Optimality criteria for regression models based on predicted variance

By HOLGER DETTE

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
holger.dette@ruhr-uni-bochum.de

AND TIMOTHY E. O'BRIEN

K-490.3.51, CIBA-GEIGY AG, Postfach CH-4002 Basel, Schweiz
tim.obrien@bluewin.ch

SUMMARY

In the context of nonlinear regression models, a new class of optimum design criteria is developed and illustrated. This new class, termed I_L-optimality, is analogous to Kiefer's Φ_k-criterion but is based on predicted variance, whereas Kiefer's class is based on the eigenvalues of the information matrix; I_L-optimal designs are invariant with respect to different parameterisations of the model and contain G- and D-optimality as special cases. We provide a general equivalence theorem which is used to obtain and verify I_L-optimal designs. The method is illustrated by various examples.

Some key words: Bayesian design; Invariance; Optimal design.

1. INTRODUCTION AND BACKGROUND

The design problem for univariate nonlinear models of the form

$$y_k = \eta(x_k, \theta) + \varepsilon_k \quad (k = 1, \ldots, n). \tag{1}$$

is to obtain an n-point design ξ to estimate some function of the p-dimensional parameter vector θ with high efficiency. We consider approximate designs here, that is designs of the form

$$\xi = \left\{ \begin{matrix} x_1, x_2, \ldots, x_r \\ \omega_1, \omega_2, \ldots, \omega_r \end{matrix} \right\},$$

where the design points or vectors $x_1, x_2, \ldots, x_r$ are distinct elements of the design space $\mathcal{X}$, for $r \geqslant p$, and the associated weights $\omega_1, \omega_2, \ldots, \omega_r$ are nonnegative real numbers which sum to unity; conversion of an optimal approximate design to a near-optimal exact one, or a design where each ω_k is of the form n_k/n for n_k an integer, is discussed in Cook & Nachtsheim (1980) and Pukelsheim & Rieder (1992).

When the errors associated with the assumed model (1) are uncorrelated Gaussian random variables with zero means and constant variance, taken without loss of generality to be equal one, the Fisher information of a design ξ is given by

$$M(\xi, \theta) = \sum_{i=1}^{r} \omega_i \frac{\partial \eta(x_i, \theta)}{\partial \theta} \frac{\partial \eta(x_i, \theta)}{\partial \theta^{\mathrm{T}}} = V^{\mathrm{T}} \Omega V. \tag{2}$$

Here V is the $r \times p$ Jacobian of η with ith row equal to the gradient $\partial \eta(x_i, \theta)/\partial \theta$ of the response function at the point x_i, and Ω is the diagonal matrix with diagonal elements $\omega_1, \omega_2, \ldots, \omega_r$. Also, the variance function (Atkinson & Donev, 1992, p. 95) of ξ is given by

$$d(x, \xi, \theta) = \frac{\partial \eta(x, \theta)}{\partial \theta^{\mathrm{T}}} M^{-1}(\xi, \theta) \frac{\partial \eta(x, \theta)}{\partial \theta}, \tag{3}$$

where $\partial \eta(x, \theta)/\partial \theta$ is of dimension $p \times 1$, and it is assumed that $|M(\xi, \theta)| \neq 0$. Since the first-order approximation to the covariance of the maximum likelihood estimator for the parameter θ is $M^{-1}(\xi, \theta)$, this function corresponds to the first-order approximation to the variance of the predicted response.

Optimal designs typically minimise some convex function of the inverse Fisher information matrix. For example, designs which minimise the logarithm of the determinant of $M^{-1}(\xi, \theta^0)$ are called locally D-optimal, those which minimise its trace are called locally A-optimal, and those which minimise its largest eigenvalue are called locally E-optimal; the term 'locally' is used here to emphasise that the corresponding design is based on an initial parameter choice, θ^0, say. Additionally, designs which minimise the supremum over all $x \in \mathscr{X}$ of $d(x, \xi, \theta^0)$ are called locally G-optimal, and the general equivalence theorem of Kiefer & Wolfowitz (1960) establishes that D-optimal designs are also G-optimal. We remark finally that the locally optimal design problem for a nonlinear regression is a particular case of a design problem with uncontrolled variables described by Fedorov & Atkinson (1988). Therefore most of the ideas and results of this paper can be transferred to that setting without any difficulty. These extensions are omitted for the sake of brevity.

Kiefer (1974) connects D-, A- and E-optimality in one class of criteria by noting that, if $\lambda_1, \ldots, \lambda_p$ are the p eigenvalues of $M^{-1}(\xi, \theta^0)$, then the locally Φ_k-optimal designs, which minimise

$$\left\{ \frac{1}{p} (\lambda_1^k + \ldots + \lambda_p^k) \right\}^{1/k}, \tag{4}$$

for $k \in (0, \infty)$, are D-optimal for $k \to 0$, A-optimal for $k = 1$ and E-optimal for $k \to \infty$. Note that, whereas D-, A- and E-optimal designs focus on the covariance matrix associated with the maximum likelihood estimation of θ, G- and I-optimal designs, which are those minimising the average variance function $\int_{\mathscr{X}} d(x, \xi, \theta^0) \, d\mu(x)$, focus on the predicted variance associated with the estimate of the response function. Here, μ is an arbitrary weight measure over $\mathscr{X}$. The diversity of optimal designs is illustrated with the following example.

