DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS WHEN THE OUTPUT IS HIGHLY CORRELATED OVER THE INPUT SPACE

by

Yong B. Lim
William Welch

Jerome Sacks
W.J. Studden

Technical Report # 97-08

Department of Statistics
Purdue University
West Lafayette, IN USA

August, 1997

*

DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS WHEN THE OUTPUT IS HIGHLY CORRELATED OVER THE INPUT SPACE

by

Yong B. Lim[1] Jerome Sacks[2]
W.J. Studden[3] William J. Welch[4]

Abstract. Computer models or codes are widespread in science and engineering. Often, the output y is deterministic, i.e., running the code twice with the same values for the inputs or explanatory variables, x, would give the same output. To construct a predictor, the deterministic function $y(x)$ can be treated as a realization from a stochastic process,

$$Y(x) = \beta' f(x) + Z(x),$$

where $\beta' f(x)$ is a polynomial regression function, and $Z(\cdot)$ is a random function with mean zero and correlation function $R(Z(w), Z(x)) = \exp(-\theta||w-x||^2)$ for two runs of the code at inputs w and x. Given n observations of the computer code, a best linear unbiased predictor (BLUP) follows. As $\theta \to 0$, we show that the asymptotic coefficients in the BLUP are weighted combination of Lagrange interpolation polynomials. Even if there are no explicit regression terms $f(x)$ in the model, asymptotically the estimation procedure can implicitly include a polynomial trend in the inputs. We consider integrated mean squared error (IMSE) of prediction when there are no regression terms in the model. The asymptotic IMSE integrated over the design region is expressed as a quadratic form. This leads to a criterion for numerically optimizing the design.

Key words and phrases: Best linear unbiased prediction, Computer code, Integrated mean squared error, Interpolation, Optimal design, Stochastic process.

---

[1]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea. Research supported in part by the National Science Foundation Grant DMS-9121554 and Basic Science Research Institute Program, Korea Ministry of Education, BSRI-96-1415.

[2]National Institute of Statistical Sciences, P. O. Box 14162, Research Triangle Park, NC 27709-4162. Research supported in part by the National Science Foundation Grant DMS-9121554.

[3]Department of Statistics, Purdue University, W. Lafayette, IN 47907. Research supported in part by the National Science Foundation Grant DMS-9626784.

[4]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L3G1, Canada. Research supported in part by the National Science Foundation Grant DMS-9121554.

1. Introduction

Experimentation via computer models or codes is becoming increasingly common throughout science and engineering. For example, Sacks, Schiller, and Welch (1989) presented applications in chemometrics; Currin, Mitchell, Morris, and Ylvisaker (1991) and Sacks, Welch, Mitchell, and Wynn (1989) gave examples in the engineering design of electronic circuits; and Chapman, Welch, Bowman, Sacks, and Walsh (1994) and Gough and Welch (1994) described sensitivity experiments for environmental models. Other examples abound.

In these applications the output y from the computer code is often deterministic, i.e., running the code twice with the same values for the inputs or explanatory variables, x, would give the same output. To provide a basis for constructing a predictor the deterministic function $y(x)$ is regarded as if it were a realization from a stochastic process,

$$Y(x) = \beta' f(x) + Z(x), \tag{1.1}$$

where $\beta' f(x)$ is a polynomial regression function and $Z(\cdot)$ is a random function with mean zero and variance σ^2.

The correlation properties of $Z(\cdot)$ are crucial to the construction and performance of a predictor. One choice, widely used in the above applications, is

$$\mathrm{Corr}(Z(w), Z(x)) \equiv R(Z(w), Z(x)) = \exp(-\sum \theta_j |w_j - x_j|^{p_j}),$$

where w_j and x_j are the values for the jth input variable for two runs at w and x, and $\theta_j \geq 0$ and $0 < p_j \leq 2$. For simplicity in the derivations below, we assume that the θ_j's are the same for all inputs. We also take $p_j = 2$, a value arising often in applications when the parameters are estimated by maximum likelihood. Thus, the correlation function simplifies to

$$R(Z(w), Z(x)) = \exp(-\theta \| w - x \|^2). \tag{1.2}$$

Model (1.1) leads to a best linear unbiased predictor (BLUP), based on n observations of the computer code (see Section 2). This predictor respects the deterministic nature of the computer code as it interpolates the observed output values.

Working with model (1.1) in various applications has suggested that the BLUP has some special asymptotic properties as $\theta \to 0$ in (1.2). In their second chemometrics application, Sacks, Schiller, and Welch (1989) fitted (1.1) with regression functions $\beta' f(x)$ of degrees 0, 1, and 2. They found that maximum likelihood estimation chose very different values of θ in the three cases. The regression of degree 0 (which gave the best prediction accuracy) had a very small estimated θ. Lucas (1996) gave an artificial example where the deterministic "output" from five input variables was a sum of bilinear interaction terms, i.e., a polynomial. In their rejoinder, Welch et al. (1996) showed that this polynomial could be predicted almost exactly if θ in (1.2) was small. Furthermore, given 32 runs, maximum likelihood estimation clearly chose small θ.

These examples suggest that the stochastic-process component, $Z(\cdot)$, in model (1.1) can compensate for omission of polynomial terms by making θ smaller when analysing the results of a computer experiment.

Consideration of asymptotic properties as $\theta \to 0$ may also have implications for design, i.e., choosing input vectors at which to run the computer model. Choosing a design to make the BLUP from model (1.1) have small integrated mean squared error (IMSE), say, is difficult in practice because θ is unknown at the design stage and hence the IMSE cannot be computed. Sacks, Schiller, and Welch (1989) carried out several robustness studies. They compared designs from different assumed values of θ and looked at their performances for various true values. The study showed that designs from small values of θ tended to do well. Robustness studies of this type are laborious to carry out, even more so if regression polynomial functions of various degrees are also considered.

The rest of the article and its main results are as follows. Section 2 fills in some details of notation for the BLUP and its mean squared error. Section 3 has the main results on properties of the BLUP as $\theta \to 0$. We show that the asymptotic coefficients in the BLUP are weighted combinations of Lagrange interpolation polynomials. Even if there are no explicit regression terms $f(x)$ in (1.1), asymptotically the estimation procedure can implicitly include a polynomial trend in the inputs. Thus, broadly speaking, model (1.1) can work as well as polynomials when a polynomial approximation is good (and potentially much better when a low-order polynomial is inadequate).

