

## Designing for Parameter Subsets in Gaussian Nonlinear Regression Models

Timothy E. O'Brien  
*Loyola University Chicago*

*Abstract:* This article presents and illustrates several important subset design approaches for Gaussian nonlinear regression models and for linear models where interest lies in a nonlinear function of the model parameters. These design strategies are particularly useful in situations where currently-used subset design procedures fail to provide designs which can be used to fit the model function. Our original design technique is illustrated in conjunction with D-optimality, Bayesian D-optimality and Kiefer's  $\Phi_k$ -optimality, and is extended to yield subset designs which take account of curvature.

*Key words:* Antagonism, AUC, Bayesian designs, curvature, parameter subsets,  $\Phi_k$ -optimality, robustness, synergy, turning-point.

### 1. Introduction

Applied researchers often find nonlinear regression models useful to describe their phenomena under study since (1) their processes can often be described in terms of compartmental models the solution of which results in nonlinear models, (2) these models typically involve a smaller number of model parameters than linear-model-approximations and (3) the nonlinear model parameters are usually easier to interpret. For example, growth is often modelled using a sigmoidal curve and can then be characterized in terms of the so-called LD<sub>50</sub>-parameter. Similarly, pharmacologists often describe the concentration of a drug in the blood stream using a given nonlinear model function, and focus typically centers on the area under the concentration curve (the total concentration received), which involves a function of the model parameters. Additional important applications of nonlinear models are given in Finney (1952), Bates and Watts (1988), Seber and Wild (1989), Atkinson *et al* (1993), and Lindsey (2001). But before a nonlinear regression model can be fit to data and the model parameters interpreted, researchers must decide on the study design, meaning, for example, that decisions must be made as to when serum concentrations are to be measured, or which doses are to be chosen. As a result, experimental design strategies for nonlinear regression models are paramount. Further, by definition, nonlinear models are

such that the information matrix depends upon (at least some) model parameters, so that some design strategies focus on so-called *local* designs, or designs based on an initial estimate of these model parameters. Readers unfamiliar with the basics of optimal design methodology are encouraged to consult the non-technical introduction provided in O'Brien and Funk (2003).

Since interest often centers on only a subset of the parameters in a non-linear regression models (for example, just the  $LD_{50}$  parameter in a sigmoidal model function), design of experiments for estimation of this parameter subset has generated much attention; see Kiefer (1961), Hunter, Hill and Henson (1969), Box (1971b), Whittle (1973), and Hill and Hunter (1974)). Current practice, as described in Seber and Wild (1989) and Atkinson and Donev (1992), typically involves choosing a design which maximizes the information of the parameter subset, or equivalently, which minimizes the generalized first-order variance of the corresponding least-squares estimate. One problem with this approach is that it often yields designs for which the information matrix corresponding to the full parameter vector is singular. As a result, several authors have extended optimal design methodology to allow for singular information matrices; see Vuchkov (1977), Silvey (1978), Pázman (1980), and Gaffke (1985). However, when the ultimate goal is to obtain designs to fit the given model function, a more important concern is that these designs provide inadequate information about the so-called nuisance parameters, and hence cannot be used to estimate the model parameters. Indeed, in some important instances, the subset designs cannot even provide a means of estimating the important parameter subset, and their usefulness is therefore quite limited in practice.

We introduce here several important design strategies which focus on the parameter subset of interest, yet which yield some information about the nuisance parameters, thereby providing designs that can be used to fit the model. Our subset design technique, originally developed for the local D-optimality design criterion, is easily adapted to make it more robust to the initial parameter estimate by using the Bayesian design approach given in Läuter (1974b) and Chaloner and Larntz (1989), to make it robust to the D-criterion by using Kiefer's so-called  $\Phi_k$  criterion (Kiefer, 1975), and to take account of curvature. These extensions are important since, for example, if an optimal design is chosen under the assumption that the model parameters are  $\theta_1^o = 0.2$  and  $\theta_2^o = 0.7$  (see the first example in the next section), but in fact the model parameters are quite different from these values, then the quality of the chosen design may indeed be very poor; this fact, also underscored in Atkinson and Donev (1992), highlights the importance of the parameter-robust design strategy discussed below. Similar arguments highlight the importance of seeking  $\Phi_k$ -optimal designs and curvature-adjusted designs when researchers desire designs that are optimal under several design criteria or

when linear approximations to nonlinear models are inadequate.

## 2. Background and Applications

In this section we review the original subset design strategy of Hill and Hunter (1974) and describe several examples for which the design strategy is useful. Additional details of the subset criterion are given in Silvey (1980) and Atkinson and Donev (1992); as mentioned above, readers interested in a gentle introduction to the concepts involved in optimal design methodology, including a simple illustration involving an information matrix and variance function, are referred to O'Brien and Funk (2003). For the Gaussian nonlinear model with model function,  $E(Y) = \eta(x, \theta)$ , the  $p \times 1$  parameter vector  $\theta$  can be partitioned as  $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ , where  $\theta_1$  is of dimension  $p_1 \times 1$ , and  $\theta_2$  is of dimension  $p_2 \times 1$  ( $p_1 + p_2 = p$ ). The subset design problem for this model function involves choosing a design with  $n$  design points, denoted  $\xi$ , to estimate the full parameter vector, yet with the subset of interest,  $\theta_2$ , estimated more precisely than the so-called nuisance parameters,  $\theta_1$ . For example, for the above nonlinear growth-curve example, the nuisance parameters include the asymptote(s) and slope parameters, whereas the parameter of interest is the LD<sub>50</sub> parameter. Here the design  $\xi$  can be written as

$$\xi = \left\{ \begin{array}{cccc} \mathbf{x}_1, & \mathbf{x}_2, & \dots, & \mathbf{x}_k \\ \omega_1, & \omega_2, & \dots, & \omega_k \end{array} \right\}$$

where the  $k$  ( $k \leq n$ ) distinct design support "points" (or vectors)  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  are elements of the design space  $\mathbf{X}$ , and the associated weights  $\omega_1, \omega_2, \dots, \omega_k$  are non-negative real numbers which sum to one. For a given  $n$ , whenever integers  $n_1, n_2, \dots, n_k$  exist such that  $\omega_j = \frac{n_j}{n}$  for  $j = 1$  to  $k$ , then  $\xi$  is said to be a discrete-weight design; otherwise, it is said to be a continuous-weight design. Thus, for example, a study design involving 10 patients receiving (drug) dose  $x_1 = 5$ mg, 20 patients receiving dose  $x_2 = 15$ mg, and 30 patients receiving dose  $x_3 = 50$ mg, denoted  $\xi = \left\{ \begin{array}{ccc} 5 & 15 & 50 \\ 1/6 & 1/3 & 1/2 \end{array} \right\}$ , illustrates the use of a discrete-weight design for  $n = 60$  total patients. Whether  $\xi$  is discrete- or continuous-weight,  $n\omega_j$  can be thought of as the "number of observations" to be taken at experimental level  $\mathbf{x}_j$ . Continuous-weight designs can then be converted into practical ones by using rounding procedures given in Pukelsheim and Rieder (1992) and Pukelsheim (1993).

