1/12, 3/12, \cdots, 9/12, 11/12\}$; (c) their combined design, a (13×2) LHD with levels $\{0, 1/12, 2/12, \cdots, 11/12, 1\}$.

we can stop the procedure to obtain a LHD that is space-filling only with respect to the important factors. We will discuss the choice for number of stages and the sample size in each stage in later subsections.

It is interesting to note that the proposed sequential design structure forms a special case of the nested LHD structure, commonly used for designing computer experiments with different levels of accuracy (Qian (2009); Xiong, Qian and Wu (2013)). Different from the many existing sequential design methods, we will show later (in Section 5.1) that even if all factors in the computer experiments are equally important, the final combined design from multiple stages of sequential MaxPro LHDs can still achieve a similar space-filling property as a Mm LHD that is optimized in a single stage. Traditional sequential design methods without the sequential LHD structure, however, tend to have a much inferior space-filling property in this case.

In addition to improving the efficiency in optimization, there is actually another important reason for us to maintain the LHD structure in sequential design. As Ba et al. (2013) has pointed out, because the simulation model can be highly nonlinear, we may fail to detect or underestimate the importance of some factors based on the limited samples from the previous stages. Instead of dropping the inert factors or assigning them limited number of levels in the subsequent stages of experiments, we prefer to still keep them with a large number of levels to protect from possibly missing active factors. The identified "less important factors" are down weighted or even ignored in computing the sequential MaxPro criterion (3.2) for optimizing the new stage of design points, but they retain n distinct levels to enable us to re-assess their importance in latter stages.

4.2. Sliced structure in fold-over LHDs to achieve flexible sample size

If the sample size of the initial first-stage LHD is $m+1$, the sample sizes for the subsequent fold-over LHDs are $m, 2m, 4m, \ldots$, which gets doubled after each stage. In this subsection, we present strategies to enable the sequential design to have flexible sample size using a sliced structure in the fold-over LHDs.

A n-run Sliced Latin Hypercube Design (SLHD) (Qian (2012)), by definition, is a special type of LHD that can be partitioned into t slices (blocks), each of which is also a LHD of h runs ($n = ht$). By requiring each fold-over LHD to have such sliced structure, we can divide up a large fold-over LHD into t slices. Then instead of running the whole fold-over LHD as a single stage, we can run one slice (or possibly a few slices) of the fold-over LHD in each stage, whose sample size can be made arbitrarily small.

Similar to a LHD, because a randomly generated SLHD may not be space-filling, Ba, Myers and Brenneman (2015) proposed the slice-wise construction method and presented an efficient algorithm to generate the optimal SLHD for any given space-filling criterion. Here we modify this algorithm in order to generate the best fold-over LHDs with sliced structure. Different from in Ba, Myers and Brenneman (2015), instead of generating and optimizing all the slices simultaneously, we propose to only generate and optimize one slice (or a few slices) at each stage. In each step, the optimization takes into account the design points in the new stage as well as all the existing fixed design points from previous stages. After completing the experiments in each stage, sensitivity indices of input factors are re-estimated and the sequential MaxPro criterion in (3.2) is re-adjusted.

It can be easily seen that given an $(m+1)$-run initial LHD, imposing the fold-over LHDs to have sliced structure enables us to use arbitrary small sample size h for each stage (as long as m is divisible by the h). In practice, an attractive scheme is to set all subsequent stages to have exactly $h = m$ design points: run the 1^{st} fold-over LHD (m runs) in a single stage, run the 2^{nd} fold-over LHD ($2m$ runs) in two stages (slices), and run the 3^{rd} fold-over LHD ($4m$ runs) in four stages (slices), etc. An illustration of such a scheme is provided in Figure 3a. We can also use even smaller slices such as $h = m/2$ (if m is even), or have different h values in different fold-over LHDs. For example, as shown in Figure 3b, given an $(m+1)$-run initial LHD, we can generate and run the first fold-over LHD in two stages (slices) each containing $m/2$ runs, and generate the second fold-over LHD in eight stages (slices) each containing $m/4$ runs, etc. Moreover, it is also

