Biometrika (xxxx), **xx**, x, *pp*. 1–10

Printed in Great Britain

Maximum Projection Designs for Computer Experiments

BY V. ROSHAN JOSEPH AND EVREN GUL

755 Ferst Drive NW, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A.

AND SHAN BA

8700 Mason-Montgomery Road, The Procter & Gamble Company, Mason, Ohio 45040, U.S.A.

SUMMARY

Space-filling properties are important in designing computer experiments. The traditional maximin and minimax distance designs only consider space-filling in the full dimensional space. This can result in poor projections onto lower dimensional spaces, which is undesirable when only a few factors are active. Restricting maximin distance design to the class of Latin hypercubes can improve one-dimensional projections, but cannot guarantee good space-filling properties in larger subspaces. We propose designs that maximize space-filling properties on projections to all subsets of factors. We call our designs maximum projection designs. Our design criterion can be computed at a cost no more than a design criterion that ignores projection properties.

Some key words: Experimental design; Gaussian process; Latin hypercube design; Screening design; Space-filling design.

1. INTRODUCTION

Space-filling designs are commonly used in deterministic computer experiments. One such design, maximin distance design (Johnson et al., 1990), can be defined as follows. Suppose we want to construct an n-run design in p factors. Let the design region be the unit hypercube $\mathcal{X}$ and let the design be $D = \{x_1, \ldots, x_n\}$, where each design point $x_i \in \mathcal{X} = [0,1]^p$. The maximin distance design tries to spread out the design points in $\mathcal{X}$ by maximizing the minimum distance between any two design points:

$$\max_{D} \min_{x_i, x_j \in D} d(x_i, x_j), \tag{1}$$

where $d(x_i, x_j)$ is the Euclidean distance between the points x_i and x_j.

The maximin distance criterion tends to place a large proportion of points at the corners and boundaries of the hypercube $[0,1]^p$, and thus, unlike Latin hypercube designs (McKay et al., 1979), maximin distance designs do not have good projection properties for each factor. Morris & Mitchell (1995) proposed to overcome this problem by searching for the maximin distance design within the class of Latin hypercube designs. They used the criterion

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

for finding the maximin Latin hypercube design, where $k > 0$ is chosen large enough to achieve maximin distance.

Although maximin Latin hypercube designs ensure good space-filling in p dimensions and uniform projections in each dimension, their projection properties in $2, \ldots, p-1$ dimensions are not known. By the effect sparsity principle, only a few factors are expected to be active. Since the active factors are unknown before the experiment, good projections to all subspaces of the factors are important. However, little research has been done in trying to find space-filling designs that ensure good projections to subspaces of the factors (Tang, 1993; Moon et al., 2011). Draguljic et al. (2012) proposed a criterion incorporating projection properties in (2):

$$\min_D \left[\frac{1}{\binom{n}{2} \sum_{q \in J} \binom{p}{q}} \sum_{q \in J} \sum_{r=1}^{\binom{p}{q}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{ \frac{q^{1/2}}{d_{qr}(x_i, x_j)} \right\}^k \right]^{1/k}, \tag{3}$$

where $d_{qr}(x_i, x_j)$ is the Euclidean distance between points x_i and x_j in the rth projection of q factors, $q \in J \subseteq \{1, 2, \ldots, p\}$. However, (3) is difficult to compute for large p, so Draguljic et al. (2012) had to focus their attention to subspaces with no more than two factors.

A uniformity measure is another type of space-filling criterion. The idea is to spread the design points in the space so that the empirical distribution of the points is uniform on $[0, 1]^p$. Hickernell (1998) proposed the centered L_2-discrepancy criterion, which ensures good projections to all subspaces. Although uniform designs are useful for approximating integrals, it is not clear if they are as good as maximin Latin hypercube designs for approximating functions. In fact, Dette & Pepelyshev (2010) have shown that placing more points in the boundaries than around the center can minimize the prediction errors from Gaussian process modeling.