For the design  $\xi$ , the  $n \times p$  Jacobian matrix  $\mathbf{V} = \frac{\partial \eta}{\partial \theta}$  can be partitioned as  $\mathbf{V} = [\mathbf{V}_1 | \mathbf{V}_2]$ , where  $\mathbf{V}_1 = \frac{\partial \eta}{\partial \theta_1}$  is of dimension  $n \times p_1$  and  $\mathbf{V}_2 = \frac{\partial \eta}{\partial \theta_2}$  is of dimension  $n \times p_2$ . Assuming Gaussian errors, the information matrix associated with the full parameter vector  $\theta$  is given by  $\mathbf{M} = \mathbf{M}(\xi, \theta) = \mathbf{V}^T \mathbf{W} \mathbf{V}$ , where  $\mathbf{W}$

$= \text{diag}\{\omega_1, \omega_2, \dots, \omega_k\}$ . Since  $\mathbf{M}$  can likewise be partitioned as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1^T \mathbf{W} \mathbf{V}_1 & \mathbf{V}_1^T \mathbf{W} \mathbf{V}_2 \\ \mathbf{V}_2^T \mathbf{W} \mathbf{V}_1 & \mathbf{V}_2^T \mathbf{W} \mathbf{V}_2 \end{bmatrix},$$

the information corresponding to the subset  $\theta_2$  in the model function  $\eta(x, \theta)$  is given by  $\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12}$ , where  $\mathbf{M}_{11}^+$  is a generalized inverse of  $\mathbf{M}_{11}$ . The subset design procedure of Hill and Hunter (1974) advocates the use of locally  $D_s(\theta_2)$ -optimal designs, or designs which maximize the subset information defined by the determinant

$$|\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12}|; \quad (2.1)$$

again, the word ‘‘locally’’ is used here to stress the fact that these designs use an initial estimate of the parameters (Chernoff, 1953). Equivalently,  $D_s(\theta_2)$ -optimal designs also minimize the determinant of the covariance matrix of the least-squares estimate of  $\theta_2$ ,

$$(\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12})^{-1}. \quad (2.2)$$

As an aside, O'Brien and Rawlings (1996) point out that we can write

$$|\mathbf{M}| = |\mathbf{M}_{11}| |\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12}| = |\mathbf{M}_{11}| |\mathbf{M}_{22}| \Delta,$$

where  $\Delta$ , which lies between zero to one, is larger the more orthogonal the hyper-planes spanned by the columns of  $\mathbf{V}_1$  and  $\mathbf{V}_2$ . It follows that  $D_s(\theta_2)$ -optimal designs make the volume of the parallelepiped formed from the columns of  $\mathbf{W}^{1/2} \mathbf{V}_2$  and the angle between the above hyper-planes large. In contrast, D-optimal designs, or designs which maximize the determinant  $|\mathbf{M}|$ , make the volume of the parallelepiped formed from the columns of  $\mathbf{W}^{1/2} \mathbf{V}_1$ , the volume of the parallelepiped formed from the columns of  $\mathbf{W}^{1/2} \mathbf{V}_2$ , and the angle between these hyper-planes large. Hill and Hunter (1974) illustrate these concepts geometrically for the case where  $p_1 = p_2 = 1$  and using only discrete-weight designs and  $n = 2$  (so that  $\mathbf{W}$  can be ignored).

As discussed in Atkinson and Donev (1992), the variance function which corresponds to (2.2) is

$$\begin{aligned} d_S(\mathbf{x}, \xi, \theta) &= d(\mathbf{x}, \xi, \theta) - d_1(\mathbf{x}, \xi, \theta) \\ &= \frac{\partial \eta(\mathbf{x})}{\partial \theta^T} \mathbf{M}^{-1}(\xi, \theta) \frac{\partial \eta(\mathbf{x})}{\partial \theta} - \frac{\partial \eta(\mathbf{x})}{\partial \theta_1^T} \mathbf{M}_{11}^{-1}(\xi, \theta) \frac{\partial \eta(\mathbf{x})}{\partial \theta_1} \end{aligned}$$

whenever both  $\mathbf{M}$  and  $\mathbf{M}_{11}$  are invertible. To show that a given design,  $\xi^*$ , is indeed  $D_s(\theta_2)$ -optimal, the General Equivalence Theorem (originally given in Kiefer and Wolfowitz (1960) and extended to the case of subset designs in Kiefer

(1961)) can be used by verifying that for all  $\mathbf{x} \in X$ ,  $d_S(\mathbf{x}, \xi^*, \theta) \leq p_2$ , and that  $d_S(\mathbf{x}_j, \xi^*, \theta) = p_2$  at each of the support points of  $\xi^*$ . Extensions of these results to handle the situation where  $\mathbf{M}$  is singular, and which involve Fréchet derivatives, are given in Silvey (1980). Computational procedures to obtain optimal designs, discussed for example in Chapter 15 of Atkinson and Donev (1992), are often based on the General Equivalence Theorem.

These theoretical results notwithstanding, it is not uncommon, however, that the  $D_s(\theta_2)$  design strategy of Hill and Hunter (1974) produces designs with less than  $p$  support points (see the following representative and important examples), and hence these designs lead to the troublesome situation in which the model cannot be fit once the design has been implemented and the data gathered. Further, although Atkinson and Bogacka (1997) consider a combined criterion combining information regarding  $\theta_1$  into the (subset) design criterion, they fail to point out that singular subset designs can indeed result and thus be useless in fitting the assumed model. In the next section, we give a modification of this subset strategy which provides designs for which the full parameter vector is estimable and which are still efficient vis-a-vis the  $D_s(\theta_2)$ -optimal design. Throughout the rest of this article, we focus on the following key examples useful in environmental and biomedical applications and the detection of drug synergy, though mindful that the techniques developed here can easily be adapted to a wide array of other applications.