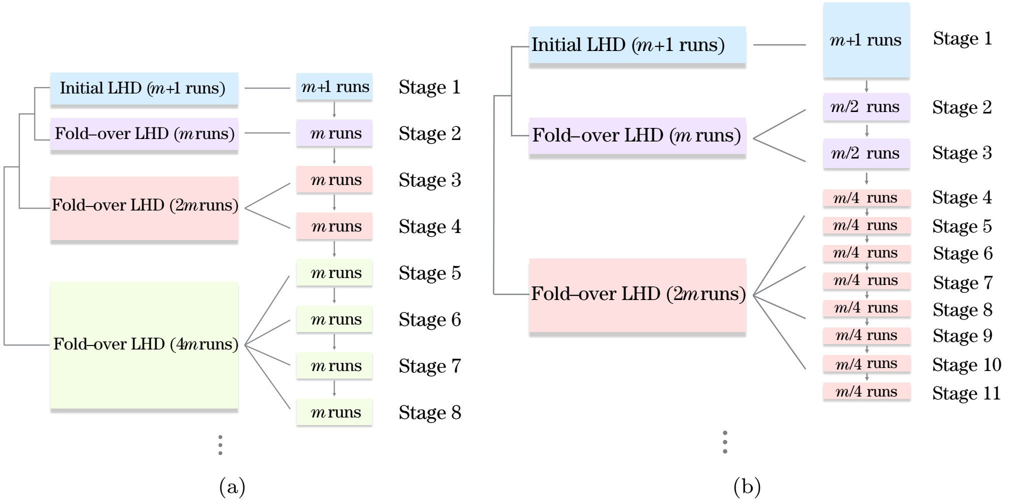


Figure 3. Illustration of (a) the standard $h = m$ scheme; (b) using different h values for different fold-over LHDs. In both cases, each stage of experiment corresponds to exactly one slice, and we generate and optimize the design matrix for one stage at a time.

possible for each stage to use more than one slice. Without loss of generality, in the rest of this paper we study only the most standard scheme ($h = m$) as shown in Figure 3a, which is approximately a batch sequential design with batch size always equal m.

Recently, Duan et al. (2016) proposed to use orthogonal arrays to construct sliced full factorial-based LHD whose slices can be used as sequential batched designs in a similar fashion. Their method was based on using non-overlapping sliceable orthogonal arrays which are less flexible than using our fold-over LHD structure. In addition, their proposed designs were not optimized by a space-filing criterion.

4.3. Sample size allocation and stopping criterion for sequential experiments

An important question to answer before we can apply the proposed sequential design method is regarding the sample size allocation. First of all, the choice of sample size for the initial MaxPro LHD usually depends on prior knowledge of the effect sparsity among factors. Suppose d factors are being studied and it is assumed that the number of active factors d_0 makes up a large percentage ($d_0 > d/2$), then we could start with a larger number of runs in the first stage (e.g., $5d$). On the other hand, if it is expected that only a smaller percentage, p,

Table 1. When all factors are equally important, comparing the minimum interpoint distance (larger the better) and the average reciprocal distance (smaller the better) for different designs in $10d+1$ runs.

	5-stage Sequential MaxPro LHD $n_1 = 2d + 1$,$n_2 = n_3$ $= n_4 = n_5 = 2d$		Single-stage Mm LHD $N_{total} = 10d + 1$		Naive 5-Stage Sequential Design $n_1 = 2d + 1$, $n_2 = n_3$ $= n_4 = n_5 = 2d$	
Dimension (d)	Min Distance	Avg Distance	Min Distance	Avg Distance	Min Distance	Avg Distance
10	0.8575036	0.9140244	0.884353	0.9181736	0.1838149	3.244135
20	1.407846	0.5907605	1.395671	0.5969556	0.2561815	2.042908

of factors are active ($d_0 \ll d/2$), or if there is no prior knowledge available for p, or if there is a tight constraint on the total number of simulation runs, one could start with a smaller run size in the first stage (e.g., $2d$). Choosing a smaller sample size in the first stage would enable more design points in subsequent stages to better focus on the sub-dimensions spanned by active factors, and thus possibly leading to smaller overall sample size. Moreover, some other practical considerations could also impact the sample size allocation. For example, if multiple simulations can be processed through parallel computing, the number of cores or computers (or a multiple of this number) is usually a good choice for the batch size. Depending on the specific settings of computer simulations, sometimes it may be more convenient to have smaller sample size in each stage but use more stages, while in some other situations it might be more practical to use larger sample size for each stage with potentially fewer total number of stages.

