The Canadian Journal of Statistics
Vol. 30, No. 1, 2002, Pages 109–126
La revue canadienne de statistique

Design and analysis of computer experiments when the output is highly correlated over the input space

Yong B. LIM, Jerome SACKS, W. J. STUDDEN & William J. WELCH

Key words and phrases: Best linear unbiased prediction; computer code; integrated mean squared error; interpolation; optimal design; stochastic process.

MSC 2000: Primary 62G08, 62K05; secondary 60G15, 60G25.

Abstract: To build a predictor, the output of a deterministic computer model or "code" is often treated as a realization of a stochastic process indexed by the code's input variables. The authors consider an asymptotic form of the Gaussian correlation function for the stochastic process where the correlation tends to unity. They show that the limiting best linear unbiased predictor involves Lagrange interpolating polynomials; linear model terms are implicitly included. The authors then develop optimal designs based on minimizing the limiting integrated mean squared error of prediction. They show through several examples that these designs lead to good prediction accuracy.

Planification et analyse d'expériences par ordinateur dont les extrants sont fortement corrélés sur l'ensemble de l'espace des intrants

Résumé : Pour bâtir un prédicteur, les extrants d'un modèle informatique déterministe appelé "code" doivent souvent être traités comme une réalisation d'un processus aléatoire indicé par les intrants de ce code. Les auteurs s'intéressent à la forme asymptotique de la fonction de corrélation gaussienne associée à un processus aléatoire dont la corrélation tend vers un. Ils montrent qu'à la limite, le meilleur prédicteur linéaire non-biaisé s'exprime en terme de polynômes d'interpolation de Lagrange; des termes de modèle linéaire sont sous-entendus. Les auteurs identifient des plans optimaux qui minimisent l'erreur quadratique moyenne intégrée de prévision. Ils montrent à travers plusieurs exemples la bonne précision des prévisions auxquelles mènent ces plans.

1. INTRODUCTION

Experimentation via computer models or codes is becoming increasingly common throughout the (pure and applied) sciences and engineering. For example, Sacks, Schiller & Welch (1989) presented applications in chemometrics; Aslett, Buck, Duvall, Sacks & Welch (1998), Currin, Mitchell, Morris & Ylvisaker (1991) and Sacks, Welch, Mitchell & Wynn (1989) gave examples in the engineering design of electronic circuits; and Chapman, Welch, Bowman, Sacks & Walsh (1994) and Gough & Welch (1994) described sensitivity experiments for environmental models.

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 Gaussian stochastic process

$$Y(x) = \beta' f(x) + Z(x), \tag{1}$$

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

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

$$\operatorname{Corr}\{Z(x), Z(x^\star)\} \equiv R\{Z(x), Z(x^\star)\} = \exp\Big(-\sum \theta_j |x_j - x_j^\star|^{p_j}\Big),$$

where x_j and $x_j^\star$ are the values for the jth input variable for two runs at x and $x^\star$, and $\theta_j \geq 0$ and $0 < p_j \leq 2$. For simplicity in the derivations below, we assume that the θ_js are the same for all inputs. For all j, we also take $p_j = 2$, a value arising often in applications when the parameters are estimated by maximum likelihood; this is known as the Gaussian correlation function. Thus, the correlation function simplifies to

$$R\{Z(x), Z(x^\star)\} = \exp\Big\{-\sum \theta(x_j - x_j^\star)^2\Big\}. \tag{2}$$

Model (1) leads to the 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) in various applications has suggested that the BLUP has some special asymptotic properties as $\theta \to 0$ in the Gaussian correlation function (2). In their second chemometrics application, Sacks, Schiller & Welch (1989) fitted model (1) with polynomial linear models $\beta' f(x)$ of degrees 0, 1, and 2. They found that maximum likelihood estimation chose very different values of θ in the three cases. The linear model of degree 0 (which gave the best prediction accuracy) had a very small estimated θ. Lucas (1996) gave an artificial example in which 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, even when $\beta' f(x)$ in (1) was of degree zero if θ was small. Furthermore, given 32 runs, maximum likelihood estimation clearly chose small values of θ.

These examples suggest that the stochastic-process component, $Z(\,\cdot\,)$, in model (1) can compensate for omission of polynomial terms by making θ smaller when analyzing 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) 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 & 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 have good relative efficiency. Robustness studies of this type are laborious to carry out, even more so when linear model polynomial functions of various degrees are also considered.

The results and organization of this article are as follows. Section 2 provides notation for the BLUP and its mean squared error. In Section 3, we develop 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 is no explicit linear model $\beta' f(x)$ in the model (1), asymptotically the estimation procedure can implicitly include a polynomial trend in the inputs. Thus, broadly speaking, model (1) can work as well as a polynomial when a low-order polynomial approximation is good (and potentially much better when a low-order polynomial is inadequate). Taking accurate prediction of the computer code as the primary objective, in Section 4 we are concerned with the mean squared error of prediction for model (1). When there are no linear model terms, it is possible to write down a fairly simple expression for the limiting MSE, and hence the limiting IMSE, as $\theta \to 0$.

We next consider experimental design. In Section 5, we express the asymptotic IMSE for the BLUP with no linear model and a given design as a quadratic form. This leads to a criterion for numerically choosing a design. We give some examples showing that the asymptotic design performs well even when the true model (1) has a moderate value of θ and polynomial linear model terms are present. Thus, an asymptotic design with no linear model 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 restricts attention to design for one-dimensional input. Asymptotically, the IMSE does not depend on the linear model $\beta' f(x)$. We show that the design points minimizing the asymptotic IMSE are roots of orthogonal polynomials.

Finally, the implications of the article are illustrated numerically in Section 7, and Section 8 contains some concluding remarks. The proofs of all the theorems are given in the Appendix.

2. THE BEST LINEAR UNBIASED PREDICTOR

Define a fixed design of n points by $t_1, \ldots, t_n$. Thus, $t_i = (t_{i1}, \ldots, t_{ik})'$ is a design point for the k input variables, whereas $x = (x_1, \ldots, x_k)'$ will represent a configuration of the input variables at which we wish to predict the unknown output $Y(x)$. For the n design points, let $F = \{f(t_1), \ldots, f(t_n)\}'$ denote the linear model function values (the "expanded design matrix"), and let $Y = (Y_1, \ldots, Y_n)'$ denote the vector of observed output values. From the Gaussian correlation function in (2), these output values have the $n \times n$ correlation matrix

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

Similarly, let

$$r_\theta(x) = \left\{\exp(-\theta\|t_1 - x\|^2), \ldots, \exp(-\theta\|t_n - x\|^2)\right\}'$$

denote the $n \times 1$ vector of correlations between $Y(x)$ and $Y_1, \ldots, Y_n$.

Based on the model in (1), we can derive a best linear unbiased predictor (BLUP), $\widehat{Y}(x) = c_\theta'(x)Y$, where $c_\theta(x) = \left\{c_1^\theta(x), \ldots, c_n^\theta(x)\right\}'$. The BLUP minimizes the mean squared error (MSE) of prediction subject to the unbiasedness constraint $F'c_\theta(x) = f(x)$. By introducing Lagrange multipliers $\lambda(x)$ for the unbiasedness constraints, we can show that the optimal coefficients $c_\theta(x)$ satisfy

$$\begin{pmatrix} 0 & F' \\ F & R_\theta \end{pmatrix} \begin{pmatrix} -\lambda(x) \\ c_\theta(x) \end{pmatrix} = \begin{pmatrix} f(x) \\ r_\theta(x) \end{pmatrix}, \tag{3}$$

and that the BLUP's MSE is

$$J_x = \mathrm{E}\left\{\widehat{Y}(x) - Y(x)\right\}^2 = 1 + c_\theta'(x)R_\theta c_\theta(x) - 2c_\theta'(x)r_\theta(x). \tag{4}$$

(See, e.g., Sacks, Schiller & Welch 1989 for details.) Without loss of generality, we have set $\sigma^2 = 1$.

