

OPTIMAL DESIGNS FOR RATIONAL MODELS

by

Zhuoqiong He  
Missouri Department  
of Conservation

and

William J. Studden  
Purdue University

Dongchu Sun  
University of Missouri - Columbia

Technical Report #95-3

Department of Statistics  
Purdue University

February 1995  
(Revised October 1995)

# OPTIMAL DESIGNS FOR RATIONAL MODELS <sup>1</sup>

BY ZHUOQIONG HE, WILLIAM J. STUDDEN AND DONGCHU SUN

*Missouri Department of Conservation, Purdue University  
and University of Missouri–Columbia*

## Abstract

In this paper, experimental designs for a rational model,  $Y = P(x)/Q(x) + \varepsilon$ , are investigated, where  $P(x) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$  and  $Q(x) = 1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q$  are polynomials and  $\varepsilon$  is a random error. Two approaches, Bayesian D-optimal and Bayesian optimal design for extrapolation are examined. The first criterion maximizes the expected increase in Shannon information provided by the experiment asymptotically, and the second minimizes the asymptotic variance of the maximum likelihood estimator (MLE) of the mean response at an extrapolation point  $x_e$ . Corresponding locally optimal designs are also discussed. Conditions are derived under which a  $p + q + 1$  point design is a locally D-optimal design. The Bayesian D-optimal design is shown to be independent of the parameters in  $P(x)$  and to be equally weighted at each support point if the number of support points is the same as the number of parameters in the model. The existence and uniqueness of the locally optimal design for extrapolation are proven. The number of support points for the locally optimal design for extrapolation is exactly  $p + q + 1$ . These  $p + q + 1$  design points are proved to be independent of the extrapolation point  $x_e$  and the parameters in  $P(x)$ . The corresponding weights are also independent of the parameters in  $P(x)$ , but depend on  $x_e$  and are not equal.

---

<sup>1</sup>He's research was supported partially by a Purdue Research Foundation research grant, Studden's research was partially supported by NSF grant DMS-91-01730, and Sun's research was partially supported by a research broad grant from the University of Missouri–System.

*AMS 1991 subject classifications.* Primary 62K05, secondary 62L99, 62F15.

*Key words and Phrases.* Nonlinear regression, rational model, Bayesian D-optimal design, Bayesian optimal design.

**1. Introduction.** Experimental design has its root in the agricultural experiments of Fisher's pioneering work at Rothamsted Experimental Station in the 1920's and 1930's, and has flourished as an integral part of statistical research and practice. Most of the design work has focused on the linear regression model due to its simplicity. However, as the amount of statistical analysis using nonlinear models increased in fields such as the chemical, biological and the clinical sciences, people became aware of the need for optimal design for these nonlinear models.

The earliest relevant nonlinear example is Fisher's (1922) work on the dilution-series problem; see Fedorov (1972, p. 121). The next major impetus took the form of various nonlinear regression models involving exponential functions for reaction rates in chemical kinetics; examples are included in Box & Lucas (1959) and Katz, Azen & Schumitzky (1981). Another important class of nonlinear regression models is the rational model, which has been widely used in agricultural and biological research; see, for example, Sparrow (1979a, b).

The nonlinear problem is intrinsically much harder than the linear problem as the information matrix depends on the unknown parameters. A common approach is to design an experiment to be efficient for a best guess of the parameter values. This approach leads to what are called "locally optimal" designs, introduced by Chernoff (1953), discussed in Fedorov (1972, Sections 1.4, 2.8 and 4.4), Silvey (1980, Sections 6 and 7), and Ford, Titterton and Kitsos (1989). Although, in practice, parameters are rarely known, Ford, Torsney and Wu (1992) note various reasons why locally optimal designs are still of interest.

A natural generalization to the locally optimal design is to use a prior distribution on the parameters rather than a single guess. An optimal design is then called a Bayesian optimal design. Although numerous articles have been written on the classical aspects of nonlinear models, such as estimation and inference, the area of nonlinear experimental designs, especially Bayesian nonlinear experimental design, started receiving attention much more recently. The literature in this area includes Tsutakawa (1972), Chaloner and Larntz (1989) and Chaloner (1993). Finding the Bayesian D-optimal design for a nonlinear model is usually quite difficult. Chaloner (1993) gives a closed form of the Bayesian D-optimal design for a one-parameter logistic model with a prior having two equally weighted points. For nondegenerate prior distributions, the examples of Bayesian D-optimal design considered so far for models with more than one parameter are found numerically and in general cannot be expressed in closed form.

It is well known that various versions of the Kiefer–Wolfowitz Equivalence Theorem play an extremely importance role in almost all the nonlinear design problems, including both theoretical and numerical approaches. Such equivalence theorems for nonlinear optimal designs can be found in Whittle (1973) and Läuter (1974), and for Bayesian nonlinear designs in Chaloner and Larntz (1989) and Pilz (1991). One usually resorts to finding an optimal design within the class of  $k$ -point designs for a particular value of  $k$  and then uses the equivalence theorem to check that the resulting

design is indeed globally optimal.

The article is organized as follows. In Section 2, we review several Bayesian optimal design criteria for the nonlinear problem, and discuss some basic properties about these designs. We will illustrate that the Bayesian D-optimal design under any prior distribution should be equally weighted at each support point if the number of support points is the same as the number of parameters in the model.

In Section 3, we consider a general rational model

$$Y = \frac{\theta_0 + \theta_1 x + \cdots + \theta_p x^p}{1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q} + \varepsilon, \quad x \in [a, b].$$

We will prove that under some suitable conditions, the Bayesian D-optimal and optimal designs for extrapolation depend only on the marginal distributions of  $(\theta_{p+1}, \dots, \theta_{p+q})$ .

In Section 4, we show that locally D-optimal designs for rational models are equally weighted on exactly  $k = p + q + 1$  support points under some mild conditions about the parameters. The relationships between rational models and polynomial models are also sought.

In Section 5, we construct the locally optimal design for extrapolation, which is supported on  $p + q + 1$  points and is independent of both the parameters  $(\theta_0, \theta_1, \dots, \theta_p)$  and the extrapolation point  $x_e$ . If  $p \geq q$ , these support points are  $(a = s_0 < s_1 < \cdots < s_{p+q-1} < s_{p+q} = b)$ , where  $(s_1, \dots, s_{p+q-1})$  are determined by the zeros of a certain polynomial. This is analogous to the result of Hoel and Levine (1964) for the ordinary polynomial case.

Finally, the Michaelis-Menten model is considered in Section 6. The analytic form of the D-Bayesian optimal two-point designs are discussed. The locally D-optimal and optimal designs for extrapolation are given and compared.

## 2. Optimal Designs for a General Nonlinear Model

*2.1. A nonlinear regression model.* To introduce the notation, consider a general nonlinear regression model of the form:

$$Y = Y(x) = f(x, \boldsymbol{\theta}) + \epsilon, \tag{1}$$

where  $Y$  is the response variable,  $x$  is an explanatory variable,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  is a  $k$  vector of unknown parameters,  $f = f(x, \boldsymbol{\theta})$  is the mean response at  $x$ , is a nonlinear function of  $\boldsymbol{\theta}$ , and  $\epsilon$  is a random error satisfying  $\mathbb{E}(\epsilon) = 0$  and  $Var(\epsilon) = \sigma^2$ . Let  $\mathcal{X}$  denote the *design space* of possible  $x$ 's, which is usually a compact subset of  $\mathbb{R}^p$ . Assume that  $\epsilon$  has a normal distribution with mean 0 and variance  $\sigma^2$ . In general,  $\sigma^2$  may depend on both  $x$  and  $\boldsymbol{\theta}$ . However, for simplicity, we assume that  $\sigma^2$  is a known constant and independent of  $x$  and  $\boldsymbol{\theta}$ . This is usually not too restrictive since our main purpose is to design the experiment concerning the parameter  $\boldsymbol{\theta}$ . Without loss of generality, we assume that  $\sigma^2 = 1$ .