**Example one.** Hill and Hunter (1974) illustrates the subset design procedure outlined above using the two-parameter intermediate product (IP2) model function (also examined in Hamilton and Watts (1985) and O'Brien (1992)), given by the equation

$$\eta(x, \theta_1, \theta_2) = \frac{\theta_1}{\theta_1 - \theta_2} \left( e^{-\theta_2 x} - e^{-\theta_1 x} \right), \quad x \geq 0.$$

This model is extremely useful in the modelling of phenomena as diverse as chemical reactions and the movement of drugs through the body (pharmacokinetics); see, for example, Atkinson and Donev (1992), Atkinson, *et al* (1993) and Lindsey (2001). To briefly highlight the *full* D-optimality criterion (which provides information regarding both model parameters), note that the local D-optimal design corresponding to the initial parameter estimates  $\theta_1^0 = 0.2$  and  $\theta_2^0 = 0.7$  is  $\xi_D = \left\{ \begin{array}{cc} 1.59 & 5.89 \\ 1/2 & 1/2 \end{array} \right\}$ . Indeed, (local) D-optimality of this design is established by noting that the corresponding variance function, graphed in Figure 1, does not exceed the line  $y = 2$  and achieves its maximum at the support points of  $\xi_D$  (viz, 1.59 and 5.89); all other designs would result in variance functions that exceed  $y = 2$  over some portion of the design region.

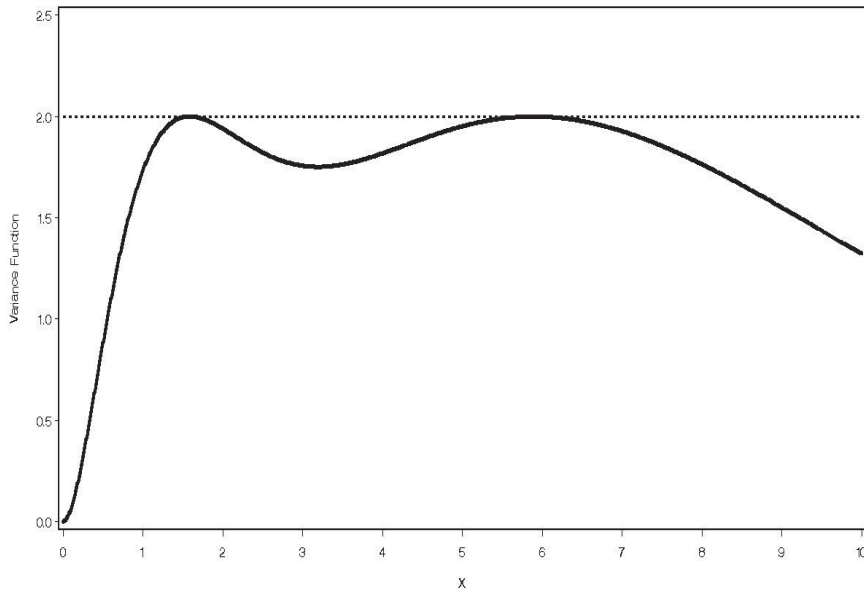


Figure 1: Variance function for IP2 model function and D-optimal design

In contrast with this design, the locally  $D_s(\theta_2)$ -optimal *subset* design for this model using the same parameter guesses places all weight at the point  $x = 6.76189$ , thereby providing no means to fit the model. For this model function, the area under the curve (AUC) — relevant in pharmacokinetic-pharmacodynamic applications — is given by  $\phi = 1/\theta_2$ . Thus, in this setting since the  $D_s(\phi)$ -optimal design also places all weight at the point  $x = 6.76189$ , even the parameter of interest ( $\phi$ ) is non-estimable using this subset design. Indeed,  $D_s(\theta_2)$ -optimality of this design is established using the extension of the General Equivalence Theorem given by Silvey (1980); interestingly, this design disagrees with the one provided in Hill and Hunter (1974), who searched only over the space of discrete-weight designs, and report a design with equal weight at  $x_1 = 6.66$  and  $x_2 = 6.87$ . Since Hill and Hunter (1974) only used a grid of points (with two decimal place accuracy) and with  $n$  fixed at two, this example also highlights the need to search for continuous-weight designs since the discrete optimization problem can evidently be unstable. Finally, lest we leave the reader with the impression that singular subset designs are more the exception to the rule than the rule itself, note that the locally  $D_s(\theta_2)$ -optimal subset design for  $(\theta_1^o, \theta_2^o) = (0.15, 0.70)$  places all the weight at  $x = 8.401$ , for  $(\theta_1^o, \theta_2^o) = (0.25, 0.70)$  places all the weight at  $x = 5.755$ , and so on. The point, therefore, is clear: subset designs can fail to provide in-

formation about even the important model parameters — at least if one uses the local D-optimality criterion.

**Example two.** Chaloner (1989) and Cook and Weisberg (1990) consider the (Gaussian) quadratic regression model, denoted here by

$$\eta(x, \theta_0, \theta_2, \phi) = \theta_0 - 2\theta_2\phi x + \theta_2x^2, \quad 0 \leq x \leq 1, \quad (2.3)$$

where focus lies in the turning point  $\phi$ . Alternatively, one could write

$$\eta(x, \theta) = \theta_0 + \theta_1x + \theta_2x^2, \quad \phi = -\frac{\theta_1}{2\theta_2}, \quad (2.4)$$

highlighting the fact that the design problem for a nonlinear function of the parameters of a linear model can often be transformed into a subset design problem for a nonlinear model. A practical application of this model occurs as a quadratic approximation in environmental science in which  $x$  corresponds to dose of ozone applied to experimental plants,  $y$  corresponds to plant yield, and for which reduction of ozone below ambient air levels actually corresponds to a drop in plant yield; see van Ewijk and Hoekstra (1994) for datasets and illustrations.

Local  $D_s(\phi)$ -optimal designs for this model, which do not require initial guesses of either  $\theta_0$  or  $\theta_2$ , have only two support points when the initial estimate  $\phi^o$  lies in the interval  $[\frac{1}{4}, \frac{3}{4}]$  and three support points otherwise; see Chaloner (1989) for a discussion and additional applications. Interestingly, some Bayesian D-optimal designs, or designs which maximize the expected log-information

$$\int \log |\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^+\mathbf{M}_{12}| p(\phi) d\phi,$$

where  $p(\phi)$  is an assumed prior for the turning point  $\phi$ , can also have only two support points for this model; indeed, Pilz (1991) gives necessary and sufficient conditions for even a one-point Bayesian design to be optimal. This is important since local optimal designs can be criticized as dependent upon the initial parameter guess, and, more importantly, since it has been inferred that this Bayesian criterion will always yield so-called robust designs, or design with extra support points that can be used to test for lack of fit (see e.g. Läuter (1974a), Atkinson and Donev (1992)). By way of illustration, Table 1 provides the local  $D_s(\phi)$ -optimal design for this model with initial estimate  $\phi^o = 0.40$  and various Bayesian  $D_s(\phi)$ -optimal designs with (discretized) uniform priors centered at this value. Note that, in contrast to the claim to the contrary in Atkinson, *et al* (1993), several of these Bayesian designs do not have enough support points to fit the assumed model function. Therefore, this example highlights that subset designs associated with the Bayesian D-optimality criterion can also be impractical.