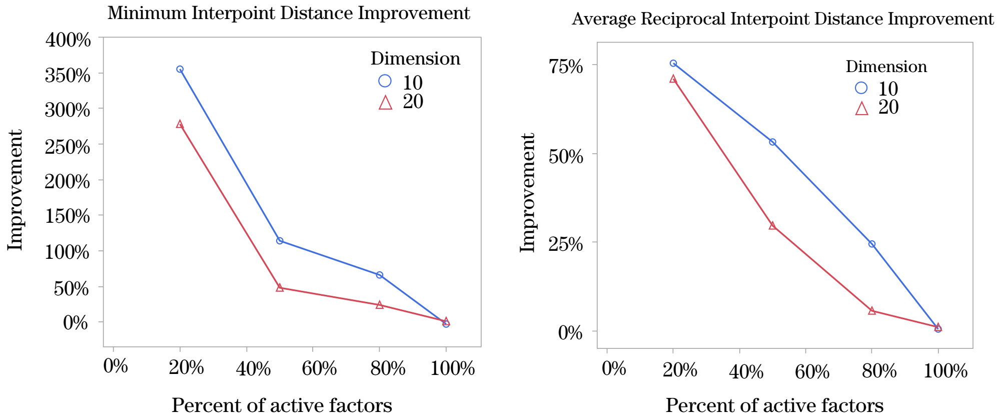
Another important question that practitioners need to address is when to stop the sequential experimentation, which determines the total sample size in the computer experiments. In practice, sometimes we can specify a threshold for satisfactory level of prediction accuracy (e.g., threshold for maximum prediction error) based on engineering domain knowledge. If this is the case, we could check the leave-one-out-cross-validation (LOOCV) error based on the fitted surrogate model and stop the sequential process if the LOOCV error is smaller than the required threshold. Alternatively, after collecting the simulation outputs in a new stage, we can use them as an independent testing dataset to evaluate the root mean square prediction error (RMSPE) of the fitted surrogate model based on training data from all previous stages. For situations where the accuracy threshold cannot be specified, a useful strategy is to monitor the percent change of LOOCV error or RMSPE of the fitted surrogate model after each stage. By quantifying the percent improvement in the model prediction accuracy after each stage, one can decide to stop the sequential experiments if the improvement has become minimal (e.g., less than 10%) after a certain stage. It can be easily seen that the total sample size determined by a sequential design framework is more sensible than using the traditional $10d$ rule because the former also considers the complexity of the true simulation model.

5. Examples

5.1. Simulation study for different percentages of important factors

To evaluate the properties of the proposed sequential design approach, we first studied the scenario where all input factors in the simulation are equally important. In this setting, the sequential design approach still has the advantage in adaptively determining the run size, but since there is no inert factor, the sequential MaxPro criterion (3.2) is no different from the traditional single-stage Mm distance criterion. Since the sequential design optimizes and fixes the design points stage by stage in a greedy way, it is interesting to study its loss in overall space-filing property compared with a single-stage global optimal LHD when all input factors are equally important. In Table 1, we compare the minimum interpoint distance and the average reciprocal distance for different designs in $10d+1$ runs ($d = 10$ and 20), where we can see that the five-stage sequential MaxPro LHDs in both cases have achieved almost the same space-filling properties as the single-stage Mm LHDs. On the contrary, Table 1 also shows the results of a naive sequential space-filling design approach, which adds and optimizes $2d$ points at a time without using the sequential LHD structure. It can clearly be seen that without the sequential LHD structure, the sequential design leads to much poorer space-filing properties in the end, due to its greedy optimization nature and the challenge in optimizing design points continuously in the high dimensional space. With the sequential LHD framework, we turn the high-dimensional continuous optimization into a more tractable combinatorial optimization problem and the resulting sequential MaxPro LHD is able to achieve space-filling properties close to a single-stage Mm LHD.

In Figure 4, we further compare the space-filling property (with respect to the active factors) of the five-stage sequential MaxPro LHD with that of the single-stage Mm LHD when the true percentage of active factors is 20%, 50%, 80% and 100%, respectively. The sequential MaxPro LHD achieves substantially



(a) (b)

Figure 4. Improvement of the five-stage sequential MaxPro LHD over the single-stage Mm LHD, in terms of (a) minimum interpoint distance; (b) average reciprocal distance.

better space-filling properties than the single-stage design when the percentage of inactive factors is high, and still performs as well as the single-stage Mm LHD when all factors are active.