Associated with an observation,  $Y$ , there is a Fisher information matrix,  $I(\boldsymbol{\theta}, x)$ , given by

$$I(\boldsymbol{\theta}, x) = \nabla f (\nabla f)^t, \quad \nabla f \equiv \nabla f(x, \boldsymbol{\theta}) = \left( \frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_k} \right)^t. \quad (2)$$

Any design can be represented by a probability measure  $\xi$  on  $\mathcal{X}$ . If  $\xi$  is supported at  $m$  points  $x_1, \dots, x_m$  with mass  $\xi(x_i) = n_i/n$  ( $\sum_{i=1}^m n_i = n$ ), the experiment takes  $n_i$  observations at  $x_i$ . The average (per observation) information matrix of a design  $\xi$  is then defined by

$$M(\xi) \equiv M(\boldsymbol{\theta}; \xi) = \int_{\mathcal{X}} I(\boldsymbol{\theta}, x) \xi(dx). \quad (3)$$

Let  $\xi_n$  denote the  $n$ -point design taking observations at the points  $\mathbf{x}_{(n)} = (x_1, \dots, x_n)$ . If  $\epsilon$  is not normally distributed but all the observations are mutually independent, then  $M^{-1}(\boldsymbol{\theta}; \xi_n)$  is proportional to the asymptotic covariance matrix for the maximum likelihood estimate for  $\boldsymbol{\theta}$ , provided that  $M^{-1}(\boldsymbol{\theta}; \xi_n)$  exists.

*2.2. The design criteria.* Bayesian optimal designs minimize the expectation of some function of  $M(\boldsymbol{\theta}; \xi)$  with respect to some prior distribution. The most commonly used criterion is

$$\phi^D(\xi) = \int_{\Theta} \log\{\det[M(\boldsymbol{\theta}; \xi)]\} d\pi(\boldsymbol{\theta}), \quad (4)$$

where  $\pi(\cdot)$  is a prior distribution of  $\boldsymbol{\theta}$ . This criterion was originally suggested by Lindley (1956) and used recently by Chaloner and Larntz (1989), Chaloner (1993), for instance. Under an asymptotic normal approximation to the posterior distribution of  $\boldsymbol{\theta}$ ,  $\phi^D(\cdot)$  approximates the expected increase in Shannon information provided by the experiment. A design  $\xi$  minimizing (4) is called a Bayesian D-optimal design. If  $\pi(\cdot)$  is degenerate, a Bayesian D-optimal design is referred to as a locally D-optimal design.

Another commonly used design is a Bayesian L-optimal design minimizing

$$\phi^L(\xi) = \int_{\Theta} \text{tr}\{L(\boldsymbol{\theta})M^{-1}(\boldsymbol{\theta}; \xi)\} d\pi(\boldsymbol{\theta}), \quad (5)$$

where  $L(\boldsymbol{\theta})$  is a symmetric, nonnegative definite,  $k$  by  $k$  matrix. If linear combinations of the  $\theta_i$ 's are of interest, then  $L(\boldsymbol{\theta})$  does not depend on  $\boldsymbol{\theta}$  and is a constant matrix. If nonlinear combinations of the  $\theta_i$ 's are of interest then  $L(\boldsymbol{\theta})$  has entries which are functions of  $\boldsymbol{\theta}$ . If  $\pi(\cdot)$  is a degenerate prior, the Bayesian L-optimal design is referred to as the locally L-optimal design.

When  $L$  is a rank one matrix of the form  $L = c c'$  for some  $k \times 1$  column vector  $c$  the corresponding designs are referred to as  $c$ -optimal.

In design problems, the experimenter often wants to know the dependence of the response on explanatory variables in those regions where observations, from the practical point of view, are impossible or extremely difficult to collect. To be specific, suppose that we need to extrapolate to a given point  $x_e \notin \mathcal{X}$ . This means estimating  $f(x_e, \boldsymbol{\theta})$ , the mean response at  $x_e$ . Let  $\hat{\boldsymbol{\theta}}$  be the MLE of  $\boldsymbol{\theta}$ . For any design  $\xi$ , the variance of the asymptotic distribution of  $f(x_e, \hat{\boldsymbol{\theta}})$  for fixed

$\theta$  is then  $\nabla f(x_e, \theta)^t M^{-1}(\theta, \xi) \nabla f(x_e, \theta)$ . This motivates a special  $c$ -optimal design criterion with  $c = \nabla f(x_e, \theta)$ . A design  $\xi$  minimizing such a criterion will be called a Bayesian  $c_e$ -optimal design. The corresponding locally optimal design will be called the  $c_e$ -locally optimal design.

The following theorem gives an upper bound for the number of support points needed for a Bayesian optimal design.

**Theorem 1** *If the prior distribution of  $\theta$  is supported on no more than  $r$  distinct points, there exists an Bayesian D-optimal design, supported at no more than  $\frac{1}{2}rk(k+1)$  points.*

**Proof.** This can be proven in a manner similar to Theorem 4 in Läuter (1974). See also Theorem 2.1 of Dette and Neugebauer (1995).  $\square$

In the case of one-parameter nonlinear models, this means that the support of the Bayesian optimal designs consists of no more points than the support of the prior distribution. In addition, there must be a locally optimal design with one support point.

**Theorem 2** *Consider a general nonlinear regression model,  $Y = f(x, \theta) + \epsilon$ , where  $x \in \mathcal{X} = [a, b]$  and  $\theta = (\theta_1, \dots, \theta_k)$ . For any fixed prior of  $\theta$ , among all the  $k$ -point designs, a Bayesian D-optimal design has equal weight on its  $k$  supporting points.*

**Proof.** See Karlin and Studden (1966) page 329.  $\square$

We should point out that for a  $k$  parameter design problem, if the Bayesian D-optimal design  $\xi$  has more than  $k$  supporting points,  $\xi$  may not be equally weighted.

### 3. Bayesian Optimal Designs.

*3.1. Preliminaries and a gradient identity.* In this section we consider a general rational regression model,

$$Y = f(x, \theta) + \epsilon, \text{ with } f(x, \theta) = P(x)/Q(x) \text{ and } x \in \mathcal{X} \equiv [a, b], \quad (6)$$

where  $P(x) = P(x, \theta_P)$  and  $Q(x) = Q(x, \theta_Q)$  are two polynomials in  $x$ , defined by

$$P(x) = \theta_0 + \theta_1 x + \dots + \theta_p x^p, \quad (7)$$

$$Q(x) = 1 + \theta_{p+1} x + \dots + \theta_{p+q} x^q. \quad (8)$$

Here  $k = p + q + 1$ ,  $\theta_P = (\theta_0, \theta_1, \dots, \theta_p)^t$ ,  $\theta_Q = (\theta_{p+1}, \dots, \theta_{p+q})^t$ , and  $\theta = (\theta_P, \theta_Q)$ . This, of course, is a special case of the general nonlinear regression model (1). We will examine the design problems for the Bayesian D-optimal design and the Bayesian  $c_e$ -optimal design. More detailed results about the corresponding locally optimal designs will be investigated in Sections 4 and 5.

In using Kiefer-Wolfowitz type equivalence theorems, a useful concept is the directional derivative of  $\xi$  in the direction  $\xi_x$ , defined by

$$d(\xi, x) = \lim_{\epsilon \downarrow 0} \frac{\phi((1-\epsilon)\xi + \epsilon\xi_x) - \phi(\xi)}{\epsilon}.$$

Here  $\xi_x$  is a point mass at  $x \in [a, b]$  and  $\phi(\xi)$  is a design criterion such that  $d(\xi, x)$  exists. From Whittle (1973) or Chaloner and Larntz (1989), a design  $\xi_0$  is a Bayesian optimal design if and only if

$$d(\xi_0, x) \leq 0, \text{ for all } x \in \mathcal{X}.$$

To obtain an expression for  $d(\xi, x)$  we first derive a simple expression for the gradient vector  $\nabla f$ . For any positive integer  $n$ , let

$$\mathbf{h}_n \equiv \mathbf{h}_n(x) = (1, x, x^2, \dots, x^n)^t. \quad (9)$$

For any integer  $i$ , ( $0 \leq i \leq p+q$ ), let  $\mathbf{b}_i$  denote the following  $p+q+1$  column vector:

$$\mathbf{b}_i = \begin{cases} \underbrace{(0, 0, \dots, 0, 1, \theta_{p+1}, \theta_{p+2}, \dots, \theta_{p+q}, 0, 0, \dots, 0)}_i^t, & \text{if } 0 \leq i \leq p, \\ -\underbrace{(0, 0, \dots, 0, \theta_0, \theta_1, \dots, \theta_p, 0, 0, \dots, 0)}_{i-p}^t, & \text{if } p+1 \leq i \leq p+q. \end{cases} \quad (10)$$

Define a  $(p+q+1)$  by  $(p+q+1)$  matrix  $\mathbf{B}$  by

$$\mathbf{B} \equiv \mathbf{B}(\boldsymbol{\theta}) = (\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{p+q})^t. \quad (11)$$

Also let

$$g_i(x) = \frac{\partial}{\partial \theta_i} f(x, \boldsymbol{\theta}) = \begin{cases} \frac{x^i}{Q(x)}, & \text{if } 0 \leq i \leq p, \\ -\frac{x^{i-p} P(x)}{Q^2(x)}, & \text{if } p+1 \leq i \leq p+q. \end{cases} \quad (12)$$

**Lemma 1** (*A Gradient Identity*) *We have the following relationship:*

$$\nabla f = \nabla f(x, \boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}) \mathbf{h}_{p+q}^t(x) / Q^2(x, \boldsymbol{\theta}_Q). \quad (13)$$

The proof of the identity is fairly straightforward and will be omitted. It is crucial in proving most of the results related to the rational model (6).

