

# involve

a journal of mathematics

Curvature measures for nonlinear regression models using  
continuous designs with applications to optimal experimental  
design

Timothy O'Brien, Somsri Jamroenpinyo and Chinnaphong Bumrungrsup

 mathematical sciences publishers

# Curvature measures for nonlinear regression models using continuous designs with applications to optimal experimental design

Timothy O'Brien, Somsri Jamroenpinyo and Chinnaphong Bumrungrsup

(Communicated by Kenneth S. Berenhaut)

We present and illustrate the methodology to calculate curvature measures for continuous designs, and extend design criteria to incorporate continuous designs. These design algorithms include quadratic design procedures, a subset design criterion, a second-order mean-square error design criterion, and a marginal curvature design methodology. A discussion of confidence intervals is also provided for continuous designs.

## 1. Introduction

When researchers have a given nonlinear regression model in mind to describe a specific process, they typically seek an experimental design to efficiently estimate the  $p$  model parameters with only negligible curvature. An additional requirement is usually that the design provide the researcher with the ability to test for the adequacy of the assumed model. Measures of curvature and nonlinearity are developed in [Beale 1960; Bates and Watts 1980; Ratkowsky 1983; Clarke 1987]; additional results and references are given in [Bates and Watts 1988; Seber and Wild 1989; Haines et al. 2004]. These measures assume that the chosen design has discrete weights, meaning that it is assumed that a sample size is fixed *a priori*, and that some fraction of these points is chosen at each of the design support points.

This paper argues that so-called continuous designs should instead be sought, and provides and demonstrates the means of calculating curvature measures for these designs. A by-product of this result is an extension to allow for continuous designs of the second-order mean-square error (MSE) design criterion in [Clarke and Haines 1995], the marginal curvature design strategy in [O'Brien 2006], and the second-order volume design criterion in [O'Brien 1992].

---

*MSC2000:* primary 53A45, 62J02; secondary 62K05.

*Keywords:* design measures, Fieller–Creasy problem, marginal curvature, model misspecification, nonlinearity, parameter subsets, robust designs.

Furthermore, most optimal design procedures provide designs with only  $p$  support points regardless of the sample size. Depending on the dispersion of the assumed prior distribution, one exception is the Bayesian procedures introduced in [Chaloner and Larntz 1989; Atkinson and Donev 1992]. O'Brien [1992] shows how discrete  $(p+1)$ -point designs with reduced curvature could be obtained. In this article, we further extend this volume design strategy to incorporate continuous and larger discrete designs; these designs are recommended since they can also be used to test for lack-of-fit of the assumed model.

## 2. Notation and terminology

A general  $n$ -point design (measure)  $\zeta$  can be written as

$$\zeta = \left\{ \begin{array}{l} s_1, \dots, s_r \\ \lambda_1, \dots, \lambda_r \end{array} \right\}.$$

Here, the  $r \leq n$  support points,  $s_1, s_2, \dots, s_r$  are elements of the design space  $X$ , and the associated design weights — the  $\lambda$ 's — are nonnegative real numbers which sum to one. When the support points are chosen from a discrete grid, we call this a discrete-point design; otherwise, it is a continuous-point design. Lattice and cyclic designs provide examples of the former type, whereas designs for dose-response curves or response surfaces are usually of the continuous variety. In addition, whenever the design weights are rational numbers proportional to  $1/n$ , we call the design a rational-mass (or rational-weight) design; otherwise, it is a continuous-mass design. Thus, even though  $\zeta$  has  $n$  *design* points (which are not necessarily distinct), it may have only  $r$  distinct *support* points. Regardless of whether  $\zeta$  is a continuous- or rational-mass design,  $n\lambda_i$  is to be thought of as the *number of observations* taken at the experimental level  $s_i$ .

In regression settings, we advocate here that continuous point-continuous mass designs be obtained as a general rule-of-thumb, at least as a starting point; these designs can later be rounded to practical designs using the methodology given in [Pukelsheim and Rieder 1992]. For simplicity, we refer to these designs as *continuous designs* for the remainder of this paper. We encourage the use of continuous designs for two reasons. First, for some design criteria, optimality of the resulting design can be verified using a variation of the general equivalence theorem. This important result, first developed for linear models in [Kiefer and Wolfowitz 1960], and extended to nonlinear models in [White 1973], is illustrated for nonlinear models in [Haines 1992; O'Brien and Funk 2003]; important extensions to these original results are given in [Pukelsheim 1993; Atkinson and Haines 1996; Dette and O'Brien 1999]. This verification that the derived design is indeed optimal can be achieved by obtaining a graph of the corresponding variance function and noting whether or not this graph lies below a certain horizontal line. Since the

theorem is valid only for continuous designs (with continuous weights), verification of optimality of other designs is not possible in general.

Our second reason for preferring continuous designs is to expedite our search for an optimal design strategy in a given setting. For example, using the model function  $\eta(x, \theta) = 1 - e^{-\theta x}$  and a specific prior distribution for  $\theta$ , Atkinson [1988] provides discrete Bayesian  $D$ -optimal designs for sample sizes of  $n = 1, 2$ , and  $3$ . The reported designs are  $\{12\}$ ,  $\{7, 27\}$ , and  $\{8, 8, 52\}$ , respectively. Instead of continuing this trend of finding designs for larger sample sizes, we simply note that the continuous Bayesian  $D$ -optimal design associates the weight  $\lambda_1 = 0.890$  with the point  $s_1 = 9.05$  and the weight  $\lambda_2 = 0.110$  with the point  $s_2 = 120.1$ . Thus, as we continue the sequence above up to the sample size  $n = 9$ , the first eight of these points would be placed nearer and nearer to  $s_1 = 9.05$ , whereas the last point approaches  $s_2 = 120.1$ . Thus, the strategy of first seeking continuous optimal designs gives us a clearer picture at the outset of what is required to achieve optimality.

