Sequential Exploration of Complex Surfaces Using Minimum Energy Designs

V. Roshan Joseph, Tirthankar Dasgupta, Rui Tuo & C. F. Jeff Wu

Sequential Exploration of Complex Surfaces Using Minimum Energy Designs

V. Roshan JOSEPH

H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(*roshan@isye.gatech.edu*)

Rui TUO

Academy of Mathematics and Systems Science
Chinese Academy of Sciences
100190 Beijing, China
(*tuorui@amss.ac.cn*)

Tirthankar DASGUPTA

Department of Statistics
Harvard University
Cambridge, MA 02138
(*dasgupta@stat.harvard.edu*)

C. F. Jeff WU

H. Milton Stewart School of Industrial
and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(*jeffwu@isye.gatech.edu*)

A new space-filling design, called minimum energy design (MED), is proposed to explore unknown regions of the design space of particular interest to an experimenter. The key ideas involved in constructing the MED are the visualization of each design point as a charged particle inside a box, and minimization of the total potential energy of these particles. It is shown through theoretical arguments and simulations that with a proper choice of the charge function, the MED can asymptotically generate any arbitrary probability density function. A version of the MED, which adaptively updates the design by "learning" about the unknown response surface sequentially, is proposed and implemented. Two potential applications of MED in simulation of complex probability densities and optimization of complex response surfaces are discussed and demonstrated with examples. This article has supplementary material online.

KEY WORDS: Experiments; Kriging; Optimization; Quasi-Monte Carlo; Sequential designs; Space-filling designs.

1.   INTRODUCTION

Many modern-day scientific experiments involve exploration of complex response surfaces. Such exploration may have different goals depending on the context and the nature of the experiment, for example, (i) prediction of the response surface, (ii) detecting regions in the design space where the value of the response exceeds a certain threshold, (iii) finding the global optimum, and (iv) finding settings which are robust to changes in uncontrollable variables. This work proposes designs suitable for exploring complex surfaces which can be easily adopted depending on the objective of the experiment.

Our initial motivation for this work comes from the nanostructure yield surface shown in the top left panel of Figure 1 (Dasgupta et al. 2008). As can be seen in that figure, the surface is complex with many local optima and regions with no yield. This makes designing an experiment to find the global optimum with respect to the two variables, pressure and temperature, a challenging task. Design points placed in the zero-yield regions are not useful because they produce no nanostructures. Thus, we would like our experimental design to place more points in the high-yield regions. Similar problems arise quite often in the global optimization of expensive black-box functions (e.g., Jones 2001). The top right panel of Figure 1 shows the plot of gap lift in an engine head and block joint sealing assembly against two variables: gasket thickness and deck face surface flatness (Joseph, Hung, and Sudjianto 2008). If the objective is to find the global optimum of the gap lift and to identify other

good locations so that the product can be optimized with respect to a secondary criterion (e.g., minimize variance for achieving robustness), then we need to spread out the design points in the optimum regions. A similar example is shown in bottom left panel of Figure 1, which depicts the thickness of a circular wafer for making computer chips (Jin, Chang, and Shi 2012). The objective was to devise a measurement strategy to understand the wafer profile with minimum number of measurements. Since the variability toward the edges is higher (a characteristic of the manufacturing process), Jin, Chang, and Shi (2012) argued that a strategy that places more points near the edges than at the center would be better. A similar goal was observed in a totally different application area, which is shown in bottom right panel of Figure 1. Here the objective is to obtain a representative sample from the probability density which allocates more points in the high probability regions than in the low probability regions.

A major feature in these examples is the spreading out of design points in the regions of interest. Keeping the probability density example in mind, we can state the objective as "sampling points which are as apart as possible, but in a way still mimic the density of the distribution." This is related to the

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/tech*.
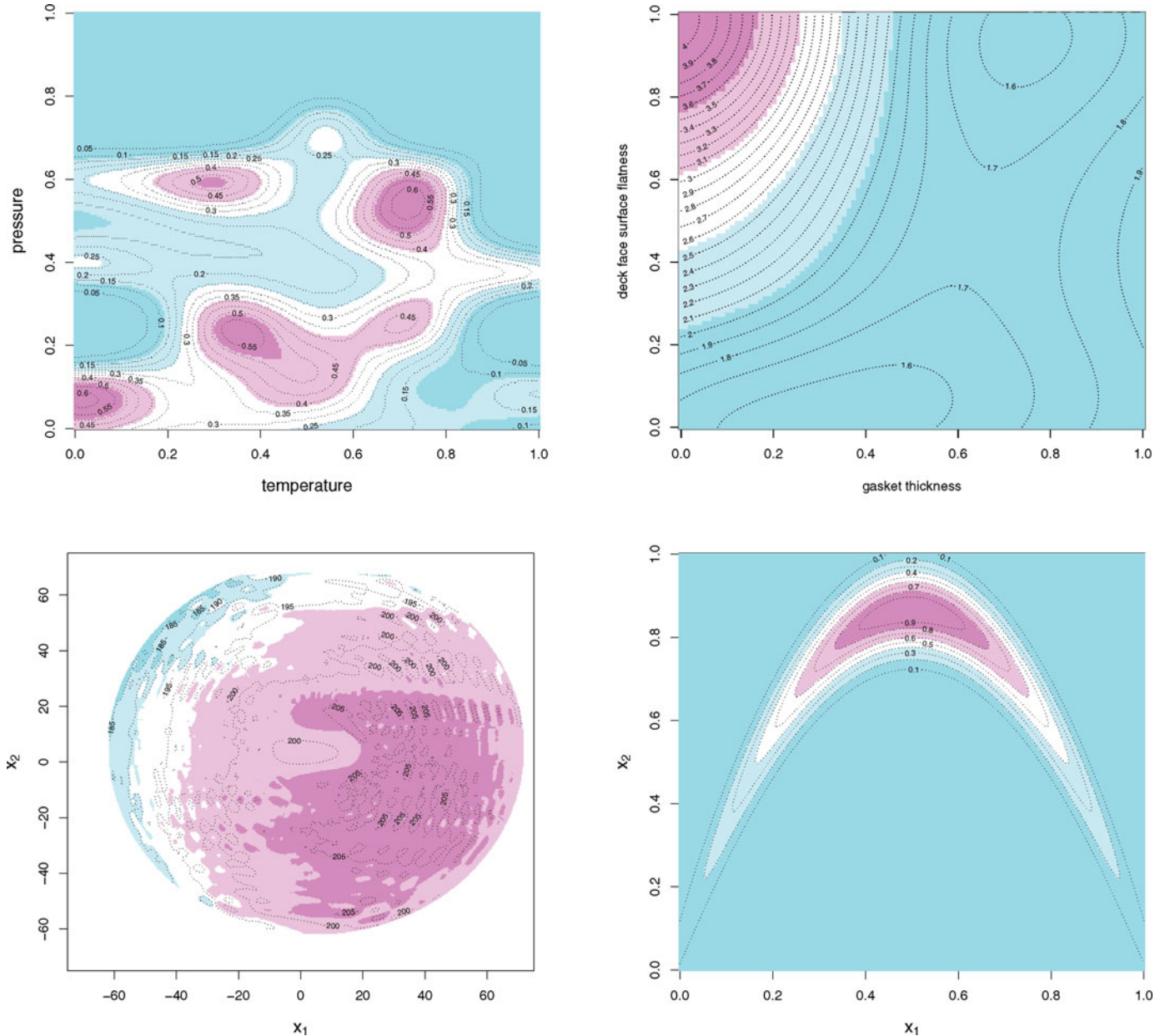
Figure 1. Motivating examples.

idea of principal points in stratified sampling (Dalenius 1950; Cox 1957; Flury 1990), quantizers in communications theory (Llyod 1957; Max 1960), and the mean squared error representative points in Quasi-Monte Carlo (see, e.g., Fang and Wang 1994). However, the ability to mimic an arbitrary distribution is unique to our design strategy. Moreover, the complex nature of the surface and expensive simulations/experiments make our approach quite different from theirs. Although many of our ideas are intertwined with the above literature, because of the other potential applications, we keep a broad perspective in the exposition of our article.

We use a physical system analogy of electrical particles inside a box to motivate the proposed design, which we call *minimum energy design*. This physical analogy helps to visualize the basic working of the design strategy and devise sequential designs in a physically interpretable way. Similar analogies have been used in the literature by Audze and Eglais (1977) and Kessels et al. (2009). Different from the earlier works, we explicitly

use the "charge" of the electrical particles, which is crucial for emulating the underlying distributions.