Section 4 is concerned with design. Taking prediction of the computer code as the primary objective, we focus on the IMSE of prediction.

In Section 5 we consider design when there are no regression terms in model (1.1). The asymptotic IMSE criterion for a given design is expressed as a quadratic form. This leads to an algorithm for numerically optimizing the design. We give some examples showing that the asymptotic design performs well even when the true model (1.1) has a moderate value of θ and polynomial regression terms are present. Thus, an asymptotic design with no regression terms may provide a robust way of designing experiments when little is known. In particular, it is a candidate for the initial design in a sequential approach.

Section 6 considers design for one-dimensional input. Asymptotically, the IMSE does not depend on the linear model $f(x)$. It is shown that the design points minimizing the asymptotic IMSE are roots of orthogonal polynomials.

The proofs of all the theorems are given in an appendix.

2. The Best Linear Unbiased Predictor

Define a fixed design of n points (sets of d-dimensional inputs) by $t_1, \ldots, t_n$, with corresponding observations $Y_1, \ldots, Y_n$. Thus, $t_i = (t_{i1}, \ldots, t_{id})'$ is a design point, whereas $x = (x_1, \ldots, x_d)'$ will represent an input at which we wish to predict the unknown output $Y(x)$. Let

$$Y = (Y_1, \ldots, Y_n)', \quad F = (f(t_1), \ldots, f(t_n))',$$

$$R_\theta = \begin{bmatrix} 1 & \exp(-\theta\|t_1 - t_2\|^2) & \ldots & \exp(-\theta\|t_1 - t_n\|^2) \\ & \ddots & & \vdots \\ \text{sym.} & & & \exp(-\theta\|t_{n-1} - t_n\|^2) \\ & & & 1 \end{bmatrix}$$

and

$$r_\theta(x) = (\exp(-\theta\|t_1 - x\|^2), \ldots, \exp(-\theta\|t_n - x\|^2))'.$$

Consider the best linear unbiased predictor (BLUP), $\hat{Y}(x) = c_\theta'(x)Y$, where $c_\theta(x) = (c_1^\theta(x), \ldots, c_n^\theta(x))'$ is the solution of

$$\begin{bmatrix} 0 & F' \\ F & R_\theta \end{bmatrix} \begin{bmatrix} -\lambda \\ c_\theta(x) \end{bmatrix} = \begin{bmatrix} f(x) \\ r_\theta(x) \end{bmatrix}. \tag{2.1}$$

(For details, see Sacks, Schiller and Welch 1989). Then, from Model (1.2), the MSE of the BLUP at x is

$$J_x = E[\hat{Y}(x) - Y(x)]^2 = 1 + c_\theta(x)'R_\theta c_\theta(x) - 2c_\theta(x)'r_\theta(x). \tag{2.2}$$

Here we set $\sigma^2 = 1$.

For given θ, apart from numerical ill-conditioning of R_θ, it is straightforward to calculate $c_\theta(x)$ from (2.1) and the corresponding minimal J_x.

3. Asymptotic Properties of the BLUP

Let

$$x^\ell = \prod_{i=1}^{d} x_i^{l_i}, \ |\ell| = \sum_{i=1}^{d} l_i, \ \ell! = \prod_{i=1}^{d} l_i! \ \text{ and } \ \begin{bmatrix} m \\ \ell \end{bmatrix} = \frac{m!}{\ell!}, \tag{3.1}$$

where $x = (x_1, \ldots, x_d)$, and $\ell = (l_1, \ldots, l_d)$ denotes the monomials of degree $|\ell|$. Further, let

$$N(d, m) = \binom{d+m}{m} \tag{3.2}$$

and

$$d_m = \binom{m+d-1}{m} = N(d,m) - N(d,m-1).$$

We order the set of monomials x^ℓ first on the degree $|\ell|$ and then any particular order within the degrees. This is the same as ordering the polynomials in d variables as

$$\underset{|\ell|=0}{1}, \quad \underset{|\ell|=1}{x_1, x_2, \ldots, x_d}, \quad \underbrace{x_1^2, \ldots, x_i x_j, \ldots,}_{|\ell|=2} \quad \text{etc.} \tag{3.3}$$

It is well known that the number of monomials of exact degree m is d_m and the number of monomials of degree at most m is $N(d, m)$.

Lagrange interpolating polynomials or functions will be used extensively in the following. For a given set of n points $t_1, \ldots, t_n \in R^d$ and n functions $u_1(x), \ldots, u_n(x)$ the corresponding Lagrange interpolating functions are given by

$$\begin{aligned} L_i(x) &= L_i(x; u_1, \ldots, u_n) \\ &= \frac{D\begin{pmatrix} u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_n \\ t_1, \ldots, t_{i-1}, x, t_{i+1}, \ldots, t_n \end{pmatrix}}{D\begin{pmatrix} u_1, \ldots, u_n \\ t_1, \ldots, t_n \end{pmatrix}} \end{aligned} \tag{3.4}$$

where

$$D\begin{pmatrix} u_1, & \ldots, & u_n \\ t_1, & \ldots, & t_n \end{pmatrix} \tag{3.5}$$

denotes the determinant of the matrix with elements $u_i(t_j)$ for $i, j = 1, \ldots, n$. If the denominator determinant is zero we formally define $L_i(x) = 0$. Note that $L_i(t_j) = 1$ if $i = j$ and 0 otherwise. Also, for any given function $h(x)$ the linear combination of $u_1, \ldots, u_n$ given by

$$h_n(x) = \sum_{i=1}^{n} h(t_i) L_i(x) \tag{3.6}$$

interpolates the function h at $x = t_i$ for $i = 1, \ldots, n$. In the following we will always be interpolating at the same set of design points $t_1, \ldots, t_n$; the functions $u_1, \ldots, u_n$ used will vary considerably.