In the following it will always be assumed that on the support sets of  $\xi$  and  $\pi$  the polynomial  $Q^2(x, \boldsymbol{\theta}_Q)$  is bounded away from zero and that  $\int \log(|B(\boldsymbol{\theta})|^2) d\pi(\boldsymbol{\theta}) < \infty$ .

**Lemma 2** *For the rational model (6), the Fisher information matrix  $I(\boldsymbol{\theta}, x)$  and the average information  $M(\boldsymbol{\theta}, \xi)$  are*

$$I(\boldsymbol{\theta}, x) = \frac{1}{Q^4} \mathbf{B} \mathbf{h}_{p+q} \mathbf{h}_{p+q}^t \mathbf{B}^t, \text{ and} \quad (14)$$

$$M(\boldsymbol{\theta}; \xi) = \mathbf{B} M_0(\boldsymbol{\theta}_Q; \xi) \mathbf{B}^t, \quad (15)$$

respectively, where

$$M_0(\boldsymbol{\theta}_Q; \xi) = \int_a^b \frac{1}{Q^4(x; \boldsymbol{\theta}_Q)} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x) \xi(dx). \quad (16)$$

**Proof.** This follows immediately from the Gradient Identity in Lemma 1.  $\square$

### 3.2. Bayesian D-optimal designs.

**Theorem 3** *The design criterion  $\phi^D(\xi)$  and the corresponding directional derivative  $d(\xi, x)$  are*

$$\phi^D(\xi) = \int_{\Theta} \log(|\mathbf{B}|^2) \pi(d\theta) + \int_{\Theta} \log(|M_0(\theta_Q; \xi)|) \pi(d\theta) \quad (17)$$

and

$$d^D(\xi, x) = \int_{\Theta} \text{tr} \left\{ M_0^{-1}(\theta_Q; \xi) \left( \frac{1}{Q^4} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x) \right) \right\} \pi(d\theta) - (p+q+1), \quad (18)$$

respectively.

**Proof.** The result follows from the definition of directional derivative and formula (1.1.34) of Fedorov (1972).  $\square$

**Corollary 1** *Under the assumptions of Theorem 2, the Bayesian D-optimal design is independent of  $\theta_P = (\theta_0, \theta_1, \dots, \theta_p)$ , the parameters in the numerator polynomial  $P(x, \theta_P)$ .*

Since the mean response is linear in  $\theta_P$ , the parameters in  $P(x)$ , one expects that the optimal design should be independent of  $\theta_P$ . This type of nonlinear regression model is known as a partially nonlinear model and has been discussed by Hill (1980). It is expected that an optimal design for a partially nonlinear model is independent of the linear parameters. Unfortunately, this is not always the case and a counter example can be found in Hill (1980). When all the parameters in a rational regression model are of interest, Corollary 1 confirms the above property for the design criterion (4). Surprisingly, if we are interested in only a subset of all the parameters in the model, a Bayesian or locally D-optimal design may depend on all the parameters in the model in general. For details, see He (1993).

**Remark 1** *If the Bayesian D-optimal design  $\xi$  has support on  $p+q+1$  points, then Theorem 2 tells us that  $\xi$  must have equal weight on these points.*

*Example 1.* In describing crop yield on the fertilizer input, Sparrow (1979a) considered the model

$$Y = (\theta_0 + \theta_1 x) / (1 + \theta_2 x) + \epsilon, \quad x \in [a, b], \quad (19)$$

where  $\theta_2 \geq 0$ ,  $0 \leq a < b < \infty$ . This is a special case of (6) when  $p = q = 1$ . It is easy to see that

$$\mathbf{B} = \begin{pmatrix} 1 & \theta_2 & 0 \\ 0 & 1 & \theta_2 \\ 0 & -\theta_0 & -\theta_1 \end{pmatrix}$$

and  $|\mathbf{B}| = \theta_0 \theta_2 - \theta_1$ .



**Proposition 1** Consider the model (19). Assume that the marginal prior distribution of  $\theta_2$  has half its mass at each of the points  $c$  and  $d$ , where  $0 < c \leq d$ , and that, on the support of the prior,  $\theta_0\theta_2 - \theta_1 \neq 0$ . Then among all three-point designs, the Bayesian D-optimal design  $\xi$  has equal weight on  $\{a, x^*, b\}$ , where

$$x^* = \frac{abcd - 1 + \sqrt{(ac+1)(ad+1)(bc+1)(bd+1)}}{(a+b)cd + c + d}. \quad (20)$$

**Proof.** For any design  $\xi$  having equal weight at points  $a \leq x_0 < x_1 < x_2 \leq b$

$$\phi^D(\xi) = C + 2g(x_0, x_1, x_2),$$

where  $C$  is some constant and

$$g(x_0, x_1, x_2) = \log \frac{(x_1 - x_0)(x_2 - x_0)(x_2 - x_1)}{(cx_0 + 1)(cx_1 + 1)(cx_2 + 1)(dx_0 + 1)(dx_1 + 1)(dx_2 + 1)}.$$

Using this it can readily be verified that  $x_0$  must be  $x_0 = a$  and then  $x_2 = b$ . The derivative with respect to  $x_1$  results in a quadratic equation of which (20) is the relevant root.  $\square$

We have found the Bayesian D-optimal design for the model (19) among all three-point designs. An open question is to verify that this design is indeed the Bayesian D-optimal design.

**3.3. Bayesian L-optimal and  $c_e$ -optimal Designs.** We first derive the formula for the general Bayesian L-optimal design.

**Theorem 4** The design criterion  $\phi^L(\xi)$  and the corresponding directional derivative  $d^L(\xi, x)$  are

$$\phi^L(\xi) = \int_{\Theta} \text{tr} \left\{ \mathbf{B}^{-1} L(\boldsymbol{\theta}) (\mathbf{B}^{-1})^t M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \right\} \pi(d\boldsymbol{\theta}) \quad (21)$$

and

$$d^L(\xi, x) = \phi^L(\xi) - \int_{\Theta} \text{tr} \left\{ \mathbf{B}^{-1} L(\boldsymbol{\theta}) (\mathbf{B}^{-1})^t M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \left( \frac{1}{Q^4} \mathbf{h}_{p+q} \mathbf{h}_{p+q}^t \right) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \right\} \pi(d\boldsymbol{\theta}), \quad (22)$$

respectively. Here  $M_0(\boldsymbol{\theta}_Q; \xi)$  is given by (16) and  $\mathbf{B}$  is given by (11).

**Proof.** This follows from Lemma 1 above and formula (1.1.33) of Fedorov (1972).  $\square$

Suppose that it is necessary to estimate  $f(x_e, \boldsymbol{\theta}) = P(x_e)/Q(x_e)$ , the mean response at a given point  $x_e \notin [a, b]$ . A  $c_e$ -optimal design is an L-optimal design with  $L(\boldsymbol{\theta}) = \nabla f(x_e, \boldsymbol{\theta}) \nabla f(x_e, \boldsymbol{\theta})^t = Q^{-4}(x_e) \mathbf{B} \mathbf{h}_{p+q}(x_e) \mathbf{h}_{p+q}^t(x_e) \mathbf{B}^t$ . Then we have the following results.