For given θ, apart from possible numerical ill-conditioning of R_θ, it is straightforward to calculate $c_\theta(x)$ from (3) and the corresponding minimal J_x from (4).

3. ASYMPTOTIC PROPERTIES OF THE BLUP

Since this paper is largely concerned with polynomial approximations, we will first present the notation to be used for the monomials comprising a polynomial function. The monomial $x_1^{d_1} \times \cdots \times x_k^{d_k}$ has integer degree $d_j \geq 0$ for input variable x_j, $j = 1, \ldots, k$. We denote this monomial by x^δ, where $\delta = (d_1, \ldots, d_k)$. Then $|\delta| = \sum_{j=1}^k d_j$ is the degree of the monomial. We will also write $\delta!$ for $d_1! \times \cdots \times d_k!$.

With k input variables, it is well known that the number of monomials of degree at most d is

$$m(d,k) = \binom{d+k}{d} \tag{5}$$

and that the number of monomials of exact degree d is

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

We order monomials by the degree d, with any arbitrary order within degree. For example, the polynomial in $k = 3$ variables has 1, 3, and 6 monomials of degree 0, 1, and 2, respectively. Up to degree $d = 2$, then, there are $m(2,3) = 10$ monomials in total, which we order

$$\underset{d=0}{1}, \quad \underbrace{x_1, x_2, x_3}_{d=1}, \quad \underbrace{x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3}_{d=2}. \tag{6}$$

The BLUP interpolates the data, and we will relate it to Lagrange interpolating polynomials or functions throughout the remainder of the paper. Given a set of n points, $t_1, \ldots, t_n$, and n functions, $u_1(x), \ldots, u_n(x)$, the corresponding Lagrange interpolating functions are

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

where

$$D\begin{pmatrix} u_1 & \ldots & u_n \\ t_1 & \ldots & t_n \end{pmatrix} \tag{7}$$

denotes the determinant of the matrix with element i, j given by $u_i(t_j)$ for $i, j = 1, \ldots, n$. If the denominator determinant is zero, we formally define $L_i(x; u_1, \ldots, u_n) = 0$. Note that the dependence of $L_i(x; u_1, \ldots, u_n)$ on $t_1, \ldots, t_n$ is suppressed for now in our notation, because we will be interpolating at the same set of design points. The functions $u_1, \ldots, u_n$ will vary considerably, however, and hence are stated explicitly.

Lagrange interpolating functions have several useful properties. At the design points $t_1, \ldots, t_n$, we have $L_i(t_j; u_1, \ldots, u_n) = 1$ if $i = j$ and 0 otherwise. Thus, for any given function $y(x)$, the predictor

$$\hat{y}(x) = \sum_{i=1}^{n} y(t_i) L_i(x; u_1, \ldots, u_n) \tag{8}$$

interpolates $y(x)$ at the design points. It is also easily shown that if $y(x) = u_i(x)$ for any i, then the interpolator (8) reproduces $u_i(x)$ exactly for all x (a property used in the proof of Theorem 3).

We will first show that the BLUP for model (1) can be written as an average of Lagrange interpolators, each involving n functions. The behaviour as $\theta \to 0$ will then be examined.

Henceforth, we assume that the linear model component $\beta' f(x)$ in model (1) is a polynomial and $f(x)$ has all monomials up to degree d_f and none of higher degree. Thus, $m_f = m(d_f, k)$ from (5) will refer to the number of monomial terms in $\beta' f(x)$. These will play a role in m_f of the n Lagrange interpolating functions. The remaining $r_f = n - m_f$ Lagrange functions will also be various monomials.

Specifically, consider the n functions

$$\exp(\theta \|x\|^2) f(x), x^{\delta_1}, \ldots, x^{\delta_{r_f}}, \tag{9}$$

where $\delta_1, \ldots, \delta_{r_f}$ define r_f arbitrary monomials. Let

$$L_i^{\theta}(x; \delta_1, \ldots, \delta_{r_f}), \qquad i = 1, \ldots, n$$

be the Lagrange functions from $t_1, \ldots, t_n$ and the functions in (9). The corresponding determinant in (7) appearing in the denominator of these Lagrange functions will be denoted by $D(\exp(\theta\|x\|^2) f(x), \delta_1, \ldots, \delta_{r_f})$ or simply as $D_\theta(\delta_1, \ldots, \delta_{r_f})$. The BLUP can be written as a weighted combination of these Lagrange functions according to the following theorem.

THEOREM 1. *The coefficient vector* $c_\theta(x) = \{c_1^\theta(x), \dots, c_n^\theta(x)\}'$ *in* (3) *defining the BLUP for model* (1) *is given by*

$$c_i^\theta(x) = \exp\{\theta(\|t_i\|^2 - \|x\|^2)\} \sum_{\delta_1 < \cdots < \delta_{r_f}} w_\theta(\delta_1, \dots, \delta_{r_f}) L_i^\theta(x; \delta_1, \dots, \delta_{r_f}), \tag{10}$$

where

$$w_\theta(\delta_1, \dots, \delta_{r_f}) \propto D_\theta^2(\delta_1, \dots, \delta_{r_f}) \prod_{j=1}^{r_f} \frac{(2\theta)^{|\delta_j|}}{\delta_j!} \tag{11}$$

and

$$\sum_{\delta_1 < \cdots < \delta_{r_f}} w_\theta(\delta_1, \dots, \delta_{r_f}) = 1.$$

Here the infinite summation is over all possible ordered sets of r_f distinct monomial terms of any degree. The ordering of the monomials is arbitrary, as in (6) for instance.

To illustrate, when we have $n = 9$ observations and a first-degree linear model in $k = 2$ input variables and then $d_f = 1$, there are $m_f = m(1, 2) = 3$ monomials, namely 1, x_1, and x_2, and $r_f = n - m_f = 6$. Theorem 1 says that the BLUP's coefficient $c_i^\theta(x)$ can be expressed as a weighted combination of Lagrange interpolating functions in $\exp(\theta\|x\|^2)(1, x_1, x_2)$ and any six monomials. In general, the BLUP $c_\theta'(x)Y$ itself is equivalent to a weighted average of interpolators of the data, with weights given in (11). Each interpolator is based on the functions in (9) for a given set of monomials defined by $\delta_1, \dots, \delta_{r_f}$.

Obtaining an expression for the limiting coefficients in (10) is complicated by the fact that $L_i^\theta(x; \delta_1, \dots, \delta_{r_f})$ and $D_\theta(\delta_1, \dots, \delta_{r_f})$ depend on θ. When there is no linear model, however, neither of these quantities involve θ, because the functions $\exp(\theta\|x\|^2)f(x)$ do not appear, and the result is nearly immediate. Note also that there are now $r_f = n$ remaining functions to choose. The weight in (11) becomes

$$w_\theta(\delta_1, \dots, \delta_n) \propto D^2(\delta_1, \dots, \delta_n) \prod_{j=1}^{n} \frac{(2\theta)^{|\delta_j|}}{\delta_j!},$$

and the leading terms in θ are obtained from the the lowest-order monomials.

Specifically, in the case of no linear model terms, suppose n is large enough to fit all monomials up to degree $d_n - 1$ but not all of degree d_n, i.e.,

$$m(d_n - 1, k) \le n < m(d_n, k). \tag{12}$$

There is a remainder of $r_n = n - m(d_n - 1, k)$ monomials from those of degree d_n. Let g_1 denote the vector of monomials of degree up to $d_n - 1$. These must always be included. The remaining r_n monomials, denoted by $\delta_1, \dots, \delta_{r_n}$, must be chosen from those of degree d_n. When $r_n = 0$, i.e., n is exactly big enough to fit all monomials up to degree $d_n - 1$, the limiting coefficients $c_i^*(x) = \lim_{\theta \to 0} c_i^\theta(x)$ are the Lagrange interpolators of degree $d_n - 1$ at t_i for $i = 1, \dots, n$.