2. MAXIMUM PROJECTION DESIGNS

When a design is projected onto a subspace, the distances between the points are calculated with respect to the factors that define the subspace. Therefore, define a weighted Euclidean distance between the points x_i and x_j as

$$d(x_i, x_j; \theta) = \left\{ \sum_{l=1}^{p} \theta_l (x_{il} - x_{jl})^2 \right\}^{1/2},$$

where $\theta_l = 1$ for the factors defining the subspace and $\theta_l = 0$ for the remaining factors. It makes sense to use weights between 0 and 1, which can be viewed as measures of importance for the factors. Let $0 \leq \theta_l \leq 1$ be the weight assigned to factor l and let $\sum_{l=1}^{p} \theta_l = 1$. Then, the criterion in (2) can be modified to

$$\min_D \phi_k(D; \theta) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(x_i, x_j; \theta)}, \tag{4}$$

where $\theta = (\theta_1, \ldots, \theta_{p-1})^T$ and $\theta_p = 1 - \sum_{l=1}^{p-1} \theta_l$. We omit the power $1/k$ in (2) because we are interested only in finite values of k. Unfortunately, we have no idea about the importance of the factors before the experiment, so there is no easy way to choose θ. One approach to overcome this is to adopt a Bayesian framework, i.e., to assign a prior distribution for θ and then to minimize the expected value of the objective function.

Assuming equal importance to all factors a priori, we take the prior distribution for θ to be

$$p(\theta) = \frac{1}{(p-1)!}, \theta \in S_{p-1}, \tag{5}$$

where $S_{p-1} = \{\theta : \theta_1, \ldots, \theta_{p-1} \geq 0, \sum_{i=1}^{p-1} \theta_i \leq 1\}$. Thus, our design criterion becomes

$$\min_D E\{\phi_k(D;\theta)\} = \int_{S_{p-1}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^k(x_i, x_j; \theta)} p(\theta) d\theta. \tag{6}$$

In general, this is not easy to evaluate. However, we can perform the integration analytically for a special case of k as shown below. All the proofs are given in the Appendix.

THEOREM 1. *If $k = 2p$, then under the prior in (5)*

$$E\{\phi_k(D;\theta)\} = \frac{1}{\{(p-1)!\}^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{\prod_{l=1}^{p} (x_{il} - x_{jl})^2}. \tag{7}$$

Thus, we propose a new criterion:

$$\min_D \psi(D) = \left\{ \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{\prod_{l=1}^{p} (x_{il} - x_{jl})^2} \right\}^{1/p}. \tag{8}$$

For any l, if $x_{il} = x_{jl}$ for $i \neq j$, then $\psi(D) = \infty$. Therefore, the design that minimizes $\psi(D)$ must have n distinct levels for each factor. Furthermore, because the denominator of (8) has products of squared distances from all the factors, no two points can get close to each other in any of the projections. Thus, the design that minimizes $\psi(D)$ tends to maximize its projection capability in all subspaces of factors. Therefore, we call the optimal design a maximum projection design.

3. OPTIMAL DESIGN CONSTRUCTION ALGORITHM

Although $\psi(D)$ in (8) is easy to compute, finding the maximum projection design by minimizing it is not an easy problem. First, the number of variables in the optimization, np, is extremely large even for moderate-sized problems. Second, the objective function has many local minima because it becomes infinite whenever $x_{il} = x_{jl}$ for any $l = 1, \ldots, p$ and $i \neq j$. Moreover, the design remains the same under the reordering of rows or columns, which produces many local minima. Thus, optimization of (8) directly using a continuous optimization algorithm can easily terminated at a local optimum.

Because a maximum projection design will have n distinct levels for each factor, it can be viewed as a Latin hypercube design, but not necessarily with equally spaced levels. Therefore, we can make use of the algorithms for the construction of optimal Latin hypercube designs such as simulated annealing (Morris & Mitchell, 1995). After obtaining the optimal maximum projection Latin hypercube design, we apply a continuous optimization algorithm to find the locally-optimal maximum projection design in the neighborhood of the optimal maximum projection Latin hypercube design. The gradient of the objective function is

$$\frac{\partial \psi^p(D)}{\partial z_{rs}} = \frac{2}{\binom{n}{2}} \sum_{i \neq r} \frac{1}{\prod_{l=1}^{p} (x_{il} - x_{rl})^2} \frac{1}{(x_{is} - x_{rs})},$$

which can be used to implement a fast derivative-based optimization algorithm.