### 3. Curvature measures for continuous designs

Under the usual assumption of uncorrelated Gaussian errors with zero mean and constant variance  $\sigma^2$  (without loss of generality taken to equal one), the Fisher information *per observation* is given by

$$\mathbf{M}(\xi, \theta) = \sum_{k=1}^r \lambda_k \frac{\partial \eta(s_k)}{\partial \theta} \frac{\partial \eta(s_k)}{\partial \theta^T} = \mathbf{V}_r^T \mathbf{\Lambda} \mathbf{V}_r,$$

where  $\mathbf{V}_r$  is the  $r \times p$  Jacobian of  $\eta$  associated with  $\xi$  with  $k$ th row equal to  $\partial \eta(s_k) / \partial \theta^T$ , and  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_r\}$ . It follows that the *total* information associated with an  $n$ -point design  $\xi$  is given by

$$n\mathbf{M}(\xi, \theta) = \sum_{k=1}^r n\lambda_k \frac{\partial \eta(s_k)}{\partial \theta} \frac{\partial \eta(s_k)}{\partial \theta^T} = \sum_{k=1}^r \frac{\partial \eta_w(s_k)}{\partial \theta} \frac{\partial \eta_w(s_k)}{\partial \theta^T} = \mathbf{V}_w^T \mathbf{V}_w.$$

In this expression,

$$\eta_w(s_k) = \sqrt{n\lambda_k} \eta(s_k) \tag{1}$$

is the weighted model function where the model weights correspond to the “number of observations” taken at the point  $s_k$ , and  $\mathbf{V}_w$  is the  $r \times p$  Jacobian of  $\eta_w$  with  $k$ th row equal to  $\partial \eta_w(s_k) / \partial \theta^T$ . Thus, for continuous designs with  $n$  observations and  $r$  support points, and for second-order curvature, we point out that variance and volume measures should be based on

$$\mathbf{V}_w = \frac{\partial \eta_w(s, \theta)}{\partial \theta} = n^{1/2} \mathbf{\Lambda}^{1/2} \mathbf{V}_r, \quad \mathbf{W}_w = \frac{\partial^2 \eta_w(s, \theta)}{\partial \theta \partial \theta^T} = n^{1/2} [\mathbf{\Lambda}^{1/2}] [\mathbf{W}_r]. \tag{2}$$

Here  $\Lambda^{1/2} = \text{diag}\{\lambda_1^{1/2}, \dots, \lambda_r^{1/2}\}$  and square bracket multiplication of arrays is defined in [Seber and Wild 1989, Appendix B]. The  $r \times p$  matrix  $V_w$  plays the role of the Jacobian matrix and the  $r \times p \times p$  array  $W_w$  that of the second derivative array; since the marginal curvature measures in [Clarke 1987] require third derivatives, the  $r \times p \times p \times p$  array of third derivatives is defined analogously.

**Example 1.** Consider the Fieller–Creasy problem highlighted in [Cook and Witmer 1985; Clarke 1987; Seber and Wild 1989; Haines et al. 2004], in which the relevant nonlinear model function is  $\eta(x, \theta_1, \theta_2) = \theta_1 x + \theta_1 \theta_2 (1 - x)$ , with  $x = 1$  for group  $A$  and  $x = 0$  for group  $B$ . Interest lies in the parameter  $\theta_2$ , which is the ratio of the group  $B$  mean over the group  $A$  mean. Here, we take  $n_A = n\lambda$  and  $n_B = n(1 - \lambda)$  for  $\lambda$  between 0 and 1;  $\lambda$  is thus the percentage of the experimental units assigned to group  $A$ . So as to adjust Wald confidence intervals to bring them more in line with likelihood intervals, [Clarke 1987] introduces marginal curvature measures  $\Gamma$  and  $\beta$ , where  $\Gamma$  assesses skewness of the interval and  $\beta$  (along with  $\Gamma$ ) quantifies excessive kurtosis. Here,  $\Gamma = \beta = 0$  is an indication of no nonlinearity, or exact coincidence between the two types of confidence intervals, as this is indeed the case for linear models. For the equal-weight case ( $\lambda = 1/2$ ), these  $\Gamma$  and  $\beta$  expressions for  $\theta_2$  are given in [Haines et al. 2004, page 565]; extensions to the general case are given by the expressions

$$\Gamma = \frac{-2\theta_2\phi^{1/2}}{\theta_1\sqrt{n_A(1+\phi\theta_2^2)}} \quad \text{and} \quad \beta = \frac{1+2\phi\theta_2^2}{n_A\theta_1^2(1+\phi\theta_2^2)},$$

where  $\phi = n_B/n_A = (1 - \lambda)/\lambda$ . For example, when  $n = n_A + n_B = 20$ ,  $\theta_1 = 0.10$  and  $\theta_2 = 1$ , the equal-weight case results in  $\Gamma = -4.4721$  and  $\beta = 15.0$ . In contrast, when  $n_A = 19$  and  $n_B = 1$  (so  $\lambda = 0.95$ ),  $\Gamma = -1.0260$  and  $\beta = 5.0263$ . This reduction in both  $\Gamma$  and  $\beta$  means that the Wald and likelihood intervals for  $\theta_2$  coincide to a greater degree in the latter situation than in the former (equal-weight) one.

Of course, both  $\Gamma$  and  $\beta$  for  $\theta_1$  are zero here since this parameter enters the model function a linear manner, so the two intervals will coincide for this parameter.

#### 4. Applications to optimal design

An important benefit of extending the definitions of curvature measures to incorporate continuous designs is that more efficient designs can often then be obtained. Important examples include finding continuous designs for the  $Q$ -optimality, subset and MSE criteria, but first we examine discrete  $Q$ -optimal designs.

**4.1. Discrete  $Q$ -optimal designs.** For nonlinear models, we have from [Hamilton and Watts 1985] the second-order volume

$$v = c|V_0^T V_0|^{-1/2}|D_0|^{-1/2}\{1 + k^2 \text{tr}(D_0^{-1}M_0)\}, \tag{3}$$

as an approximation to the likelihood-based confidence region for the parameter vector  $\theta$ . In this expression,  $c$  and  $k$  are constants with respect to the chosen design, matrices are evaluated at some initial guess  $\theta_0$  for  $\theta$ . Also,  $V_0 = V(\theta_0)$  is the  $n \times p$  Jacobian matrix whereas  $M_0$  and  $D_0 = I_p - B_0$  are associated with parameter-effects and intrinsic curvature, respectively. Whereas Hamilton and Watts [1985] obtain only designs with  $n = p$ , O'Brien [1992] shows how designs with  $p + 1$  support points can be obtained to minimize the volume in (3), designs which are called discrete  $Q$ -optimal designs.

Specifically, the  $QR$  decomposition of  $V_0$  is

$$V_0 = Q_0 R_0 = [U_0 | N_0] R_0 = U_0 L_0^{-1},$$

so that the columns of  $U_0$  form an orthonormal basis for the tangent plane to  $\eta(\theta)$  at  $\theta = \theta_0$ , and the columns of  $N_0$  form an orthonormal basis for the space orthogonal to this tangent plane. Since the residual vector is always orthogonal to the tangent plane at the maximum likelihood estimate, we can write

$$\epsilon_0 = N_0 \alpha_0, \tag{4}$$

where  $\alpha_0$  is a vector of dimension  $(n - p) \times 1$ . Considering only the case  $n = p + 1$ , O'Brien [1992] chooses the scalar  $\alpha_0$  equal to  $\sigma_0$  since in general the expected squared length

$$E(\epsilon_0^T \epsilon_0) \approx E(\epsilon^T \epsilon) - E[(\theta - \theta_0)^T R_0^T R_0 (\theta - \theta_0)] \approx n\sigma^2 - p\sigma^2. \tag{5}$$

We extend this result here for discrete designs with  $n = p + s$  and  $s \geq 2$ .