The article is organized as follows. In the following section, we develop the MED using the physical analogy mentioned in the previous paragraph, explore its connections with similar designs, and present some theoretical and heuristic arguments to provide insights to these connections and properties of MED. In Section 3, we develop the sequential version of MED and propose an algorithm for its implementation. Two potential applications, one in the simulation of complex probability densities, and the other related to optimization of complex response surfaces, are discussed in Section 4. Section 5 presents some concluding remarks and directions for future research.

2.　MINIMUM ENERGY DESIGNS

In this section, we first present the main idea of minimum energy designs and then study its asymptotic properties. We then

present a computational algorithm to construct the proposed designs and illustrate it with a few simple examples. High dimensional and more complex problems are considered in Section 4.

2.1   Energy Criterion

We start with a motivation from physics and explain how it drives the proposed design procedure. Consider a box containing some charged particles. If they have the same sign for the charge, then they will repel each other and occupy positions inside the box so as to minimize the total potential energy. *Here, the box is the experimental region, each position taken by the charged particles is a design point, and the charge represents the experimental response.* Consequently, the experimental design consists of all the positions occupied by the particles. Because this design is obtained by minimizing the potential energy, we call it *minimum energy design* (MED). This follows from Thomson's theorem in electrodynamics (Zhou 1999, p. 59). Because the particles repel and try to be as away as possible from each other, as will be seen later, MED has close connections to the maximin distance designs (Johnson, Moore, and Ylvisaker 1990). The box representing the experimental region can be in a high-dimensional space. Let there be p factors. For simplicity, we scale the experimental range of each factor into [0, 1], so that the box under consideration is $[0, 1]^p$.

Because the "repulsion principle" mentioned in the previous paragraph requires all the particles to have the same sign for the charge, without loss of generality, assume that the particle charge is positive. Note that if the response is unrestricted, it can be converted to a positive variable for the purpose of constructing the design (see Section 4.2). Let $q(\boldsymbol{x}_i)$ be the charge of the particle at the ith design point $\boldsymbol{x}_i$. Then the potential energy between the ith and jth particle is proportional to $q(\boldsymbol{x}_i)q(\boldsymbol{x}_j)/d(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ denotes the Euclidean distance between the two points. We can take the proportionality constant to be 1. Thus, the total potential energy for n charged particles is given by

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q(\boldsymbol{x}_i)q(\boldsymbol{x}_j)}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}. \tag{1}$$

MED can be obtained by minimizing E with respect to the design $\boldsymbol{D} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$.

Now the different objectives stated in the introduction can be achieved by judiciously choosing the charge of the particles. For example, if we are interested in the high-yield regions of the nanostructures, we could take $q(\boldsymbol{x}) = 1 - y(\boldsymbol{x})$, where $y(\boldsymbol{x})$ denotes the yield at point $\boldsymbol{x}$ which is in [0, 1]. By doing this, higher charges will be assigned to particles in low-yield region, which causes the particles to repel from those regions and occupy positions in the high-yield regions. Careful choice of the charge function will be discussed soon.

We first consider the case where equal charges are assigned to the particles. Without loss of generality, let $q(\boldsymbol{x}_i) = 1$ for all $i = 1, \ldots, n$. Then our criterion reduces to

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

which can be compared with some of the existing criteria in the literature. Audze and Eglais (1977) proposed to find designs by minimizing the force among the particles:

$$\min_{\boldsymbol{D}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)},$$

which uses a power of 2 in the denominator instead of 1. Bates, Sienz, and Langley (2003) have studied its applications for finding optimal Latin hypercube designs (LHD). Morris and Mitchell (1995) proposed the following criterion

$$\min_{\boldsymbol{D}} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(\boldsymbol{x}_i, \boldsymbol{x}_j)} \right\}^{1/k}, \tag{2}$$

for finding maximin LHDs. This is because as $k \to \infty$, the criterion in (2) reduces to

$$\max_{\boldsymbol{D}} \min_{i,j} d(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{3}$$

which is the maximin distance criterion proposed by Johnson, Moore, and Ylvisaker (1990). Morris and Mitchell (1995) further showed that for large but finite values of k, (2) produces nearly maximin distance designs with the smallest index, where index is the number of pairs of points with minimum distance. In a recent work, Dette and Pepelyshev (2010) considered a family of criteria similar to (2), but for small values of k in the range of [0, 1]. It is easy to show that as $k \to 0$, the criterion reduces to

$$\min_{\boldsymbol{D}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} -\log d(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

which is called the logarithmic potential criterion. In another recent work, Kessels et al. (2009) proposed the criterion

$$\min_{\boldsymbol{D}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{1}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} + d^2(\boldsymbol{x}_i, \boldsymbol{x}_j) \right),$$

where the second term, representing the potential energy of a spring attached to each particle, is added to the potential energy of the electrical particles to remove the box constraints. Their design, called minimum potential design, is available in the popular statistical software package JMP.

Motivated by the foregoing works, we propose a generalized version of our MED given by

$$\min_{\boldsymbol{D}} \text{GE}_k = \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{q(\boldsymbol{x}_i)q(\boldsymbol{x}_j)}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} \right)^k \right\}^{1/k}. \tag{4}$$

In contrast to Dette and Pepelyshev (2010), we are more interested in larger values of k in the range of $[1, \infty)$ as in Morris and Mitchell (1995). The extreme case of $k \to \infty$ gives the criterion

$$\max_{\boldsymbol{D}} \min_{i,j} \frac{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}{q(\boldsymbol{x}_i)q(\boldsymbol{x}_j)}. \tag{5}$$

Following an example from Johnson, Moore, and Ylvisaker (1990), an interesting interpretation can be given for the MED produced by the above criterion. Consider the example of finding the best locations to start gas stations in a region. A maximin distance design will try to look for locations that will minimize

the competition among the gas stations, whereas by choosing the charges to be inversely proportional to the population density, the MED will look for locations that are highly populated as well as to minimize the competition among the gas stations. Thus, our designs will tend to avoid the less-populated regions to start the gas stations, which is meaningful.

2.2 Limiting Distribution and Charge Function

Our choice for the charge function is guided by the limiting distribution of the design points generated by the MED. That is, we choose the charge function so that we can get the desired distribution of the points.

We need the following definition to explain the results. For a design $\boldsymbol{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, define its index to be the number of pairs $(\boldsymbol{x}_l, \boldsymbol{x}_m)$, $1 \leq l < m \leq n$, with the smallest value of $d(\boldsymbol{x}_i, \boldsymbol{x}_j)/(q(\boldsymbol{x}_i)q(\boldsymbol{x}_j))$ over all $i \neq j$. Denote the index of $\boldsymbol{D}$ by $IN(\boldsymbol{D})$. We are particularly interested in the MEDs with the smallest index, because such designs enjoy some nice theoretical properties. Let $\boldsymbol{D}^* = \{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*\}$ be the MED produced by (5) with the smallest index, and $\mathscr{B}$ be the Borel σ-algebra over $\mathcal{X} = [0, 1]^p$. Define the probability measures on $(\mathcal{X}, \mathscr{B})$:

$$\mathcal{P}_n(A) = \frac{\#\{\boldsymbol{x}_i^* : 1 \leq i \leq n, \boldsymbol{x}_i^* \in A\}}{n}, \text{ for any } A \in \mathscr{B}. \quad (6)$$

We have the following result on the limiting measure of $\mathcal{P}_n$ as $n \to \infty$. All the proofs are given in the online supplementary material file associated with this article.

Theorem 1. If $q(\boldsymbol{x}) \equiv 1$, $\mathcal{P}_n$ converges to the uniform distribution over $\mathcal{X}$.

Theorem 1 establishes the fact that a uniform distribution can be obtained as the limiting distribution of the design points by choosing the particle charges to be equal. This theorem forms the basis of our main result to be discussed next. Further, this theorem is interesting in its own right, because it makes a three-way connection between MED, maximin distance designs, and uniform designs (see, e.g, Fang, Li, and Sudjianto 2006), and places the MED criterion on a firm theoretical footing.

Now we want to study what happens when $q(\boldsymbol{x})$ is not identical to 1. If $q(\boldsymbol{x}) \neq 1$, then the points will accumulate more in the regions where the charge is less. Thus, intuitively, the limiting measure of $\mathcal{P}_n$ will be inversely proportional to the charge function $q(\boldsymbol{x})$. A stronger result is stated below. Because only a heuristic proof is given, we refer to it as "result."

Result 1. If $1/q$ is differentiable over $\mathcal{X}$ and bounded away from 0, there exists a probability measure $\mathcal{P}$ such that $\mathcal{P}_n$ converges to $\mathcal{P}$. Moreover, $\mathcal{P}$ has a density f over $\mathcal{X}$ with $f(\mathbf{x}) \propto 1/q^{2p}(\mathbf{x})$.