4. NUMERICAL RESULTS

In this section, we compare the performances of maximum projection designs and maximum projection Latin hypercube designs and other popular space-filling designs such as maximin Latin hypercube designs, uniform designs, and generalized maximin Latin hypercube designs (Dette & Pepelyshev, 2010), on nine simulation settings with $p = 5, 10, 20$ and $n = 5p, 10p, 20p$. Because the conclusions from all the nine cases are similar, we report results only for the 100-run 10-factor design. The maximin Latin hypercube design is constructed using the R package SLHD (Ba, 2013), the uniform design is constructed using the software JMP, and the generalized maximin Latin hypercube design is constructed using the arc-sine transformation of the maximin Latin hypercube design (Dette & Pepelyshev, 2010).

Designs are compared on three space-filling criteria: a maximin distance measure, a minimax distance measure, and a uniformity measure. These three measures are summarized for each sub-dimensional projection. For a given sub-dimension q, a measure is computed for all possible projections and the worst case is used for comparison.

Consider the maximin criterion in (1). A better maximin measure in projection dimension q that also incorporates the maximin index of the design is

$$\mathrm{Mm}_q = \min_{r=1,\ldots,\binom{p}{q}} \left\{ \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d_{qr}^{2q}(x_i, x_j)} \right\}^{-1/(2q)},$$

where $d_{qr}(x_i, x_j)$ is the Euclidean distance between points x_i and x_j in the rth projection of dimension q. This measure is plotted in Figure 1. As expected, the maximin Latin hypercube and generalized maximin Latin hypercube designs have the largest Mm_q values in the full design space. On the other hand the maximum projection and maximum projection Latin hypercube designs outperform the others even when the projection dimension is reduced by one.

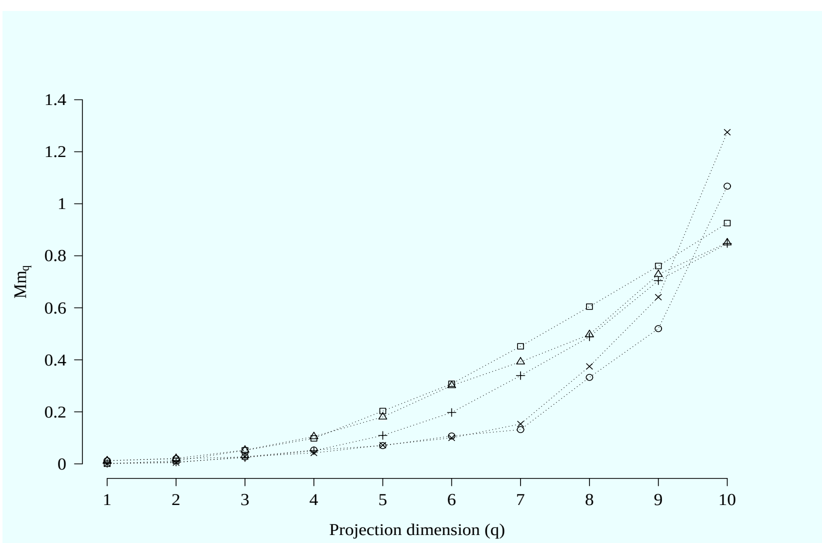

Fig. 1. Plot of Mm_q (larger-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), uniform (pluses), and generalized maximin Latin hypercube (crosses) designs.