Whenever  $s = 2$  (i.e.,  $n = p + 2$ ), we can write  $\alpha_0 = \sqrt{2}\sigma_0 \begin{pmatrix} \sin \phi \\ \cos \phi \end{pmatrix}$ , for some  $\phi$  between 0 and  $\pi/2$ . With this choice for  $\alpha_0$ , we keep its expected length identical to that of  $\epsilon_0$  by (4) and (5) since  $N_0$  is an orthonormal matrix. With this choice of  $\epsilon_0$ , we can calculate the volume in (3)—subject to knowledge of  $\phi$ . We thus define a  $(p + 2)$ -point to be locally  $Q$ -optimal if it minimizes the expected volume

$$E[v(\phi)] = \int_0^{\pi/2} v(\phi) d\phi. \tag{6}$$

By extension, when  $n = p + s$  for  $s \geq 3$ , we use hyperspherical coordinates for the  $s$ -vector  $\alpha_0$ :

$$\alpha_0 = \sqrt{s}\sigma_0 \begin{pmatrix} \sin \phi_1 & \sin \phi_2 & \sin \phi_3 & \cdots & \sin \phi_{s-1} \\ \cos \phi_1 & \sin \phi_2 & \sin \phi_3 & \cdots & \sin \phi_{s-1} \\ & \cos \phi_2 & \sin \phi_3 & \cdots & \sin \phi_{s-1} \\ & & & & \vdots \\ & & & & \cos \phi_{s-1} \end{pmatrix}, \tag{7}$$

for  $0 \leq \varphi_1, \varphi_2, \dots, \varphi_{s-1} \leq \pi/2$ . In this case, a discrete locally  $Q$ -optimal design minimizes the expected volume

$$\int_0^{\pi/2} \cdots \int_0^{\pi/2} v(\phi_1, \dots, \phi_{s-1}) d\phi_1 \cdots d\phi_{s-1},$$

for  $v = v(\varphi_1, \varphi_2, \dots, \varphi_{s-1})$  given in (3), with  $\epsilon_0$  in (4) and  $\alpha_0$  defined in (7). We illustrate this design strategy with the following example.

**Example 2.** To illustrate, consider the two-parameter intermediate product (IP2) model function

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} \{ \exp(-\theta_2 x) - \exp(-\theta_1 x) \}, \tag{8}$$

for  $\theta_1 > \theta_2 > 0$  and  $x > 0$ . This model function is used extensively in chemical kinetics and pharmacology. Using the initial parameter values  $\theta_0^T = (0.7, 0.2)$  and  $\sigma_0 = 0.1$ , the 2-point discrete locally  $Q$ -optimal design takes one observation at each of  $s_1 = 1.04$  and  $s_2 = 5.59$ , the 3-point discrete locally  $Q$ -optimal design takes one observation at each of  $s_1 = 1.02$ ,  $s_2 = 4.72$ , and  $s_3 = 6.81$ , and the 4-point discrete locally  $Q$ -optimal design takes one observation at the points  $s_1 = 1.00$ ,  $s_2 = 1.23$ ,  $s_3 = 5.35$ , and  $s_4 = 6.73$ . This last design is obtained using the expected volume design expression given in (6). Table 1 contains the design points for the 2-point, 3-point and 4-point discrete locally  $Q$ -optimal designs for the intermediate product (IP2) function and the initial estimates  $\theta_0^T = (0.7, 0.2)$  and  $\sigma_0 = 0, 0.05, 0.10, 0.15$ , and  $0.20$ .

Not unexpectedly, the discrete  $Q$ -optimal designs for  $\sigma_0 = 0$ , which coincide exactly with the local  $D$ -optimal designs, match exactly those given in [Box and Lucas 1959]. What may be surprising, though, is that for this example, whenever discrete  $Q$ -optimal designs were sought with five or more support points, the resulting designs had only four support points. Similar situations were observed

$\sigma$	2-pt design		3-pt design			4-pt design			
	$s_1$	$s_2$	$s_1$	$s_2$	$s_3$	$s_1$	$s_2$	$s_3$	$s_4$
0.00	1.23	6.86	1.23	6.86	6.86	1.23	1.23	6.86	6.86
0.05	1.15	6.36	1.15	5.79	7.11	1.12	1.26	6.18	6.97
0.10	1.04	5.59	1.02	4.72	6.81	1.00	1.23	5.35	6.72
0.15	0.94	4.88	0.91	4.24	6.24	0.90	1.18	4.74	6.34
0.20	0.81	4.05	0.82	3.92	5.52	0.82	1.13	4.34	5.91

**Table 1.** Design points associated with 2-, 3-, and 4-point discrete locally  $Q$ -optimal designs for the IP2 model function.

for other model functions. Thus, we conclude that in general, this discrete design strategy often produces designs with one or two extra design points.

**4.2. Continuous  $Q$ -optimal designs.** In (2), we showed how the first-order Jacobian matrix and second-order derivative array are extended to allow for continuous designs, and these are easily used in the calculation of the second-order volume approximation given in (3); designs which minimize this volume are called continuous  $Q$ -optimal designs.

**Example 2** (continued). To illustrate, consider again the two-parameter intermediate product model given in expression (8) again with initial parameter estimates  $\theta_0^T = (0.7, 0.2)$  and  $\sigma_0 = 0.1$ . In this case, the continuous locally  $Q$ -optimal design with  $r = 3$  points associates the weights  $\lambda = 0.46, 0.28, 0.26$  with the points  $s = 1.06, 5.02, 6.73$ , respectively. This latter design represents a 3.9% volume reduction in terms of (3) relative to the discrete 3-point  $Q$ -optimal design given in Section 4.1, and is therefore preferred. Table 2 contains the continuous locally  $Q$ -optimal designs with  $n = 3$  for this model for  $\sigma_0 = 0, 0.05, 0.10, 0.15$ , and  $0.20$ . As pointed out above, techniques to obtain practical designs from design measures are discussed in [Pukelsheim and Rieder 1992].

In analogous manner to the performance of this model function with the discrete 5-point design criterion discussed above, continuous locally  $Q$ -optimal designs for  $n > 3$  collapsed to those given in Table 2 for this model function.

Of course, this volume design strategy does not always yield designs with extra support points. One obvious counterexample is the Fieller–Creasy problem in which the dummy value takes on only two values — one for each of the two

$\sigma$	$\xi$		
0.00	$s = 1.23$ $\lambda = 0.50$	6.86 0.50	
0.05	$s = 1.17$ $\lambda = 0.49$	4.79 0.04	6.62 0.47
0.10	$s = 1.06$ $\lambda = 0.46$	5.02 0.28	6.78 0.26
0.15	$s = 0.96$ $\lambda = 0.43$	4.69 0.47	7.84 0.10
0.20	$s = 0.82$ $\lambda = 0.37$	4.06 0.58	10.48 0.05

**Table 2.** Continuous locally  $Q$ -optimal designs for the IP2 model function.



groups. Nonetheless, use of the criterion does reemphasize the importance of using continuous designs.