**Corollary 2** The Bayesian  $c_e$ -optimal design criterion and the corresponding directional derivative are

$$\phi_e(\xi) = \phi_e(\xi, x_e) = \int_{\Theta} \frac{1}{Q^4(x_e)} \mathbf{h}_{p+q}^t(x_e) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \mathbf{h}_{p+q}(x_e) \pi(d\boldsymbol{\theta}) \quad (23)$$

and

$$d_e(\xi, x) = \phi_e(\xi) - \int_{\Theta} \frac{1}{Q^4(x_e) Q^4(x)} \left\{ \mathbf{h}_{p+q}^t(x_e) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) \mathbf{h}_{p+q}(x) \right\}^2 \pi(d\boldsymbol{\theta}), \quad (24)$$

respectively.

**Corollary 3** *If  $\mathbf{B}^{-1}L(\boldsymbol{\theta})(\mathbf{B}^{-1})^t$  does not depend on  $\boldsymbol{\theta}_P$ , then the Bayesian design criterion  $\phi^L(\xi)$  does not depend on  $\boldsymbol{\theta}_P$ . In particular, the Bayesian  $c_e$ -optimal design does not depend on the marginal distribution of  $\boldsymbol{\theta}_Q$ .*

Although the Bayesian D-optimal design itself does not depend on  $\boldsymbol{\theta}_P$ , the criterion functional  $\phi^D(\xi)$  does depend on  $\boldsymbol{\theta}_P$ . We have given a condition where  $\phi^L(\xi)$  itself is independent of  $\boldsymbol{\theta}_P$ . Here is another example where we can apply the corollary. Assume that we are interested in  $g(\boldsymbol{\theta}) = \theta_p/\theta_{p+q}$ , the ratio of the coefficients of the highest orders in the two polynomials  $P$  and  $Q$ . Consider  $g(\hat{\boldsymbol{\theta}})$ , where  $\hat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$ . When  $\boldsymbol{\theta}$  is given, the asymptotic variance of  $g(\hat{\boldsymbol{\theta}})$  is  $\nabla g^t M^{-1}(\boldsymbol{\theta}, \xi) \nabla g$ , where

$$\nabla g = \theta_{p+q}^{-2} \underbrace{(0, \dots, 0, \theta_{p+q}, 0, \dots, 0, -\theta_p)^t}_p \underbrace{\quad}_{q-1}. \quad (25)$$

Thus, for the corresponding Bayesian L-optimal design,  $L(\boldsymbol{\theta}) = \nabla g \nabla g^t$ . It turns out that

$$\nabla g = \mathbf{B} \frac{1}{\theta_{p+q}^2} (0, \dots, 0, 1)^t. \quad (26)$$

Clearly,  $\mathbf{B}^{-1}L(\boldsymbol{\theta})(\mathbf{B}^{-1})^t$  does not depend on  $\boldsymbol{\theta}_P$ . In fact, the Bayesian design criterion is

$$\int_{\boldsymbol{\Theta}} \frac{1}{\theta_{p+q}^4} (0, \dots, 0, 1) M_0^{-1}(\boldsymbol{\theta}_Q; \xi) (0, \dots, 0, 1)^t \pi(d\boldsymbol{\theta}),$$

which is indeed independent of  $\boldsymbol{\theta}_P$ . It is also of interesting to note that this L-optimal criterion is a limiting case of the  $c_e$ -optimal criterion. Details are given at the end of Section 5.

#### 4. Locally D-optimal Designs.

*4.1 Existence.* It is known that the locally D-optimal design for a simple polynomial regression of degree  $p$  has exactly  $p + 1$  supporting points. The following theorem tells us the number of supporting points for a general rational model (6).

**Theorem 5** *Consider the general rational model (6). Assume that  $|\mathbf{B}| \neq 0$ . Then the maximum of  $\log |M(\boldsymbol{\theta}, \xi)|$  with respect to the set of all probability measures is attained by a measure  $\xi^*$  which is concentrated on exactly  $p+q+1$  points with equal weights  $1/(p+q+1)$ , if any one of the following conditions holds:*

1.  $Q(x)$  does not vanish for all  $x \in [a, b]$ , and  $\frac{\partial^{2(p+q)+1} Q^4(x)}{\partial x^{2(p+q)+1}}$ , as a function of  $x$ , has no zero on the open interval  $(a, b)$  (which implies that  $p < q$ ).
2.  $Q(x)$  does not vanish for all  $x \in [a, b]$  and  $p \geq q$ .
3.  $\{1, \frac{1}{Q^4}, \frac{x}{Q^4}, \frac{x^2}{Q^4}, \dots, \frac{x^{2(p+q)}}{Q^4}\}$  is a  $T$ -system on  $[a, b]$ .

**Proof.** By Theorem 2, we have that

$$\log |M(\boldsymbol{\theta}, \xi)| = \log |\mathbf{B}|^2 + \log \det \left[ \int_a^b \frac{1}{Q^4} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}(x)^t d\xi(x) \right]. \quad (27)$$

Note that the first term in the right hand side of the above equation is independent of  $\xi$ , and hence can be ignored. Also note that  $\log(\cdot)$  is a monotonely increasing function. Therefore, the problem of maximizing  $\log |M(\xi)|$  among  $\xi \in \Xi$  is equivalent to that of

$$\max_{\xi} \det \left[ \int_a^b \frac{1}{Q^4(x)} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}(x)^t d\xi(x) \right]. \quad (28)$$

If we let  $w(x) = 1/Q^4(x)$ , we conclude that the D-optimal design  $\xi^*$  is concentrated on exactly  $p + q + 1$  points by using a theorem in Karlin and Studden (1966, Page 33). Furthermore, from Theorem 2,  $\xi$  should be equally weighted on its supporting points.  $\square$

*Example 1 (continued).* Again consider the model(19). Assume that  $|\mathbf{B}| = \theta_0\theta_2 - \theta_1 \neq 0$ . By Part 2 of Theorem 5, the locally D-optimal design is equally weighted on three points. If we let  $c = d = \theta_2$  in Proposition 1, we find the locally D-optimal design  $\xi^*$  to be the three-point design with equal weights  $1/3$  at  $\{a, x^*, b\}$ , where

$$x^* = \frac{a + b + 2ab\theta_2}{2 + (a + b)\theta_2}. \quad (29)$$

Note that  $x^*$  given by (29) is monotone decreasing in  $\theta_2 \in [0, \infty]$ , ranging from the arithmetic mean to the harmonic mean of  $a$  and  $b$ .

*4.2. Relationship with polynomial regression.* It is known that the D-optimal design for a polynomial regression model is invariant under a change of interval. Suppose we have two linear models

$$\text{Model 1 : } Y = \theta_0 + \theta_1 x + \cdots + \theta_p x^p + \epsilon, \quad a \leq x \leq b,$$

$$\text{Model 2 : } Y = \theta_0^* + \theta_1^* z + \cdots + \theta_p^* z^p + \epsilon, \quad 0 \leq z \leq 1,$$

Let  $\{x_0, x_1, \dots, x_k\}$  and  $\{z_0, z_1, \dots, z_k\}$  be the supporting points of the D-optimal designs for *Model 1* and *Model 2*, respectively. Then, we have

$$x_j = a + (b - a)z_j, \quad j = 0, \dots, p. \quad (30)$$

This result is actually a special case of Lemma 3 below. For a rational model, the locally D-optimal design depends on the parameter in  $Q(x)$ . We have the following results.

**Lemma 3** *Suppose that the following two models are used to describe the same experiment using the transformation  $x = a + (b - a)z$ ;*

$$\text{Model 1 : } Y = \frac{\theta_0 + \theta_1 x + \cdots + \theta_p x^p}{1 + \theta_{p+1} x + \cdots + \theta_{p+q} x^q} + \epsilon, \quad a \leq x \leq b,$$

$$\text{Model 2 : } Y = \frac{\theta_0^* + \theta_1^* z + \cdots + \theta_p^* z^p}{1 + \theta_{p+1}^* z + \cdots + \theta_{p+q}^* z^q} + \epsilon, \quad 0 \leq z \leq 1.$$

Then we have the following one-to-one relationship.

$$\theta_j^* = \frac{(b-a)^j}{Q(a)} \sum_{i=j}^q \theta_i \binom{i}{j} a^{i-j}, \quad j = 0, \dots, p, \quad (31)$$

$$\theta_{p+j}^* = \frac{(b-a)^j}{Q(a)} \sum_{i=j}^q \theta_{p+i} \binom{i}{j} a^{i-j}, \quad j = 1, \dots, q, \quad (32)$$

where  $Q(a) = 1 + \sum_{i=1}^q \theta_{p+i} a^i$ .