Now consider the minimax distance criterion (Johnson et al., 1990):

$$\min_{D} \max_{x \in \mathcal{X}} d(x, D),$$

where $d(x, D) = \min_{x_i \in D} d(x, x_i)$. Its computation for a given projection dimension q is cumbersome because we need to search over the whole space $[0, 1]^q$ to find the point having maximum distance to the nearest point in the design. Therefore, we approximate the distance by sampling a large number N_q of points from $[0, 1]^q$. We use the union of 3^q factorial design with levels $\{0, 0.5, 1\}$ and 2^{16} run Sobol sequence for the sample. A better minimax measure that also incorporates the minimax index of the design is

$$\mathrm{mM}_q = \max_{r=1,\ldots,\binom{p}{q}} \max_{u \in \mathcal{X}_q} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_{qr}^{2q}(u, x_i)} \right\}^{-1/(2q)},$$

where $\mathcal{X}_q$ is the set of sample points with size $N_q = 3^q + 2^{16}$. This measure is plotted in Figure 2. There is no clear winner in this case.

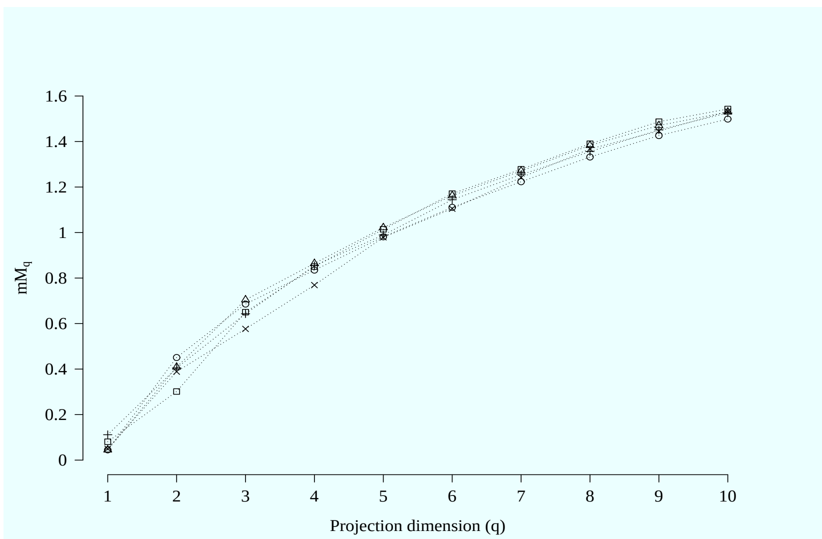

Fig. 2. Plot of mM_q (smaller-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), uniform (pluses), and generalized maximin Latin hypercube (crosses) designs.

Finally, consider the centered L_2-discrepancy measure, CL_2, defined in Hickernell (1998). The maximum values of CL_2 among the q-dimensional projections are shown in Figure 3. As expected, the uniform design performs best under this criterion because it is obtained by minimizing CL_2. The performance of the maximum projection design is significantly worse for the CL_2 criterion compared to the uniform design, but much better than that of the generalized maximin Latin hypercube design. This is because both maximum projection and generalized maximin Latin hypercube designs favor more points towards the boundaries than at the center, which affects the overall uniformity of the points in the design space. Interestingly, the maximum projection Latin hypercube design has much better uniformity than the maximum projection design, over all dimensions. This is surprising because the Latin hypercube restriction only makes the spacing of the levels equal, but somehow it improves the uniformity in all subspaces. We believe that uniformity is not as important as the maximin and minimax distance measures in a computer