Thus, to obtain a desired density $f(\boldsymbol{x})$, we choose the charge function to be

$$q(\boldsymbol{x}) = \frac{1}{\{f(\boldsymbol{x})\}^{1/(2p)}}. \quad (7)$$

This choice agrees with our intuition that the charges should be inversely proportional to the function values (when large function values are desired). Moreover, Result 1 gives us the right value of the power to be used for a nonnegative function

so that we get the desired density in the limit as the sample size grows. The result can be made to work approximately for densities whose domain is different from $[0, 1]^p$ through some modifications. For $\boldsymbol{x} \in \mathbb{R}^p$, find a box that encloses most of the probability mass of $f(\boldsymbol{x})$. Since the box size can be made very large, the MED samples will approximately follow the distribution $f(\boldsymbol{x})$. For a nonrectangular region $\mathcal{X}$ such as the circular wafer shown in Figure 1, find a box that covers $\mathcal{X}$, rescale it to $[0, 1]^p$, and change $f(\boldsymbol{x})$ to $f(\boldsymbol{x}) + \epsilon$. Now for $\epsilon > 0$ small enough, the MED samples can approximate $f(\boldsymbol{x})$. In practical implementation, these modifications are not necessary. One can use unconstrained optimization for finding MED if $\boldsymbol{x} \in \mathbb{R}^p$ and use a constrained optimization with $\boldsymbol{x} \in \mathcal{X}$ if $\mathcal{X}$ is nonrectangular.

2.3 A Greedy Algorithm

Finding the MED is a computationally difficult problem. We can adapt the optimal design algorithms used in the literature such as simulated annealing (Morris and Mitchell 1995) and stochastic evolutionary algorithm (Jin, Chen, and Sudjianto 2005) for our purpose. See Fang, Li, and Sudjianto (2006, ch. 4) for a review of some useful optimization algorithms. Different from theirs, we propose a one-point-at-a-time greedy algorithm for the generation of our designs because it serves as a motivation and justification for developing the sequential design in the next section. This is described below.

Suppose we want to generate an N-point MED $\boldsymbol{D}_N$. We start with a "good" point $\boldsymbol{x}_1$, and generate $\boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$ sequentially. Suppose we have already generated n points using the criterion in (4). Then the $(n + 1)$th point is obtained by

$$\begin{aligned}
\boldsymbol{x}_{n+1} &= \arg\min_{\boldsymbol{x}} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{q(\boldsymbol{x}_i)q(\boldsymbol{x}_j)}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} \right)^k \right. \\
&\qquad \left. + \sum_{i=1}^{n} \left(\frac{q(\boldsymbol{x}_i)q(\boldsymbol{x})}{d(\boldsymbol{x}_i, \boldsymbol{x})} \right)^k \right\}^{1/k} \\
&= \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} \left(\frac{q(\boldsymbol{x}_i)q(\boldsymbol{x})}{d(\boldsymbol{x}_i, \boldsymbol{x})} \right)^k \\
&= \arg\min_{\boldsymbol{x}} \frac{1}{f^{k/2p}(\boldsymbol{x})} \sum_{i=1}^{n} \frac{1}{f^{k/2p}(\boldsymbol{x}_i)d^k(\boldsymbol{x}_i, \boldsymbol{x})}. \quad (8)
\end{aligned}$$

Such an algorithm can get stuck in a local optimum, but we had great success with it as long as the starting point is good. We choose the starting point by

$$\boldsymbol{x}_1 = \arg\min_{\boldsymbol{x}} q(\boldsymbol{x}) = \arg\max_{\boldsymbol{x}} f(\boldsymbol{x}). \quad (9)$$

Based on an earlier version of our article, a heuristic proof for the convergence of this algorithm was provided by Wang (2011) in his Ph.D. thesis. He showed that the design points converge to the target distribution as long as $k > p$.

To illustrate the algorithm, first consider the case with $q(\boldsymbol{x}) = 1$. According to Theorem 1, the asymptotic distribution of MED points should be uniform. A large value of k in (8) is needed to achieve uniformity, but at the same time, a very large value of k can cause numerical instability in the algorithm. In our simulations (not reported here), we found $k = 4p$ to be a good compromise choice, which will be used throughout this
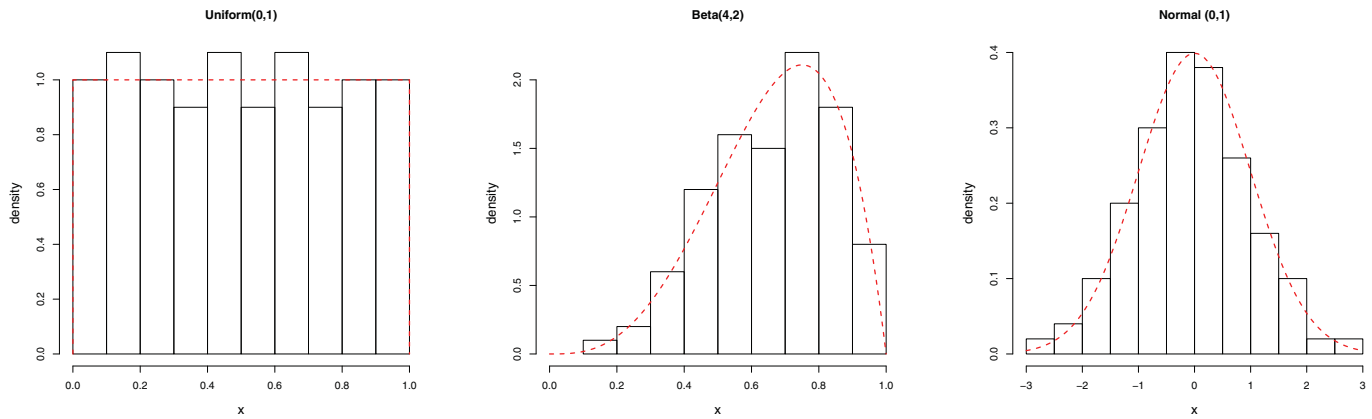
Figure 2. Histograms of 100 points generated by MED. The true density curves are plotted in dashed lines.

article. The left panel in Figure 2 shows the histogram of 100 points generated by the greedy algorithm. Although the exact optimum {0, 1/99, . . . , 1} is missed, the approximation looks reasonable.

Now consider the case of nonuniform distributions. The middle and right panels in Figure 2 shows the histogram of 100 points generated by MED from a Beta (4,2) distribution and that of a standard normal distribution (Q-Q plots are given in the online supplementary material file). We can see that by choosing $q(x) = 1/\sqrt{f(x)}$, the MED is able to approximately reproduce the true distributions. Note that the support of normal is not in [0, 1] and, therefore, more care is needed to generate the MED points. To generate the 100 points, we first chose 1000 equally spaced candidate points in [−3, 3]. In general, it is not easy to identify the correct range for the candidate points. In such cases, one can use an unconstrained continuous optimization algorithm for obtaining each MED point, but it is computationally more challenging because of the multimodality of the objective function.

The first 25 MED points for a two-dimensional uniform distribution is shown in Figure 3(a). We can see that it is a full factorial 5^2 design, which can be viewed as a quasirandom sample for the two-dimensional uniform distribution. Here 16 out of the 25 points are on the boundary and only 9 points are inside the square [0, 1]². This may not be an attractive design if our

interest is in approximating functions inside [0, 1]². One option to overcome this boundary problem is to shrink the design region. As shown in Lemma 2 in the Appendix, a lower bound for the minimum distance of an N-point maximin distance design is $1/(N^{1/p} + 1)$. Thus, instead of generating points in [0, 1]^p, we can generate points in