In the case of no linear model, then, the limiting coefficients of the BLUP involve only the Lagrange interpolators

$$L_i(x; g_1, \delta_1, \dots, \delta_{r_n}),$$

as summarized by the following corollary.

COROLLARY 1. *If model* (1) *has no linear model terms* $\beta' f(x)$, *then the limiting coefficients* $c_i^*(x)$ *of the BLUP are given by*

$$c_i^*(x) = \sum_{\substack{\delta_1 < \cdots < \delta_{r_n} \\ |\delta_1| = \cdots = |\delta_{r_n}| = d_n}} w(\delta_1, \dots, \delta_{r_n}) L_i(x; g_1, \delta_1, \dots, \delta_{r_n}), \tag{13}$$

where

$$w(\delta_1, \ldots, \delta_{r_n}) \propto D^2(g_1, \delta_1, \ldots, \delta_{r_n}) \prod_{j=1}^{r_n} \frac{1}{\delta_j!} \tag{14}$$

and

$$\sum_{\substack{\delta_1 < \cdots < \delta_{r_n} \\ |\delta_1| = \cdots = |\delta_{r_n}| = d_n}} w(\delta_1, \ldots, \delta_{r_n}) = 1.$$

Thus, even with no explicit linear model functions $f(x)$ in the model (1), we see that the limiting BLUP involves the functions $g_1(x)$. We are tacitly assuming that the design points $t_1, \ldots, t_n$ are such that at least one of the determinants $D(g_1, \delta_1, \ldots, \delta_{r_n})$ is nonzero so that the weights $w(\delta_1, \ldots, \delta_{r_n})$ in (14) are not all zero.

Example 1 (No linear model). Suppose there are $k = 2$ explanatory variables and $n = 4$ observations. The three monomials

$$g_1 = (\delta_1, \delta_2, \delta_3) = \{(0,0), (1,0), (0,1)\}$$

(i.e., 1, x_1, and x_2) of degree up to 1 can be fitted. In addition, we take $r_n = 1$ of the following monomials of degree $d_n = 2$:

$$\delta_4^{(20)} = (2,0), \quad \text{or} \quad \delta_4^{(02)} = (0,2), \quad \text{or} \quad \delta_4^{(11)} = (1,1)$$

(i.e., x_1^2, or x_2^2, or $x_1 x_2$). The weight function (11) in Theorem 1 has

$$D^2(g_1, \delta_4) \frac{(2\theta)^0}{1} \frac{(2\theta)^1}{1} \frac{(2\theta)^1}{1} \frac{(2\theta)^2}{\delta_4!} = D^2(g_1, \delta_4) \frac{16\theta^4}{\delta_4!}$$

as the leading term in θ for these three possibilities for (g_1, δ_4); it is $O(\theta^5)$ for any other set of monomials. Let

$$D_{20} = D(1, x_1, x_2, x_1^2), \quad D_{02} = D(1, x_1, x_2, x_2^2), \quad \text{and} \quad D_{11} = D(1, x_1, x_2, x_1 x_2)$$

denote the determinants in (7) with the indicated functions evaluated at the four two-dimensional design points t_1, t_2, t_3, and t_4. Also,

$$\delta_4^{(20)}! = \delta_4^{(02)}! = 2 \quad \text{and} \quad \delta_4^{(11)}! = 1.$$

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

$$c_i^*(x) = \frac{D_{20}^2}{T_1} L_i(x; 1, x_1, x_2, x_1^2) + \frac{D_{02}^2}{T_1} L_i(x; 1, x_1, x_2, x_2^2) + \frac{2D_{11}^2}{T_1} L_i(x; 1, x_1, x_2, x_1 x_2),$$

where $T_1 = D_{20}^2 + D_{02}^2 + 2D_{11}^2$, which could be obtained directly from Corollary 1.

When linear model polynomial terms $\beta' f(x)$ are present, the expression for the limiting $c_i^*(x)$ is still of the form (13) in Corollary 1. Recall that $f(x)$ is assumed to include all monomials up to a certain degree. Possibly positive limiting weights can result only when we choose $\delta_1, \ldots, \delta_{r_f}$ in Theorem 1 from the lowest-degree monomials not in $f(x)$. These weights are more complicated, however, than those given in (14). Example 2 illustrates the additional complications.

Example 2 (Linear model present). Let $n = 4$ and $k = 2$, as in Example 1. Now, however, assume that $\beta' f(x)$ is the constant model of degree 0, so that the number of monomials in $f(x)$ is $m_f = 1$, and $r_f = n - m_f = 3$. Noting that

$$D(\exp(\theta||x||^2), x^{\delta_1}, x^{\delta_2}, x^{\delta_3}) = D(1, x^{\delta_1}, x^{\delta_2}, x^{\delta_3}) + \theta D(||x||^2, x^{\delta_1}, x^{\delta_2}, x^{\delta_3}) + O(\theta^2), \quad (15)$$

the leading term in (11) involves θ^4 and occurs when

(i) $|\delta_1| = |\delta_2| = 1$ and $|\delta_3| = 2$ in the first term on the right of (15) or

(ii) $|\delta_1| = 0$ and $|\delta_2| = |\delta_3| = 1$ in the second term.

(The monomial with $|\delta| = 0$ would duplicate 1 if included in the first term and give $D = 0$.) Case (i), with three possibilities for δ_3, leads to D_{20}, D_{02}, and D_{11} of Example 1. In case (ii), the weight in (11) is proportional to

$$[\theta D(||x||^2, 1, x_1, x_2)]^2 \frac{(2\theta)^0}{1} \frac{(2\theta)^1}{1} \frac{(2\theta)^1}{1} = 4\theta^4 (D_{20} + D_{02})^2,$$

since $||x||^2 = x_1^2 + x_2^2$ and

$$D(||x||^2, 1, x_1, x_2) = D(x_1^2, 1, x_1, x_2) + D(x_2^2, 1, x_1, x_2) = D_{20} + D_{02}.$$

Thus

$$\begin{aligned} c_i^*(x) \quad = \quad & \frac{D_{20}^2}{T_2} L_i(x; 1, x_1, x_2, x_1^2) + \frac{D_{02}^2}{T_2} L_i(x; 1, x_1, x_2, x_2^2) + \frac{2D_{11}^2}{T_2} L_i(x; 1, x_1, x_2, x_1 x_2) \\ & + \frac{(D_{20} + D_{02})^2/2}{T_2} L_i(x; 1, x_1, x_2, x_1^2 + x_2^2), \end{aligned} \quad (16)$$

where $T_2 = T_1 + (D_{20} + D_{02})^2/2$ and $T_1 = D_{20}^2 + D_{02}^2 + 2D_{11}^2$ as in Example 1. Note that the last interpolator can be written as

$$L_i(x; 1, x_1, x_2, x_1^2 + x_2^2) = \frac{D_{20}}{D_{20} + D_{02}} L_i(x; 1, x_1, x_2, x_1^2) + \frac{D_{02}}{D_{20} + D_{02}} L_i(x; 1, x_1, x_2, x_2^2),$$

so that (16) is a weighted combination of the first three interpolators, as in Example 1. The weights differ, however. For instance, the weight for $L_i(x; 1, x_1, x_2, x_1^2)$ is now

$$\frac{D_{20}^2 + \frac{1}{2}(D_{20} + D_{02})^2 \dfrac{D_{20}}{D_{20} + D_{02}}}{T_2} = \frac{D_{20}^2 + \frac{1}{2} D_{20}(D_{20} + D_{02})}{T_1 + \frac{1}{2}(D_{20} + D_{02})^2},$$

versus D_{20}^2/T_1 in Example 1.