**Proof.** The result follows from some simple algebra.  $\square$

**Theorem 6** Consider the two models in Lemma 3 for the same experiment. Assume that the locally  $D$ -optimal design using Model 2 has its support on the  $k$  points  $z_1, \dots, z_k$ . Then the locally  $D$ -optimal design using Model 1 with parameters given by (32) has its support on the  $k$  points,  $x_i = a + (b-a)z_i, i = 1, \dots, k$ .

**Proof.** This result follows from simple invariance considerations in linear models. For completeness we spell out a direct proof. Let  $\mathbf{B}$  and  $\mathbf{B}^*$  be the matrices defined by (11) for Model 1 and Model 2, respectively. For any design  $\xi$  under Model 1, it follows from (17) that

$$\begin{aligned} \phi(\xi) &= \log(|\mathbf{B}|^2) + \log\left(\det\left\{\int_a^b \frac{1}{Q^4(x)} \mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x) \xi(dx)\right\}\right) \\ &= \log(|\mathbf{B}|^2) + \log\left(\det\left\{\int_0^1 \frac{\mathbf{h}_{p+q}(a+(b-a)z) \mathbf{h}_{p+q}^t(a+(b-a)z)}{Q^4(a+(b-a)z)} d\xi(a+(b-a)z)\right\}\right), \end{aligned}$$

where  $\mathbf{h}_{p+q}(x)$  is defined by (9). It is clear that  $Q(a+(b-a)z) = Q(a)Q^*(z)$ , where  $Q^*(z) = 1 + \theta_{p+1}^* z + \dots + \theta_{p+q}^* z^{p+q}$ . Since  $\mathbf{h}_{p+q}(a+(b-a)z)$  is equal to

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a & b-a & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a^{p+q} & \binom{p+q}{1}(b-a)a^{p+q-1} & \binom{p+q}{2}(b-a)^2 a^{p+q-2} & \dots & (b-a)^{p+q} \end{pmatrix} \mathbf{h}_{p+q}(z),$$

we then have

$$\begin{aligned} \phi(\xi) &= \log(|\mathbf{B}|^2) - \log(Q^4(a)) + (p+q)(p+q+1) \log(b-a) \\ &\quad + \log\left(\det\left\{\int_0^1 \frac{\mathbf{h}_{p+q}(z) \mathbf{h}_{p+q}^t(z)}{[Q^*(z)]^4} d\xi(a+(b-a)z)\right\}\right), \end{aligned}$$

Noting that the analog of the design criterion under Model 2 is

$$\phi^*(\xi^*) = \log(|\mathbf{B}^*|^2) + \log\left(\det\left\{\int_0^1 \frac{1}{[Q^*(z)]^4} \mathbf{h}_{p+q}(z) \mathbf{h}_{p+q}^t(z) d\xi^*(z)\right\}\right),$$

We obtain the result immediately.  $\square$

**4.3. Weak Convergence.** In model (6), if  $\boldsymbol{\theta}_Q \rightarrow \mathbf{0}_q$  the mean response goes to a  $p^{\text{th}}$  degree ordinary polynomial regression model, where  $\mathbf{0}_q$  is the zero vector in  $\mathbb{R}^q$ . One might think that the locally D-optimal design for the model (6) goes to the locally D-optimal design for a  $p^{\text{th}}$  degree ordinary polynomial regression model. This is not quite correct. The answer is a special case of the following general theorem about weak continuity of the locally D-optimal design.

**Theorem 7** *Consider the rational model (6). Assume that  $Q(x, \boldsymbol{\theta})$  does not vanish on  $[a, b]$  for any  $\boldsymbol{\theta}$  under consideration. In this case the locally D-optimal design exists. Let  $\xi(\boldsymbol{\theta}_Q)$  be the locally D-optimal design (which depends only on  $\boldsymbol{\theta}_Q$ ). Then  $\xi(\boldsymbol{\theta}_Q)$  is weakly continuous in  $\boldsymbol{\theta}_Q$  in the following sense. For any sequence  $\{\boldsymbol{\theta}_Q^m\}_{m \geq 1}$ , assume that  $\boldsymbol{\theta}_Q^m \rightarrow \boldsymbol{\theta}_Q^0$ , and the locally D-optimal design  $\xi(\boldsymbol{\theta}_Q^0)$  is unique, then  $\xi(\boldsymbol{\theta}_Q^m) \xrightarrow{w} \xi(\boldsymbol{\theta}_Q^0)$ .*

**Proof.** For the first result, it is enough to prove that every weakly convergent subsequence of  $\{\xi(\boldsymbol{\theta}_Q^m)\}_{m \geq 1}$  converges to the same measure  $\xi(\boldsymbol{\theta}_Q^0)$ . In fact, let  $\{\xi(\boldsymbol{\theta}_Q^{m_j})\}_{j \geq 1}$  be any subsequence of  $\{\xi(\boldsymbol{\theta}_Q^m)\}_{m \geq 1}$  so that  $\xi(\boldsymbol{\theta}_Q^{m_j}) \xrightarrow{w} \xi^*$  for some measure  $\xi^*$ . It is easy to show that

$$\begin{aligned} & \text{tr} \left[ \left( \int_a^b \frac{\mathbf{h}_{p+q}(s) \mathbf{h}_{p+q}^t(s)}{Q^4(s; \boldsymbol{\theta}_Q^{m_j})} \xi(\boldsymbol{\theta}_Q^{m_j})(ds) \right)^{-1} \left( \frac{\mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x)}{Q^4(x; \boldsymbol{\theta}_Q^{m_j})} \right) \right] - (p+q+1) \\ & \rightarrow \text{tr} \left[ \left( \int_a^b \frac{\mathbf{h}_{p+q}(s) \mathbf{h}_{p+q}^t(s)}{Q^4(s; \boldsymbol{\theta}_Q^0)} \xi^*(ds) \right)^{-1} \left( \frac{\mathbf{h}_{p+q}(x) \mathbf{h}_{p+q}^t(x)}{Q^4(x; \boldsymbol{\theta}_Q^0)} \right) \right] - (p+q+1), \end{aligned} \quad (33)$$

as  $j \rightarrow \infty$  for each  $x \in [a, b]$ . Since each term in the left-hand side is negative for each  $x \in [a, b]$ , so is the right-hand side. Note that the right-hand side is the directional derivative of  $\xi^*$  in the direction  $\xi_x$ , so we know that  $\xi^*$  is the locally D-optimal design for  $\boldsymbol{\theta}_Q^0$ , which is unique by assumption. Therefore  $\xi^* = \xi(\boldsymbol{\theta}_Q^0)$ . The desired results then follow.  $\square$

**Corollary 4** *Under the assumptions of Theorem 7, if  $\boldsymbol{\theta}_Q^m \rightarrow \mathbf{0}_q$ , then  $\xi(\boldsymbol{\theta}_Q^m) \xrightarrow{w} \xi_{p+q}$ , where  $\mathbf{0}_q$  is the zero vector in  $\mathbb{R}^q$  and  $\xi_{p+q}$  is the D-optimal design for a  $(p+q)^{\text{th}}$  ordinary polynomial regression model.*

**Proof.** If  $\boldsymbol{\theta}_Q^0 = \mathbf{0}_q$ , then  $\xi(\boldsymbol{\theta}_Q^0)$  is the D-optimal design for the  $(p+q)^{\text{th}}$  degree polynomial regression model and is indeed unique from Guest (1958).  $\square$

*Example 1 (continued).* As  $\theta_2 \rightarrow 0$ , (29) goes to the arithmetic mean of  $a$  and  $b$  so that the locally D-optimal design  $\xi^*$  tends to the design equally weighted at  $\{a, (a+b)/2, b\}$ , which is the D-optimal design for the polynomial model  $Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \epsilon$ ,  $x \in \mathcal{X} = [a, b]$ . This agrees with the conclusion in Corollary 4.