$$\left[\frac{0.5}{N^{1/p} + 1}, 1 - \frac{0.5}{N^{1/p} + 1} \right]^p .$$

This helps in reducing the L_2 discrepancy measure (L_2) (see Fang and Wang 1994, p. 34) from 0.0397 to 0.0234 showing that the overall uniformity of the points within [0, 1]² has been improved. However, the design (not shown here) is still a full factorial in the reduced region, which again is not the best design for approximating smooth functions in [0, 1]². For comparison, a Sobol sequence with 25 points shown in Figure 3(b) has a much smaller $L_2 = 0.0168$. One possible way to improve the uniformity of the MED is to use a larger low-discrepancy sequence as the candidate set for finding the MED. For example, the MED obtained using a 100-point Sobol sequence as the candidate set is shown in Figure 3(c). It has a smaller $L_2 = 0.0157$ as well as a much larger interpoint distance (0.1482) compared to that of the 25-point Sobol (0.0883).
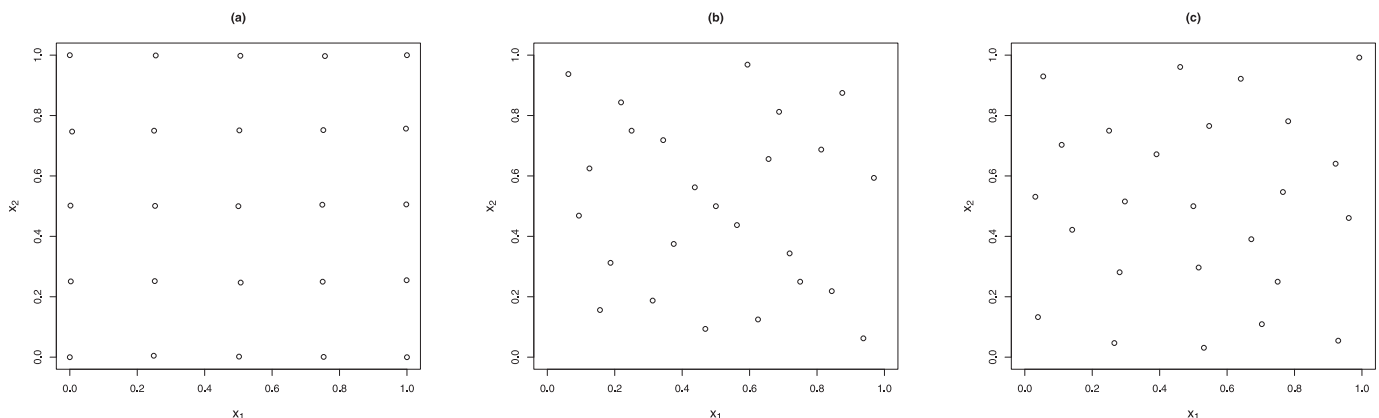

Figure 3. Two-dimensional uniform distribution. (a) 25-point MED, (b) 25-point Sobol sequence, and (c) 25-point MED obtained from a 100-point Sobol sequence as the candidate set.

3. SEQUENTIAL MINIMUM ENERGY DESIGNS

Direct application of MED is impossible in situations where the true response function $f(\boldsymbol{x})$ is unknown or its evaluation is computationally expensive. For example, in physical experiments, the underlying response surface $f(\boldsymbol{x})$ is unknown. In computer experiments, even though the computer model is known, it is often an expensive black-box code making the optimization in (8) impossible to perform. Here we propose a sequential strategy to implement MED in such situations. The key idea is to "learn" about the underlying surface sequentially, and implement the MED accordingly. For this purpose, we adapt the one-point-at-a-time greedy algorithm described in the previous section. Suppose we have already generated an n-point design $\boldsymbol{D}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and obtained the observations $\boldsymbol{y}^{(n)} = (y_1, \ldots, y_n)'$, where $y_i = f(\boldsymbol{x}_i)$. We can use some statistical techniques such as regression or kriging to estimate the underlying response surface. Denote it by $\hat{f}^{(n)}(\boldsymbol{x})$. Let $\hat{q}^{(n)}(\boldsymbol{x}) = 1/\{\hat{f}^{(n)}(\boldsymbol{x})\}^{1/2p}$. Then, the next design point is given by

$$\boldsymbol{x}_{n+1} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} \left(\frac{\hat{q}^{(n)}(\boldsymbol{x}_i)\hat{q}^{(n)}(\boldsymbol{x})}{d(\boldsymbol{x}_i, \boldsymbol{x})} \right)^k. \quad (10)$$

We call the design generated by this algorithm as *sequential minimum energy design* (SMED).

A major problem with this approach is that we cannot ensure the starting point to be good because we cannot find it as in (9) and thus, the whole sequence of points can go astray. One approach to overcome this problem is to start with an n_0-run initial design ($n_0 \ll N$) and discard points that have small function values, say less than a threshold Δ. This is similar to a "burn-in" procedure in a Markov chain Monte Carlo (MCMC) algorithm except that we reject only the bad points. A rule of thumb for choosing the threshold is $\Delta = 0.01 \max_i\{\hat{f}^{(n_0)}(\boldsymbol{x}_i)\}$, that is, 1% of the maximum observed value from the initial design. We can decrease this threshold as the sample size grows, a strategy that will allow the design to explore the low (or tail) regions as well. Therefore, we choose

$$\Delta_n = 0.01 \left(\frac{n_0}{n} \right)^{1/p} \max_i\{\hat{f}^{(n)}(\boldsymbol{x}_i)\}.$$

Let $I_n = \{i : \hat{f}^{(n)}(\boldsymbol{x}_i) > \Delta_n\}$. Thus, SMED becomes

$$\boldsymbol{x}_{n+1} = \arg\min_{\boldsymbol{x}} \sum_{i \in I_n} \left(\frac{\hat{q}^{(n)}(\boldsymbol{x}_i)\hat{q}^{(n)}(\boldsymbol{x})}{d(\boldsymbol{x}_i, \boldsymbol{x})} \right)^k, \quad (11)$$

for $n = n_0 + 1, \ldots, N$.

The initial design can be a fractional factorial design for physical experiments or a space-filling design for computer experiments. In fact, SMED can be made as a fully sequential procedure by using an MED as the initial design. For doing this, let

$$\hat{q}^{(n)}(\boldsymbol{x}) = \frac{1}{\{\hat{f}^{(n)}(\boldsymbol{x})\}^{\gamma_n}}. \quad (12)$$

Now, we can take $\gamma_n = 0$ for $n = 1, \ldots, n_0$ to generate the initial design as described at the end of Section 2 and then use $\gamma_n = 1/(2p)$ for $n > n_0$. One could also consider increasing γ_n slowly from 0 to $1/(2p)$ for $n = 1, \ldots, n_0$ to get some adaptivity from the beginning.

4. APPLICATIONS

As discussed in the introduction, we envisage several applications for the proposed MED and SMED. Here we demonstrate two of the potential applications.

4.1 Simulations From Complex Probability Densities

Because MED can reconstruct the underlying distribution, we believe that a major application for MED is in obtaining a representative sample from a desired density. The standard approach to obtain a representative sample is to first generate a uniform sample and then use the inverse probability transform method. However, such an approach cannot be used when the density is complex, which is common in Bayesian applications. In fact, in most Bayesian applications, the posterior is known only up to a normalizing constant and thus, it is impossible to use the inverse probability transform method to obtain the samples. This does not cause a problem for MED because the normalizing constant does not affect the optimization in (8). Moreover, simulating from a computationally expensive probability density is a challenging task, which can be efficiently solved using the proposed SMED algorithm.

Consider, for example, the two-dimensional probability density with banana-shaped contours discussed in Haario, Saksman, and Tamminen (2001) given by

$$f(\boldsymbol{x}) \propto \exp \left\{ -\frac{1}{2}\frac{x_1^2}{100} - \frac{1}{2}\left(x_2 + 0.03x_1^2 - 3 \right)^2 \right\}.$$

Using (9), we place the first sample at the mode. Now 49 more points are added one-by-one using the greedy algorithm (8) and are shown in Figure 4 (the region $[0, 1]^2$ corresponds to $[-20, 20] \times [-10, 5]$ in the actual scale). We used the
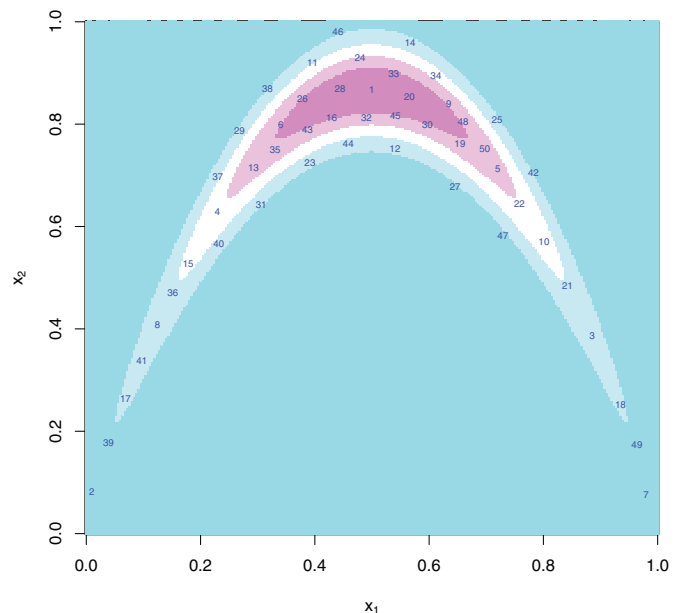

Figure 4. 50-point MED in the banana-shaped probability density example.

generalized simulated annealing algorithm using the R package *GenSA* (Gubian et al. 2012) for the optimization. We can see that the samples are well spread out and are placed at the high-probability regions of the density.