Table 1: Local and Bayesian  $D_2(\phi)$ -optimal designs for example 2.

$p(\phi)$	Bayesian D-optimal Design
$\phi^o = 0.40$	$\begin{Bmatrix} 0 & 0.8 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.35, 0.45]$	$\begin{Bmatrix} 0 & 0.88 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.30, 0.50]$	$\begin{Bmatrix} 0 & 0.97 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.25, 0.55]$	$\begin{Bmatrix} 0.08 & 1 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.20, 0.60]$	$\begin{Bmatrix} 0.18 & 1 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.10, 0.70]$	$\begin{Bmatrix} 0.38 & 1 \\ 0.5 & 0.5 \end{Bmatrix}$
Uniform on $[0.02, 0.78]$	$\begin{Bmatrix} 0 & 0.5 & 1 \\ 0.019 & 0.5 & 0.481 \end{Bmatrix}$

This and the previous example highlight situations in which interest focuses on a (linear or nonlinear) function  $\phi = \phi(\theta)$ , and where  $\phi$  maps  $\mathbb{R}^p$  onto  $\mathbb{R}^s$  for  $s \leq p$ . When  $\phi$  is linear, we write  $\phi = \mathbf{A}^T \theta$ , and if in addition  $s = p$  and  $\mathbf{A}$  is nonsingular, then  $\phi$  corresponds to a reparameterization of the original model function. Whether  $\phi$  is linear or nonlinear, the covariance matrix of the least-squares estimate of  $\phi$  is proportional to  $\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A}$ , where  $\mathbf{A}^T = \frac{\partial \phi}{\partial \theta}$  is  $s \times p$  of rank  $s$ . Designs which minimize the determinant of

$$\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A}, \quad (2.5)$$

called  $D_A$ -optimal designs, were developed in Sibson (1974); when  $s = 1$  — as in the previous examples — these designs are called  $c_\theta$ -optimal (see Atkinson, *et al* (1993)). Finally, for the case where  $s = p_2$  and  $\phi = \theta_2$  (so that  $\mathbf{A}^T = [\mathbf{0}, \mathbf{I}_{p_2}]$ , equation (2.5) reduces to equation (2.2), and it is thus seen that  $D_A$ -optimality indeed generalizes  $D_s$ -optimality.

**Example three.** An important concern in biomedical research is the testing for synergy between similar drugs since then an enhanced therapeutic effect is realized when the drugs are used in combination; see Tallarida (2000) for biomedical applications and O'Brien (2004) for additional examples and references. When the couple  $(a, b)$  represents the amount of compounds or drugs A and B present in a given medium (e.g., the body), the effective dose (or concentration) present in the medium can be modelled statistically by

$$d = a + \theta_4 b + \theta_5 \sqrt{\theta_4 a b}, \quad (2.6)$$

where  $\theta_4$  is the potency of compound B relative to compound A, and  $\theta_5$ , called the coefficient of synergy in Finney (1952), measures the degree of antagonism



or synergy between the two compounds. Gerig, Blum and Meier (1989) relates the effective dose of two phenolic acids to the expected growth of cucumber seedlings ( $\mu = E(Y)$ ) using the Gaussian distribution, the identity link ( $\mu = \eta$ ) and the expression (2.6) with the three-parameter log-logistic (LL3) sigmoidal model function,

$$\eta = \frac{\theta_1}{1 + \left(\frac{d}{\theta_2}\right)^{\theta_3}}. \quad (2.7)$$

This model function is also useful in detecting interactions when studying the effect of air pollutants on crop yield (see O'Brien (1993)) and the efficacy of the antiviral drugs such as AZT and ddI (see Machado and Robinson (1994)). Instead of focusing on the LD<sub>50</sub> parameter ( $\theta_2$ ), the goal here is first to obtain a subset design focusing on the coefficient of synergy  $\theta_5$  so as to assess whether the given compounds act synergistically (i.e., enhance one another) or act antagonistically (counter-act one another), and to then measure the degree of such synergy or antagonism.

Using the parameter estimates given in Gerig, Blum and Meier (1989) — viz,  $\theta_2^o = 0.194$ ,  $\theta_3^o = 2.72$ ,  $\theta_4^o = 0.626$ ,  $\theta_5^o = -0.826$  — the locally  $D_s(\theta_5)$ -optimal design for this model, which is independent of the value of  $\theta_1$ , is

$$\left\{ \begin{array}{ccc} (0.165) & (0.194) & (0) \\ (0.264) & 0 & (0.310) \\ 1/2 & 1/4 & 1/4 \end{array} \right\} \quad (2.8)$$

Interestingly, the dose at each of these design points is identical ( $d = 0.194$ ) — and identical to the initial estimate of the LD<sub>50</sub> — and so use of this design would mean that the model function (2.7) could not be fit to the data. Equally interesting is the fact that the locally  $D_s(\theta_4, \theta_5)$ -optimal design - or subset D-optimal design focusing on the subset  $\begin{pmatrix} \theta_4 \\ \theta_5 \end{pmatrix}$  - has the same design support points as in (2.8) but each with the weights  $\omega_1 = \omega_2 = \omega_3 = 1/3$ .

Further, so as to illustrate that the above singularity problem associated with subset designs can arise when other subset design criteria are used, let  $\alpha_1, \dots, \alpha_{p_2}$  be the  $p_2$  eigenvalues associated with the covariance matrix in (2.2). Then designs which minimize the average

$$\left\{ \frac{\alpha_1^k + \dots + \alpha_{p_2}^k}{p_2} \right\}^{1/k}, \quad k \in (0, \infty), \quad (2.9)$$

are called (locally) subset- $\Phi_k$  optimal (see Kiefer (1975)); then,  $D_s$ -,  $A_s$ - and  $E_s$ -optimal designs are obtained as special cases for  $k \rightarrow 0^+$ ,  $k = 1$ , and  $k \rightarrow \infty$ , respectively. For the similar compounds example, each of the locally  $D_s(\theta_4, \theta_5)$ -,  $A_s(\theta_4, \theta_5)$ - and  $E_s(\theta_4, \theta_5)$ -optimal designs has the support points given above,  $\mathbf{x}_1 =$

$\begin{pmatrix} 0.165 \\ 0.264 \end{pmatrix}$ ,  $\mathbf{x}_2 = \begin{pmatrix} 0.194 \\ 0 \end{pmatrix}$ , and  $\mathbf{x}_3 = \begin{pmatrix} 0 \\ 0.310 \end{pmatrix}$ , but the weights are  $\omega_1 = \omega_2 = \omega_3 = \frac{1}{3}$  for the  $D_s$  criterion,  $\omega_1 = 0.406$ ,  $\omega_2 = \omega_3 = 0.297$  for the  $A_s$  criterion, and  $\omega_1 = 1/2$ ,  $\omega_2 = \omega_3 = 1/4$  for the  $E_s$  criterion. Hence, we note that as  $k$  increases in (2.9), subset designs focusing on the subset  $\begin{pmatrix} \theta_A \\ \theta_s \end{pmatrix}$  tend to put more weight on the interior point  $\mathbf{x}_1$  and less weight where each factor (phenolic acid, drug, pollutant, etc.) is varied in isolation.