**Example 1** (continued). For the Fieller-Creasy exercise, recall that  $\lambda$  is the proportion taken from group  $A$ , and the remaining  $1 - \lambda$  is the proportion taken from group  $B$ . It can be readily shown that since  $n = p$  here,  $\mathbf{D}_0$  is identity matrix,  $\mathbf{I}_2$ , the volume in (3) is proportional to

$$\lambda^{-1/2}(1 - \lambda)^{-1/2}\{1 + z/\lambda\}, \tag{9}$$

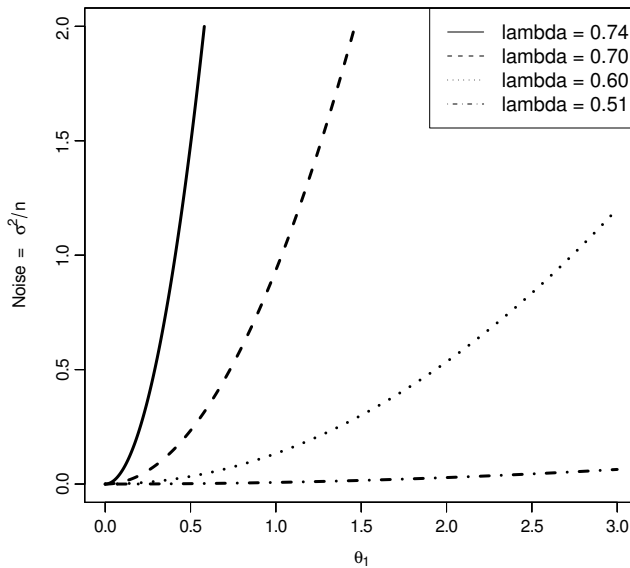
where

$$z = \frac{\sigma^2 \chi_{2,1-\alpha}^2}{4n\theta_1^2}.$$

Since the  $D$ -optimal design is obtained by taking  $\sigma = 0$  in expression (9), this design is thus the equal-weight one; this design ignores curvature and minimizes only the generalized variance of  $\theta$ . When  $\sigma > 0$ , the  $Q$ -optimal design is obtained by choosing

$$\lambda^* = \{1 - 4z + (16z^2 + 16z + 1)^{1/2}\}/4.$$

Contours of this optimal choice of the weight for group  $A$  are plotted in Figure 1 as a function of levels of noise  $= \sigma^2/n$  and  $\theta_1$ .



**Figure 1.** Contour plots of optimal  $\lambda$  as a function of Noise  $= \sigma^2/n$  and  $\theta_1$  for the Fieller-Creasy problem. Contour plots are of the form noise  $= c_\lambda \theta_1^2$ , where  $c_\lambda$  is a constant depending upon  $\lambda$ . The horizontal line noise  $= 0$  corresponds to  $\lambda = 1/2$ , whereas the vertical line  $\theta_1 = 0$  corresponds to  $\lambda = 3/4$ .

In general,  $\lambda^*$  increases (from its lower bound of  $1/2$ ) with the noise level, and its maximum value is  $3/4$ . This result is in line with the observation in Section 3 that the marginal curvature measures decrease as  $\lambda$  increases. Nevertheless, for values of  $\lambda$  above  $3/4$ , the generalized variance becomes exceedingly large.

**4.3. Continuous optimal designs for parameter subsets.** When the  $p$ -dimensional parameter vector is written  $\theta^T = (\theta_1^T, \theta_2^T)$  in situations where, without loss of generality, the  $p_1$ -vector  $\theta_1$  is the parameter (subset) vector of interest and the  $p_2$ -vector  $\theta_2$  are nuisance parameters, so that  $p = p_1 + p_2$ , then an optimal design can be sought using a subset design criteria. For example, Hill and Hunter [1974] have developed discrete subset  $D$ -optimal designs and given geometric interpretations of these designs. Atkinson and Donev [1992] extended this criterion to allow for continuous designs but did not give the geometric results; we provide them here.

*Geometric aspects of continuous  $D_s(\theta_1)$ -optimal designs.* We partition the  $r \times p$  Jacobian  $V_r$  as

$$V_r = [V_1 \mid V_2], \tag{10}$$

so  $V_1$  is  $r \times p_1$  and corresponds to  $\theta_1$  and  $V_2$  is  $r \times p_2$  and corresponds to  $\theta_2$ . Then the Fisher information matrix can be written

$$M(\xi, \theta) = V_r^T \Lambda V_r = \begin{bmatrix} V_1^T \Lambda V_1 & V_1^T \Lambda V_2 \\ V_2^T \Lambda V_1 & V_2^T \Lambda V_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

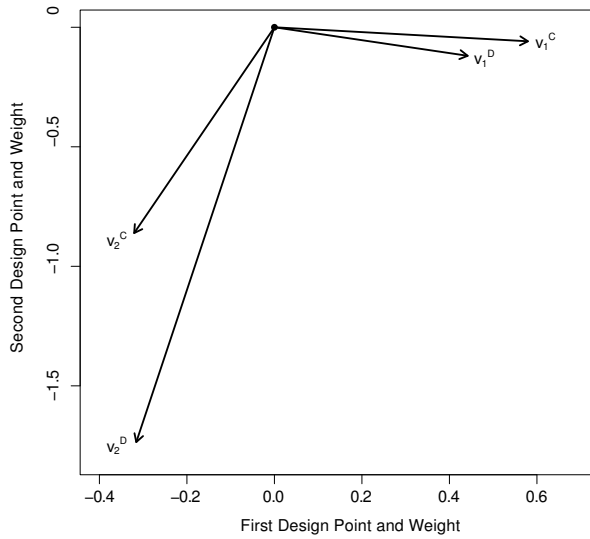
and (local)  $D_s(\theta_1)$ -optimal designs are those which maximize

$$|M|/|M_{22}| = |M_{11} - M_{12}M_{22}^{-1}M_{21}|. \tag{11}$$

provided  $M_{22}$  is nonsingular. In line with the above definitions, when the sample size  $n$  is set *a priori* and the design weights are scalar multiples of  $1/n$ , then the design is called discrete (mass); otherwise it is continuous (mass).

The case where  $V_1 = v_1$  and  $V_2 = v_2$  are column vectors (i.e., when  $p_1 = p_2 = 1$ ) is of particular interest since it is easy to visualize the underlying objectives of the optimal design strategies. For example, Box and Lucas [1959] point out that in this instance the  $D$ -optimality criterion seeks designs which maximize the product of three terms: a term which captures the length of the vector  $v_1$ , a term which captures the length of  $v_2$ , and a term which increases with the angle between these two vectors. Hill and Hunter [1974] point out that the subset criterion in (11) chooses designs to maximize only the product of the first and third of these terms.