An important application of MED is that it can be used as the experimental design for DoIt (Joseph 2012, 2013). DoIt is mainly developed for approximating computationally expensive posterior densities. With the use of MED, DoIt can also be used for efficiently approximating inexpensive posteriors because MED can generate good space-filling designs in the high-probability regions without the need of any model fitting. Figure 3 in the online supplementary material file shows the DoIt approximation with the 50 MED samples generated earlier. This approximation is much better than the 100-run maximin LHD-based DoIt approximation shown in Figure 2 of Joseph (2013).

A major advantage of MED over the QMC sampling techniques is that it does not require the specification of an accurate region for exploration, which is usually unknown in most Bayesian applications. For example, suppose we use a larger region $[-40, 40] \times [-20, 10]$ for exploring the banana-shaped density. The MED samples are plotted in Figure 5. We can see that only a few points fall outside the high-probability region, whereas if we use a Sobol sequence, then 45 out of 50 points fall outside the high-probability region. This advantage of MED over QMC sampling increases as the number of dimensions increases because the region occupied by a density of highly correlated variables inside a hypercube can be negligibly small in higher dimensions. However, this advantage is not easy to realize in practice because generating MED samples in higher dimensions is hampered by the computational complexity of the global optimization required in (8).

To understand more about the effect of number of dimensions, consider the problem of simulating from a multivariate normal distribution. Let the mean be $(0, \ldots, 0)'$ and
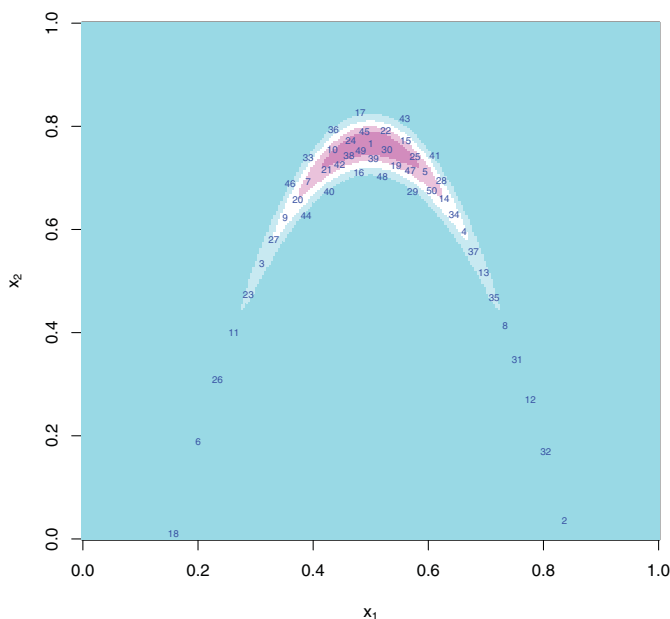

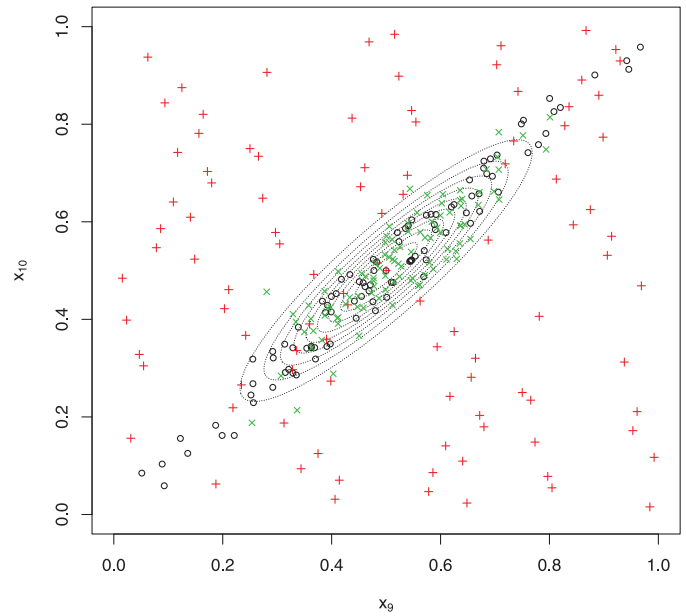Figure 6. A two-dimensional projection of 100 MC ($\times$'s), QMC (+'s), and MED (o's) samples of a 10-dimensional multivariate normal distribution.

variance-covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$, where $\sigma_{ij} = \rho^{|i-j|}$. Let $\rho = 0.9$ and $[-4, 4]^p$ be the region of exploration. Figure 4 in the online supplementary material file shows the time taken to generate $N = 10p$ and $N = 20p$ MED samples as a function of p in a 3.4 GHz desktop computer (the maximum number of iterations in *GenSA* is fixed at $n + 10p$, for $n = 1, \ldots, N$). Through curve fitting we find that the time increases approximately at the rate of $p^{1.5}N^{2.25}$. If $N \propto p$, then the time increases at the rate of $p^{3.75}$, which limits the applicability of MED in very high dimensions. To check the quality of the samples, we have generated MC and QMC samples (Sobol) of $N = 10p$ for $p = 10$. Figure 6 shows one of the two-dimensional projections of the samples. We can see that MED and MC samples nicely capture the high-probability regions of the density, whereas most of the QMC samples are wasted. Compared to MC simulation, MED has ceratin advantages. First, MED can be used for obtaining samples from any complex densities, whereas MC samples can be generated only for some standard densities. Of course, more advanced sampling techniques such as MCMC can be used to generate samples from complex densities but those samples are not as efficient as the MC samples because of their dependency. Second, since the MED points are well spread out, they can represent the probability density with much fewer samples than required by MC/MCMC samples. This helps in reducing the cost of subsequent predictions and optimizations using the simulated samples. Third, the advantage of MED over MC/MCMC increases as the probability density becomes more and more expensive to evaluate (Fielding, Nott, and Liong 2011) because we can use the SMED algorithm to generate samples. We now discuss the last point in more detail.

To illustrate the use of SMED, now suppose the probability density is expensive to evaluate. Because each evaluation is expensive, we can neither start at the mode nor can we find the design points by directly minimizing the energy criterion. We

Figure 5. MED with a larger region of exploration in the banana-shaped probability density example.

need to start somewhere in the design region and approximate the density using an easy-to-evaluate metamodel. The approximation can be refined as more points are added to the design. Here we choose a Gaussian process (GP) model (or kriging) for constructing the metamodel. Because the approximated density should be nonnegative, we fit the following stationary GP model (ordinary kriging) after taking a logarithmic transformation on the unnormalized density:

$$\log f(\boldsymbol{x}) \sim \mathrm{GP}(\mu, \sigma^2 R(\cdot)),$$

where the correlation function is defined as $\mathrm{cor}(\log f(\boldsymbol{x}_i), \log f(\boldsymbol{x}_j)) = R(\boldsymbol{x}_i - \boldsymbol{x}_j)$. If we observe the data $\boldsymbol{y}^{(n)} = (y_1, \ldots, y_n)'$ at the locations $\boldsymbol{D}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, where $y_i = \log f(\boldsymbol{x}_i)$, then the metamodel is given by (see Santner, Williams, and Notz 2003)

$$\hat{f}^{(n)}(\boldsymbol{x}) = \exp\left\{\hat{\mu}^{(n)} + \boldsymbol{r}^{(n)}(\boldsymbol{x})' \boldsymbol{R}_{(n)}^{-1}(\boldsymbol{y}^{(n)} - \hat{\mu}^{(n)} \boldsymbol{1}_n)\right\},$$

where $\boldsymbol{r}^{(n)}(\boldsymbol{x})$ is a vector of length n with ith element $R(\boldsymbol{x} - \boldsymbol{x}_i)$, $\boldsymbol{R}_{(n)}$ is the $n \times n$ correlation matrix with ijth element $R(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $\boldsymbol{1}_n$ is a vector of 1's, and $\hat{\mu}^{(n)} = \boldsymbol{1}_n' \boldsymbol{R}_{(n)}^{-1} \boldsymbol{y}^{(n)} / \boldsymbol{1}_n' \boldsymbol{R}_{(n)}^{-1} \boldsymbol{1}_n$. For the correlation function, we choose the popular Gaussian correlation function given by