Before leaving this example, we provide one final generalization of the original  $D_s(\theta_2)$ -optimality criterion, and which connects the  $D_A$  criterion of (2.5) and Kiefer's subset- $\Phi_k$  criterion. Denoting the  $s$  eigenvalues of  $\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A}$  in (2.5) by  $\tau_1, \dots, \tau_s$ , designs which minimize the sum  $\{(\tau_1^k + \dots + \tau_s^k)/s\}^{1/k}$  for  $k \in (0, \infty)$ , called  $\Phi_{A,k}$ -optimal here, provide subset-optimal designs for the  $s$  functions,  $\phi = \phi(\theta)$ . Even in this most general of settings, singular subset designs do result from the subset design procedures.

These mathematical details notwithstanding, this example demonstrates that subset designs for a *whole class* of subset design criteria (subset- $\Phi_{A,k}$  optimality) can be inadequate for estimating the model parameters, and are thus impractical in many applied settings. Also, the previous example highlights that the Bayesian parameter-robust strategy does not provide enough design support points, so a more direct design approach is needed, one which provides some information regarding all model parameters.

### 3. A First-Order Design Criterion

We now provide an extension of the subset criterion so as to obtain designs which can be used to fit the assumed model. The problem with seeking designs to maximize the determinant-information in (2.1) is that these designs occasionally provide no information regarding the nuisance parameters  $\theta_1$  in the model function  $\eta(\mathbf{x}, \theta) = \eta(\mathbf{x}, \theta_1, \theta_2)$ . Alternatively, for  $0 \leq \beta \leq 1$ , the objective function

$$\psi_\theta(\xi, \beta) = \frac{1 - \beta}{p_1} \log |\mathbf{M}_{11}| + \frac{\beta}{p_2} \log |\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12}|$$

is a weighted sum of (1) the log information of  $\theta_1$  when  $\theta_2$  is assumed known,  $\mathbf{M}_{11}$ , and (2) the log information of  $\theta_2$  when  $\theta_1$  is unknown,  $\mathbf{M}_{22} - \mathbf{M}_{21} \mathbf{M}_{11}^+ \mathbf{M}_{12}$ . This objective function includes the  $D_s(\theta_2)$ -optimality and the D-optimality criterion functions as special cases for the values  $\beta = 1$  and  $\beta = \frac{p_2}{p}$ , respectively. Thus, for a given choice of  $\beta$  in the interval  $[p_2/p, 1]$  and initial parameter estimate  $\theta^o$ , we define designs which maximize  $\psi_{\theta^o}(\xi, \beta)$  to be locally  $D_\beta$ -optimal designs. This objective function is analogous to the weighted average used in Atkinson and Donev (1992) for model nesting in linear models, and was proposed in O'Brien (1993) for singular subset designs and used in Atkinson and Bogacka (1997) to allow for a degree of model uncertainty. Extensions of this criterion to include a

Bayesian prior  $p(\theta)$ , via use of the criterion function  $\int \psi_\theta(\xi, \beta) p(\theta) d\theta$ , and to generalize the  $D_A$  criterion and Kiefer's subset- $\Phi_k$  criterion discussed Example 3 above are straightforward. Finally, the variance function (White (1973), Atkinson and Donev (1992)) corresponding to this objective function can be used to ensure optimality of a candidate  $D_\beta$ -optimal design by invoking the relevant version of the General Equivalence Theorem.

The design strategy suggested here involves finding an array of  $D_\beta$ -optimal designs for various choices of  $\beta$  (typically near 0.90) and then selecting a final design based on a reasonably small value of the subset information loss (SIL). Since the  $D_s(\theta_2)$ -optimal design,  $\xi_s^*$ , optimizes  $|\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^+\mathbf{M}_{12}| = |\mathbf{M}^{22}|$ , the subset information loss for an arbitrary design  $\xi$  relative to  $\xi_s^*$  is defined here as

$$\left( 1 - \left[ \frac{|\mathbf{M}^{22}(\xi)|}{|\mathbf{M}^{22}(\xi_s^*)|} \right]^{1/p_2} \right) \times 100\%$$

These ideas are illustrated with the examples introduced in the previous Section.

Table 2:  $D_\beta$ -optimal designs and subset information loss for the Intermediate Product (IP2) model.

$\beta$	Optimal design	SIL(%)
1.00	$\begin{Bmatrix} 6.76 \\ 1 \end{Bmatrix}$	0.0
0.99	$\begin{Bmatrix} 1.37 & 6.73 \\ 0.01 & 0.99 \end{Bmatrix}$	1.0
0.95	$\begin{Bmatrix} 1.40 & 6.61 \\ 0.06 & 0.94 \end{Bmatrix}$	4.9
0.90	$\begin{Bmatrix} 1.43 & 6.49 \\ 0.11 & 0.89 \end{Bmatrix}$	9.8
0.85	$\begin{Bmatrix} 1.45 & 6.39 \\ 0.16 & 0.84 \end{Bmatrix}$	14.6
0.50	$\begin{Bmatrix} 1.59 & 5.89 \\ 0.5 & 0.5 \end{Bmatrix}$	48.8

**Example one continued.** For the IP2 model function and initial parameter choices given in Hill and Hunter (1974) and used above ( $\theta_1^o = 0.2$  and  $\theta_2^o = 0.7$ ), locally  $D_\beta$ -optimal designs and subset information loss (SIL) values are given in Table 2 for  $\beta = 1, 0.99, 0.95, 0.90, 0.85$  and  $0.50$ . For this example,  $\beta = 1$  corresponds to the (local)  $D_s(\theta_2)$ -optimal design and  $\beta = 0.50$  corresponds to the full D-optimal design, and these designs are provided only for the purpose of

comparison. Our recommendation here would be to use the compromise design corresponding to  $\beta = 0.90$  since the information loss associated with this design is only 9.8% and since this design's having two support points ensures that the assumed model function can be fit to the data. Note too that this design is especially useful if the final sample size can be chosen to be a multiple of 9 (since the weight of  $x = 1.43$  is approximately  $1/9$ ), although other rounding procedures for continuous-weight designs are discussed at length in Pukelsheim and Rieder (1992). Our suggestion, then, might be the use of a design with 2 runs at  $x = 1.43$  and 16 runs at  $x = 6.49$ , and although minimal (9.8%) information is lost regarding the subset parameter of interest, this design can indeed be used to fit the IP2 model function.