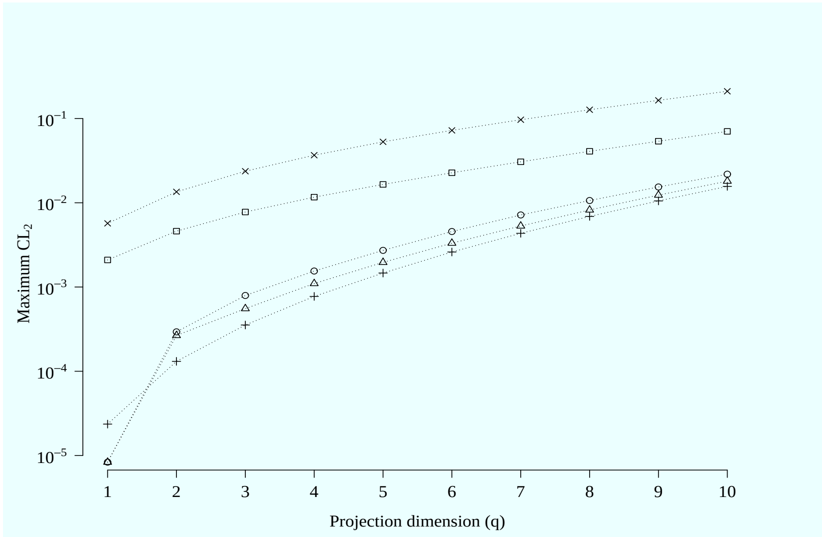

Fig. 3. Plot of maximum CL_2 (smaller-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), uniform (pluses), and generalized maximin Latin hypercube (crosses) designs.

experiment because the primary objective of a computer experiment is approximation and not integration. Thus, the poor performance of maximum projection designs under the uniformity measure is not of great concern.

5. Gaussian Process Modeling

The space-filling designs discussed earlier are model-independent, which allows the experimenter to fit a wide variety of models to the data. On the other hand, better designs can be developed for a specified model class. In computer experiments, Gaussian process modeling or kriging is widely used for approximating the response surface (Sacks et al., 1989). The ordinary kriging model is $Y(x) = \mu + Z(x)$, where $x \in \mathbb{R}^p$, μ is the overall mean, and $Z(x)$ is a stationary Gaussian process with mean zero and covariance function $\sigma^2 R(\cdot)$. A popular choice for $R(\cdot)$ is the Gaussian correlation function,

$$R(x_i - x_j; \alpha) = \exp\left\{-\sum_{l=1}^{p} \alpha_l (x_{il} - x_{jl})^2\right\}, \alpha_l \in (0, \infty), l = 1, \ldots, p. \tag{9}$$

The maximum entropy design (Shewry & Wynn, 1987) is obtained by maximizing the determinant $|R(\alpha)|$, where $R(\alpha)$ is the correlation matrix with (i, j) element $R(x_i - x_j; \alpha)$. A major drawback with such a model-based design is that we need to specify the value of α for finding the optimal design, which is unknown before conducting the experiment. One can use a guessed value of α for finding the optimal design, but the designs can be very poor if the guess is badly wrong. One way to mitigate this problem is to assign a prior distribution for α and optimize the expected value of the objective function. Let us assume a noninformative prior for α:

$$p(\alpha) \propto 1, \alpha \in \mathbb{R}_+^p. \tag{10}$$

THEOREM 2. *For the Gaussian correlation function and the noninformative prior for α in* (10)*, maximum projection designs minimize $E\{\sum_{i=1}^n \sum_{j\neq i} R_{ij}^{\gamma}(\alpha)\}$ for any $\gamma > 0$.*

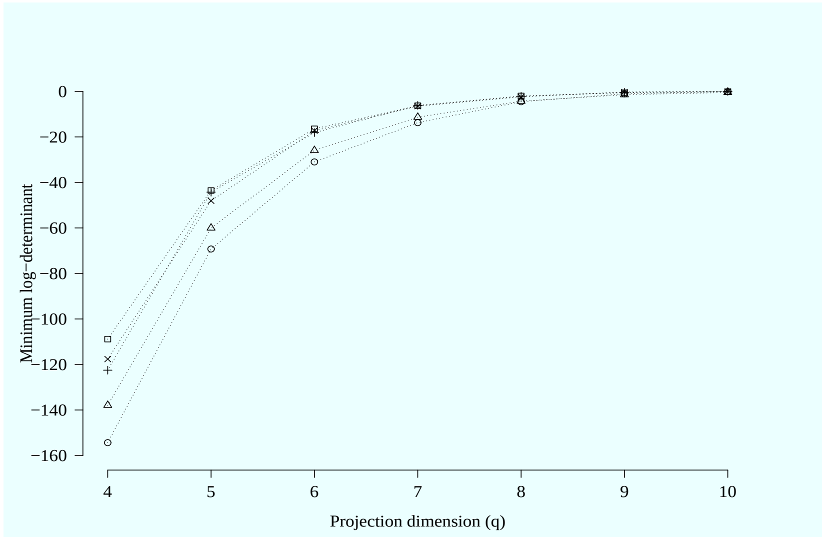

Fig. 4. Plot of minimum log-determinant (larger-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), maximum entropy design (pluses), and generalized maximin Latin hypercube (crosses) designs.