We will assume that the polynomial part $\beta' f(x)$ in (1.1) is an arbitrary polynomial of degree s, so the number of functions is $r = N(d, s)$. We further define m and r_m such that

$$N(d, m-1) \leq n < N(d, m) \qquad \text{and} \qquad r_m = n - N(d, m-1). \tag{3.7}$$

Clearly, $s \leq m - 1$. Let

$$e^{\theta \|x\|^2} f(x), x^{\ell_1}, \ldots, x^{\ell_q} \tag{3.8}$$

denote n functions, where $q = n - r$ and $\ell_1, \ldots, \ell_q$ are arbitrary and let

$$L_i^{\theta}(x; \ell_1, \ldots, \ell_q) \tag{3.9}$$

be the corresponding Lagrange functions using $t_1, \ldots, t_n$ and the functions in (3.8).

The corresponding determinant in (3.5) appearing in the denominator of the Lagrange function in (3.9) will be denoted simply as $D_\theta(l_1, \ldots, l_q)$

Theorem 3.1 The coefficient vector $c'_\theta(x) = (c_1^\theta(x), \ldots, c_n^\theta(x))$ of the BLUP for the model (1.1) is given by

$$c_i^\theta(x) = e^{\theta(\|t_i\|^2 - \|x\|^2)} \sum_{\ell_1 < \ldots < \ell_q} w_\theta(\ell_1, \ldots, \ell_q) L_i^\theta(x; \ell_1, \ldots, \ell_q), \tag{3.10}$$

where

$$w_\theta(\ell_1, \ldots, \ell_q) \propto D_\theta^2(\ell_1, \ldots, \ell_q) \prod_{i=1}^{q} \frac{(2\theta)^{|\ell_i|}}{\ell_i!} \tag{3.11}$$

and

$$\sum_{\ell_1 < \ldots < \ell_q} w_\theta(\ell_1, \ldots, \ell_q) = 1.$$

Here the infinite summation is over all possible combinations of selecting q monomial terms of any degree. For example, when we have a first degree linear regression model with $n = 9$ observations in $d = 2$ input variables, then $r = N(2,1) = 3$, and $q = n - r = 6$. Theorem 1 says that the coefficient $c_i^\theta(x)$ is expressed as a weighted combination of the Lagrange interpolator in $e^{\theta\|x\|^2}(1, x_1, x_2)$ and any 6 monomials. Note that Theorem 3.1 says that $c_\theta'(x)Y$ is equivalent to interpolating the data with the functions in (3.8) for a given set $\ell_1, \ldots, \ell_q$ and weighting or averaging these with the weights given in (3.11).

The limiting behavior of $c_i^\theta(x)$ as $\theta \to 0$ can be obtained from Theorem 3.1. It can be seen that the limit of w_θ in (3.11) is possibly positive only when $x^{\ell_1}, \ldots, x^{\ell_q}$ are chosen in such a way that the functions in (3.8) for $\theta = 0$ consist of all monomials of degree up to $m - 1$ and choices of r_m monomials from those of degree m. Thus, in the case $r_m = 0$, i.e., $n = N(d, m-1)$, the limiting coefficients $c_i^*(x) = \lim_{\theta \to 0} c_i^\theta(x)$ are the Lagrange interpolators of degree $m - 1$ at t_i for $i = 1, \ldots, n$ regardless of the linear model $f(x)$.

Taking a limit in (3.10) is complicated by the fact that θ is involved in the determinants $D_\theta(\ell_1, \ldots, \ell_q)$ using the functions in (3.8). For the no linear model case, when $r = 0$ the limit $c_i^*(x)$ is nearly immediate. Recalling (3.7), let g_1 denote the vector of monomials of degree up to $m - 1$. Also, let $\ell_1, \ldots, \ell_{r_m}$ be such that $|\ell_i| = m$ for $i = 1, \ldots, r_m$, whereupon

$$L_i(x; g_1, \ell_1, \ldots, \ell_{r_m})$$

is the Lagrange interpolator using g_1 and $x^{\ell_1}, \ldots, x^{\ell_{r_m}}$.

Corollary 3.2 If the linear model $\beta' f(x)$ is absent, then the limiting $c_i^*(x)$ are given by

$$c_i^*(x) = \sum_{\substack{\ell_1 < \ldots < \ell_{r_m} \\ |\ell_i| = m}} w(\ell_1, \ldots, \ell_{r_m}) L_i(x; g_1, \ell_1, \ldots, \ell_{r_m}), \tag{3.12}$$

where

$$w(\ell_1, \ldots, \ell_{r_m}) \propto D^2(g_1, \ell_1, \ldots, \ell_{r_m}) \prod_{i=1}^{r_m} \begin{bmatrix} m \\ \ell_i \end{bmatrix} \tag{3.13}$$

and

$$\sum_{\substack{\ell_1 < \ldots < \ell_{r_m} \\ |\ell_i| = m}} w(\ell_1, \ldots, \ell_{r_m}) = 1.$$

The limiting $c_i^*(x)$ is a weighted combination of the Lagrange interpolators using all the monomials of degree up to $m-1$ and, if necessary, all possible choices of r_m monomials from those of degree m. The weights depend on r_m and $\ell_1, \ldots, \ell_{r_m}$. Thus, even for the no linear model case, asymptotically the optimal estimation procedure implicity includes a deterministic polynomial trend in x. We are tacitly assuming that the design points $t_1, \ldots, t_n$ are such that at least one of the determinants $D(g_1, \ell_1, \ldots, \ell_{r_m})$ is nonzero so that the weights $w(\ell_1, \ldots, \ell_{r_m})$ in (3.13) are not all zero. When the linear polynomial regression model $\beta' f(x)$ is present the expression for the limiting $c_i^*(x)$ is still of the form (3.12); however the weights are more complicated than those given in (3.13). Examples 3.1 and 3.2 below illustrate the additional complications.