4. ASYMPTOTIC MEAN SQUARED ERROR OF PREDICTION

We first show that the MSE of prediction J_x in (4) can be re-expressed in terms of the accuracy of interpolation of monomials of all degrees. Suppose we interpolate the monomial x^δ using (8), the functions in (9), and the design points $t_1, \ldots, t_n$. Let $\mathcal{I}_\theta(x, \delta; \delta_1, \ldots, \delta_{r_f})$ denote this interpolator of x^δ, i.e.,

$$\mathcal{I}_\theta(x, \delta; \delta_1, \ldots, \delta_{r_f}) = \sum_{i=1}^n t_i^\delta L_i^\theta(x; \delta_1, \ldots, \delta_{r_f}). \quad (17)$$

There are an infinite number of such interpolators of x^δ, generated by the choices for $\delta_1, \ldots, \delta_{r_f}$. Theorem 2 expresses the MSE of prediction in terms of the accuracy of interpolating x^δ with a weighted average of its possible interpolators, averaged over all possible monomials x^δ.

THEOREM 2. *The MSE of the BLUP for model* (1) *is*

$$J_x = \exp(-2\theta\|x\|^2) \sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta\}^2$$

where

$$\bar{\mathcal{I}}_\theta(x,\delta) = \sum_{\delta_1 < \cdots < \delta_{r_f}} w_\theta(\delta_1, \ldots, \delta_{r_f}) \mathcal{I}_\theta(x, \delta; \delta_1, \ldots, \delta_{r_f}), \tag{18}$$

and $w_\theta(\delta_1, \ldots, \delta_{r_f})$ *is given by* (11) *in Theorem 1.*

Theorem 2 leads to a convenient expression for the limiting MSE as $\theta \to 0$ when no linear model terms are present. We now let $\mathcal{I}^*(x,\delta)$ denote the interpolator of the function x^δ using the limiting BLUP coefficients $c_i^*(x)$ in (13), i.e.,

$$\mathcal{I}^*(x,\delta) = \sum_{i=1}^n t_i^\delta c_i^*(t_i). \tag{19}$$

The next theorem shows that the leading term in θ of the limiting MSE is $O(\theta^{d_n})$, where d_n is defined in (12). Furthermore, it involves only interpolators $\mathcal{I}^*(x,\delta)$ of x^δ for monomials δ of exact degree d_n.

THEOREM 3. *As* $\theta \to 0$, *the limiting MSE of the BLUP for model* (1) *with no linear model terms is*

$$J_x = (2\theta)^{d_n} \sum_{\delta:|\delta|=d_n} \frac{\{\mathcal{I}^*(x,\delta) - x^\delta\}^2}{\delta!} + O(\theta^{d_n+1}).$$

When linear model terms are present, a similar result holds, but the coefficient of θ^{d_n} in J_x is considerably more complicated and will not be given here.

5. ASYMPTOTIC DESIGN (NO LINEAR MODEL TERMS)

We use the integrated MSE (IMSE) of prediction as a design criterion. Integration is over the region of interest and may be weighted. The IMSE criterion seeks the design that minimizes

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

where $\omega(x)$ is a given weight function.

In practice, θ is unknown, but the robustness studies of Sacks, Schiller & Welch (1989) suggested that small values of θ should be used to choose an experimental design. Thus, we will now consider the limiting J_{IMSE} as $\theta \to 0$. We will rewrite the asymptotic expression for J_x in Theorem 3 in the form

$$J_x = h(x; t_1, \ldots, t_n)\theta^{d_n} + O(\theta^{d_n+1}).$$

Let $\bar{h}(t_1, \ldots, t_n)$ denote the integral of $h(x; t_1, \ldots, t_n)$ with respect to $\omega(x)$. The design problem is then to minimize $\bar{h}(t_1, \ldots, t_n)$ in

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

over the design points $t_1, \ldots, t_n$.

When model (1) has no linear model terms, 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 optimal asymptotic designs. Recall from Corollary 1 that $g_1(x)$ is the vector of monomials of degree at most $d_n - 1$. Similarly, let $g_2(x)$ denote the vector of $m(d_n, k)$ monomials of exact degree d_n. Also, define Δ_2 to be the diagonal matrix with elements $1/\delta!$ in the same order as the $m(d_n, k)$ monomials in g_2, and let

$$G_1 = \{g_1(t_1), \ldots, g_1(t_n)\}' \quad \text{and} \quad G_2 = \{g_2(t_1), \ldots, g_2(t_n)\}'. \tag{20}$$

THEOREM 4. *When model* (1) *has no linear model terms, the BLUP's limiting coefficient vector* $c^*(x)$ *is the solution of the following constrained minimization problem:*

$$\min_{c(x)}\{G_2'c(x) - g_2(x)\}'\Delta_2\{G_2'c(x) - g_2(x)\} \quad \textit{subject to} \quad G_1'c(x) = g_1(x), \tag{21}$$

and the leading term $h(x; t_1, \ldots, t_n)$ *of* J_x, *except for the factor* 2^m, *is the quadratic form*

$$h(x; t_1, \ldots, t_n) = \{G_2'c^*(x) - g_2(x)\}'\Delta_2\{G_2'c^*(x) - g_2(x)\}. \tag{22}$$

We now use Theorem 4 to write the leading term $h(x; t_1, \ldots, t_n)$ of J_x, and hence the leading term of J_{IMSE}, in a form convenient for numerical optimization. Use of Lagrange multipliers produces the equations

$$G_2\Delta_2G_2'c^*(x) - G_2\Delta_2g_2(x) - G_1\lambda(x) = 0 \tag{23}$$

and

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

or

$$\begin{pmatrix} 0 & G_1' \\ G_1 & G_2\Delta_2G_2' \end{pmatrix}\begin{pmatrix} -\lambda(x) \\ c^*(x) \end{pmatrix} = \begin{pmatrix} g_1(x) \\ G_2\Delta_2g_2(x) \end{pmatrix}. \tag{24}$$

Premultiplying (23) by $c^{*\prime}(x)$, and then noting $g_1(x) = G_1'c^*(x)$, we obtain

$$c^{*\prime}(x)G_2\Delta_2G_2'c^*(x) = c^{*\prime}(x)G_2\Delta_2g_2(x) + g_1'(x)\lambda(x).$$

Substituting this into (22) gives

$$\begin{aligned} h(x; t_1, \ldots, t_n) &= g_2'(x)\Delta_2g_2(x) + g_1'(x)\lambda(x) - g_2'(x)\Delta_2G_2'c^*(x) \\ &= g_2'(x)\Delta_2g_2(x) \\ &\quad - \begin{pmatrix} g_1'(x) & g_2'(x)\Delta_2G_2' \end{pmatrix}\begin{pmatrix} 0 & G_1' \\ G_1 & G_2\Delta_2G_2' \end{pmatrix}^{-1}\begin{pmatrix} g_1(x) \\ G_2\Delta_2g_2(x) \end{pmatrix}. \end{aligned}$$

Therefore, the leading term of J_{IMSE}, found by integrating $h(x; t_1, \ldots, t_n)$ with respect to $\omega(x)$, is

$$\begin{aligned} \bar{h}(t_1, \ldots, t_n) &= \text{trace}\,(\Delta_2M_{22}) \\ &\quad -\text{trace}\left\{\begin{pmatrix} 0 & G_1' \\ G_1 & G_2\Delta_2G_2' \end{pmatrix}^{-1}\begin{pmatrix} M_{11} & M_{12}\Delta_2G_2' \\ G_2\Delta_2M_{21} & G_2\Delta_2M_{22}\Delta_2G_2' \end{pmatrix}\right\}, \end{aligned} \tag{25}$$

where

$$M_{ij} = \int g_i(x)g_j'(x)\omega(x)\,dx, \quad i, j = 1, 2 \tag{26}$$

is a matrix containing various moments of the degree d_n polynomial model with respect to the weight function $\omega(x)$. Often the blocks of M_{ij} are patterned; for example, when $\omega(x)$ is uniform over $[-1/2, 1/2]^k$, all the odd moments are zero.