## 5. Locally $c_\epsilon$ -optimal Designs.

The problem of extrapolation with a rational response model follows the result for the ordinary polynomial to some extent. The polynomial case was first derived by Hoel and Levine (1964)

who showed that the support of the optimal extrapolation design does not depend on the point of extrapolation  $x_e$ . Again consider model (6). From (23), if we define  $\mathbf{c} = \mathbf{h}_{p+q}(x_e)/Q^2(x_e)$ , then the optimal design  $\xi$  minimizing  $\min_{\xi} \phi_e(\xi; x_e)$  is the optimal design minimizing  $\mathbf{c}^t M_0^{-1}(\boldsymbol{\theta}_Q, \xi) \mathbf{c}$ .

The solution depends on the existence of a certain oscillating polynomial described in the following lemma. Denote a class of functions on  $[a, b]$  by  $\mathcal{W} = \{W(x) = \sum_{i=0}^{p+q} a_i x^i / Q^2(x) \mid a_i \in \mathbb{R}, x \in [a, b]\}$  and define the determinant by

$$H \begin{pmatrix} 0, & 1, & \dots, & p+q \\ x_0, & x_1, & \dots, & x_{p+q} \end{pmatrix} \equiv \begin{vmatrix} \frac{1}{Q^2(x_0)} & \frac{1}{Q^2(x_1)} & \dots & \frac{1}{Q^2(x_{p+q})} \\ \frac{x_0}{Q^2(x_0)} & \frac{x_1}{Q^2(x_1)} & \dots & \frac{x_{p+q}}{Q^2(x_{p+q})} \\ \vdots & \vdots & & \vdots \\ \frac{x_0^{p+q}}{Q^2(x_0)} & \frac{x_1^{p+q}}{Q^2(x_1)} & \dots & \frac{x_{p+q}^{p+q}}{Q^2(x_{p+q})} \end{vmatrix}. \quad (34)$$

**Lemma 4** *Assume that  $Q(x) = Q(x, \theta_Q) \neq 0$  for any  $x \in [a, b]$ . Then there exists a unique  $W^*(x) = \sum_{i=0}^{p+q} a_i^* x^i / Q^2(x) \in \mathcal{W}$  satisfying the following properties:*

1.  $|W^*(x)| \leq 1$ ;
2. *there are  $p+q+1$  points  $a \leq s_0 < s_1 < \dots < s_{p+q} \leq b$ , such that  $W^*(s_i) = (-1)^{p+q-i}$ ,  $i = 0, 1, \dots, p+q$ . Furthermore, when  $U(x) \equiv 1 \in \mathcal{W}$ , then  $s_0 = a$  and  $s_{p+q} = b$ , and equality occurs in part 1 only for  $x = s_0, s_1, \dots, s_{p+q}$ .*

**Proof.** The result follows from Theorem II. 10.1 of Karlin and Studden (1966) and the fact that  $\frac{1}{Q^2(x)} \mathbf{h}_{p+q}(x) = \frac{1}{Q^2(x)} (1, x, \dots, x^{p+q})^t$  forms a T-system on  $[a, b]$ .  $\square$

For  $s_0, s_1, \dots, s_{p+q}$  given by Lemma 4, define the following determinants:

$$D_i = H \begin{pmatrix} 0, & \dots, & i-1, & i, & \dots, & p+q-1, & p+q \\ s_0, & \dots, & s_{i-1}, & s_{i+1}, & \dots, & s_{p+q}, & x^e \end{pmatrix}, \quad (35)$$

where  $i = 0, \dots, p+q$ . We know that  $D_i > 0$  if  $x_e > b$ , and the sign of  $D_i$  will be  $(-1)^{p+q}$  if  $x_e < a$ . We further let

$$p_i = |D_i| / \sum_{j=0}^{p+q} |D_j|, \quad i = 0, \dots, p+q. \quad (36)$$

Then we have the following theorem for the locally  $c_e$ -optimal design.

**Theorem 8** *Assume that  $Q(x) \neq 0$  for any  $x \in \{x_e\} \cup [a, b]$ . Then*

(a) *For any design  $\xi$ ,*

$$\phi_e(\xi, x_e) \geq [W^*(x_e)]^2. \quad (37)$$

(b) *Equality occurs in (37) if and only if  $\xi = \xi^*$  which puts mass  $p_i$  at the point  $s_i$ ,  $i = 0, 1, \dots, p+q$ .*

**Proof.** The proof follows the same arguments as in the proof of Theorem 2.2 of Studden (1968) and is therefore omitted.  $\square$

**Remark 2** Assume that the conditions of Theorem 8 hold. If  $p \geq q$ , then

(a)  $Q^2(x)W^*(x)$  is the  $(p+q)$ th orthogonal polynomial with respect to

$$\frac{1}{\sqrt{(b-x)(x-a)} Q^4(x)};$$

(b) the two boundary supporting points of the  $c_e$ -optimal design are  $s_0 = a$  and  $s_{p+q} = b$ ;

(c) the middle supporting points  $(s_1, \dots, s_{p+q-1})$  for the  $c_e$ -optimal design are the roots of the  $(p+q-1)$ st orthogonal polynomial  $u_{p+q-1}(x)$  with respect to

$$\frac{\sqrt{(b-x)(x-a)}}{Q^4(x)}.$$

**Proof.** Part (a) follows from Krein and Nudelman (1977, Pages 363-364). Part (b) follows from Theorem 2.2 of Studden (1968) and Part (c) follows from Lau (1983, Page 126).  $\square$

We know that the locally  $c_e$ -optimal design, say  $\xi_e$ , is independent of  $\theta_P$ , the parameters in  $P(x)$ . From Lemma 4 and Theorem 8, the  $p+q+1$  support points  $(s_0, s_1, \dots, s_{p+q})$  of  $\xi_e$  do not depend on  $x_e$  (this agrees with the result for  $c_e$ -optimal designs under a polynomial regression model). In contrast, the weights at  $(s_0, s_1, \dots, s_{p+q})$  for  $\xi_e$  depend on  $x_e$  and are not equally weighted. Recall that, in Theorem 5, we showed that the locally D-optimal design,  $\xi_D$  say, is equally weighted on  $p+q+1$  points. However, we do not know the location of these  $p+q+1$  points. In practice, instead of finding  $\xi_D$ , one may use the design  $\xi_e^{ew}$  which puts equal weight on the  $p+q+1$  support points of  $\xi_e$ . Such a  $\xi_e^{ew}$  is often “close” to  $\xi_D$ , as the following example illustrates.

*Example 1 (continued).* Suppose that we want to extrapolate the experimental dependence to a given point  $x_e > b$  for model (19), i.e., to estimate,  $f(x_e, \theta) = (\theta_0 + \theta_1 x_e)/(1 + \theta_2 x_e)$ , the mean response at  $x_e$ . Since  $p \geq q$ , from Remark 2, we know that the locally  $c_e$ -optimal design has three supporting points  $\{a, s_1, b\}$ , where  $s_1$  is the root of the linear function orthogonal to  $\sqrt{(b-x)(x-a)}/(1 + \theta_2 x)^4$ . Therefore, we have

$$s_1 = \left\{ \int_a^b \frac{x \sqrt{(b-x)(x-a)}}{(1 + \theta_2 x)^4} dx \right\} / \left\{ \int_a^b \frac{\sqrt{(b-x)(x-a)}}{(1 + \theta_2 x)^4} dx \right\}. \quad (38)$$

Furthermore, by Theorem 8, the weights for the optimal design at the three points  $\{a, s_1, b\}$  are  $\{p_0, p_1, p_2\}$ , where,  $p_i = C_i/(C_0 + C_1 + C_2)$ ,  $i = 0, 1, 2$ :  $C_0 = (1 + \theta_2 a)^2(b - s_1)(x_e - b)(x_e - s_1)$ ,  $C_1 = (1 + \theta_2 s_1)^2(b - a)(x_e - b)(x_e - a)$ , and  $C_2 = (1 + \theta_2 b)^2(s_1 - a)(x_e - s_1)(x_e - a)$ .

Let us compare numerical values of the two middle design points,  $x^*$  (defined by (29) for the locally D-optimal design) and  $s_1$  (defined by (38) for the locally  $c_e$ -optimal design). We chose  $a = 0$  and  $b = 1$ . Then numerical values of  $x^*$  and  $s_1$  for various values of  $\theta_2$  are given in Table 1. The weights  $p_i$  for the locally  $c_e$ -optimal design are also given. It is seen that  $x^*$  and  $s_1$  are identical up to  $10^{-8}$ .