Example 3.1 Let $n = 4$ and $d = 2$. Since $N(2,1) < 4 < N(2,2)$, then $m = 2$ and $r_m = 1$. Suppose we assume that no linear model is present. Then the leading term (in θ) in the weight function in (3.11) is θ^4 when

$$|\ell_1| = 0, |\ell_2| = |\ell_3| = 1, \text{ and } |\ell_4| = 2;$$

i.e., when we have the polynomials $1, x_1, x_2$ of degree at most one, and one of the quadratic terms x_1^2, x_1x_2, x_2^2. Let

$$D_1 = D(1, x_1, x_2, x_1^2), D_2 = D(1, x_1, x_2, x_2^2), \text{ and } D_3 = D(1, x_1, x_2, x_1x_2)$$

denote the determinants in (3.5) with the indicated functions evaluated at the four two-dimensional design points t_1, t_2, t_3, t_4. Recalling (3.1) we have

$$\begin{bmatrix} 2 \\ (2,0) \end{bmatrix} = 1, \begin{bmatrix} 2 \\ (0,2) \end{bmatrix} = 1 \text{ and} \begin{bmatrix} 2 \\ (1,1) \end{bmatrix} = 2.$$

Thus the limiting $c_i^*(x)$ is a weighted combination of the polynomial interpolators using three sets of functions and is given by

$$\begin{aligned} c_i^*(x) &= \frac{D_1^2}{D_1^2 + D_2^2 + 2D_3^2} \; L_i(x; 1, x_1, x_2, x_1^2) \\ &+ \frac{D_2^2}{D_1^2 + D_2^2 + 2D_3^2} \; L_i(x; 1, x_1, x_2, x_1^2) \\ &+ \frac{2D_3^2}{D_1^2 + D_2^2 + 2D_3^2} \; L_i(x; 1, x_1, x_2, x_1x_2). \end{aligned}$$

Example 3.2 Let $n = 4$ and $d = 2$ as in Example 3.1. Now, however, assume that $\beta' f(x)$ is the constant model of degree $s = 0$. In this case the number of functions in $f(x)$ is $r = 1$ and $q = n - r = 3$. Noting that

$$D(e^{\theta||x||^2}, x^{\ell_1}, x^{\ell_2}, x^{\ell_3}) = D(1, x^{\ell_1}, x^{\ell_2}, x^{\ell_3})$$
$$+\theta D(||x||^2, x^{\ell_1}, x^{\ell_2}, x^{\ell_3}) + O(\theta^2),$$

the leading term in (3.11) is now when $|l_1| = |l_2| = 1$ and $|l_3| = 2$ or when $|l_1| = |l_2| = 1$ and $|l_3| = 0$. In the latter case, the weight is proportional to

$$D^2(e^{\theta||x||^2}, x_1, x_2, 1)(2\theta)(2\theta) = 2^2\theta^4 D^2(||x||^2, x_1, x_2, 1) + O(\theta^5)$$
$$= 2^2\theta^4(D_1 + D_2)^2 + O(\theta^5).$$

Thus

$$c_i^*(x) = \quad S^{-1}\{D_1^2 L_i(x; 1, x_1, x_2, x_1^2) \; + \; D_2^2 L_i(x; 1, x_1, x_2, x_2^2)$$

$$+ \; 2D_3^2 L_i(x; 1, x_1, x_2, x_1 x_2) \; + \; \frac{(D+D_2)^2}{2} L_i(x; 1, x_1, x_2, x_1^2 + x_2^2)\} \quad (3.14)$$

where S is the sum of the four coefficients of the L_i terms.

Note that the last interpolator is

$$L_i(x; 1, x_1, x_2, x_1^2 + x_2^2)$$
$$= (D_1 + D_2)^{-1}(D_1 L_i(x; 1, x_1, x_2, x_1^2) + D_2 L_i(x; 1, x_1, x_2, x_2^2)),$$

so that (3.14) can be written as as weighted combination of the first three interpolators as in the no model case; however the weights will change. The weight for $L_i(x; 1, x_1, x_2, x_1^2)$ is

$$\frac{D_1^2 + \frac{D_1}{D_1+D_2}\frac{1}{2}(D_1 + D_2)^2}{D_1^2 + D_2^2 + 2D_3^2 + \frac{1}{2}(D_1 + D_2)^2}.$$

4. Asymptotic Design for Minimizing Integrated Mean Squared Error of Prediction

The MSE of prediction, J_x, at x can be written explicitly as a weighted sum of squares of residuals of weighted interpolators to each monomial of any degree.

Let $W_\theta(x, \ell; \ell_1, \ldots, \ell_q)$ denote the interpolator (see (3.6)) of x^ℓ using the functions in (3.8) and the design points $t_1, \ldots, t_n$.

Theorem 4.1 The MSE of the BLUP for the model (1.1) is

$$J_x = e^{-\theta\|x\|^2} \sum_{\ell} \frac{(2\theta)^{|\ell|}}{|\ell|!} \begin{bmatrix} |\ell| \\ \ell \end{bmatrix} \{\overline{W}_\theta(x;\ell) - x^\ell\}^2 \tag{4.1}$$

where

$$\overline{W}_\theta(x;\ell) = \sum_{\ell_1 < \cdots < \ell_q} w_\theta(\ell_1, \ldots, \ell_q) W_\theta(x, \ell; \ell_1, \ldots, \ell_q).$$

¿From Theorem 4.1, the MSE of prediction, for the no model case, can be described as $\theta \to 0$. We now let $W^*(x;\ell)$ denote the interpolator to the function x^ℓ, using the limiting $c_i^*(x)$.

Theorem 4.2 For the no model case,

$$J_x = \theta^m \{ \frac{2^m}{m!} \sum_{\ell : |\ell| = m} \begin{bmatrix} m \\ \ell \end{bmatrix} (W^*(x;\ell) - x^\ell)^2 \} + O(\theta^{m+1}).$$

When the polynomial model part is present the coefficient of θ^m in J_x in Theorem 4.2 is considerably more complicated and will not be given here.

We use the integrated MSE (IMSE) of prediction as a design criterion for the prediction problem. Integrating is over the region of interest and may be weighted. The IMSE criterion would then seek the design that minimizes

$$J_{IMSE} = \int J_x \phi(x) dx,$$

where $\phi(x)$ is a suitable weight function. Since θ is unknown we may write

$$J_{IMSE} = h(t_1, \ldots, t_n)\theta^m + O(\theta^{m+1})$$

by integrating the asymptotic expression for J_x in Theorem 4.2. The resulting problem is then to minimize $h(t_1, \ldots, t_n)$ with respect to the design points $t_1, \ldots, t_n$.