Thus, the design problem reduces to finding $t_1^*, \ldots, t_n^*$ that minimize (25). The algorithm for implementing this optimization has the following steps.

1. Find d_n from (12) to determine the monomials that will be implicitly included.
2. Generate the quantities that do not depend on $t_1, \ldots, t_n$: the diagonal matrix Δ_2, the moment matrices M_{ij} in (26), and hence trace $(\Delta_2 M_{22})$ in (25).
3. Apply an optimization algorithm to find the optimal design points $t_1^*, \ldots, t_n^*$ minimizing $\bar{h}(t_1, \ldots, t_n)$ in (25). For each $t_1, \ldots, t_n$ considered by the algorithm:
   3.1 Generate G_1 and G_2 in (20).
   3.2 Calculate $\bar{h}(t_1, \ldots, t_n)$.

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

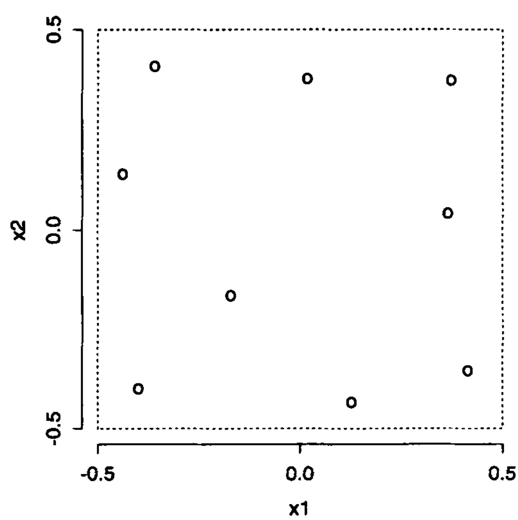

FIGURE 1: Optimal asymptotic design for two explanatory variables and nine runs.

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) were constructed for constant (β_0) or first-order ($\beta_0 + \beta_1 x_1 + \beta_2 x_2$) 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 relative efficiencies of the asymptotic design. Efficiency remains high when θ is not small. Moreover, although the asymptotic design does not explicitly include any linear model terms, efficiency is maintained when a first-order linear model should be included.

TABLE 1: Percent efficiency of the optimal asymptotic 18-run design for four explanatory variables. Efficiency is relative to the design optimal for a given linear model and value of θ.

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

In contrast, the designs optimal for large θ have very poor efficiency when θ is small. Using $\bar{h}(t_1, \ldots, t_n)$ in (25) as the criterion, relative to the optimal asymptotic design, the designs for $\theta = 10$ have efficiencies of .3% for the constant model and .1% for the first-order model.

6. ASYMPTOTIC DESIGN FOR ONE-DIMENSIONAL INPUT

In the case of only one input variable, the optimal asymptotic design can be written down almost immediately. Let $\mathcal{I}_n^*(x)$ denote the asymptotic interpolator of x^n using the n functions $1, x, \ldots, x^{n-1}$ and the n design points $t_1, \ldots, t_n$. In this one-dimensional case, Theorem 3 leads to a simple expression for the asymptotic integrated mean squared error, viz.

$$J_{\text{IMSE}} \propto \theta^n \int \{\mathcal{I}_n^*(x) - x^n\}^2 \omega(x)\, dx + O(\theta^{n+1}),$$

regardless of the explicit polynomial terms $\beta' f(x)$ in (1). The design problem then is to choose $t_1, \ldots, t_n$ such that

$$\int \{\mathcal{I}_n^*(x) - x^n\}^2 \omega(x)\, dx$$

is minimized. It is well known that the minimum of this integral occurs when $P_n(x) = \mathcal{I}_n^*(x) - x^n$ is proportional to the polynomial of degree n orthogonal with respect to $\omega(x)$. As $\mathcal{I}_n^*(x)$ 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 $\omega(x)$.

Suppose, for instance, that $\omega(x)$ is proportional to $(1/2-x)^a(1/2+x)^b$. It is well known that the Jacobi polynomials $P_n^{(a,b)}(x)$ are orthogonal on $[-1/2, 1/2]$ with respect to $\omega(x)$ (see, e.g., Ghizzetti & Ossicini 1970, p. 58). When $\omega(x)$ is uniform $(a = b = 0)$, then $P_n(x) = P_n^{(0,0)}(x)$ is the Legendre polynomial of degree n, recursively defined by

$$P_0(x) = 1, \quad P_1(x) = 2x$$

and

$$nP_n(x) = (2n-1)2xP_{n-1}(x) - (n-1)P_{n-2}(x)$$

for $n \geq 2$.

7. NUMERICAL ILLUSTRATION OF THE RESULTS

The results in Section 3 show that the BLUP interpolator with the Gaussian correlation function (2) becomes a polynomial as $\theta \to 0$. Thus, if the true function $y(x)$ is a low-order polynomial, the BLUP should be able to predict it very accurately for sufficiently small θ. To illustrate this, let $y(x)$ be a polynomial of two-dimensional input, with terms up to degree 5:

$$\begin{aligned} y(x_1, x_2) &= 9 + \frac{5}{2}x_1 - \frac{35}{2}x_2 + \frac{5}{2}x_1x_2 + 19x_2^2 - \frac{15}{2}x_1^3 - \frac{5}{2}x_1x_2^2 \\ &\quad - \frac{11}{2}x_2^4 + x_1^3x_2^2, \quad 0 \leq x_1, x_2 \leq 1. \end{aligned} \tag{27}$$

The contour and perspective plots in Figure 2 show that this function is moderately nonlinear and representative of the fairly complex but smooth functions that often arise in computer experiments.

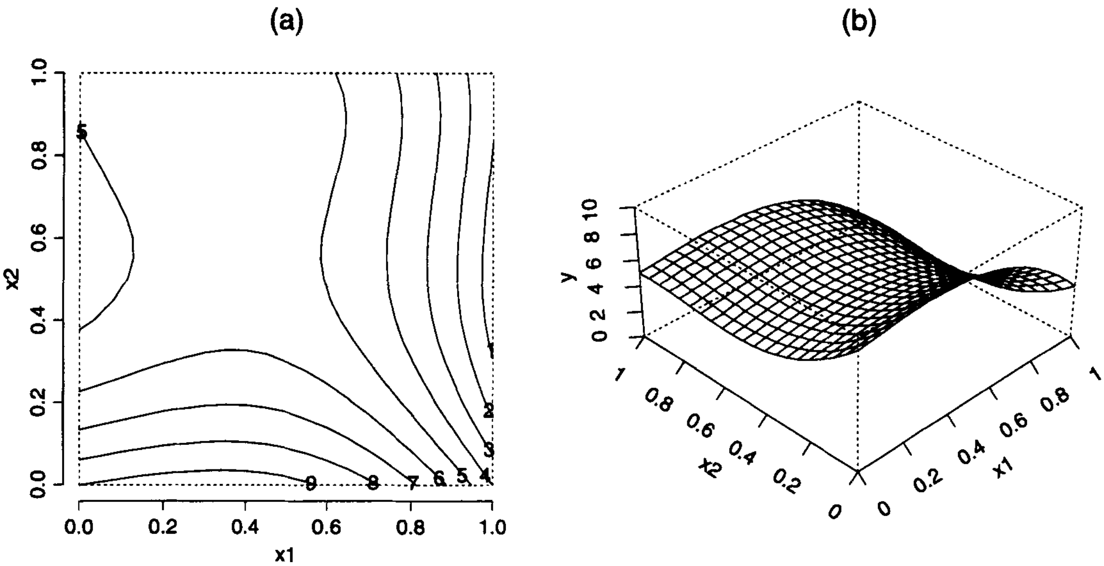

FIGURE 2: Polynomial function $y(x)$ in (27): (a) Contour plot and (b) Perspective plot.