$$R(\boldsymbol{h}) = \exp\left\{-\sum_{i=1}^{p} \theta_i h_i^2\right\}.$$

The correlation parameters can be estimated using a profile likelihood method (Santner, Williams, and Notz 2003). However, if a fully sequential design algorithm is used or if n_0 is small, then the maximum likelihood estimates may not exist. Therefore, we use a proper prior on the correlation parameters. We assume

$$\log \theta_i^{(n)} \stackrel{\text{iid}}{\sim} N\left(\frac{\log n}{2p}, n^{1/p}\right),$$

for $i = 1, \ldots, p$. This choice is based on our experience that a larger θ_i is needed for larger n. The variance is also increased with n because the optimal values of the correlation parameters can vary a lot, depending on the distribution of the points. Thus, $\log \boldsymbol{\theta}^{(n)}$ is estimated by minimizing

$$n \log \hat{\sigma}^2 + |\boldsymbol{R}_{(n)}| + \frac{1}{n^{1/p}} \sum_{i=1}^{p}\left(\log \theta_i - \frac{\log n}{2p}\right)^2,$$

where $\hat{\sigma}^2 = (\boldsymbol{y}^{(n)} - \hat{\mu}^{(n)} \boldsymbol{1}_n)' \boldsymbol{R}_{(n)}^{-1}(\boldsymbol{y}^{(n)} - \hat{\mu}^{(n)} \boldsymbol{1}_n)/n$.

We generated an initial design of $n_0 = 20$ runs using a maximin LHD (Morris and Mitchell 1995). Another 30 points were generated using the SMED algorithm and are shown as circles in Figure 7 along with the 20-run initial points. For comparison, we also generated 30 points using the sequential design method in DoIt (Joseph 2013), that are also shown in the same figure as crosses. We can see that the SMED is a much better design which places more points in the high-probability regions. To quantify the improvement, we fitted the DoIt approximation using both the designs and computed the error $\widehat{\sqrt{f(\boldsymbol{x})}} - \sqrt{f(\boldsymbol{x})}$. The mean squared error computed over a 100×100 grid for the approximation using SMED (0.0051) is found to be much less than that of the other sequential design method (0.0080). We repeated this process 100 times using a random LHD as the
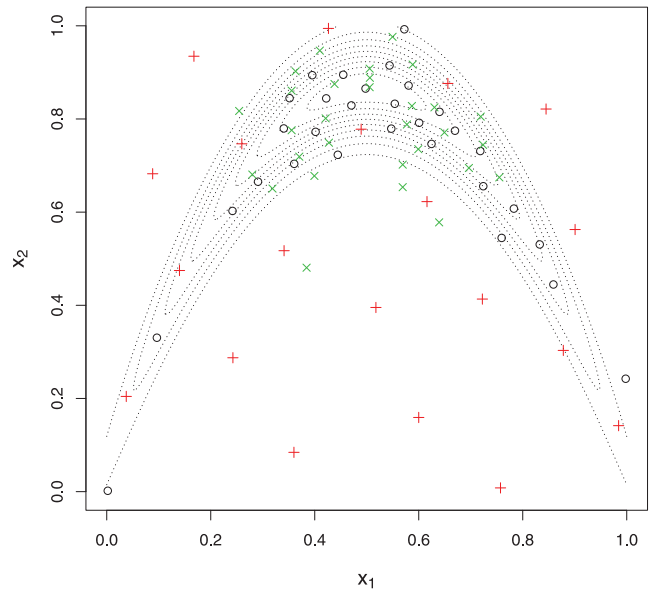

Figure 7. Comparison of SMED (o's) with the sequential design of Joseph (2013) (×'s). The initial points generated using a maximin LHD are shown as +'s.

initial design. The ratio of the mean squared error of Joseph's sequential design to that of the SMED has a median (Q_2) of 2.74 ($Q_1 = 1.80$ and $Q_3 = 4.57$), which clearly shows the superiority of SMED.

4.2 Exploration and Optimization of Expensive Black-Box Functions

Global optimization is an important but difficult problem. Many algorithms can be used to search for global optima such as simulated annealing and genetic algorithm. However, these algorithms require numerous evaluations of the function which can become costly if the objective function is expensive to evaluate. A popular approach for the optimization of expensive functions is the kriging-based expected improvement (EI) algorithm (Jones, Schonlau, and Welch 1998). See Jones (2001) for an entertaining review. Because SMED generates points with large function values and also with large interpoint distances, the points can become a dense set in the region of exploration and thus, can be useful for global optimization.

Instead of tuning the SMED for the purpose of global optimization, we focus on a slightly different objective. Suppose the objective is not only to find a global optimum, but also to find several good points that can serve as alternatives to the global optimum. This situation arises quite often in multiobjective optimization. Here is a real situation that we encountered during our consulting experience. A company wanted to optimize the design of their product using a computer model. However, the design engineers were not solely interested in the global optimum because the final decision about the product design would eventually be made by marketing people by looking at the aesthetics of the product and choosing the one that would be most appealing to the customers. Because the customer's taste was not well-defined and no mathematical formula could be derived
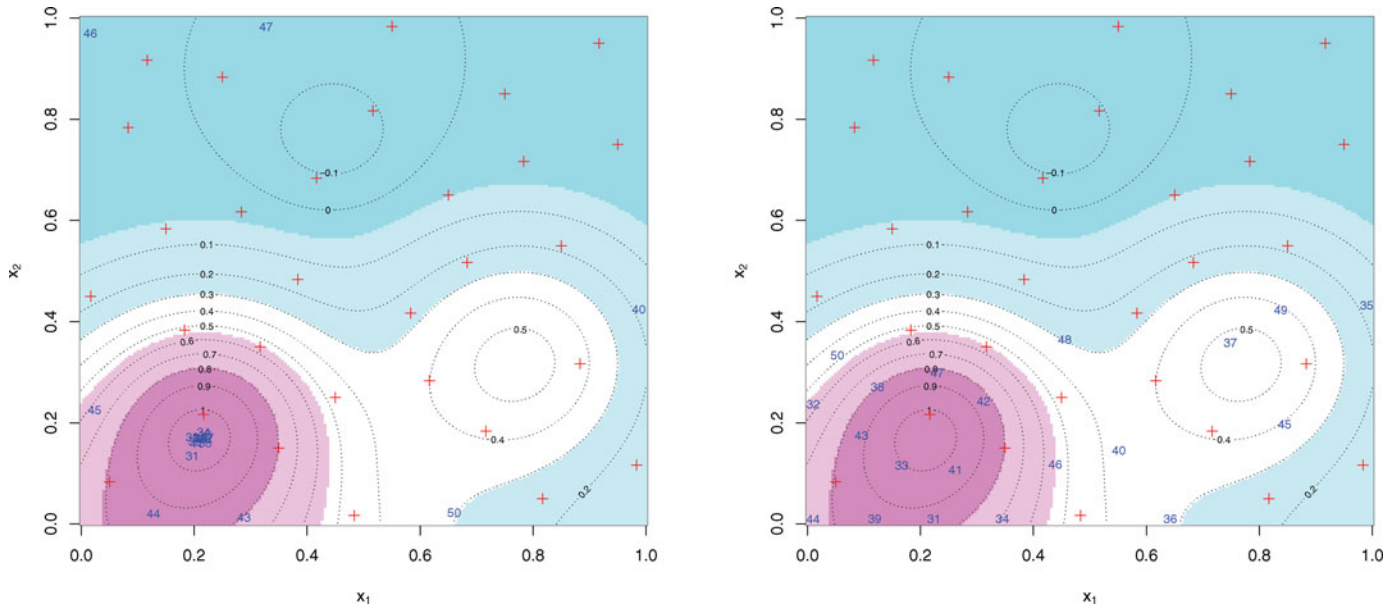
Figure 8. EI algorithm (left) and SMED (right) in the Franke's test function example.

on the beauty of the product, the only option for the engineers was to produce several prototypes and show them to the marketing people. In creating these prototypes, they wanted to choose products that would optimize the quality measures and also make them look different, so that they could present a wide choice of products to the marketing people. Clearly, SMED has the potential of offering a good solution to this problem.

Without loss of generality, define the optimization problem as to maximize $f(\boldsymbol{x})$ for $\boldsymbol{x}$ in some bounded region $\mathcal{X}$, which is taken to be $[0, 1]^p$. If the objective is minimization, we can use $-f(\boldsymbol{x})$ in place of $f(\boldsymbol{x})$. We cannot directly use SMED on $f(\boldsymbol{x})$ because the function values can be negative. Therefore, we define an auxiliary function:

$$g(\boldsymbol{x}) = f(\boldsymbol{x}) - f_{\min},$$

where $f_{\min} = \min_{\mathcal{X}} f(\boldsymbol{x})$. Clearly, $g(\boldsymbol{x}) \geq 0$ and has the same shape and maximum as $f(\boldsymbol{x})$. Now SMED can be applied on $g(\boldsymbol{x})$. However, for expensive functions, $f_{\min}$ cannot be obtained easily. Therefore, we proceed as before in approximating $f(\boldsymbol{x})$ sequentially.