5. Asymptotic Design with No Regression Terms

In the case of no linear model, the following theorem shows that the limiting $c^*(x)$ is the same as the solution of a constrained minimization problem and provides a basis for a numerical algorithm for constructing asymptotic optimal designs. Recall from Corollary 3.2 that $g_1(x)$ is the vector of monomials of degree at most $m-1$. Let $g_2(x)$ denote the vector of d_m monomials of exact degree m. Also define

$$G_1' = (g_1(t_1), \ldots, g_1(t_n)) \text{ and } G_2' = (g_2(t_1), \ldots, g_2(t_n)).$$

Further denote by D_2 the d_m x d_m diagonal matrix with diagonal elements $\begin{bmatrix} m \\ \ell \end{bmatrix}$, as ℓ ranges over $|\ell| = m$.

Theorem 5.1 For the no model case the limiting $c^*(x)$ is the same as the solution from the following constrained minimization problem:

$$\min_{c(x)} (G_2'c(x) - g_2(x))' D_2 (G_2'c(x) - g_2(x))$$

$$\text{subject to } \; G_1'c(x) = g_1(x), \tag{5.1}$$

and the leading term $h(x; t_1, \ldots, t_n)$ of J_x except for the factor $\frac{2^m}{m!}$ is given by

$$h(x; t_1, \ldots, t_n) = (G_2'c^*(x) - g_2(x))' \; D_2(G_2'c^*(x) - g_2(x)). \tag{5.2}$$

Use of Lagrange multipliers produces the equations

$$G_2D_2G_2'c^*(x) - G_2D_2g_2(x) - G_1\lambda(x) = 0 \tag{5.3}$$

$$G_1'c^*(x) - g_1(x) = 0$$

or

$$\begin{bmatrix} 0 & G_1' \\ G_1 & G_2D_2G_2' \end{bmatrix} \begin{bmatrix} -\lambda(x) \\ c^*(x) \end{bmatrix} = \begin{bmatrix} g_1(x) \\ G_2D_2g_2(x) \end{bmatrix}. \tag{5.4}$$

Premultiplying $c^*(x)$ in the first equation of (5.3), and then noting $g_1(x) = G_1'c^*(x)$, we get

$$c^*(x)'G_2D_2G_2'c^*(x) = c^*(x)'G_2D_2g_2(x) + g_1(x)'\lambda(x). \tag{5.5}$$

Substituting (5.5) into (5.2), we get

$$h(x: t_1, \ldots, t_n) = g_2(x)'D_2g_2(x) + g_1(x)'\lambda(x) - g_2(x)'D_2G_2'c^*(x)$$

$$= g_2(x)'D_2g_2(x) - [g_1(x)' \;\; g_2(x)'D_2G_2'] \begin{bmatrix} 0 & G_1' \\ G_1 & G_2D_2G_2' \end{bmatrix}^{-1} \begin{bmatrix} g_1(x) \\ G_2D_2g_2(x) \end{bmatrix}.$$

Therefore, the leading term $\int h(x : t_1, \ldots, t_n)\phi(x)dx$ of J_{IMSE} is expressed as a quadratic form

$$\int (G_2'c^*(x) - g_2(x))'D_2(G_2'c^*(x) - g_2(x))\phi(x)dx$$

$$= TrD_2M_{22} - Tr \begin{bmatrix} 0 & G_1' \\ G_1 & G_2D_2G_2' \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{I} & 0 \\ 0 & G_2D_2 \end{bmatrix} M \begin{bmatrix} \mathrm{I} & 0 \\ 0 & D_2G_2' \end{bmatrix}, \quad (5.6)$$

where

$$\begin{aligned} M &= \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \\ &= \begin{bmatrix} \int g_1(x)g_1(x)'\phi(x)dx & \int g_1(x)g_2(x)'\phi(x)dx \\ \int g_2(x)g_1(x)'\phi(x)dx & \int g_2(x)g_2(x)'\phi(x)dx \end{bmatrix} \end{aligned}$$

is the moment matrix of the degree m polynomial model w.r.t. the weight function $\phi(x)$. Thus the design problem reduces to finding $T^* = \{t_1^*, \ldots, t_n^*\}$ which minimizes (5.6).

Noting that $\int g_i(x)g_j(x)'\phi(x)dx$ is a moment matrix of the weight function $\phi(x)$, it may be a patterned matrix. For the uniform $\phi(x)$ over $[-\frac{1}{2}, \frac{1}{2}]^d$, all the odd moments are zero.

In summary, the algorithm for an asymptotic optimal design is as follows;
Step 1. Find m such that $N(d, m-1) \leq n < N(d, m)$.

Step 2. Generate

$$\begin{bmatrix} 0 & G_1' \\ G_1 & G_2D_2G_2' \end{bmatrix}$$

and its inverse.

Step 3. Generate the moment matrix M for the degree m polynomial model w.r.t. the weight function $\phi(x)$.

Step 4. Calculate

$$h(t_1, \ldots, t_n) = \mathrm{trace}(D_2M_{22}) - \mathrm{trace}\left(\begin{bmatrix} 0 & G_1' \\ G_1 & G_2D_2G_2' \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{I} & 0 \\ 0 & G_2D_2 \end{bmatrix} M \begin{bmatrix} \mathrm{I} & 0 \\ 0 & D_2G_2' \end{bmatrix} \right).$$

	θ	
Linear model	1	10
Constant	99.6	96.1
First order	94.6	96.3

Table 1: Percent efficiences of the asymptotic design relative to the optimal design for various linear models and values of θ.

Step 5. Apply an optimization algorithm to find $T^* = \{t_1^*, \ldots, t_n^*\}$ which minimizes $h(t_1, \ldots, t_n)$.

To illustrate the qualitative features of asymptotic optimal designs, Figure 1 shows the design constructed by the above algorithm for two explanatory variables and nine design points. It can be seen that the design is space-filling. Compared with the minimax or maximin designs produced by Johnson, Moore, and Ylvisaker (1990), however, it is slightly concentrated around the edges of the design space, to reduce the MSE of prediction where it is largest.