In other words, maximum projection designs minimize the expected sum of off-diagonal elements of the correlation matrix. Although this result is not directly related to the maximum entropy criterion, there seems to be some connections. An application of Hadamard's inequality and Gershgorin's theorem gives the following bounds on $|R(\alpha)|$:

$$\prod_{s=1}^{n}\left\{1-\sum_{j\neq i_s} R_{i_s j}(\alpha)\right\}_+ \leq |R(\alpha)| \leq 1,$$

where $i_1,\ldots,i_n \in \{1,\ldots,n\}$, which need not be distinct and $\{x\}_+ = \max\{x,0\}$. Thus, minimizing the off-diagonal elements of the $R(\alpha)$ matrix tends to increase the lower bound on the $|R(\alpha)|$, and the upper bound is achieved when all of the off-diagonal elements are 0. Therefore, the maximum projection designs are expected to perform well on the maximum entropy criterion. The nice thing is that we do not need to specify any value for the correlation parameters to obtain maximum projection designs.

To compare the performance of maximum projection design with maximum entropy design, we generated a 100-run 10-factor maximum entropy design using the software JMP, where the correlation parameter α_l is set to be equal to 5 for $l=1,\ldots,10$. Now we compute the minimum determinant among the q-dimensional projections, $\min_r \log|R^{q,r}(\alpha)|$, where $R^{q,r}$ is the correlation matrix calculated for the rth q-dimensional projection. As seen in Figure 4, the maximum projection design is better than the maximum entropy design in lower dimensional projections and has comparable performance in higher dimensional projections.

The ordinary kriging predictor is

$$\hat{y}(x) = \hat{\mu}(\alpha) + r(x;\alpha)^T R^{-1}(\alpha)\{y-\hat{\mu}(\alpha)1_n\},$$

where $\hat{\mu}(\alpha) = 1_n^T R^{-1}(\alpha)y / 1_n^T R^{-1}(\alpha)1_n$, $r(x;\alpha)$ is an $n\times 1$ vector with ith element $R(x - x_i;\alpha)$, $y=(y_1,\ldots,y_n)^T$ is the experimental data, and 1_n is a vector of 1's having length n. We

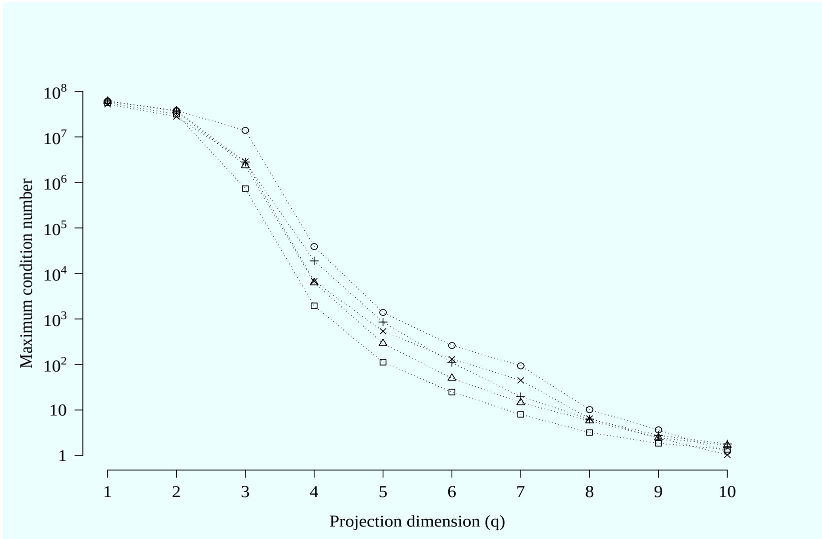

Fig. 5. Plot of maximum condition number (smaller-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), uniform (pluses), and generalized maximin Latin hypercube (crosses) designs.