5.2. OTL circuit function

Ben-Ari and Steinberg (2007) described an output transformerless (OTL) push-pull circuit (Chen et al. (1983), Wu, Mao and Ma (1990)), whose midpoint voltage (V_m) is given by:

$$V_m(x) = \frac{(V_{b1}+0.74)\beta(R_{c2}+9)}{\beta(R_{c2}+9)+R_f} + \frac{11.35R_f}{\beta(R_{c2}+9)+R_f} + \frac{0.74R_f\beta(R_{c2}+9)}{(\beta(R_{c2}+9)+R_f)R_{c1}},$$

where

$$V_{b1} = \frac{12R_{b2}}{R_{b1}+R_{b2}}.$$

The input variables and their usual ranges are resistance b1: $R_{b1} \in [50, 150]$ (K-Ohms), resistance b2: $R_{b2} \in [25, 70]$ (K-Ohms), resistance f: $R_f \in [0.5, 3.0]$ (K-Ohms), resistance c1: $R_{c1} \in [1.2, 2.5]$ (K-Ohms), resistance c2: $R_{c2} \in [0.25, 1.2]$ (K-Ohms) and current gain: $\beta \in [50, 300]$ (Amperes).

For this example, we generated sequential MaxPro LHDs in five stages, with each stage consisting of about $2d$ runs ($n_1 = 12 + 1$, $n_2 = n_3 = n_4 = n_5 = 12$). For comparison, a single-stage Mm LHD containing 61 runs was also generated by JMP®. As shown in Figure 5, after the completion of the second stage, the 25-run sequential MaxPro LHD has been able to achieve a RMSPE smaller than

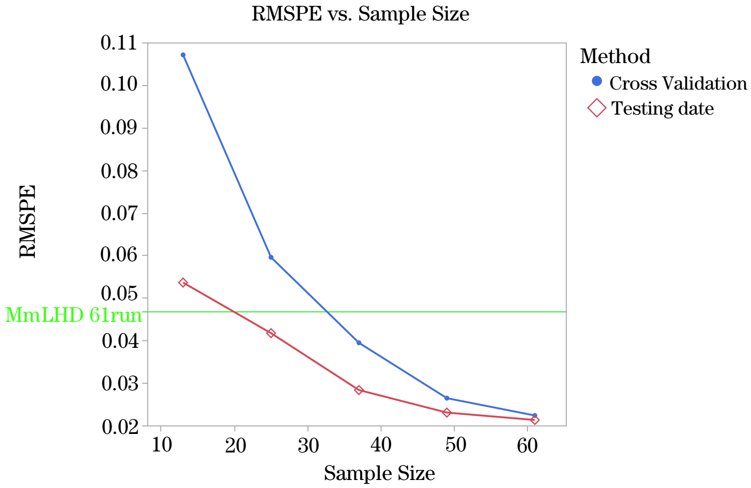


Figure 5. The root mean square prediction error of the five-stage sequential MaxPro LHD compared to that of the single-stage 61-run Mm LHD. The testing RMSPE is evaluated based on 600 independently generated data points in the design space.

that of the traditional single-stage Mm LHD in 61 runs.

The higher efficiency of the sequential MaxPro LHD over the traditional single-stage Mm LHD can be easily explained by the fact that the six-dimension OTL circuit function is actually dominated by the two most important factors, resistance b1 and resistance b2. Because the sequential MaxPro LHD incorporated the factor importance information from the previous stages and adjusted the subsequent stage of design points to focus more on b1 and b2, it required much smaller sample size to cover this important two-dimensional subspace. Figure 6 shows the projections of the 61-run sequential MaxPro LHD and the 61-run single-stage Mm LHD onto the subspace spanned by b1 and b2 which are important in emulating the OTL circuit function. We can clearly see that the sequential MaxPro LHD is much more space-filling.

5.3. Wing weight function

In this example, we consider a 10-dimension function which models a light aircraft wings weight as (Forrester, Sobester and Keane (2008))

$$f(X) = 0.036 S_w^{0.758} W_{fw}^{0.0035} \left(\frac{A}{\cos^2(\Lambda)}\right)^{0.6} q^{0.006} \lambda^{0.04} \left(\frac{100 t_c}{\cos(\Lambda)}\right)^{-0.3} (N_z W_{dg})^{0.49} \\ + S_w W_p,$$