We also investigated the efficiency of the optimal asymptotic design for a larger example with four explanatory variables and 18 runs. The optimal designs for model (1.1) were constructed for constant (β_0) or first-order linear models and $\theta = 1$ or 10, i.e., moderately large values. Under each of these four scenarios, the IMSE for the asymptotic design can be compared with the optimal IMSE. Table 1 gives the efficiencies of the asymptotic design. Efficiency remains high when θ is not small. Moreover, although the asymptotic design is independent of the linear model, efficiency is maintained when a first-order linear model should be included.

In constrast, the designs optimal for large θ have very poor efficiency when θ is small. Compared with the value of the criterion in Step 4 for the optimal asymptotic design, the designs for $\theta = 10$ have efficiencies of 0.3% for the constant model and 0.1% for the first-order model.

6. Asymptotic Design for One Dimensional Input

Let $W_n(x)$ denote the interpolator of x^n using the n functions $1, x, \ldots, x^{n-1}$

and the n design points $t_1, \ldots, t_n$. When $d = 1$ the integrated mean squared error is given asymptotically by

$$J_{IMSE} \propto \theta^n \int [W_n(x) - x^n]^2 \phi(x) dx + O(\theta^{n+1}), \tag{6.1}$$

regardless of the polynomial part $\beta' f(x)$ in (1.1). The design problem then is to choose $t_1, \ldots, t_n$ such that

$$\int [W_n(x) - x^n]^2 \phi(x) dx$$

is minimized. It is well-known that this is minimized when $P_n(x) = W_n(x) - x^n$ is proportional to the polynomial of degree n orthogonal with respect to $\phi(x)$. Since W_n interpolates x^n at t_i, we have $P_n(t_i) = 0$ implying that $t_1, \ldots, t_n$ must be the zeros of the orthogonal polynomial of degree n with respect to $\phi(x)$.

Suppose $\phi(x)$ is proportional to $(\frac{1}{2} - x)^a (\frac{1}{2} + x)^b$. It is well known that the Jacobi polynomials $P_n^{(a,b)}(x)$ are orthogonal on $[-\frac{1}{2}, \frac{1}{2}]$ with respect to $\phi(x)$ (e.g., Ghizzetti and Ossicini 1970). When $\phi(x)$ is uniform ($a = b = 0$), then $P_n(x) = P_n^{(0,0)}(x)$ is the Legendre polynomial of degree n, recursively defined by

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= 2x \\ &\vdots \\ qP_q(x) &= (2q-1)2xP_{q-1}(x) - (q-1)P_{q-2}(x) \quad (q \geq 2). \end{aligned}$$

Appendix

Proof of Theorem 3.1: Recalling (2.1), $c_\theta(x)$ is the solution of

$$\begin{bmatrix} 0 & F' \\ F & R_\theta \end{bmatrix} \begin{bmatrix} -\lambda \\ c_\theta(x) \end{bmatrix} = \begin{bmatrix} f(x) \\ r_\theta(x) \end{bmatrix}. \tag{A.1}$$

Let

$$A = \begin{bmatrix} 0 & f(t_1) & \dots & f(t_n) \\ f'(s_1) & & & \\ \vdots & & R_\theta(s_i, t_j) & \\ f'(s_n) & & & \end{bmatrix}$$

which is simply the matrix on the left-hand side of (A.1) where we have designated the design points by $s_1, \dots, s_n$ in the rows to facilitate notation below. The coefficients $c_\theta(x)$ are obtained using Cramer's Rule so that

$$c_i^\theta(x) = \frac{|B|}{|A|},$$

where B is obtained from A by replacing the appropriate column in A by the right-hand side of (A.1).

Theorem 3.1 will follow by showing that

$$|A| = e^{-\theta \sum_{i=1}^n |s_i|^2 - \theta \sum_{i=1}^n |t_i|^2} (-1)^{r^2} \times \tag{A.2}$$

$$\sum_{\ell_1 < \dots < \ell_q} \prod_{i=1}^q \frac{(2\theta)^{|\ell_i|}}{|\ell_i|!} \begin{bmatrix} |\ell_i| \\ \ell_i \end{bmatrix} D\begin{pmatrix} f_\theta(x), & x^{\ell_1}, & \dots, x^{\ell_q} \\ t_1 & \dots & t_n \end{pmatrix} D\begin{pmatrix} f_\theta(x), & x^{\ell_1}, & \dots, x^{\ell_q} \\ s_1 & \dots & s_n \end{pmatrix},$$

where $f_\theta(x) = e^{\theta||x||^2} f(x)$.

To obtain $c_i^\theta(x)$ in Theorem 3.1, we replace t_i by x in $|A|$ and then divide and multiply each term by $D\begin{pmatrix} f_\theta(x), & x^{\ell_1}, & \dots, x^{\ell_q} \\ s_1 & \dots & s_n \end{pmatrix}$.

To show (A.2) we first write

$$R_\theta(s,t) = e^{-\theta||s||^2 - \theta||t||^2} Q_\theta(s,t),$$

where

$$\begin{aligned} Q_\theta(s,t) &= e^{2\theta(s,t)} \\ &= \sum_{k=0}^\infty \frac{(2\theta)^k}{k!} (s,t)^k \end{aligned}$$

$$= \sum_{\ell} \frac{(2\theta)^{|\ell|}}{|\ell|!} \begin{bmatrix} |\ell| \\ \ell \end{bmatrix} s^{\ell} t^{\ell}, \tag{A.3}$$

and then

$$|A| = e^{-\theta \sum_1^n \|s_i\|^2 - \theta \sum_1^n \|t_i\|^2} |C|,$$

where

$$C = \begin{bmatrix} 0 & f_\theta(t_1) & \dots & f_\theta(t_n) \\ f_\theta'(s_1) & & & \\ \vdots & & Q(s_i, t_j) & \\ f_\theta'(s_n) & & & \end{bmatrix}.$$

We first expand the $(n + r) \times (n + r)$ determinant $|C|$ by a Laplace expansion using the last n columns. Since the $r \times r$ upper left block is all zeros, the first r rows must be selected and hence

$$|C| = \sum_{i_1 < \dots < i_q} \begin{vmatrix} f_\theta(t_1) & \dots & f_\theta(t_n) \\ Q_\theta(s_{i_1}, t_1) & \dots & Q_\theta(s_{i_1}, t_n) \\ \vdots & & \\ Q_\theta(s_{i_q}, t_1) & \dots & Q_\theta(s_{i_q}, t_n) \end{vmatrix} (-1)^{\sum_1^r (r+i_j^*) + \sum_1^r j} \begin{vmatrix} f_\theta'(s_{i_1}^*) \\ \vdots \\ f_\theta'(s_{i_r}^*) \end{vmatrix}, \tag{A.4}$$

where $\{s_{i_1}^*, \dots, s_{i_r}^*\}$ is the complementary set to $\{s_{i_1}, \dots, s_{i_q}\}$ and i_j^* is the row position of $s_{i_j}^*$.