Table 1. The middle design points  $x^*$  and  $s_1$  and the weights  $(p_1, p_2, p_3)$  for the model

$$Y = (\theta_0 + \theta_1 x)/(1 + \theta_2 x) + \epsilon, \quad x \in [0, 1].$$

$\theta_2$	$x^*$	$s_1$	$x_e$	$(p_1, p_2, p_3)$
0	0.4999999991	0.5000000000	2	(0.1765, 0.4706, 0.3529)
			20	(0.2437, 0.4998, 0.2565)
			200	(0.2494, 0.5000, 0.2506)
1	0.3333333341	0.3333333333	2	(0.1220, 0.3902, 0.4878)
			20	(0.1719, 0.4662, 0.3619)
			200	(0.1760, 0.4702, 0.3538)
10	0.0833333313	0.0833333333	2	(0.0373, 0.1426, 0.8201)
			20	(0.0615, 0.2264, 0.7121)
			200	(0.0636, 0.2333, 0.7031)
100	0.0098039039	0.0098039216	2	(0.0048, 0.0192, 0.9759)
			20	(0.0090, 0.0356, 0.9554)
			200	(0.0094, 0.0372, 0.9534)

Before concluding this section, let us state a relationship between the locally  $c_e$ -optimal design and locally L-optimal design when  $L(\theta) = \nabla g^t \nabla g$  and  $\nabla g$  is given by (25). This arises from estimating  $\theta_p/\theta_{p+q}$ . It is of interest that the supporting points for these two locally optimal designs are identical.

**Theorem 9** *Assume that  $Q(x) \neq 0$  for any  $x \in [a, \infty)$  and  $\theta_{p+q} \neq 0$ . Then the supporting points for the locally  $c_e$ -optimal design are the same as those of the locally L-optimal design when  $L(\theta) = \nabla g^t \nabla g$  and  $\nabla g$  is given by (25).*

**Proof.** From Theorem 8, the  $c_e$ -optimal design for estimating  $P(x_e)/Q(x_e)$  is independent of  $x_e$ . The optimal design for estimating  $P(x_e)/Q(x_e)$  is invariant under multiplication by a constant. Let  $x_e \rightarrow \infty$ ; then  $x_e^{q-p} P(x_e)/Q(x_e) \rightarrow \theta_p/\theta_{p+q}$ . The result follows immediately.  $\square$

## 6. The Michaelis–Menten model.

*6.1. The model and the D-optimal design.* For many physical and biological phenomena, saturation functions are often described by a rectangular hyperbola. The Michaelis-Menten model

$$Y = \frac{\theta_1 x}{\theta_2 + x} + \epsilon, \quad 0 \leq x \leq b \leq \infty, \quad (39)$$

is frequently used to model the speed of enzyme kinetic reactions, where  $Y$  is the reaction velocity,  $x$  the concentration of substrate,  $\theta_1 > 0$  the maximum velocity of the reaction, and  $\theta_2 \geq 0$  the half-saturation constant, since the mean response is half-maximal at  $\theta_2$ . Note that this model is not a special case of model (6).



Currie (1982) concluded that the locally D-optimal design is obtained by taking half the observations at the point  $\theta_2$  and the other half as high as possible, namely at  $b$ . We will see that this is only approximately true when  $b$  is very large. Define

$$\tilde{\mathbf{B}} = \begin{pmatrix} \theta_2 & 1 \\ -\theta_1 & 0 \end{pmatrix}, \quad \tilde{\mathbf{h}} \equiv \tilde{\mathbf{h}}(x) = \frac{1}{(\theta_2 + x)^2} \begin{pmatrix} x \\ x^2 \end{pmatrix}. \quad (40)$$

Then  $\nabla f(x, \boldsymbol{\theta}) = \tilde{\mathbf{B}}\tilde{\mathbf{h}}(x)$  and  $I(\boldsymbol{\theta}, x) = \nabla f \nabla f^t = \tilde{\mathbf{B}}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^t \tilde{\mathbf{B}}^t$ .

**Theorem 10** Consider model (39). (a) Assume that the marginal prior distribution of  $\theta_2$  has 1/2 mass at each of the points  $c$  and  $d$ . Among all two-point designs, the Bayesian optimal design  $\xi$  puts equal weight on two points,  $x^*$  and  $b$ , where

$$x^* = \frac{\sqrt{cd(b+c)(b+d)} - c}{c+d+b}. \quad (41)$$

(b) The locally D-optimal design,  $\xi^*$  is the equally weighted two-point design at  $\{x^*, b\}$ , where

$$x^* = \theta_2 b / (2\theta_2 + b). \quad (42)$$

**Proof.** Part (a) can be proved similarly to Proposition 1. For Part (b), by an argument similar to Theorem 5, we can show that the locally D-optimal design is equally supported on two points. These two points can be found from Part (a) by letting  $c = d = \theta_2$ .  $\square$

**6.2. The locally  $c_e$ -optimal design.** Suppose that we want to extrapolate the model to  $x_e > b$ , i.e., to estimate  $f(x_e, \boldsymbol{\theta}) = \theta_1 x_e / (\theta_2 + x_e)$ . The design criterion is

$$\phi_e(\xi; x_e) = \nabla f(x_e, \boldsymbol{\theta})^t M^{-1}(\boldsymbol{\theta}, \xi) \nabla f(x_e, \boldsymbol{\theta}), \quad M(\boldsymbol{\theta}, \xi) = \int_{\mathcal{X}} I(\boldsymbol{\theta}, x) \xi(dx).$$

It is easy to see that

$$\phi_e(\xi; x_e) = \tilde{\mathbf{h}}^t(x_e) M_0^{-1}(\theta_2; \xi) \tilde{\mathbf{h}}(x_e), \quad M_0^{-1}(\theta_2; \xi) = \int_{\mathcal{X}} \tilde{\mathbf{h}}(x) \tilde{\mathbf{h}}^t(x) \xi(dx).$$

**Theorem 11** The  $c_e$ -optimal design at  $x_e$  is the two point design at  $z^*$  and  $b$  with weights  $p^*$  and  $1 - p^*$ , respectively, where

$$z^* = \frac{(\sqrt{2} - 1)\theta_2^2 + \sqrt{2}\theta_2 b}{\theta_2^2 + 4\theta_2 b + 2b^2} b, \quad (43)$$

$$p^* = \frac{b(x_e - b)(\theta_2 + z^*)^2}{b(x_e - b)(\theta_2 + z^*)^2 + z^*(x_e - z^*)(\theta_2 + b)^2}. \quad (44)$$

**Proof.** This is similar to Theorem 8.  $\square$

The results give some comparisons for the two locally optimal designs.

**Proposition 2** For any given  $\theta_2$ , let  $z^*$ ,  $p^*$  and  $x^*$  be given by (43), (44) and (42), respectively.

Then (a)  $\lim_{b \rightarrow \infty} z^* = \frac{\sqrt{2}}{2}\theta_2$ ; (b)  $\lim_{b \rightarrow \infty} x^* = \theta_2$ ; (c) the lower locally D-optimal design point is always smaller than the optimal design point for extrapolation, i.e.,  $z^* < x^*$ , for any  $b > 0$ ; (d) for fixed  $b > 0$ ,  $\lim_{x_e \downarrow b} p^* = 0$  and  $\lim_{x_e \uparrow \infty} p^* = b \left( \frac{\sqrt{2}}{2} + 1 \right)^2 \theta_2^2 / \left[ b \left( \frac{\sqrt{2}}{2} + 1 \right)^2 \theta_2^2 + \frac{\sqrt{2}}{2} \theta_2 (\theta_2 + b)^2 \right]$ .

**Proof.** Parts (a) and (b) are obvious. For Part (c), it is enough to check that  $[(\sqrt{2} - 1)\theta_2 + \sqrt{2}b](b + 2\theta_2) \leq 2b^2 + 4b\theta_2 + \theta_2^2$ . Part (d) follows from part (a) and some algebra.  $\square$

An example of the graph of the supporting points  $x^*$  and  $z^*$  is showed in Figure 1. For the graph, we chose  $\theta_2 = 50$ . We see that both  $x^*$  and  $z^*$  are monotonically increasing in  $b$  and satisfy  $z^* < x^* < \theta_2$ .

**Acknowledgements.** Most of the work was done while Z. He was a graduate student at Purdue University and D. Sun was visiting the Department of Statistics, Purdue University. The authors would like to thank the Department of Statistics for its hospitality. The authors are also indebted to Professor Lawrence D. Brown, an associate editor and two referees for constructive comments on an earlier version of this paper.