Suppose again we choose a GP model to approximate the expensive computer model: $f(\boldsymbol{x}) \sim \mathrm{GP}(\mu, \sigma^2 R(\cdot))$. As before, based on the evaluations using an initial space-filling design containing n_0 points, we have the approximation

$$\hat{f}^{(n_0)}(\boldsymbol{x}) = \hat{\mu}^{(n_0)} + \boldsymbol{r}^{(n_0)}(\boldsymbol{x})' \boldsymbol{R}_{(n_0)}^{-1} \left(\boldsymbol{y}^{(n_0)} - \hat{\mu}^{(n_0)} \mathbf{1}_{n_0} \right).$$

Note that if the computer model is noisy or if the data are from a physical experiment, then a nugget term (see, e.g., Joseph 2006) should be added to the foregoing kriging predictor. Now we can obtain $\hat{f}^{(n_0)}_{\min}$ by minimizing $\hat{f}^{(n_0)}(\boldsymbol{x})$. Thus, we obtain $\hat{g}^{(n_0)}(\boldsymbol{x}) = \hat{f}^{(n_0)}(\boldsymbol{x}) - \hat{f}^{(n_0)}_{\min}$, which can be used in the SMED algorithm. At each $n = n_0 + 1, \ldots, N$, $\hat{f}^{(n)}(\boldsymbol{x})$ and $\hat{f}^{(n)}_{\min}$ can be updated to give the new $\hat{g}^{(n)}(\boldsymbol{x})$.

For illustration, consider Franke's two-dimensional function (see, e.g., Fasshauer 2007)

$$\begin{aligned}
f(\boldsymbol{x}) &= \frac{3}{4} \exp\left\{ -\frac{1}{4}(9x_1 - 2)^2 - \frac{1}{4}(9x_2 - 2)^2 \right\} \\
&+ \frac{3}{4} \exp\left\{ -\frac{1}{49}(9x_1 + 1)^2 - \frac{1}{10}(9x_2 + 1)^2 \right\} \\
&+ \frac{1}{2} \exp\left\{ -\frac{1}{4}(9x_1 - 7)^2 - \frac{1}{4}(9x_2 - 3)^2 \right\} \\
&- \frac{1}{5} \exp\{ -(9x_1 - 4)^2 - (9x_2 - 7)^2 \}.
\end{aligned}$$

We start with a maximin LHD of $n_0 = 30$ points. The minimum value is estimated as $\hat{f}^{30}_{\min} = -0.095$. Now SMED can be applied on the auxiliary function. The next 20 points are shown in the right panel of Figure 8. Although the global optimum $\boldsymbol{x}^* = (0.21, 0.17)$ is missed, there are many points close to the optimum. In fact, if we maximize $\hat{f}^{50}(\boldsymbol{x})$, we obtain approximately the optimum $\boldsymbol{x}^*$. This is because there are many points around the optimum and thus f is well approximated by $\hat{f}$ in that region.

For comparison, the EI algorithm is used to generate 20 points using the same 30-run initial design. The points are shown in the left panel of Figure 8. We can see that EI works beautifully and identifies the global optimum successfully. In fact, there are several points clustered around the global optimum. However, this clustering can be perceived as a weakness because it cannot provide as many distinct alternatives as SMED. For example, the SMED contains 13 points for which the function values are higher than 0.5 but less than 1, whereas only 3 points are available in the design generated by the EI algorithm. Moreover, the clustering can cause severe numerical instability in the kriging predictor (Haaland and Qian 2011), leading to a poor approximation of the overall surface. In fact, when we used the R package *DiceOptim* (Ginsbourger, Picheny, and Roustant 2013), it could not add more than four points to the initial design due to numerical problems.
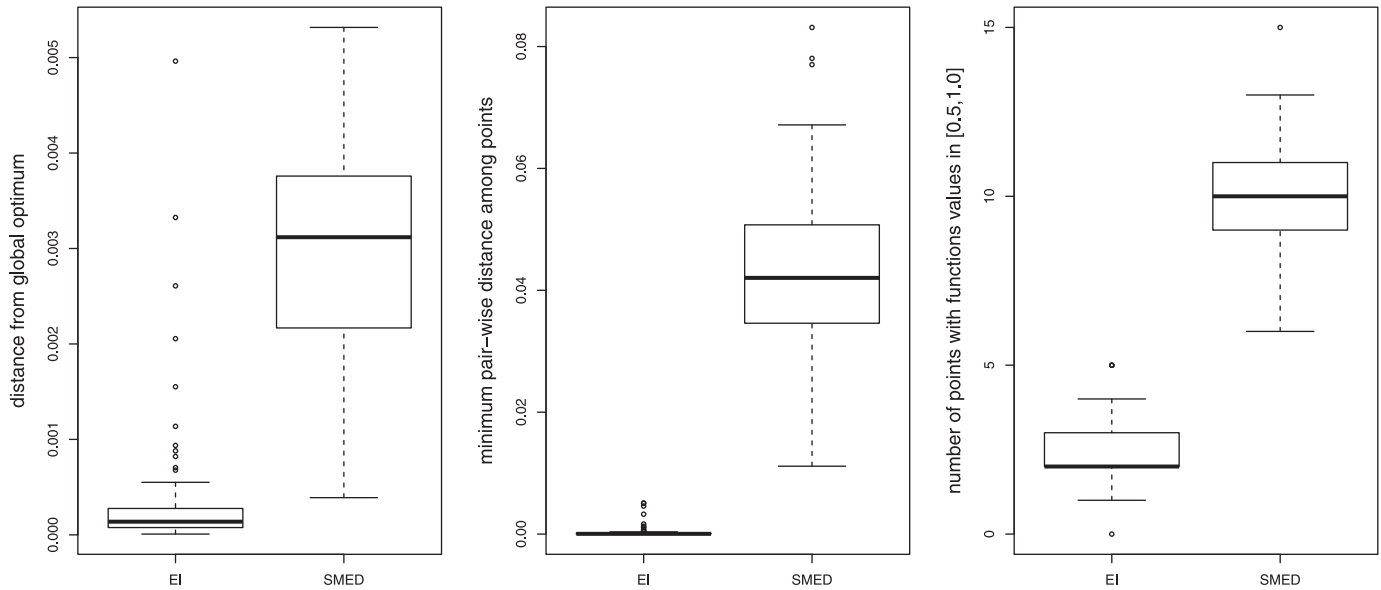
Figure 9. Comparison of EI algorithm and SMED.

We have repeated the above comparison 100 times using randomly generated LHDs as the initial designs. Three performance measures are computed for the EI algorithm and SMED: (i) the Euclidean distance of the optimum found by the method from the global optimum, (ii) minimum pair-wise distance among the 50 points, and (iii) number of points among the 20 sequentially generated points with function values more than 0.5 but less than 1.0. The boxplots of the performance measures are shown in Figure 9. As expected, the EI algorithm shows superior performance over SMED in terms of finding the global optimum. However, the performance of SMED is not bad in the sense that it could locate the global optimum within a radius of 0.005. On the other hand, SMED is superior to EI in terms of the minimum distance. In fact, the minimum distance among the SMED points is about 1000 times larger than that of EI. This will have a significant effect on numerical stability of kriging and the quality of approximation of the whole surface. Clearly, the number of points identified by SMED with high function values are much larger than that of EI. Thus, we believe that the points generated by SMED can be more useful than those generated by the EI algorithm in many problems, especially in multiobjective optimization.

5. CONCLUDING REMARKS

We have proposed a new space-filling design for exploration of complex response surfaces, with particular sensitivity to situations where specific regions of the design space are of special interest. The proposed minimum energy design (MED) is based on ideas drawn from the physical system analogy of visualizing design points as charged particles in a box, and minimizing the total potential energy inside the box. The key aspect which makes MED a novel approach different from similar criteria-based designs is its ability to adapt to different types of response surfaces by choosing the charge function inversely proportional to the function of interest. A theoretical result associated with

MED provides us with insights into connections between two well-known designs—maximin distance designs and uniform designs. We have also heuristically argued that under regularity conditions and proper choice of the charge function, the MED can generate any arbitrary probability density function in the limit. We have developed a sequential algorithm for adaptive implementation of MED and demonstrated two potential applications in simulation and optimization through implementation of the algorithm.