Using the expansion for $Q_\theta(s,t)$ in (A.3) and a slight extension of the Basic Composition Formula (see Karlin 1968, page 17) we may write

$$\begin{vmatrix} f_\theta(t_1) & \dots f_\theta(t_n) \\ Q_\theta(s_{i_1}, t_1) & \dots Q_\theta(s_{i_1}, t_n) \\ \vdots & \vdots \\ Q_\theta(s_{i_q}, t_1) & \dots Q_\theta(s_{i_q}, t_n) \end{vmatrix}$$

$$= \sum_{\ell_1 < \dots < \ell_q} \begin{vmatrix} f_\theta(t_1) & \dots & f_\theta(t_n) \\ t_1^{\ell_1} & \dots & t_n^{\ell_1} \\ \vdots & & \vdots \\ t_1^{\ell_q} & \dots & t_n^{\ell_q} \end{vmatrix} \begin{vmatrix} s_{i_1}^{\ell_1} & \dots s_{i_q}^{\ell_1} \\ \vdots & \vdots \\ s_{i_1}^{\ell_q} & \dots s_{i_q}^{\ell_q} \end{vmatrix} \prod_{i=1}^{q} \frac{(2\theta)^{|\ell_i|}}{|\ell_i|!} \begin{bmatrix} |\ell_i| \\ \ell_i \end{bmatrix}. \tag{A.5}$$

The expression in (A.5) is substituted in (A.4) and the summations are interchanged. The inner sum, now over $i_1 < \ldots < i_q$ can be summed using a Laplace expansion to give (A.2).

Corollary 3.2 follows from Theorem 3.1 by noting that in (3.10) and (3.11) the functions $e^{\theta||x||^2} f(x)$ do not appear in $L_i^\theta(x; \ell_1, \ldots, \ell_q)$ nor $D_\theta(\ell_1, \ldots, \ell_q)$ and that $q = n$.

Proof of Theorem 4.1: Recalling (A.3), we may write

$$R_\theta = D^{-1}(\sum_\ell \mu_\theta(\ell)\pi_\ell \pi'_\ell)D^{-1}$$

and

$$r_\theta(x) = e^{-\theta||x||^2} D^{-1} \sum_\ell \mu_\theta(\ell)\pi_\ell x^\ell,$$

where

$$\mu_\theta(\ell) = \frac{(2\theta)^{|\ell|}}{|\ell|!} \begin{bmatrix} |\ell| \\ \ell \end{bmatrix}, \ \pi'_\ell = (t_1^\ell, \ldots, t_n^\ell),$$

and D is a diagonal matrix with elements $e^{\theta||t_i||^2}$ for $i = 1, \ldots, n$. ¿From (3.10)

$$c_\theta(x) = e^{-\theta||x||^2} D \sum_{\ell_1 < \ldots < \ell_q} w_\theta(\ell_1, \ldots, \ell_q) L_\theta(x; \ell_1, \ldots, \ell_q), \qquad (A.7)$$

where

$$L'_\theta(x; \ell_1, \ldots, \ell_q) = (L_1^\theta(x; \ell_1, \ldots, \ell_q), \ldots, L_n^\theta(x; \ell_1, \ldots, \ell_q)).$$

Then $c'_\theta(x) r(x)$

$$\begin{aligned}
&= e^{-2\theta||x||^2} \sum_\ell \mu_\theta(\ell) x^\ell \sum_{\ell_1 < \ldots < \ell_q} w_\theta(\ell_1, \ldots, \ell_q) \pi'_\ell L_\theta(x; \ell_1, \ldots, \ell_q) \\
&= e^{-2\theta||x||^2} \sum_\ell \mu_\theta(\ell) x^\ell \left[\overline{W}_\theta(x; \ell)\right] \\
&= e^{-2\theta||x||^2} \sum_\ell \mu_\theta(\ell) x^\ell \left[\overline{W}_\theta(x; \ell) - x^\ell + x^\ell\right] \\
&= 1 + e^{-2\theta||x||^2} \sum_\ell \mu_\theta(\ell) x^\ell \left[\overline{W}_\theta(x; \ell) - x^\ell\right], \qquad (A.8)
\end{aligned}$$

and

$$\begin{aligned} c'_\theta R_\theta c_\theta &= e^{-2\theta\|x\|^2} \sum_\ell \mu_\theta(\ell) \left[\overline{W}_\theta(x;\ell)\right]^2 \qquad (A.9) \\ &= e^{-2\theta\|x\|^2} \sum_\ell \mu_\theta(\ell) \left[\overline{W}_\theta(x;\ell) - x^\ell + x^\ell\right]^2 \\ &= 1 + 2e^{-2\theta\|x\|^2} \sum_\ell \mu_\theta(\ell) x^\ell \left[\overline{W}_\theta(x;\ell) - x^\ell\right] \\ &+ \; e^{-2\theta\|x\|^2} \sum_\ell \mu_\theta(\ell) \left[\overline{W}_\theta(x;\ell) - x^\ell\right]^2. \end{aligned}$$

Combining (A.8), (A.9) and (2.2) the result follows.

Proof of Theorem 4.2: The result in Theorem 4.2 is nearly immediate from the expression for J_x in Theorem 4.1.