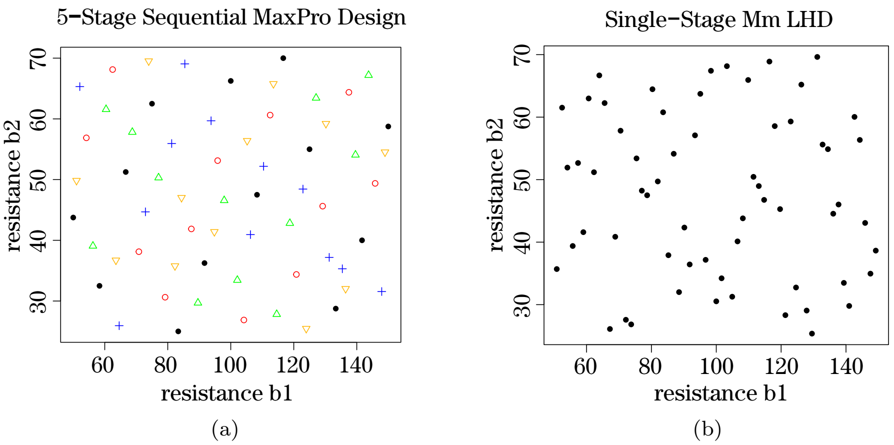


Figure 6. Design projections onto the two-dimensional subspace spanned by resistance b1 and resistance b2: (a) 5-Stage Sequential MaxPro LHD; (b) Single-Stage Mm LHD.

where the 10 inputs are: wing area $S_w \in [150, 200]$ (ft^2), weight of fuel in the wing $W_{fw} \in [220, 300]$ (lb), aspect ratio $A \in [6, 10]$, quarter-chord sweep $\Lambda \in [-10, 10]$ $(degrees)$, dynamic pressure at cruise $q \in [16, 45]$ (lb/ft^2), taper ratio $\lambda \in [0.5, 1.0]$, aerofoil thickness to chord ratio $t_c \in [0.08, 0.18]$, ultimate load factor $N_z \in [2.5, 6.0]$, flight design gross weight $W_{dg} \in$ [1,700, 2,500] (lb) and paint weight $W_p \in [0.025, 0.08]$ (lb/ft^2).

Sensitivity analyses of this true function shows that five of its input factors are quite important while the other five inputs are not very active/influential in determining the wing weight. Suppose we did not know any prior information about this function and tried to generate space-filling designs to approximate the wing weight function by a surrogate model. Sequential MaxPro LHDs in five stages were generated with each stage consisting about $2d$ runs ($n_1 = 20 + 1$, $n_2 = n_3 = n_4 = n_5 = 20$) to compare with a single-stage 101-run Mm LHD generated by JMP®. For each design, a composite Gaussian process (CGP) model (Ba and Joseph (2012)) was fitted and its RMSPE was calculated based on 1000 independent testing points (in this example, CGP model was used because it consistently yields smaller RMSPE than the stationary Gaussian process model). Figure 7 shows that a 61-run sequential MaxPro LHD (after completing the third stage) has yielded a RMSPE smaller than that of the traditional single-stage Mm LHD in 101 runs.

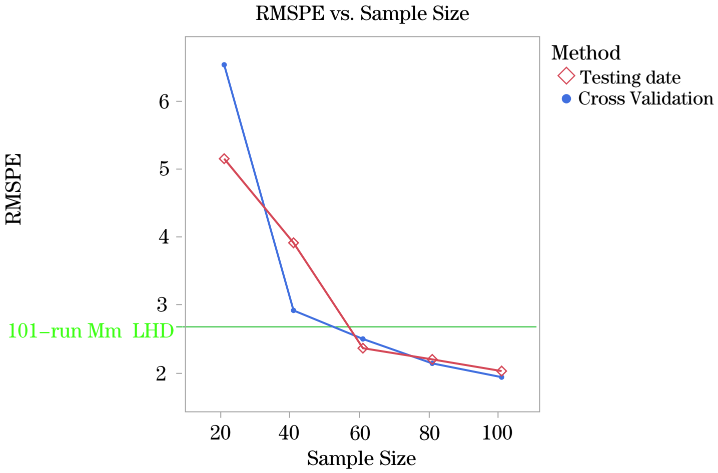


Figure 7. The RMSPE of the five-stage sequential MaxPro LHD compared to that of the single-stage 101-run Mm LHD for the Weight Wing Function.

6. Conclusions

In this paper, we develop a flexible sequential design framework that can accommodate multiple responses and quickly screen out inert factors so that the final design is space-filling with respect to the active factors. This results in a more efficient use of resources and allows the potential opportunity for fewer overall simulations. Different from many existing sequential design approaches which focus on capturing specific features in the design space such as maxima or minima, this new design framework enables one to build high-fidelity surrogate model for the entire design space. The new sequential design is model independent which can be considered as an extension of the single-stage MaxPro design, and the proposed sequential LHD structure also allows the experiments to have flexible batch size, eliminating the need to prescribe/fix the total sample size. The paper demonstrates that this sequential design approach achieves substantially better space-filling properties than a single-stage design when inactive factors are present, while performing as well as the single-stage design when all factors are active. Using this sequential design framework can achieve prediction quality that is comparable or even superior to the traditional single-stage design in fewer simulation runs.