With $n = 21$ design points, we could fit, say by least squares, a linear model with all 21 monomials up to degree 5 [put $m(5,2)$ in (5)]. In this way the function could be predicted without error. For example, the 21-run design in Figure 3(a) is a Latin hypercube sample (McKay, Conover & Beckman 1979). Latin hypercubes are a popular class of designs for computer experiments.

Without explicitly assuming a functional form, we can also achieve essentially a perfect fit here with the same design and data. Suppose we take model (1) with no trend terms, i.e.,

$$Y(x) = \beta_0 + Z(x).$$

When we predict $y(x)$ at the 21×21 grid $\{0, 1/20, \ldots, 20/20\}^2$ of x_1 and x_2 values, we obtain the first column of root mean squared error (RMSE) values in Table 2. It is seen that the RMSE decreases with smaller θ. At $\theta = .05$, the RMSE is .0026; this is .03% of the range in the true function values. Smaller values of θ should predict with even better accuracy, but $\theta = .05$ was about the smallest value that avoided ill-conditioning.

Even better results are obtained in conjunction with the optimal 21-run asymptotic design generated by the algorithm in Section 5 and shown in Figure 3(b). The second column of RMSE values in Table 2 show the asymptotic design performs uniformly better here than the Latin hypercube. At $\theta = .05$, the RMSE is only .003% of the range in the true values.

In practice, we would not know that $y(x)$ is a moderate-degree polynomial and hence that a small value of θ leads to the best prediction accuracy. With the data from the asymptotic design, maximum likelihood gives estimates $\hat{\theta}_1 = .055$ and $\hat{\theta}_2 = .079$ (x_1 and x_2 are allowed to have different θ values). These fairly small estimates lead to the RMSE value of .0003 reported in Table 2, i.e., a highly accurate predictor of the polynomial function.

TABLE 2: Root mean squared error of prediction for various values of θ in the BLUP with coefficients given by (3). The polynomial function is in (27) and the nonpolynomial function is in (28).

	RMSE of prediction			
	Polynomial function		Nonpolynomial function	
θ for predicting	Latin hypercube	Asymptotic design	Latin hypercube	Asymptotic design
.05	.0026	.0002	.151	.053
.1	.0053	.0004	.122	.052
1.0	.0513	.0070	.103	.050
10.0	.3906	.2342	.414	.251
MLE	.0064	.0003	.162	.081

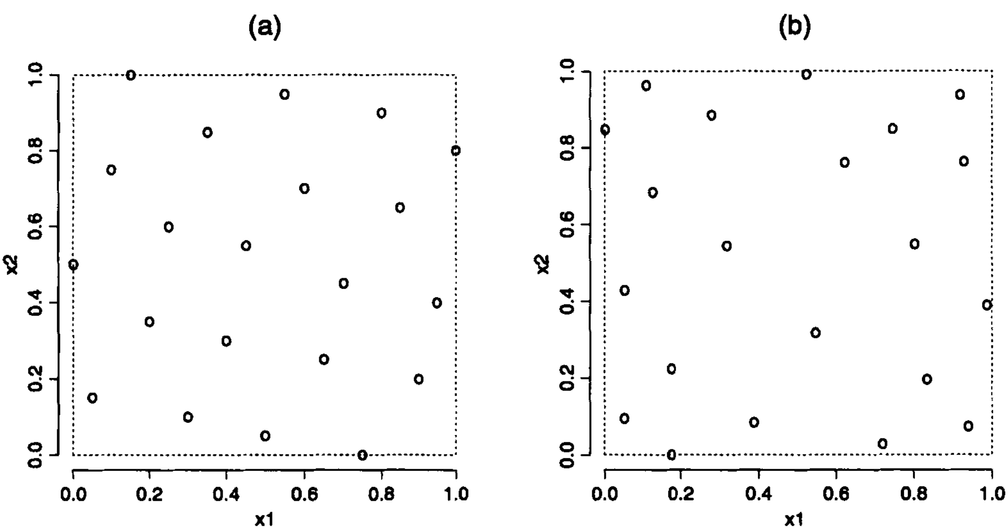

FIGURE 3: Designs for two explanatory variables and 21 runs: (a) Latin hypercube sample and (b) Optimal asymptotic design.

Also in practice, a computer model need not be well approximated by a low-degree polynomial. To look at the implications of our results in this situation, consider a second, nonpolynomial function,

$$y(x_1, x_2) = \frac{\{30 + 5x_1 \sin(5x_1)\}\{4 + \exp(-5x_2)\} - 100}{6}, \quad 0 \leq x_1, x_2 \leq 1. \tag{28}$$

Other than some rescaling, this is the same function used by Welch, Buck, Sacks, Wynn, Mitchell & Morris (1992) to illustrate the type of nonlinear behaviour that might occur in a computer model. Figure 4 gives contour and perspective plots of this function. The features are qualitatively similar to those of the polynomial (27). The two functions also have similar ranges of y values, facilitating comparison of their prediction RMSE values. [The coefficients of the polynomial (27) were chosen with these aims in mind.]

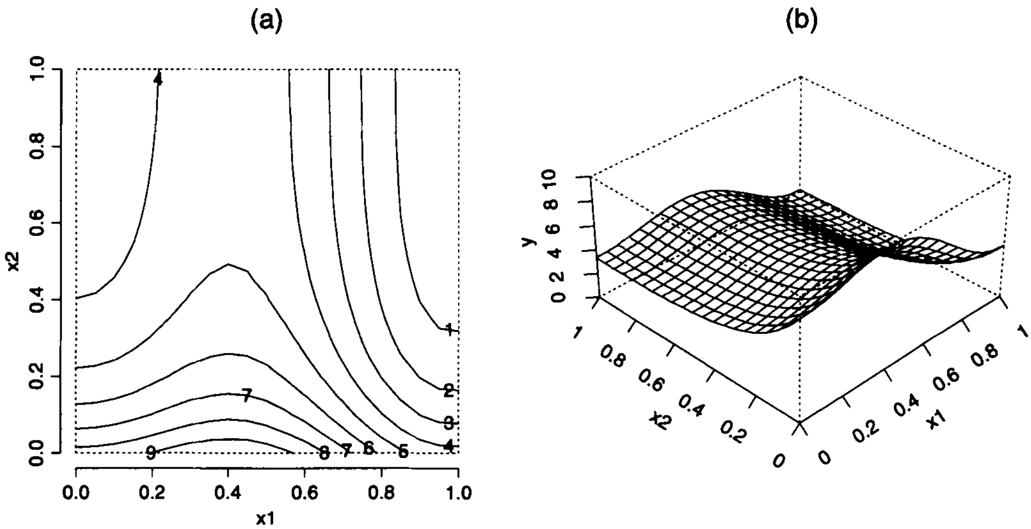

FIGURE 4: Nonpolynomial function $y(x)$ in (28): (a) Contour plot and (b) Perspective plot.