Whenever uncertainty exist concerning the initial parameter guesses, the Bayesian  $D_\beta$ -optimal design for  $\beta = 0.90$  should be obtained for this model, for example, with  $\theta_1$  and  $\theta_2$  independently and uniformly distributed over the region  $[0.10, 0.30] \times [0.50, 0.90]$ . This Bayesian design associates the weights 0.21 and 0.79 with the support points 1.62 and 6.57, respectively, and results in a moderate shift from the local design proposed in the previous paragraph. Further, since the subset information loss of this Bayesian design relative to the Bayesian design for  $\beta = 1.00$  is only 2.1%, this Bayesian  $D_{0.90}$ -optimal design is indeed recommended in this setting. Therefore, our final suggestion might be to take 2 measurements at  $x = 1.62$  and 18 measurements at  $x = 6.57$ ; this final design differs slightly from the local  $D_\beta$ -optimal design given above.

**Example two continued.** Whereas the (singular) subset design for the quadratic model with turning point  $\phi^o = 0.50$  associates the weight  $\omega = 1/2$  with each of the points  $x = 0$  and  $x = 1.00$ , the  $D_\beta$ -optimal design,

$$\left\{ \begin{array}{ccc} 0 & 1/2 & 1 \\ (1 + \beta)/4 & (1 - \beta)/2 & (1 + \beta)/4 \end{array} \right\}$$

is nonsingular for  $\beta \in (\frac{1}{3}, 1)$  since it has three support points. It turns out also that this  $D_\beta$ -optimal designs has a subset information loss of  $\frac{(1-\beta)}{2} \times 100\%$ , so for example the  $D_{0.80}$ -optimal design,  $\left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 0.45 & 0.10 & 0.45 \end{array} \right\}$ , has a SIL of only 10%, and is thus recommended here. This latter design would take for example 4 runs at  $x = 1/2$  and 18 runs at each of  $x = 0$  and  $x = 1$  in a study involving  $n = 40$  total runs.

**Example three continued.** Table 3 gives the locally  $D_\beta(\theta_5)$ -optimal designs for the LL2 model function,  $\eta = 1/[1 + (d/\theta_2)^{\theta_3}]$  with effective dose  $d = a + \theta_4 b + \theta_5 \sqrt{\theta_4 a b}$  for  $\beta = 1.00, 0.90,$  and  $0.25$ , with the corresponding effective

concentrations; note that  $\beta = 1.00$  corresponds to the singular  $D_s$ -optimal design given in (2.8) and  $\beta = 0.25$  corresponds to the full D-optimal design for which equal information is obtained regarding each of the model parameters. We use the (simplified) LL2 model function here (with upper asymptote set to unity) instead of the LL3 model function used in the previous Section since designs for the LL3 model only require that additional runs (20% of the total) be placed at the origin. We observe that  $D_\beta(\theta_5)$ -optimal designs for this model have three support points on each of two isoboles, or curves of equal dose, and as  $\beta \rightarrow 1^-$ , these two isoboles become coincidental, whence the singularity of the subset design. This can be observed in Table 3 since as  $\beta$  increases from 0.25 to 0.90, the isoboles move together from effective doses 0.148 and 0.252 to effective doses 0.176 and 0.213; finally, as  $\beta$  becomes unity, these isoboles become identical, and insufficient information is provided to fit the model function. In situations where the primary purpose of the experiment is to detect synergy or antagonism, our recommendation is to use the  $D_{0.90}(\theta_5)$ -optimal design since this design is efficient for estimating the synergy parameter ( $\theta_5$ ) and also provides the opportunity to fit the model to the data. This latter design is well-approximated by associating weights  $1/8$  with each of the four points on the vertical and horizontal axes (corresponding to the phenolic acids in isolation) and weights  $2/8$  with each of the two center points, (0.150,0.240) and (0.181,0.290); again the information loss is minimal (less than 10%) and this design can be used to fit the model.

Table 3:  $D_\beta$ -optimal designs for the log-logistic (LL2) synergy model.

$\beta$	Optimal design	Effective dose
1.00	$\left\{ \begin{matrix} \begin{pmatrix} 0.165 \\ 0.264 \\ 1/2 \end{pmatrix} & \begin{pmatrix} 0.194 \\ 0 \\ 1/4 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.310 \\ 1/4 \end{pmatrix} \end{matrix} \right\}$	$d = 0.194$ only
0.90	$\left\{ \begin{matrix} \begin{pmatrix} 0.150 \\ 0.240 \\ 0.246 \end{pmatrix} & \begin{pmatrix} 0.181 \\ 0.290 \\ 0.246 \end{pmatrix} & \begin{pmatrix} 0.176 \\ 0 \\ 0.127 \end{pmatrix} & \begin{pmatrix} 0.213 \\ 0 \\ 0.127 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.281 \\ 0.127 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.340 \\ 0.127 \end{pmatrix} \end{matrix} \right\}$	$d = 0.176$ and $0.213$
0.25	$\left\{ \begin{matrix} \begin{pmatrix} 0.127 \\ 0.201 \\ 1/6 \end{pmatrix} & \begin{pmatrix} 0.214 \\ 0.345 \\ 1/6 \end{pmatrix} & \begin{pmatrix} 0.148 \\ 0 \\ 1/6 \end{pmatrix} & \begin{pmatrix} 0.253 \\ 0 \\ 1/6 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.238 \\ 1/6 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.405 \\ 1/6 \end{pmatrix} \end{matrix} \right\}$	$d = 0.1486$ and $0.252$