**Example 2** (continued). For the IP2 model setting described above, note that the discrete  $D_s(\theta_1)$ -optimal design associates that weight  $\lambda = 1/2$  with each of the points  $s_1 = 1.172$  and  $s_2 = 7.441$ , and the column vectors are plotted in Figure 2 and denoted  $v_1^D$  and  $v_2^D$  respectively.

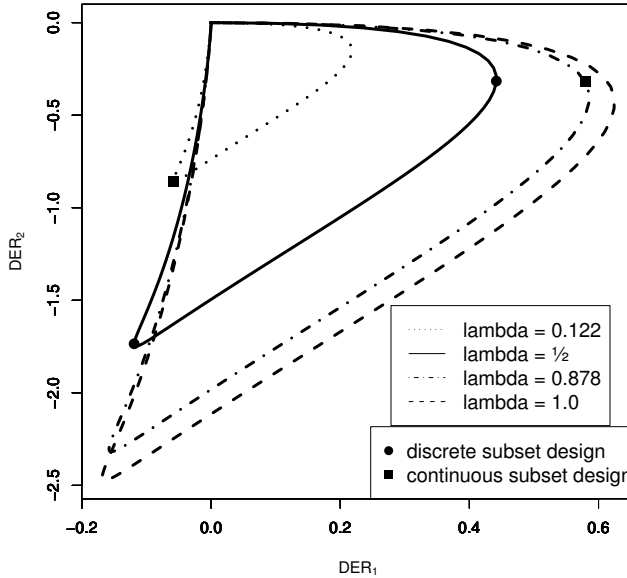


**Figure 2.** IP2 model and visualization of  $V_1$  and  $V_2$  for the  $D_s(\theta_1)$ -optimal (subset) continuous (C) and discrete (D) designs.

If we take  $n = 2$ , this is the same result we obtain if we use the  $V_W$  matrix in (2): precisely what is used to derive analogous results for the continuous subset design. This latter design associates the weight  $\lambda_1 = 0.878$  with the point  $s_1 = 0.994$ , and the weight  $\lambda_2 = 0.122$  with the point  $s_2 = 7.122$ ; the corresponding vectors, denoted  $v_1^C$  and  $v_2^C$ , are also plotted in Figure 2. In terms of the measure in (11), the efficiency of the discrete design with respect to the continuous one is only 0.6549 — that is, about 65%. Since  $1/0.6549 = 1.53$ , this means that the discrete design needs to be 53% larger than the continuous one in order to yield the same amount of information as the continuous design. Geometrically, as we shift from the discrete design ( $v_1^D$  and  $v_2^D$ ) to the continuous one ( $v_1^C$  and  $v_2^C$ ) in Figure 2, the small reduction in orthogonality between  $v_1$  and  $v_2$  in the figure results in a significant lengthening of the  $v_1$  vector.

Before we leave this example, let's now extend the concept of *design locus* presented in [Hill and Hunter 1974, page 430] and [Atkinson and Donev 1992, page 200] to allow for continuous designs. Whereas the geometric strategy in Figure 2 is to focus on the columns of  $V_W$ , the design loci, graphed in Figure 3, focus on the rows of this matrix.

For this example, the solid curve in the center of Figure 3 corresponds to the discrete design and is traced out as  $x$  ranges from 0 to infinity; the two solid circles on this design locus correspond to the discrete design support points given above. In contrast, the outer dot-dashed curve is the design locus for  $\lambda_1 = 0.878$  and the inner dotted curve for  $\lambda_2 = 0.122$ , with the corresponding continuous design points



**Figure 3.** Design loci plots of  $DER_2 = \partial \eta_w / \partial \theta_2 = \sqrt{\eta \lambda} \partial \eta / \partial \theta_2$  versus  $DER_1 = \partial \eta_w / \partial \theta_1 = \sqrt{\eta \lambda} \partial \eta / \partial \theta_1$  for  $\lambda = 0.122, 1/2, 0.878, 1.0$ . Points correspond to the discrete subset design (filled with circles on the  $\lambda = 1/2$  design locus) and the continuous subset design (filled with squares on the  $\lambda = 0.122$  and  $0.878$  design loci).

indicated on the respective loci (darkened squares). For completeness, the design locus corresponding to  $\lambda = 1$  — which represents the outermost possible design locus — is plotted in the figure with the dashed curve.

*Continuous  $D_2(\theta_1)$ -optimal designs with reduced curvature.* The results presented in Sections 4.1 and 4.2 underscore the importance of seeking optimal designs for nonlinear models with reduced curvature in addition to smaller variance, and we now consider this issue in the context of subset designs. As mentioned above, three design criteria which take account of nonlinearity are the volume criterion of [Hamilton and Watts 1985; O’Brien 1992], the second-order MSE criterion of [Clarke and Haines 1995], and the marginal curvature criterion of [O’Brien 2006]. The last of these criteria is well suited for subset designs and interested readers are referred to see [O’Brien 2006]; an extension of the volume criterion to subset designs is not as straightforward, and is not considered here. Rather, we underscore here the use of continuous designs with the MSE criterion.

Employing the second-order variance approximation given in [Clarke 1980] and the second-order bias approximation given in [Box 1971], Clarke and Haines [1995] illustrate the use of discrete subset second-order minimum MSE designs.

We call these designs (local)  $D_2(\theta_1)$ -optimal here. This criterion is well suited to incorporating continuous design as is demonstrated in the following illustration.

**Example 2** (continued). For the IP2 model setting described above, discrete and continuous  $D_s(\theta_1)$ -optimal designs are given in the previous section, however we point out that these designs do not attempt to reduce nonlinearity. In contrast, the continuous  $D_2(\theta_1)$ -optimal design, which associates the weight  $\lambda_1 = 0.825$  with the point  $s_1 = 0.834$  and the weight  $\lambda_2 = 0.175$  with the point  $s_2 = 5.706$ , does result in lower curvature measures. We return to the assessment of curvature in situations such as the present one in the Discussion. Interestingly, in comparing this continuous subset design with the first-order ( $D_s(\theta_1)$ -optimal) one, although the design weights remain essentially unchanged, the design support points shift downward towards the origin.

### 5. Applications to confidence intervals and regions

Important confidence regions for model parameters associated with nonlinear models using discrete designs are the Wald, likelihood-based and second-order regions of the form  $\{\theta \text{ such that LHS} \leq \rho_a^2\}$ . The left hand sides (LHS's) in this expression are as follows:

$$\begin{aligned} \text{Wald : LHS} &= (\theta - \theta_*)^T V_*^T V_* (\theta - \theta_*), \\ \text{Likelihood : LHS} &= S(\theta) - S(\theta_*), \\ \text{Second-order : LHS} &= \tau^T (\mathbf{I}_p - \mathbf{B}) \tau. \end{aligned} \tag{12}$$

The first two of these regions are discussed in [Seber and Wild 1989]. The last of these regions, developed in [Hamilton et al. 1982], leads directly to the second-order volume design criterion given in [Hamilton and Watts 1985] since the given volume approximates the volume of this second-order region. In these expressions,  $\theta_*$  is the maximum likelihood estimate of  $\theta$ ,  $S(\theta) = \Sigma[y_k - \eta_k(\theta)]^2$  is the usual sum-of-squares function,  $\mathbf{B}$  is as in (3), and

$$\tau = U^T [\eta(\theta) - \eta(\theta_*)]$$

for  $U$  discussed in the paragraph following (3). Also, based on [Seber and Wild 1989, page 261], we take  $\rho_a^2 = \sigma^2 \chi_p^2$  since  $\sigma$  is assumed to be known. With the definitions given in Section 3 we now extend these regions to allow for continuous designs.