We again include no trend terms in model (1). Data are generated from (28), using the 21-run designs in Figure 3, and prediction is at the same 21×21 grid. The third and fourth columns of RMSE values in Table 2 show that prediction is less accurate than for the polynomial (27), but still fairly good: The best RMSE of prediction values are about .5% of the range in y. For both designs, as θ becomes smaller and the BLUP approaches a low-degree polynomial, there is no reason why accuracy should improve now. The best value of θ appears to be a moderate value of about 1, particularly for the Latin hypercube design. Interestingly, though, the optimal asymptotic design again substantially outperforms the Latin hypercube. Also curious is the fact that, in conjunction with the asymptotic design, the BLUP does fairly well for small values of θ. At $\theta = .05$, prediction accuracy is not quite optimal, but better than that of the BLUP with maximum likelihood estimates ($\hat{\theta}_1 = 2.54$ and $\hat{\theta}_2 = 2.24$).

8. CONCLUSIONS

The results presented here follow from some special properties of the Gaussian correlation function in (2). A power series expansion of the exponential in the Gaussian correlation function generates polynomial terms. [This is a critical step in the proof of Theorem 1; see (31).] The low-degree terms dominate as $\theta \to 0$. In the case of no linear model terms in the model (1), the BLUP is an interpolator using basis functions from the correlation function. Thus, the BLUP is also asymptotically a polynomial. Under these conditions, if $y(x)$ is a polynomial, it can be approximated essentially without error.

To realize this result for polynomial $y(x)$, a small value of θ has to be selected. In a numerical example in Section 7, maximum likelihood estimation does indeed choose a fairly small value. This empirical work is backed up by analysis of the likelihood function by Huang (2000), who gave examples for one-dimensional x and some more general results for equally spaced one-dimensional designs. These results suggest that if $y(x)$ is a polynomial, the likelihood is maximized when $\theta \to 0$, provided the design has more than twice as many data points as the degree of the polynomial. For instance, if $y(x)$ is quadratic in x, then $\theta \to 0$ is the maximum likelihood estimate if there are at least $n = 5$ points in the design.

The theoretical relative efficiencies in Table 1 show that the asymptotic optimal design performs very well, even when θ takes a moderately large value. In contrast, designs optimal for moderately large θ perform poorly in the asymptotic setting. These theoretical results are supported by the numerical root mean squared errors of prediction in Table 2. The optimal asymptotic design performs well for both the polynomial and nonpolynomial test functions. The algorithm for constructing optimal asymptotic designs is practical for problems with up to about 10 explanatory variables and 100 runs. Further research is needed to handle larger designs.

APPENDIX A: PROOFS

Proof of Theorem 1. Recalling (3), $c_\theta(x)$ is found by solving

$$\begin{pmatrix} 0 & F' \\ F & R_\theta \end{pmatrix} \begin{pmatrix} -\lambda(x) \\ c_\theta(x) \end{pmatrix} = \begin{pmatrix} f(x) \\ r_\theta(x) \end{pmatrix}. \tag{29}$$

Let

$$A = \left(\begin{array}{c|ccc} 0 & f(t_1) & \cdots & f(t_n) \\ \hline f'(s_1) & & & \\ \vdots & & R_\theta(s_i, t_j) & \\ f'(s_n) & & & \end{array}\right),$$

which is simply the first matrix in (29) with the design points denoted by $s_1, \ldots, s_n$ in the rows to distinguish rows and columns below. The coefficients $c_\theta(x)$ are obtained using Cramér'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 (29).

The key step is to show that

$$\begin{aligned} |A| \;=\; & \exp\left(-\theta \sum_{j=1}^{n} \|s_j\|^2 - \theta \sum_{j=1}^{n} \|t_j\|^2\right)(-1)^{m_f^2} \sum_{\delta_1 < \cdots < \delta_{r_f}} \prod_{j=1}^{r_f} \frac{(2\theta)^{|\delta_j|}}{\delta_j!} \\ & \times D\begin{pmatrix} f_\theta(x) & x^{\delta_1} & \ldots & x^{\delta_{r_f}} \\ t_1 & & \ldots & t_n \end{pmatrix} D\begin{pmatrix} f_\theta(x) & x^{\delta_1} & \ldots & x^{\delta_{r_f}} \\ s_1 & & \ldots & s_n \end{pmatrix}, \end{aligned} \tag{30}$$

where $f_\theta(x) = \exp(\theta\|x\|^2)f(x)$. With this result, $c_i^\theta(x)$ in Theorem 1 follows when t_i is replaced by x in $|A|$ and then each term is divided and multiplied by the determinant D.

To show (30), we first write

$$R_\theta(s,t) = \exp(-\theta\|s\|^2 - \theta\|t\|^2)Q_\theta(s,t),$$

where

$$Q_\theta(s,t) = \exp(2\theta s't) = \sum_{h=0}^{\infty} \frac{(2\theta)^h}{h!}(s't)^h = \sum_{\delta} \frac{(2\theta)^{|\delta|}}{\delta!} s^\delta t^\delta. \tag{31}$$

Note that each value of h in the first sum generates terms in the second sum with $|\delta| = h$, and $s^\delta t^\delta$ occurs $|\delta|!/\delta!$ times. Then,

$$|A| = \exp\left(-\theta \sum_{j=1}^{n} \|s_j\|^2 - \theta \sum_{j=1}^{n} \|t_j\|^2\right)|C|,$$

where

$$C = \left(\begin{array}{c|ccc} 0 & f_\theta(t_1) & \cdots & f_\theta(t_n) \\ \hline f'_\theta(s_1) & & & \\ \vdots & & Q_\theta(s_i, t_j) & \\ f'_\theta(s_n) & & & \end{array} \right).$$

We first expand the determinant of the $(m_f + n) \times (m_f + n)$ matrix C by a Laplace expansion using the last n columns. Since the $m_f \times m_f$ upper-left block is all zeros, only selections involving all of the first m_f rows plus $r_f = n - m_f$ of the remaining rows contribute to the expansion. Hence,

$$|C| = \sum_{i_1 < \cdots < i_{r_f}} \begin{vmatrix} f_\theta(t_1) & \cdots & f_\theta(t_n) \\ Q_\theta(s_{i_1}, t_1) & \cdots & Q_\theta(s_{i_1}, t_n) \\ \vdots & & \vdots \\ Q_\theta(s_{i_{r_f}}, t_1) & \cdots & Q_\theta(s_{i_{r_f}}, t_n) \end{vmatrix} (-1)^{\sum_{j=1}^{m_f}(m_f + i_j^*) + \sum_{j=1}^{m_f} j} \begin{vmatrix} f'_\theta(s^*_{i_1}) \\ \vdots \\ f'_\theta(s^*_{i_{m_f}}) \end{vmatrix}, \quad (32)$$

where $\{s^*_{i_1}, \ldots, s^*_{i_{m_f}}\}$ is the complementary set to $\{s_{i_1}, \ldots, s_{i_{r_f}}\}$ and i_j^* is the row position of $s^*_{i_j}$. For convenience below, we are summing over the selected row and column indices but using the complementary indices to determine the parity.

Using the expansion for $Q_\theta(s,t)$ in (31) and a slight extension of the Basic Composition Formula (see Karlin 1968, p. 17), we may write the first determinant in the right-hand side of (32) as

$$\sum_{\delta_1 < \cdots < \delta_{r_f}} \begin{vmatrix} f_\theta(t_1) & \cdots & f_\theta(t_n) \\ t_1^{\delta_1} & \cdots & t_n^{\delta_1} \\ \vdots & & \\ t_1^{\delta_{r_f}} & \cdots & t_n^{\delta_{r_f}} \end{vmatrix} \begin{vmatrix} s_{i_1}^{\delta_1} & \cdots & s_{i_{r_f}}^{\delta_1} \\ \vdots & & \vdots \\ s_{i_1}^{\delta_{r_f}} & \cdots & s_{i_{r_f}}^{\delta_{r_f}} \end{vmatrix} \prod_{j=1}^{r_f} \frac{(2\theta)^{|\delta_j|}}{\delta_j!}.$$

In (32) we can also interchange the summations. The inner sum, now over $i_1 < \cdots < i_{r_f}$, can be summed using a Laplace expansion to give (30).