#### 4. A Second-Order Design Criterion

In this section, we briefly discuss an extension of the above results to account for model nonlinearity; readers unfamiliar with the basic concepts of curvature are advised to consult the non-technical overviews given in Ratkowsky (1983) and O'Brien and Wang (1996). Note that the  $D_\beta$ -optimal designs discussed above

equivalently minimize the objective function

$$\gamma_1(\xi, \beta, \theta) = \frac{1 - \beta}{p_1} \log |\mathbf{M}_{11}^{-1}| + \frac{\beta}{p_2} \log |(\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^+\mathbf{M}_{12})^{-1}|,$$

where  $\mathbf{M}_{11}^{-1}$  is proportional to the first-order covariance matrix corresponding to  $\theta_1$  in the model  $\eta(\theta_1, \theta_2)$  when  $\theta_2$  is known and  $(\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^+\mathbf{M}_{12})^{-1}$  is proportional to the first-order covariance matrix of  $\theta_2$  in the same model when  $\theta_1$  is unknown. In contrast, O'Brien and Rawlings (1996) explores second-order MSE designs based on the second-order covariance matrix,  $\mathbf{S} = \sigma^2\mathbf{K}\mathbf{K}^T + \sigma^4\mathbf{K}(\mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3)\mathbf{K}^T$ , given in Clarke (1980), and the second-order bias vector,  $\mathbf{b} = -\frac{1}{2}\sigma^2\mathbf{K}\sum\mathbf{a}_{kk}^T$ , given in Box (1971a); further details of these expressions are given in Seber and Wild (1989). For the partitions of  $\mathbf{S}$  and  $\mathbf{b}$ ,  $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$  and  $\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$  corresponding to  $\theta_1$  and  $\theta_2$  respectively, a natural extension of the above  $\gamma_1(\xi, \beta, \theta)$  objective function is

$$\gamma_2(\xi, \beta, \theta) = \frac{1 - \beta}{p_1} \log \frac{|\mathbf{S} + \mathbf{b}\mathbf{b}^T|}{|\mathbf{S}_{22} + \mathbf{b}_2\mathbf{b}_2^T|} + \frac{\beta}{p_2} \log |\mathbf{S}_{22} + \mathbf{b}_2\mathbf{b}_2^T|,$$

and we define designs that minimize  $\gamma_2(\xi, \beta, \theta)$  to be locally  $D_{2\beta}$ -optimal and those that minimize  $\int \gamma_2(\xi, \beta, \theta) p(\theta) d\theta$  to be Bayesian  $D_{2\beta}$ -optimal.

In contrast with first-order designs,  $D_{2\beta}$ -optimal designs depend on the noise level  $\sigma^* = \sigma/\sqrt{r}$  where  $\sigma$  is the model standard deviation and  $r$  ( $r \geq 1$ ) is the number of replications used of the design. Indeed the connection between these criteria is that  $D_{2\beta}$ -optimal designs for  $\sigma = 0$  are equivalent to D-optimal designs. Thus, although D-optimal designs for situations where  $\sigma^o = 0.05$  and for those where  $\sigma^o = 0.25$  are identical, the corresponding  $D_{2\beta}$ -optimal designs may differ appreciably. More precisely, the difference between  $D_{\beta}$ - and  $D_{2\beta}$ -optimal designs is an increasing function of the noise level and the degree of nonlinearity (or curvature) associated with the model function and design. One measure of curvature is the root mean square parameter-effects curvature measure ( $\text{RMS}_\theta$ ) introduced in Beale (1960) and highlighted in Bates and Watts (1980, 1988), and is used in the next example to highlight the lower curvature and bias associated with a locally  $D_{2\beta}$ -optimal design; those unfamiliar with these nonlinearity measures can nonetheless appreciate the corresponding curvature reduction associated with the  $D_{2\beta}$ -optimal design.

**Example one continued.** Table 4 gives the optimal designs, subset information loss, squared length of the bias vector and  $\text{RMS}_\theta$  curvature associated with the locally and Bayesian  $D_{0.90}$ -optimal designs given above and with the locally  $D_{2\beta}$ -optimal design for  $\beta = 0.90$  using  $\sigma^o = 0.05$ . Thus, this example highlights

the high parameter-effects curvature ( $\text{RMS}_\theta$  curvature) associated with the local  $D_{0.90}$ -optimal design and the substantial shift in moving from the first- to the second-order local design. That the local second-order designs is preferable here to the first-order one is witnessed by noting that the associated bias vector is 85% shorter and the  $\text{RMS}_\theta$  curvature is 52% less. It is interesting to point out that for this example, the Bayesian first-order design has relatively low bias and curvature measures as well. Our final recommendation for a study involving  $n = 18$  runs would be to place 6 runs at  $x = 1.75$  and 12 runs at  $x = 6.20$ .

Table 4: First- and second-order local and Bayesian optimal designs, subset information loss, squared length of bias vector, and  $\text{RMS}_\theta$  curvature for the Intermediate Product (IP2) model.

Design Criterion	Optimal Design	SIL (%)	$\mathbf{b}^T \mathbf{b}$	$\text{RMS}_\theta$ curvature
local $D_{0.90}$ -optimality	$\begin{pmatrix} 1.43 & 6.49 \\ 0.11 & 0.89 \end{pmatrix}$	9.8	0.14	1.66
Bayesian $D_{0.90}$ -optimality	$\begin{pmatrix} 1.62 & 6.57 \\ 0.21 & 0.79 \end{pmatrix}$	2.1	0.04	0.90
local $D_{2\beta}$ -optimality	$\begin{pmatrix} 1.75 & 6.20 \\ 0.32 & 0.68 \end{pmatrix}$	4.1	0.02	0.79

## 5. Discussion

Although the subset design procedure give in Hill and Hunter (1974) focusses on estimating the subset  $\theta_2$  more precisely than the nuisance subset  $\theta_1$ , numerous important examples demonstrate that the corresponding design fails to allow for estimation of all the model parameters, and in some examples (as in the AUC parameter for the IP2 model) cannot even be used to estimate the model subset parameters. In contrast, the first- and second-order design strategies presented here more directly address the issue of providing a great deal of information about the parameters of interest, yet allow for estimation of all model parameters and result in only a modest loss in subset information. The Bayesian extension is often essential since Bayesian designs have been found to be quite robust to misspecification of the initial parameter estimate and prior distribution; details are provided in Arumugham (1992). Also, as mentioned above, Bayesian optimal designs can often have “extra” (greater than the number of parameters) support points which in turn can be used to check the adequacy of the assumed model function should this be an important issue to the researcher. As has been argued in Hamilton and Watts (1985), nonlinear design criteria should take account of curvature

as well as provide for efficient estimation of the model parameters, and the example discussed in the previous section highlights the substantial difference between first- and second-order designs. Thus, in situations where interest focusses on a subset of the parameters of a nonlinear regression model, our recommendation is to obtain the Bayesian  $D_\beta$ -optimal design for  $\beta$  near 0.90 and to calculate an associated curvature measure (such as RMS or marginal curvature). If this latter curvature measure is high, the Bayesian  $D_{2\beta}$ -optimal design is recommended instead. Interested readers are encouraged to download SAS programs to find (and verify) these designs from the author's webpage, [www.math.luc.edu/~tobrien/](http://www.math.luc.edu/~tobrien/).