As demonstrated in Section 4.1, the time required to obtain the MED samples explodes as the dimension gets very large. For example, generating 100 samples in 10 dimensions took about 3 min but it took more than 3 hr to generate 300 samples in 30 dimensions. Of course, efficient programming and changing from R to C++ environment can drastically reduce this computational time. However, it can still be a problem if we need to generate samples of larger size. Therefore, developing an efficient optimization algorithm that is specifically designed for the MED objective function, especially in high dimensions, is an important problem for future investigation.

Another topic mentioned by a reviewer is about incorporating the uncertainties of the approximate model $\hat{f}^{(n)}(\boldsymbol{x})$ in the SMED algorithm. The current implementation neglects these uncertainties and, therefore, SMED can miss some important regions of the response surface if the initial design is poor. We hope to address this important issue in a future work.

APPENDIX: TWO LEMMAS

Two lemmas are stated in this appendix which will be used for proving Theorem 1. Lemma 1 describes an important property of the maximin distance designs with the smallest index (note that when $q(\boldsymbol{x}) \equiv 1$, an MED with $k = \infty$ becomes a maximin distance design.) Lemma 2 gives lower and upper bounds of the minimum distance for n-point designs. The proof of the lemmas are given in the supplementary file.

First we introduce the following notation. Given a design D, define the minimum distance over a subset $A \subset \mathcal{X}$ by

$$d_*(A, D) := \min\{d(x_i, x_j) : x_i, x_j \in A \cap D, x_i \neq x_j\},$$

if $A \cap D$ has at least two points. Let $d_*(D) = d_*(\mathcal{X}, D)$. It is easily seen that

$$d_*(D) = \min(d_*(A, D), d_*(A^c, D), d(D \cap A, D \cap A^c)),$$
(A.1)

for *any* A so that both $A \cap D$ and $A^c \cap D$ have at least two points. We call $x \in D$ a *critical point* if there exists $x' \in D$ with $x' \neq x$ such that $d(x, x') = d_*(D)$. Let $B(a, r)$ be the open ball centered at a with radius r.

Lemma A.1. Suppose D is an n-point maximin distance design with the smallest index and D' is an n'-point design with $n' > n$ and $d_*(D') = d_*(D)$. Then $IN(D') > IN(D)$ holds.

Lemma A.2. Suppose $D = \{x_1, \ldots, x_n\}$ is a design over Ω. Let $d_0 = \min_{i \neq j} d(x_i, x_j)$. Then

1. If $\Omega = [0, 1]^p$, and D is a maximin distance design, then

$$\frac{1}{n^{1/p} + 1} < d_0.$$

2. If $\Omega = [0, 1]^p$, for any design D,

$$d_0 < \frac{2(\Gamma(p/2 + 1))^{1/p}}{n^{1/p}\pi^{1/2} - 2(\Gamma(p/2 + 1))^{1/p}},$$

provided that $n > (2/\sqrt{\pi})^p \Gamma(p/2 + 1)$, where Γ is the Gamma function.

3. Suppose $\Omega = [0, 1]^p$ and D is a maximin distance design. For any $B(a, r) \subset \Omega$, let $N_B = \text{card}(D \cap B(a, r))$. Then for sufficiently large n,

$$N_B \geq \left(\frac{2r - 2d_0}{\sqrt{p}d_0} - 1\right)^p.$$

4. If $\Omega = cl(B(0, r_1) - B(0, r_2))$, $0 < r_2 < r_1$, where cl denotes the closure of a set and the two balls are p-dimensional, then for sufficiently large n and any design D,

$$nd_0^p < (2r_1 + d_0)^p - (2r_2 - d_0)^p.$$

SUPPLEMENTARY MATERIALS

Proofs and Figures: The proofs of Theorem 1, Result 1, and the two lemmas stated in the Appendix are in the file supplemtary.pdf. It also contains some figures.
R codes: The R codes can be downloaded as a .zip file.

ACKNOWLEDGMENTS

[Received December 2012. Revised January 2014.]

REFERENCES

Audze, P., and Eglais, V. (1977), "New Approach for Planning Out of Experiments" (Riga: Zinatne Publishing House (in Russian)), *Problems of Dynamics and Strengths*, 35, 104–107. [65,66]

Bates, S. J., Sienz, J., and Langley, D. S. (2003), "Formulation of the Audze–Eglais Uniform Latin Hypercube Design of Experiments," *Advances in Engineering Software*, 34, 493–406. [66]

Cox, D. R. (1957), "Note on Grouping," *Journal of the American Statistical Association*, 52, 543–547. [65]

Dalenius, T. (1950), "The Problem of Optimum Stratification," *Skandinavisk Aktuarietidskrift*, 33, 203–213. [65]

Dasgupta, T., Ma, C., Joseph, V. R., Wang, Z. L., and Wu, C. F. J. (2008), "Statistical Modeling and Analysis for Robust Synthesis of Nanostructures," *Journal of the American Statistical Association*, 103, 594–603. [64]

Dette, H., and Pepelyshev, A. (2010), "Generalized Latin Hypercube Design for Computer Experiments," *Technometrics*, 52, 421–429. [66]

Fang, K.-T., and Wang, Y. (1994), *Number-Theoretical Methods in Statistics*, Boca Raton, FL: Chapman & Hall. [65,68]

Fang, K.-T., Li, R., and Sudijanto, A. (2006), *Design and Modeling for Computer Experiments*, Boca Raton, FL: Chapman & Hall. [67]

Fasshauer, G. (2007), *Meshfree Approximation Methods With MATLAB*, Singapore: World Scientific Publishers. [72]

Fielding, M., Nott, D. J., and Liong, S.-Y. (2011), "Efficient MCMC Schemes for Computationally Expensive Posterior Distributions," *Technometrics*, 53, 16–28. [70]

Flury, B. (1990), "Principal Points," *Biometrika*, 77, 33–41. [65]

Gubian, S., Xiang, Y., Suomela, B., and Hoeng, J. (2012), *GenSA: R Functions for Generalized Simulated Annealing*, R Package Version 1.1.2. [70]

Ginsbourger, D., Picheny, V., and Roustant, O. (2013), *DiceOptim: Kriging-Based Optimization for Computer Experiments*, R Package Version 1.4. [72]

Haaland, B., and Qian, P. Z. G. (2011), "Accurate Emulators for Large-Scale Computer Experiments," *The Annals of Statistics*, 39, 2974–3002. [72]

Jin, R., Chang, C. -J., and Shi, J. (2012), "Sequential Measurement Strategy for Wafer Geometric Profile Estimation," *IIE Transactions*, 44, 1–12. [64]

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. [67]

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131–148. [66]

Jones, D. R. (2001), "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization*, 21, 345–383. [64,71]

Jones, D. R., Schonlau, M., and Welch, W. J. (1998), "Efficient Global Optimization of Expensive Black Box Functions," *Journal of Global Optimization*, 13, 455–492. [71]

Joseph, V. R. (2006), "Limit Kriging," *Technometrics*, 48, 458–466. [72]

——— (2012), "Bayesian Computation Using Design of Experiments-Based Interpolation Technique" (with discussion), *Technometrics*, 209–242. [70]

——— (2013), "A Note on Nonnegative DoIt Approximation," *Technometrics*, 55, 103–107. [70,71]

Joseph, V. R., Hung, Y., and Sudjianto, A. (2008), "Blind Kriging: A New Method for Developing Metamodels," *ASME Journal of Mechanical Design*, 130, 031102–031108. [64]

Kessels, R., Jones, B., Goos, P., and Vanderbrook, M. (2009), "An Efficient Algorithm for Constructing Bayesian Optimal Choice Designs," *Journal of Business and Economic Statistics*, 27, 279–291. [66]

Lloyd, S. P. (1957), "Least Squares Quantization in PCM's," Bell Telephone Laboratories Article, Murray Hill, NJ. [65]

Max, J. (1960), "Quantizing for Minimum Distortion," *IRE Transactions on Information Theory*, IT-6, 7–12. [65]

Morris, M. D., and Mitchell, T. J. (1995), "Exploratory Designs for Computer Experiments," *Journal of Statistical Planning and Inference*, 43, 381–402. [65,66,67,71]

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer. [71]

Wang, Y. (2011), *Asymptotic Theory for Decentralized Sequential Hypothesis Testing Problems and Sequential Minimum Energy Design Algorithm*, Ph.D. thesis, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA. [67]

Zhou, S. (1999), *Electrodynamics of Solids and Microwave Superconductivity*, New York: Wiley. [66]