Proof of Theorem 2. Recalling (31), we may write

$$R_\theta = E^{-1} \sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \mathcal{T}_\delta \mathcal{T}_\delta' E^{-1}$$

and

$$r_\theta(x) = \exp(-\theta \|x\|^2) E^{-1} \sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \mathcal{T}_\delta x^\delta,$$

where

$$\mathcal{T}_\delta = (t_1^\delta, \ldots, t_n^\delta)',$$

and E is a diagonal matrix with elements $\exp(\theta \|t_i\|^2)$ for $i = 1, \ldots, n$.

From (10),

$$c_\theta(x) = \exp(-\theta \|x\|^2) E \sum_{\delta_1 < \cdots < \delta_{r_f}} w_\theta(\delta_1, \ldots, \delta_{r_f}) L_\theta(x; \delta_1, \ldots, \delta_{r_f}),$$

where

$$L_\theta(x; \delta_1, \ldots, \delta_{r_f}) = \left\{ L_1^\theta(x; \delta_1, \ldots, \delta_{r_f}), \ldots, L_n^\theta(x; \delta_1, \ldots, \delta_{r_f}) \right\}',$$

and by (17) and (18), we have

$$\mathcal{T}_\delta{}' \sum_{\delta_1<\cdots<\delta_{r_f}} w_\theta(\delta_1,\ldots,\delta_{r_f}) L_\theta(x;\delta_1,\ldots,\delta_{r_f}) = \bar{\mathcal{I}}_\theta(x,\delta).$$

Therefore,

$$\begin{aligned} c_\theta'(x)r(x) &= \exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} x^\delta \bar{\mathcal{I}}_\theta(x,\delta) \\ &= \exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} x^\delta \{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta + x^\delta\} \\ &= 1+\exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} x^\delta \{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta\}, \end{aligned} \quad (33)$$

and

$$\begin{aligned} c_\theta' R_\theta c_\theta &= \exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \{\bar{\mathcal{I}}_\theta(x,\delta)\}^2 \\ &= \exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta + x^\delta\}^2 \\ &= 1+\exp(-2\theta\|x\|^2)\sum_\delta \frac{(2\theta)^{|\delta|}}{\delta!} \left[2x^\delta\{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta\} + \{\bar{\mathcal{I}}_\theta(x,\delta) - x^\delta\}^2\right]. \end{aligned} \quad (34)$$

Combining (4), (33), and (34), the result follows.

Proof of Theorem 3. The result is nearly immediate from the expression for J_x in Theorem 2. We simply note that the limiting coefficients in (13) always include the monomials g_1 of degree up to $d_n - 1$ as interpolating functions. Hence, for x^δ up to degree $d_n - 1$, the interpolator (17) and its average in (18) reproduce x^δ exactly and make no contribution to J_x.

Proof of Theorem 4. It is easily seen from the definition of G_2 in (20) that $G_2'c^*(x)$ generates the interpolators $\mathcal{I}^*(x,\delta)$ in (19) for all δ with $|\delta| = d_n$. Hence, the expression in (21) with $c(x) = c^*(x)$ is equal to the coefficient of θ^{d_n} in Theorem 3 except for the factor 2^{d_n}. Therefore, it suffices to show that the solution to (21) gives the correct $c^*(x)$, which is given in (13). This result follows from (24) and via steps analogous to the proof of Theorem 1,

$$\begin{aligned} \begin{vmatrix} 0 & G_1' \\ G_1 & G_2\Delta_2 G_2' \end{vmatrix} &= (-1)^{m_1^2} \sum_{\substack{\delta_1<\cdots<\delta_{r_n} \\ |\delta_1|=\cdots=|\delta_{r_n}|=d_n}} \prod_{j=1}^{r_n} \frac{d_n!}{\delta_j!} \\ &\quad \times D\begin{pmatrix} g_1(x) & x^{\delta_1} & \ldots & x^{\delta_{r_n}} \\ t_1 & & \ldots & t_n \end{pmatrix} D\begin{pmatrix} g_1(x) & x^{\delta_1} & \ldots & x^{\delta_{r_n}} \\ s_1 & & \ldots & s_n \end{pmatrix}. \end{aligned}$$

Again, the last rows are indexed by $s_1,\ldots,s_n$. Also, $m_1 = m(d_n - 1, k)$ denotes the number of columns of G_1.

ACKNOWLEDGEMENTS

The research of Lim, Sacks, and Welch was supported in part by National Science Foundation Grant DMS-9121554. Lim also received funding from the Basic Science Research Institute Program, BSRI-96-1415, Korea Ministry of Education. Welch also received support from the Natural Sciences and Engineering Research Council of Canada. Studden's research was funded in part by National Science Foundation Grant DMS-9626784. We thank a referee for suggestions which greatly improved the presentation.

REFERENCES

R. Aslett, R. J. Buck, S. G. Duvall, J. Sacks & W. J. Welch (1998). Circuit optimization via sequential computer experiments: design of an output buffer. *Applied Statistics*, 47, 31–48.

W. L. Chapman, W. J. Welch, K. P. Bowman, J. Sacks & J. E. Walsh (1994). Arctic sea ice variability: Model sensitivities and a multidecadal simulation. *Journal of Geophysical Research*, 99, 919–935.

C. Currin, T. Mitchell, M. Morris & D. Ylvisaker (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.

A. Ghizzetti & A. Ossicini (1970). *Quadrature Formulae.* Academic Press, New York & Birkhäuser, Basel.

W. A. Gough & W. J. Welch (1994). Parameter space exploration of an ocean general circulation model using an isopycnal mixing parameterization. *Journal of Marine Research*, 52, 773–796.

W. Huang (2000). *Properties of Parameters in a Stochastic Process Model for Computer Experiments.* Unpublished M. Math. thesis, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.

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

S. Karlin (1968). *Total Positivity*, Volume I. Stanford University Press.

J. M. Lucas (1996). Comments on "Screening, predicting, and computer experiments" [by W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell & M. D. Morris (1992)]. *Technometrics*, 38, 197–199.

M. D. McKay, W. J. Conover & R. J. Beckman (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.

J. Sacks, S. B. Schiller & W. J. Welch (1989). Designs for computer experiments. *Technometrics*, 31, 41–47.

J. Sacks, W. J. Welch, T. J. Mitchell & H. P. Wynn (1989). Design and analysis of computer experiments (with discussion). *Statistical Science*, 4, 409–435.

W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell & M. D. Morris (1992). Screening, predicting, and computer experiments. *Technometrics*, 34, 15–25.

W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, M. D. Morris & M. Schonlau (1996). Reply to "Comments on 'Screening, predicting, and computer experiments' " [by J. M. Lucas (1996)]. *Technometrics*, 38, 199–203.

Received 1 October 1997
Accepted 20 July 2001

Yong B. LIM: yblim@mm.ewha.ac.kr
Department of Statistics, Ewha Woman's University
Seoul 120-750, Korea

Jerome SACKS: sacks@niss.org
National Institute of Statistical Sciences
P. O. Box 14006, Research Triangle Park, NC 27709-4006, USA

W. J. STUDDEN: studden@stat.purdue.edu
Department of Statistics, Purdue University
West Lafayette, IN 47907, USA

William J. WELCH: wjwelch@uwaterloo.ca
Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1