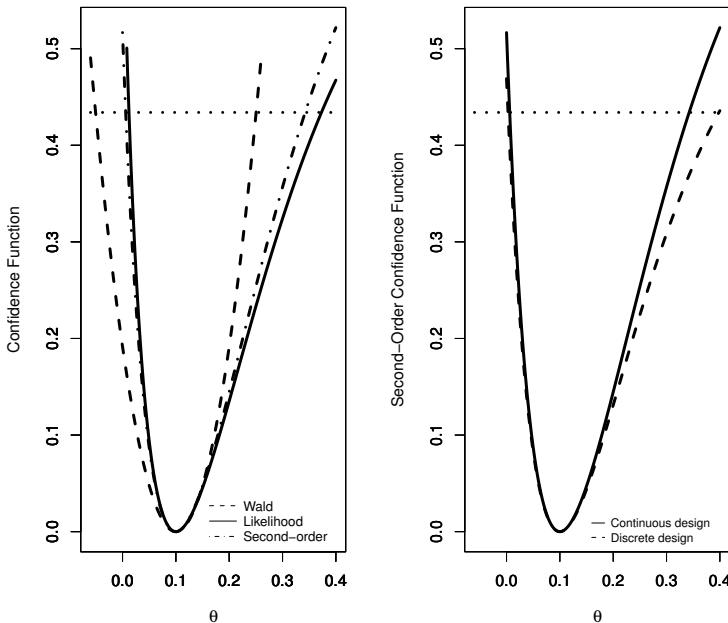
Specifically, we use the weighted Jacobian matrix given in (2) for the Wald region, the weighted model function in (1) in the calculations of the sum-of-squares functions used in the likelihood-based region, and the weighted Jacobian matrix and Hessian array in (2) in the calculations for the second-order region. The following example provides an illustration.

**Example 3.** Using the one-parameter simple exponential (SE1) model function,  $\eta(x, \theta) = e^{-\theta x}$ , and initial parameter estimates  $\theta_0 = 0.10$  and  $\sigma_0 = 0.40$ , the two-point discrete  $Q$ -optimal design associates the weight  $\lambda = 1/2$  with each of the points  $s = 5.10$  and  $7.34$ , whereas the continuous  $Q$ -optimal design associates the weight  $\lambda_1 = 0.926$  with the point  $s_1 = 5.49$  and the weight  $\lambda_2 = 0.074$  with the point  $s_2 = 25.92$ . Then, for this continuous design, the left-hand sides (LHS) in (12), that is, the so-called confidence functions, are given as follows:

$$\begin{aligned} \text{Wald : LHS} &= 19.163(\theta - 0.10)^2, \\ \text{Likelihood : LHS} &= (0.85 - 1.36e^{-5.49\theta})^2 + (-0.37 - 0.39e^{-25.9\theta})^2 - 0.40^2, \\ \text{Second-order : LHS} &= 1.314(1.341e^{-5.49\theta} + 0.066e^{-25.9\theta} - 0.780)^2. \end{aligned}$$

These confidence functions are plotted for this continuous design in the left panel of Figure 4 along with the horizontal cut-line  $\rho_a^2 = \sigma^2 \chi_{1,0.90}^2 = 0.40^2 2.71 = 0.434$ .

The corresponding confidence intervals are thus the intervals on the cut-line between the sides of the respective confidence functions. The inadequacy of the



**Figure 4.** Confidence functions for the SE1 (one-parameter simple exponential) model function. In the left panel, plot of the three confidence interval methods in (12). In the right panel, plot of the second-order (Hamilton and Watts) function for the continuous and discrete  $Q$ -optimal designs. Horizontal cut-line corresponds to 90% confidence.

Wald interval is readily apparent here. Also, note the considerable coincidence here between the second-order and likelihood functions (and thus confidence intervals).

In the right panel of Figure 4, the second-order confidence functions are plotted for the continuous and discrete  $Q$ -optimal designs. That the confidence interval for the continuous design is shorter (by about 15%) than the confidence interval for the discrete design attests to the efficiency of the former design and underscores the importance of using continuous designs.

## 6. Discussion

This article demonstrates how simple yet important adjustments to the model function in (1) and to the associated Jacobian and other matrices given in (2) can be made so as to permit the use of continuous designs in curvature measures, optimal design procedures and the calculation of confidence intervals. The gains of continuous designs over discrete designs, wherein the final sample size is assumed fixed *a priori*, are often significant. As a result, we recommend that continuous designs always be obtained and examined at least as a starting point.

As mentioned above in Section 4, several design criteria exist which can be used to obtain designs with reduced curvature. The  $Q$ -optimality criterion, used extensively in this paper, provides designs that minimize the second-order approximation to the volume of the likelihood confidence region. The MSE criterion of [Clarke and Haines 1995] provides designs that minimize a second-order approximation of the mean-square error of the least-squares parameter estimate. The MC (marginal curvature) approach of [O'Brien 2006] yields designs which minimize the first-order variance approximation and the marginal curvature of [Clarke 1987] for a specified parameter or set of parameters. The latter two of these criteria are readily adapted to yield efficient designs for parameter subsets, as demonstrated on p. 327. Clearly more work needs to be performed exploring which of these three criteria is "best". It may well be the case, however, that each criterion does well only for the respective curvature measure (but not the others), and how one determines which is best may have more to do with how one defines curvature or nonlinearity.

A clear advantage of the volume criterion developed in [O'Brien 1992] and extended here is that this design strategy usually yields designs with extra support points which can be used to test for model misspecification. The same beneficial result is associated with the design strategy in [O'Brien 2006], which focuses on geometric designs. Govaerts [1996] correctly calls an optimal design's having only  $p$  support points a "big limitation" in practical settings. This is especially important when one keeps in mind George Box's comment that "all models are wrong but some are useful" [Box 1979]. These arguments clearly lead us to prefer the  $Q$ -optimality and MC design approaches discussed here.