## Acknowledgements

This work was funded by grants from The Dow Chemical Company (U.S.), the Deutscher Akademischer Austauschdienst (Germany) and the U.S. National Science Foundation's International Postdoctoral Fellows Program, and by financial support from the Katholieke Universiteit Leuven (KUL), and was completed while the author was on sabbatical visit at KUL in Belgium. The author is grateful for important comments from the Editor and Reviewers.

## References

- Arumugham, T. (1992). *Curvature and Experimental Design for the Weibull Model*, Ph.D. thesis, North Carolina State University, Raleigh, North Carolina.
- Atkinson, A. C. and Bogacka, B. (1997) Compound  $D$ - and  $D_S$ -optimum designs for determining the order of a chemical reaction, *Technometrics* **39**, 347-356.
- Atkinson, A. C., Chaloner, K., Herzberg, A. M. and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model, *Biometrics* **49**, 325-337.
- Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Clarendon Press.
- Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion), *Journal of the Royal Statistical Society, Series B* **42**, 1-25.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*. Wiley.
- Beale, E. M. L. (1960). Confidence regions in non-linear estimation (with discussion), *Journal of the Royal Statistical Society, Series B* **22**, 41-88.
- Box, M. J. (1971a). Bias in nonlinear estimation (with discussion), *Journal of the Royal Statistical Society, Series B* **33**, 171-201.
- Box, M. J. (1971b). An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model, *Biometrika* **58**, 149-153.



- Chaloner, K. (1989). Bayesian design for estimating the turning point of a quadratic regression, *Communications in Statistics: Theory and Methods* **18**, 1385-1400.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments, *Journal of Statistical Planning and Inference* **21**, 191-208.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters, *Annals of Mathematical Statistics* **24**, 586-602.
- Clarke, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model, *Journal of the Royal Statistical Society, Series B* **42**, 227-237.
- Cook, R. D. and Weisberg, S. (1990). Confidence curves in nonlinear regression, *Journal of the American Statistical Association* **85**, 544-551.
- Finney, D. J. (1952). *Probit Analysis*. Cambridge University Press.
- Gaffke, N. (1985). Directional derivatives of optimality criteria at singular matrices in convex design theory, *Statistics* **16**, 373-388.
- Gerig, T. M., Blum, U. and Meier, K. (1989). Statistical analysis of the joint inhibitory action of similar compounds, *Journal of Chemical Ecology* **15**, 2403-2412.
- Hamilton, D. C. and Watts, D. G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models, *Technometrics* **27**, 241-250.
- Hill, W. J. and Hunter, W. G. (1974). Design of experiments for subsets of parameters, *Technometrics* **16**, 425-434.
- Hunter, W. G., Hill, W. J. and Henson, T. L. (1969). Designing experiments for precise estimation of all or some of the constants in a mechanistic model, *Canadian Journal of Chemical Engineering* **47**, 76-80.
- Kiefer, J. (1961). Optimum designs in regression problems II, *Annals of Mathematical Statistics* **32**, 298-325.
- Kiefer, J. (1975). Optimal design: variation in structure and performance under change of criterion, *Biometrika* **62**, 277-288.
- Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems, *Canadian Journal of Mathematics* **12**, 363-366.
- Läuter, E. (1974a). Experimental design in a class of models, *Mathematische Operationsforschung und Statistik* **5**, 379-398.
- Läuter, E. (1974b). A method of designing experiments for nonlinear models, *Mathematische Operationsforschung und Statistik* **5**, 625-636.
- Lindsey, J. K. (2001). *Nonlinear Models in Medical Statistics*. Oxford University Press.
- Machado, S. G. and Robinson, G. A. (1994). A direct, general approach based on isobolograms for assessing the joint action of drugs in pre-clinical experiments, *Statistics in Medicine* **13**, 2289-2309.
- O'Brien, T. E. (1992). A note on quadratic designs for nonlinear regression models, *Biometrika* **79**, 847-849.

- O'Brien, T. E. (1993). *Design Strategies for Nonlinear Regression Models*, Ph.D. Dissertation, North Carolina State University, Raleigh, North Carolina.
- O'Brien, T. E. (2004). Modelling and design to detect interaction of insecticides, herbicides and other similar compounds. In *Proceedings of the 15th. Conference on Applied Statistics in Agriculture* (Edited by D. Johnson and G. Milliken), 303-321. Kansas State University Press.
- O'Brien, T. E. and Funk, G. M. (2003). A gentle introduction to optimal design for regression models, *The American Statistician* **57**, 265-267.
- O'Brien, T. E. and Rawlings, J. O. (1996). A non-sequential design procedure for parameter estimation and model discrimination in nonlinear regression models, *Journal of Statistical Planning and Inference* **55**, 77-93.
- O'Brien, T. E. and Wang, Y. (1996). Using SAS software to assess and adjust for nonlinearity in nonlinear regression models. *SUGI 21 Proc.*, SAS Institute, 1274-1283.
- Pázman, A. (1980). Singular experimental designs (standard and Hilbert-space approaches), *Mathematische Operationsforschung und Statistik* **11**, 137-149.
- Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*, Wiley.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley.
- Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs, *Biometrika* **79**, 763-770.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker.
- Seber, G. A. F. and Wild, C.J. (1989). *Nonlinear Regression*. Wiley.
- Sibson, R. (1974).  $D_A$ -optimality and duality. In *Progress in Statistics, vol. II* (Edited by Gani, J., Sarkadi, K and Vincze, I), *Proc. 9th European Meeting of Statisticians*, Budapest, 1972, Amsterdam: North-Holland.
- Silvey, S. D. (1978). Optimum design measures with singular information matrices, *Biometrika* **65**, 553-559.
- Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall.
- Tallarida, R. J. (2000). *Drug Synergism and Dose-Effect Data Analysis*. Chapman and Hall.
- van Ewijk, P. H. and Hoekstra, J. A. (1994). Curvature measures and confidence intervals for the linear logistic model, *Applied Statistics* **43**, 477-487.
- Vuchkov, I. N. (1977). A ridge-type procedure for design of experiments, *Biometrika* **64**, 147-150.
- White, L. V. (1973). An extension of the general equivalence theorem to nonlinear models, *Biometrika* **60**, 345-348.

---

Whittle, P. (1973). Some general points in the theory of optimal experimental design,  
*Journal of the Royal Statistical Society, Series B* **35**, 123-130.

Received September 26, 2003; accepted March 22, 2004.

Timothy E. O'Brien  
Department of Mathematics & Statistics  
Loyola University Chicago  
Chicago, Illinois 60626 USA  
tobrien@math.luc.edu