#### REFERENCES

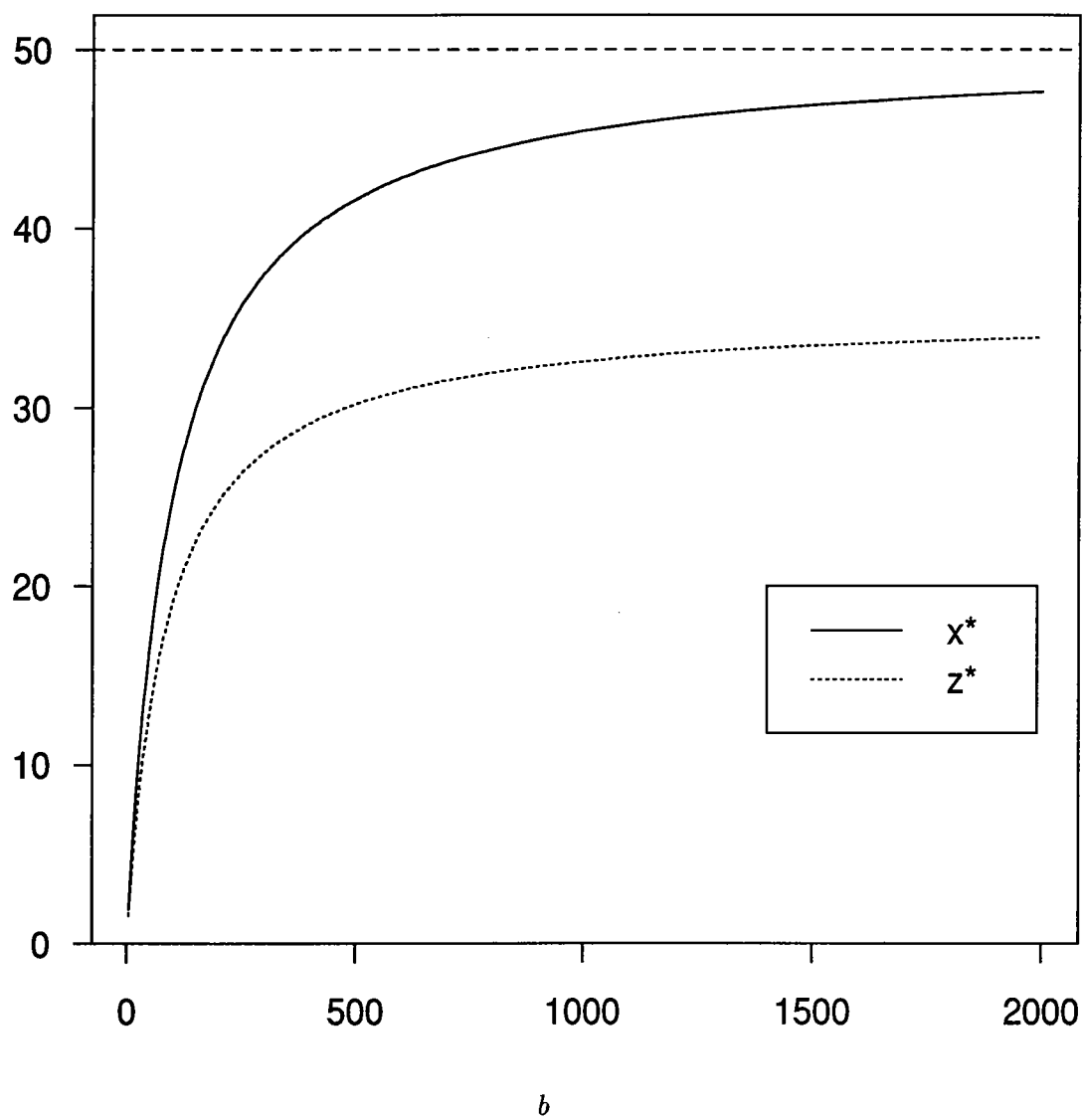
- Box, G.E.P. and Lucas, H.L. (1959). Design of experiments in non-linear situations. *Biometrika* **49**, 77-90.
- Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal Of Statistical Planning And Inference*. **37**, 229-235.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian experimental design applied to logistic regression experiments. *Journal Of Statistical Planning And Inference* **21**, 191-208.
- Chernoff, H. (1953). Local optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.
- Currie, D.J. (1982). Estimating Michaelis-Menten Parameters: Bias, Variance And Experimental Design. *Biometrics* **38**, 907-919.
- Dette, H.(1990). A generalization of  $D$ - and  $D_1$ -optimal design in polynomial regression. *Annals of Statistics* **18**(4), 1784-1804.
- Dette, H. and Neugebauer, H.-M. (1995). Bayesian optimal one point designs, to appear in *Journal Of Statistical Planning And Inference*.
- Fisher, R.A.(1922). On the mathematical foundation of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Ser. A* **22**, 309-268.
- Fedorov, V.V.(1972). *Theory of optimal experiment*. New York: Academic Press.
- Ford, I., Titterton, D.M., and Kitsos C.P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49-60.
- Ford, I., Torsney, B. and Wu, C.F.J.(1992). The use of a canonical form in the construction of locally optimal designs for nonlinear problems. *J. R. Statist. Soc. B* **54**, 5669-583.
- Guest, P. G. (1958). The spacing of observations in polynomial regression, *Ann. Math. Stat.*, **29**, 294-299.
- He, Z. (1993). Optimal designs for rational models. Ph. D. dissertation, Purdue University.
- Hill, P.D.H. (1980). D-optimal designs for partially nonlinear regression models. *Technometrics* **22**, 275-276.
- Hoel, P.G. and Levine, A. (1964). Optimal spacing and weighting in polynomial prediction. *Ann. Math. Statist.* **35**, 1553-1560.
- Karlin, S. and Studden, W.J. (1966). *Tchebycheff Systems: with application in analysis and statistics*, Interscience Publishers, A Division of John Wiley & Sons, New York.

- Katz, D., Azen, S.P., and Schumitzky, A. (1981). Bayesian approach to the analysis of nonlinear models: implementation and evaluation. *Biometrika* **37** 137-142.
- Läuter, E.(1974). Experimental design in a class of models. *Math. Operations forch. Statist. Ser. Statist.* **5**, 379-398.
- Lindley, D.V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* **27** 986-1005.
- Pilz, J. (1991). Bayesian estimation and experimental design in linear regression models, Wiley, New York.
- Silvey, S.D.(1980). *Optimal design*, London: Chapman & Hall.
- Sparrow, P.E.(1979a). Nitrogen response curves of spring barley. *Journal of Agricultural Science, Cambridge* **92**, 307-317.
- Sparrow, P.E.(1979b). The comparison of five response curves for representing the relationship between the annual dry-matter yield of grass herbage and fertilizer nitrogen. *Journal of Agricultural Science, Cambridge* **93**, 513-520.
- Studden, W.J. (1968). Optimal Designs on Tchebycheff Points. *Ann. Math. Statist.* **39**, 1435-1447.
- Tsutakawa, R.K. (1972). Design of an experiment for bioassay. *J. Amer. Statist. Assoc.* **67**, 584-590.
- Whittle, P. (1973). Some general points in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **35**, 123-130.

Z. HE  
 FISHERY AND WILDLIFE RESEARCH CENTER  
 MISSOURI DEPARTMENT OF CONSERVATION  
 1110 SOUTH COLLEGE AVENUE  
 COLUMBIA, MO 65201

W. J. STUDDEN  
 DEPARTMENT OF STATISTICS  
 PURDUE UNIVERSITY  
 MATHEMATICAL SCIENCES BUILDING  
 WEST LAFAYETTE, IN 47907-1399

D. SUN  
 DEPARTMENT OF STATISTICS  
 UNIVERSITY OF MISSOURI-COLUMBIA  
 323 MATHEMATICAL SCIENCES BUILDING  
 COLUMBIA, MO 65211



**Figure 1. Comparison of Lower Supporting Points  $x^* = b\theta_2/(2\theta_2 + b)$  and  $z^* = b[(\sqrt{2} - 1)\theta_2^2 + \sqrt{2}\theta_2 b]/[\theta_2^2 + 4\theta_2 b + 2b^2]$ , for Fixed  $\theta_2 = 50$**