## References

- [Atkinson 1988] A. C. Atkinson, "Recent developments in the methods of optimum and related experimental designs", *Internat. Statist. Rev.* **56**:2 (1988), 99–115. MR 89g:62123 Zbl 0646.62064
- [Atkinson and Donev 1992] A. C. Atkinson and A. N. Donev, *Optimum experimental designs*, Oxford Statistical Science Series **8**, Clarendon Press, 1992. Zbl 0829.62070
- [Atkinson and Haines 1996] A. C. Atkinson and L. M. Haines, "Designs for nonlinear and generalized linear models", pp. 437–475 in *Design and analysis of experiments*, edited by R. Ghosh and C. Rao, Handbook of Statist. **13**, North-Holland, Amsterdam, 1996. MR 1492576 Zbl 0910.62070
- [Bates and Watts 1980] D. M. Bates and D. G. Watts, "Relative curvature measures of nonlinearity", *J. Roy. Statist. Soc. Ser. B* **42**:1 (1980), 1–25. MR 81e:62070 Zbl 0455.62028
- [Bates and Watts 1988] D. M. Bates and D. G. Watts, *Nonlinear regression analysis and its applications*, Wiley, New York, 1988. MR 92f:62002 Zbl 0728.62062
- [Beale 1960] E. M. L. Beale, "Confidence regions in non-linear estimation", *J. Roy. Statist. Soc. Ser. B* **22** (1960), 41–88. MR 22 #10055 Zbl 0096.13201
- [Box 1971] M. J. Box, "Bias in nonlinear estimation", *J. Roy. Statist. Soc. Ser. B* **33** (1971), 171–201. MR 47 #4376 Zbl 0232.62029
- [Box 1979] G. E. P. Box, "Robustness in the strategy of scientific model building", pp. 201–236 in *Robustness in statistics* (Triangle Park, NC, 1978), edited by R. L. Launer and G. N. Wilkinson, Academic Press, New York, 1979.
- [Box and Lucas 1959] G. E. P. Box and H. L. Lucas, "Design of experiments in non-linear situations", *Biometrika* **46**:1/2 (1959), 77–90. MR 21 #949 Zbl 0086.34803
- [Chaloner and Larntz 1989] K. Chaloner and K. Larntz, "Optimal Bayesian design applied to logistic regression experiments", *J. Statist. Plann. Inference* **21**:2 (1989), 191–208. MR 90f:62241 Zbl 0666.62073
- [Clarke 1980] G. P. Y. Clarke, "Moments of the least squares estimators in a nonlinear regression model", *J. Roy. Statist. Soc. Ser. B* **42**:2 (1980), 227–237. MR 81j:62117
- [Clarke 1987] G. P. Y. Clarke, "Marginal curvatures and their usefulness in the analysis of nonlinear regression models", *J. Amer. Statist. Assoc.* **82**:399 (1987), 844–850. MR 88k:62104 Zbl 0623.62062
- [Clarke and Haines 1995] G. P. Y. Clarke and L. M. Haines, "Optimal design for models incorporating the Richards function", pp. 61–66 in *Statistical modelling* (Innsbruck, 1995), edited by G. U. H. Seeber et al., Lecture Notes in Statistics **104**, Springer, New York, 1995.
- [Cook and Witmer 1985] R. D. Cook and J. A. Witmer, "A note on parameter-effects curvature", *J. Amer. Statist. Assoc.* **80**:392 (1985), 872–878. MR 819586
- [Dette and O'Brien 1999] H. Dette and T. E. O'Brien, "Optimality criteria for regression models based on predicted variance", *Biometrika* **86**:1 (1999), 93–106. MR 2000a:62171 Zbl 1101.62357
- [Govaerts 1996] B. Govaerts, "Discussion of the papers by Atkinson, and Bates *et al*", *J. Royal Statistical Soc. B* **58**:1 (1996), 104.
- [Haines 1992] L. M. Haines, "Optimal design for inverse quadratic polynomials", *South African Statist. J.* **26**:1 (1992), 25–41. MR 93j:62189 Zbl 0763.62042
- [Haines et al. 2004] L. M. Haines, T. E. O'Brien, and G. P. Y. Clarke, "Kurtosis and curvature measures for nonlinear regression models", *Statist. Sinica* **14**:2 (2004), 547–570. MR 2059296 Zbl 1045.62065



- [Hamilton and Watts 1985] D. C. Hamilton and D. G. Watts, "A quadratic design criterion for precise estimation in nonlinear regression models", *Technometrics* **27**:3 (1985), 241–250. MR 797562 Zbl 0598.62083
- [Hamilton et al. 1982] D. C. Hamilton, D. G. Watts, and D. M. Bates, "Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions", *Ann. Statist.* **10**:2 (1982), 386–393. MR 84i:62082 Zbl 0537.62045
- [Hill and Hunter 1974] W. J. Hill and W. G. Hunter, "Design of experiments for subsets of parameters", *Technometrics* **16** (1974), 425–434. MR 50 #11655 Zbl 0311.62045
- [Kiefer and Wolfowitz 1960] J. Kiefer and J. Wolfowitz, "The equivalence of two extremum problems", *Canad. J. Math.* **12** (1960), 363–366. MR 22 #8616 Zbl 0093.15602
- [O'Brien 1992] T. E. O'Brien, "A note on quadratic designs for nonlinear regression models", *Biometrika* **79** (1992), 847–849.
- [O'Brien 2006] T. E. O'Brien, "Robust optimal design with reduced curvature", in *Proceedings of the American Statistical Association, Biopharmaceutical Section* (Alexandria, VA), 2006.
- [O'Brien and Funk 2003] T. E. O'Brien and G. M. Funk, "A gentle introduction to optimal design for regression models", *Amer. Statist.* **57**:4 (2003), 265–267. MR 2037854
- [Pukelsheim 1993] F. Pukelsheim, *Optimal design of experiments*, Wiley, New York, 1993. MR 94k:62124 Zbl 0834.62068
- [Pukelsheim and Rieder 1992] F. Pukelsheim and S. Rieder, "Efficient rounding of approximate designs", *Biometrika* **79**:4 (1992), 763–770. MR 1209476
- [Ratkowsky 1983] D. A. Ratkowsky, *Nonlinear regression modeling: A unified practical approach*, Marcel Dekker, New York, 1983. Zbl 0572.62054
- [Seber and Wild 1989] G. A. F. Seber and C. J. Wild, *Nonlinear regression*, Wiley, New York, 1989. MR 90j:62004 Zbl 0721.62062
- [White 1973] L. V. White, "An extension of the general equivalence theorem to nonlinear models", *Biometrika* **60** (1973), 345–348. MR 48 #5290 Zbl 0262.62037

Received: 2010-05-19      Accepted: 2010-06-13

tobrien@math.luc.edu      *Department of Mathematics and Statistics,  
Loyola University, 1032 W. Sheridan Road,  
Chicago, IL 60660-1537, United States*  
<http://webpages.math.luc.edu/~tobrien/home.html>

huajam@grad.sci.tu.ac.th      *Department of Mathematics and Statistics,  
Thammasat University — Rangsit Campus,  
Klongluang, Pathumthani 12121, Thailand*

chinnaphong@mathstat.sci.tu.ac.th      *Department of Mathematics and Statistics,  
Thammasat University — Rangsit Campus,  
Klongluang, Pathumthani 12121, Thailand*