can see that the predictor involves the inverse of the correlation matrix at the optimal value of α. The optimal value of α is estimated using likelihood or cross validation based methods. For example, the maximum likelihood estimate can be obtained by minimizing

$$\log|R(\alpha)| + n\log\hat{\sigma}^2(\alpha)$$

with respect to α, where $\hat{\sigma}^2(\alpha) = \{y - \hat{\mu}(\alpha)1_n\}^T R^{-1}(\alpha)\{y - \hat{\mu}(\alpha)1_n\}/n$. This again requires the computation of $R^{-1}(\alpha)$, but now for many values of α. The matrix inverse operation can become difficult and unstable at some values of α, so it will be good to use an experimental design that avoids this for all values of α. The instability of a matrix inverse can be assessed using its condition number. Figure 5 shows the maximum of condition numbers among the q-dimensional projections after adding a small nugget, 10^{-6}, to the diagonals of $R(\alpha)$, where $\alpha_l = 5$ for $l = 1, \ldots, 10$. We can see that maximum projection design has superior performance over the other designs.

The maximum prediction variance in the design space can be used as another criterion for evaluating a design. For ordinary kriging, the prediction variance is proportional to

$$1 - r(x;\alpha)^T R^{-1}(\alpha) r(x;\alpha) + \frac{\{1 - r(x;\alpha)^T R^{-1}(\alpha)1_n\}^2}{1_n^T R^{-1}(\alpha)1_n}.$$

The maximum prediction variance among the q-dimensional projections can now be approximated using the same set of N_q points used earlier in approximating the minimax measure and is plotted in Figure 6. The maximum projection design and the generalized maximin Latin hypercube design are the two winners on this performance measure.

Considering all the Gaussian process-based criteria and the space-filling criteria except the uniformity measure, the maximum projection design seems to be an attractive choice for deterministic computer experiments. However, if uniformity is also important in a particular application, then the maximum projection Latin hypercube design seems to be a good compromise choice.

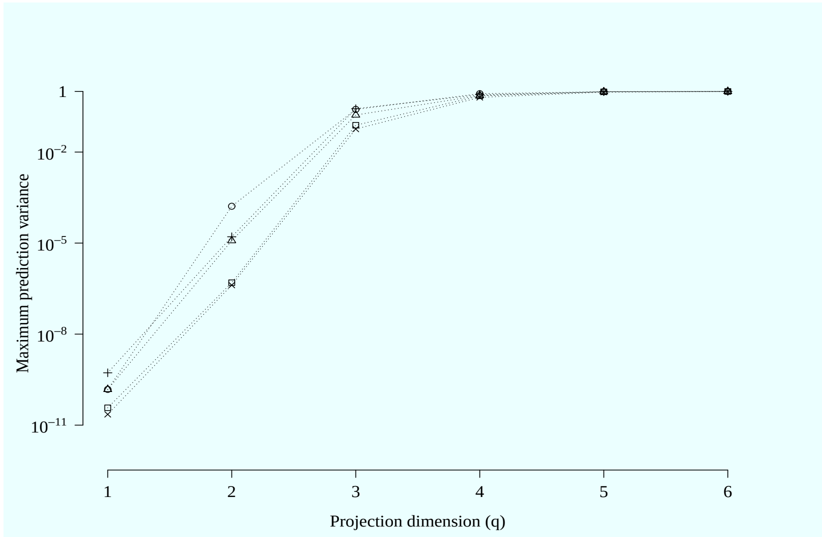

Fig. 6. Plot of maximum prediction variance of ordinary kriging (smaller-the-better) against q for maximum projection (squares), maximum projection Latin hypercube (triangles), maximin Latin hypercube (circles), uniform (pluses), and generalized maximin Latin hypercube (crosses) designs.