Proof of Theorem 5.1: It is easily seen that the expression in (5.1) with $c(x) = c^*(x)$ is equal to the coefficient of θ^m in Theorem 4.2 except for the factor $2^m/m!$. Therefore it suffices to show that the solution to (5.1) gives the correct $c^*(x)$ which is given in (3.12). This result follows from (5.4) and arguments analogous to those used in the proof of Theorem 3.1. That is, from the proof of Theorem 3.1,

$$\begin{aligned} &\begin{bmatrix} 0 & G'_1 \\ G_1 & G_2 D_2 G'_2 \end{bmatrix} \\ &= (-1)^{n_0^2} \sum_{\ell_1 < \ldots < \ell_{r_m}} \prod_{i=1}^{r_m} \begin{bmatrix} m \\ \ell_i \end{bmatrix} D\begin{pmatrix} g_1(x), x^{\ell_1}, \ldots, x^{\ell_{r_m}} \\ t_1, \ldots, t_n \end{pmatrix} D\begin{pmatrix} g_1(x), x^{\ell_1}, \ldots, x^{\ell_{r_m}} \\ s_1, \ldots, s_n \end{pmatrix}. \end{aligned}$$

As in the proof of Theorem 3.1 we have indexed the last rows by $s_1, \ldots, s_n$ and denoted the number of columns of G_1 by $n_0 = N(d, m-1)$. The solution for $c^*(x)$ follows as in the proof of Theorem 3.1.

References

Chapman, W.L., Welch, W.J., Bowman, K.P., Sacks, J., and Walsh, J.E. (1994). Arctic Sea Ice Variability: Model Sensitivities and a Multidecadal Simulation. Journal of Geophysical Research, 99, 919–935.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association*, **86**, 953–963.

Gough, W.A. and Welch, W.J. (1994). Parameter Space Exploration of an Ocean General Circulation Model Using an Isopycnal Mixing Parameterization. *Journal of Marine Research*, **52**, 773–796.

Ghizzetti, A. and Ossicini, A. (1970). *Quadrature Formulae.* Birkhauser Verlag, Basel.

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

Karlin, S. (1968). The Total Positivity, Stanford University Press, Stanford.

Lucas, J.M. (1996). Comments on Computer Experiments. *Technometrics*, **38**, 197–199.

Sacks, J., Schiller, S.B. and Welch, W.J. (1989). Designs for Computer Experiments. *Technometrics* **31**, 41–47.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and Analysis of Computer Experiments. *Statist. Sci.* **4**, 409–423

Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Morris, M.D., and Schonlau, M. (1996), Response to James M. Lucas. *Technometrics*, **38**, 199–203.

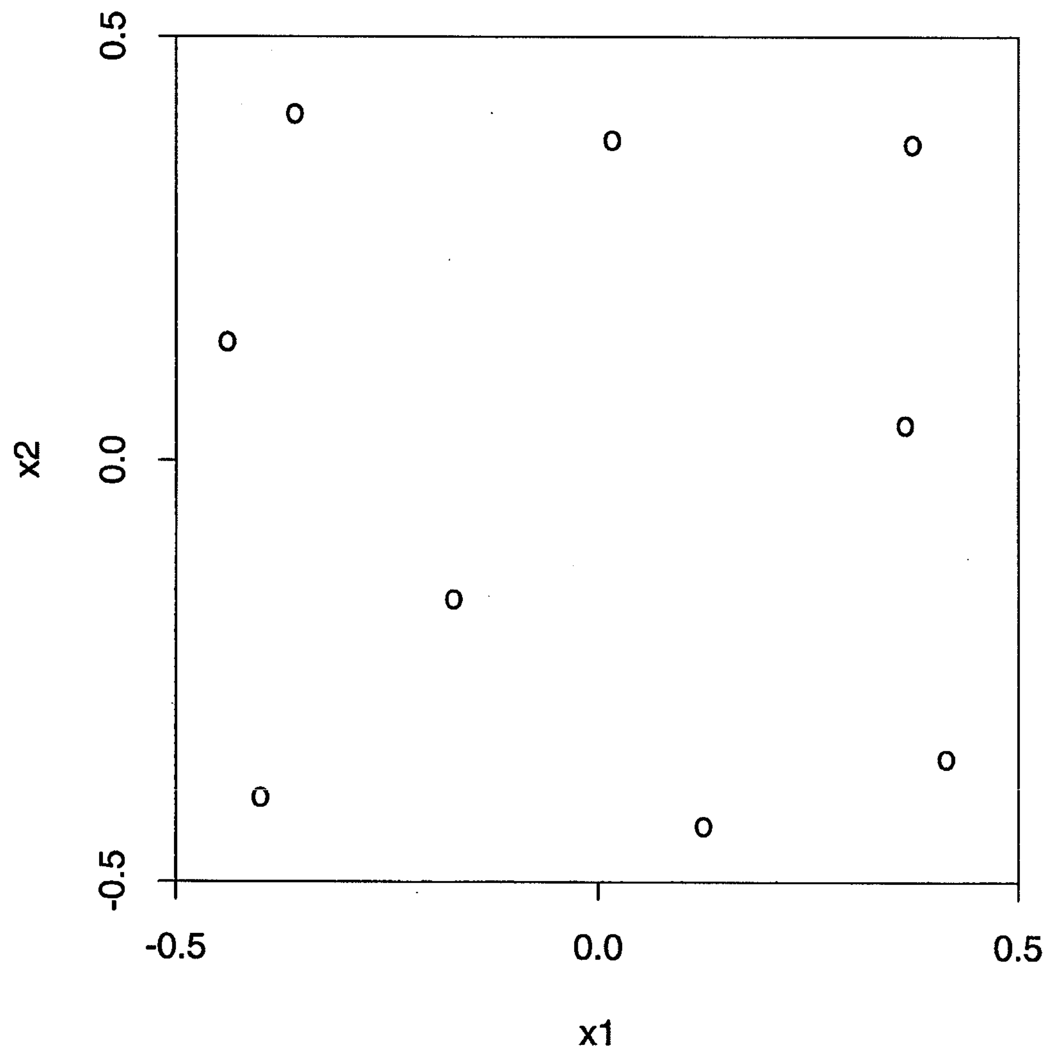

Figure 1: Asymptotic optimal design for two explanatory variables and nine design points.