# involve

pjm.math.berkeley.edu/involve

## EDITORS

### MANAGING EDITOR

Kenneth S. Berenhaut, Wake Forest University, USA, berenhs@wfu.edu

### BOARD OF EDITORS

John V. Baxley	Wake Forest University, NC, USA baxley@wfu.edu	Chi-Kwong Li	College of William and Mary, USA ckli@math.wm.edu
Arthur T. Benjamin	Harvey Mudd College, USA benjamin@hmc.edu	Robert B. Lund	Clemson University, USA lund@clemson.edu
Martin Bohner	Missouri U of Science and Technology, USA bohner@mst.edu	Gaven J. Martin	Massey University, New Zealand g.j.martin@massey.ac.nz
Nigel Boston	University of Wisconsin, USA boston@math.wisc.edu	Mary Meyer	Colorado State University, USA meyer@stat.colostate.edu
Amarjit S. Budhiraja	U of North Carolina, Chapel Hill, USA budhiraj@email.unc.edu	Emil Minchev	Ruse, Bulgaria eminchev@hotmail.com
Pietro Cerone	Victoria University, Australia pietro.cerone@vu.edu.au	Frank Morgan	Williams College, USA frank.morgan@williams.edu
Scott Chapman	Sam Houston State University, USA scott.chapman@shsu.edu	Mohammad Sal Moslehian	Ferdowsi University of Mashhad, Iran moslehian@ferdowsi.um.ac.ir
Jem N. Corcoran	University of Colorado, USA corcoran@colorado.edu	Zuhair Nashed	University of Central Florida, USA znashed@mail.ucf.edu
Michael Dorff	Brigham Young University, USA mdorff@math.byu.edu	Ken Ono	University of Wisconsin, USA ono@math.wisc.edu
Sever S. Dragomir	Victoria University, Australia sever@matilda.vu.edu.au	Joseph O'Rourke	Smith College, USA orourke@cs.smith.edu
Behrouz Emamizadeh	The Petroleum Institute, UAE bemamizadeh@pi.ac.ae	Yuval Peres	Microsoft Research, USA peres@microsoft.com
Errin W. Fulp	Wake Forest University, USA fulp@wfu.edu	Y.-F. S. Pétermann	Université de Genève, Switzerland petermann@math.unige.ch
Andrew Granville	Université Montréal, Canada andrew@dms.umontreal.ca	Robert J. Plemmons	Wake Forest University, USA plemmons@wfu.edu
Jerrold Griggs	University of South Carolina, USA griggs@math.sc.edu	Carl B. Pomerance	Dartmouth College, USA carl.pomerance@dartmouth.edu
Ron Gould	Emory University, USA rg@mathcs.emory.edu	Bjorn Poonen	UC Berkeley, USA poonen@math.berkeley.edu
Sat Gupta	U of North Carolina, Greensboro, USA sgupta@uncg.edu	James Propp	U Mass Lowell, USA jpropp@cs.uml.edu
Jim Haglund	University of Pennsylvania, USA jhaglund@math.upenn.edu	József H. Przytycki	George Washington University, USA przytyck@gwu.edu
Johnny Henderson	Baylor University, USA johnny_henderson@baylor.edu	Richard Rebarber	University of Nebraska, USA rrebarbe@math.unl.edu
Natalia Hritonenko	Prairie View A&M University, USA nahritonenko@pvamu.edu	Robert W. Robinson	University of Georgia, USA rwr@cs.uga.edu
Charles R. Johnson	College of William and Mary, USA crjohnso@math.wm.edu	Filip Saidak	U of North Carolina, Greensboro, USA f.saidak@uncg.edu
Karen Kafadar	University of Colorado, USA karen.kafadar@cudenver.edu	Andrew J. Sterge	Honorary Editor andy@ajsterge.com
K. B. Kulasekera	Clemson University, USA kk@ces.clemson.edu	Ann Trenk	Wellesley College, USA atrenk@wellesley.edu
Gerry Ladas	University of Rhode Island, USA gladas@math.uri.edu	Ravi Vakil	Stanford University, USA vakil@math.stanford.edu
David Larson	Texas A&M University, USA larson@math.tamu.edu	Ram U. Verma	University of Toledo, USA verma99@msn.com
Suzanne Lenhart	University of Tennessee, USA lenhart@math.utk.edu	John C. Wierman	Johns Hopkins University, USA wierman@jhu.edu

## PRODUCTION

Silvio Levy, Scientific Editor

Sheila Newbery, Senior Production Editor

Cover design: ©2008 Alex Scorpan

See inside back cover or <http://pjm.math.berkeley.edu/involve> for submission instructions.

The subscription price for 2010 is US \$100/year for the electronic version, and \$120/year (+\$20 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94704-3840, USA.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW™ from Mathematical Sciences Publishers.

PUBLISHED BY  
 mathematical sciences publishers  
<http://www.mathscipub.org>

A NON-PROFIT CORPORATION

Typeset in L<sup>A</sup>T<sub>E</sub>X

Copyright ©2010 by Mathematical Sciences Publishers

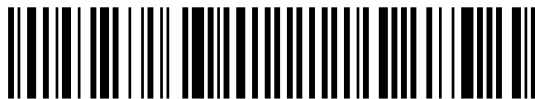
# involve

2010

vol. 3

no. 3

Gracefulness of families of spiders	241
PATRICK BAHLS, SARA LAKE AND ANDREW WERTHEIM	
Rational residuacity of primes	249
MARK BUDDEN, ALEX COLLINS, KRISTIN ELLIS LEA AND STEPHEN SAVIOLI	
Coexistence of stable ECM solutions in the Lang–Kobayashi system	259
ERICKA MOCHAN, C. DAVIS BUENGER AND TAMAS WIANDT	
A complex finite calculus	273
JOSEPH SEABORN AND PHILIP MUMMERT	
$\zeta(n)$ via hyperbolic functions	289
JOSEPH D'AVANZO AND NIKOLAI A. KRYLOV	
Infinite family of elliptic curves of rank at least 4	297
BARTOSZ NASKRECKI	
Curvature measures for nonlinear regression models using continuous designs with applications to optimal experimental design	317
TIMOTHY O'BRIEN, SOMSRI JAMROENPINYO AND CHINNAPHONG BUMRUNGSUP	
Numerical semigroups from open intervals	333
VADIM PONOMARENKO AND RYAN ROSENBAUM	
Distinct solution to a linear congruence	341
DONALD ADAMS AND VADIM PONOMARENKO	
A note on nonresidually solvable hyperlinear one-relator groups	345
JON P. BANNON AND NICOLAS NOBLETT	



1944-4176(2010)3:3;1-E