ACKNOWLEDGEMENT

This research is supported by U.S. Department of Energy Advanced Scientific Computing Research Award and Innovative Manufacturing Initiative Award.

APPENDIX 1

Proof of Theorem 1

Proof. For $k = 2p$, we have

$$E\{\phi_k(D;\theta)\} = \frac{1}{(p-1)!}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\int_{S_{p-1}}\left\{\sum_{l=1}^{p}\theta_l(x_{il}-x_{jl})^2\right\}^{-p} d\theta.$$

Let

$$Q_p(p,a) = \int_{S_{p-1}}\left\{\sum_{l=1}^{p-1}\theta_l d_l + (1-\sum_{l=1}^{p-1}\theta_l)a\right\}^{-p} d\theta.$$

For $a \neq d_{p-1}$, integrating with respect to θ_{p-1},

$$Q_p(p,a) = \frac{1}{(p-1)(a-d_{p-1})}\left\{Q_{p-1}(p-1,d_{p-1}) - Q_{p-1}(p-1,a)\right\}. \tag{A1}$$

Assume

$$Q_{p-1}(p-1,a) = \frac{1}{(p-2)!d_1\dots d_{p-2}a}. \tag{A2}$$

Then, from (A1), $Q_p(p,a) = 1/\{(p-1)!d_1\dots d_{p-2}d_{p-1}a\}$. This result holds for all a including $a = d_{p-1}$. Since $Q_2(2,a) = 1/(d_1 a)$, by mathematical induction, (A2) is true for all p. Now the result follows because $Q_p(p,d_p) = 1/\{(p-1)!d_1\dots d_p\}$. □

Proof of Theorem 2

Proof. For any $\gamma > 0$, we have

$$
\begin{aligned}
E(\sum_{i=1}^{n}\sum_{j\neq i} R_{ij}^{\gamma}) &= \sum_{i=1}^{n}\sum_{j\neq i} E\left\{\prod_{l=1}^{p} e^{-\gamma\alpha_l (x_{il}-x_{jl})^2}\right\} \\
&= \sum_{i=1}^{n}\sum_{j\neq i} \left\{\prod_{l=1}^{p}\int_0^{\infty} e^{-\gamma\alpha_l (x_{il}-x_{jl})^2} d\alpha_l\right\} \\
&= \frac{1}{\gamma^p}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\prod_{l=1}^{p}(x_{il}-x_{jl})^2},
\end{aligned}
$$

which is minimized by a maximum projection design. □

REFERENCES

BA, S. (2013). *SLHD: Maximin-Distance (Sliced) Latin Hypercube Designs*. R package version 1.2-1, http://CRAN.R-project.org/package=SLHD

DETTE, H. & PEPELYSHEV, A. (2010). Generalized Latin hypercube design for computer experiments. *Technometrics* **52**, 421–429.

DRAGULJIC, D., SANTNER, T. J. & DEAN, A. M. (2012). Non-collapsing space-filling designs for bounded non-rectangular regions. *Technometrics* **54**, 169–178.

JOHNSON, M., MOORE, L. & YLVISAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26**, 131–148.

HICKERNELL, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67**, 299–322.

MCKAY, M. D., BECKMAN, R. J. & CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.

MOON, H., DEAN, A. M. & SANTNER, T. J. (2011). Algorithms for generating maximin Latin hypercube and orthogonal designs. *J. Statist. Theory & Practice* **5**, 81–98.

MORRIS, M. D. & MITCHELL, T. J. (1995). Exploratory designs for computational experiments. *J. Statist. Plann. Inference* **43**, 381–402.

SACKS, J., WELCH, W. J., MITCHELL, T. J. & WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4**, 409–423.

SHEWRY, M. C. & WYNN, H. P. (1987). Maximum entropy sampling. *J. Appl. Statist.* **14**, 165–170.

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Am. Statist. Assoc.* **88**, 1392–1397.