Acknowledgment

The authors thank the editor, an associate editor, and two referees for their valuable comments and suggestions. The research of the third author was supported by NSAF (Grant No. U1430125).

References

Ba, S., Jain, N., Joseph, V. R. and Singh, R. K. (2013). Integrating analytical models with finite element models: An application in micromachining. *Journal of Quality Technology*, **45**, 200–212.

Ba, S. and Joseph, V. R. (2012). Composite Gaussian process models for emulating expensive functions. *Annals of Applied Statistics*, **6**, 1838–1860.

Ba, S., Myers, W. R. and Brenneman, W. A. (2015). Optimal sliced Latin hypercube designs. *Technometrics*, **57**, 479–487.

Ben-Ari, E. N. and Steinberg, D. M. (2007). Modeling data from computer experiments: an empirical comparison of kriging with mars and projection pursuit regression. *Quality Engineering*, **19**, 327–338.

Chen, G. Y, Wang, D. Q., Jian, J. B. and Zhang, L. T. (1983). Mid-point voltage of an OTL circuit. *Three-Stage Design of Experiments with Known Transfer Functions* (edited by the Committee on Three-Stage Design, Chinese Applied Statistics Society), 63–74 (in Chinese).

Chen, W., Jin, R. and Sudjianto, A. (2005). Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. *Journal of Mechanical Design*, **127**, 875–886.

Duan, W., Ankenman, B. E., Sanchez, M. S. and Sanchez, P. J. (2016). Sliced full factorial-based Latin hypercube designs as a framework for a batch sequential design algorithm. *Technometrics*. to appear.

Forrester, A., Sobester, A. and Keane, A. (2008). *Engineering Design Via Surrogate Modelling: a practical guide*. Wiley.

Gramacy, R. B. and Lee, H. K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics*, **51**, 130–145.

Jin, R., Chen, W. and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, **134**, 268–287.

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

Joseph, V., Gul, E. and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, **102**, 371–380.

Joseph, V. R. and Hung, Y. (2008). Orthogonal-maximin Latin hypercube designs. *Statistica Sinica*, **18**, 171–186.

Lam, C. and Notz, W. (2008). Sequential adaptive designs in computer experiments for response surface model fit. *Statistics and Applications*, **6**, 207–233.

Li, W. and Lin, D. K. J. (2003). Optimal foldover plans for two-level fractional factorial designs. *Technometrics*, **45**, 142–149.

Loeppky, J., Moore, L. and Williams, B. (2010). Batch sequential designs for computer experiments. *Journal of Statistical Planning and Inference*, **140**, 1452–1464.

Loeppky, J. L., Sacks, J., S. and Welch, W. J. (2009). Choosing the sample size of a computer experiment: a practical guide. *Technometrics*, **51**, 366–376.

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.

Moon, H., Dean, A. M. and Santner, T. J. (2012). Two-stage sensitivity-based group screening in computer experiments. *Technometrics*, **54**, 376–387.

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

Qian, P. Z. (2009). Nested Latin hypercube designs. *Biometrika*, **96**, 957–970.

Qian, P. Z. G. (2012). Sliced Latin hypercube designs. *Journal of the American Statistical Association*, **107**, 393–399.

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

Saltelli, A., Chan, K. and Scott, E. (2000). *Sensitivity Analysis.* John Wiley & Sons, Chichester.

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

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

Wu, C. F. J. and Hamada, M. S. (2009). *Experiments: Planning, Analysis, and Optimization (2nd ed.).* Wiley, New York.

Wu, C. F. J., Mao, S. S. and Ma, F. S. (1990). SEL: A search method based on orthogonal arrays. *Statistical Design and Analysis of Industrial Experiments* (S. Ghosh, ed.), 279–310. Marcel Dekker.

Xiong, S., Qian, P. Z. G. and Wu, C. F. J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, **55**, 37–46.

Quantitative Sciences, The Procter & Gamble Company, Mason, OH 45040, USA.

E-mail: ba.s@pg.com

Quantitative Sciences, The Procter & Gamble Company, Mason, OH 45040, USA.

E-mail: myers.wr@pg.com

Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing, 100190, China.

E-mail: dianpengwang@outlook.com

(Received March 2016; accepted January 2017)