Example 1. The two-parameter model function given by

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} (e^{-\theta_2 x} - e^{-\theta_1 x}) \quad (\theta_1, \theta_2 > 0, \ \theta_1 > \theta_2) \tag{5}$$

is often used to model the amount of intermediate product present in a compartmental model at time x; see Atkinson et al. (1993) and Box & Lucas (1959). Suppose that a researcher requires an optimal design over the design space $\mathscr{X} = [0, 20]$ with initial parameter estimates $\theta_1^0 = 0.70$ and $\theta_2^0 = 0.20$. The Elfving set (Elfving, 1952) is depicted in Fig. 1. It is defined as the convex hull of $-\mathscr{G} \cup \mathscr{G}$, where $\mathscr{G}$ is the curve of the gradient

$$\mathscr{G} = \left\{ \frac{\partial}{\partial \theta} \eta(x, \theta) \bigg|_{\theta = \theta_0} ; x \in [0, 20] \right\}.$$

The locally D-/G-, A-, E- and I-optimal designs for the uniform weight measure are given in Table 1.
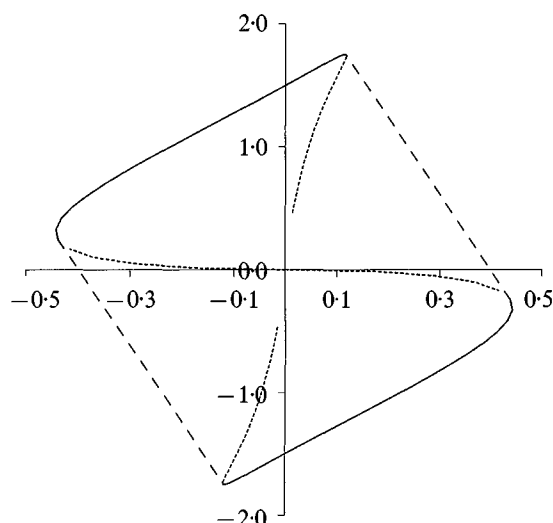

Fig. 1. The Elfving set for the intermediate product model, where $\theta^0 = (0\cdot7, 0\cdot2)$. The dotted and solid lines show the curves $\{\pm d\eta(x, \theta)/d\theta|_{\theta=\theta_0}; x \in [0, 20]\}$.

Table 1. *Locally optimal designs in the intermediate product model* (5) *with respect to various optimality criteria*

Criterion	I	A $(k = 1)$	E $(k = \infty)$	D/G $(k = 0)$
Optimal design	$\begin{Bmatrix} 1\cdot311 & 6\cdot768 \\ 0\cdot328 & 0\cdot672 \end{Bmatrix}$	$\begin{Bmatrix} 1\cdot094 & 7\cdot010 \\ 0\cdot770 & 0\cdot230 \end{Bmatrix}$	$\begin{Bmatrix} 0\cdot994 & 7\cdot122 \\ 0\cdot847 & 0\cdot153 \end{Bmatrix}$	$\begin{Bmatrix} 1\cdot229 & 6\cdot858 \\ 0\cdot5 & 0\cdot5 \end{Bmatrix}$

Note that, as k in (4) increases from zero through one to infinity, the design support points become further apart and the weight on the lower support point increases. If we move from the D-/G-optimal design to the I-optimal design, the opposite behaviour occurs.

2. I_L-OPTIMALITY

Assume that the regression function $\eta(., \theta)$ can also be defined on a set $\mathscr{Z}$ which may be larger than the design space $\mathscr{X}$. For a point $z \in \mathscr{Z}$ the variance of the predicted response is given by

$$\frac{\partial \eta(z, \theta)}{\partial \theta^{\mathrm{T}}} M^-(\xi, \theta) \frac{\partial \eta(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta^0},$$

where $M^-(\xi, \theta)$ denotes a generalised inverse of $M(\xi, \theta)$ and it is assumed that $\eta(z, \theta)$ is estimable by the design ξ, that is

$$\frac{\partial \eta(z, \theta)}{\partial \theta} \bigg|_{\theta=\theta^0} \in \mathrm{Range}\{M(\xi, \theta^0)\}. \tag{6}$$

In order to achieve a transparent presentation we assume that the experimenter is inter-

ested in prediction of the response at at least p points $z_1, \ldots, z_p$ such that the corresponding vectors

$$\left\{ \frac{\partial \eta(z_j, \theta)}{\partial \theta} \bigg|_{\theta = \theta^0} \right\}_{j=1}^{p}$$

are linearly independent. By the range inclusion (6), a design for predicting the response function at these points must have a nonsingular information matrix. Finally, we assume that μ is a measure on $\mathscr{Z}$ with $\mathrm{supp}(\mu) \supset \{z_1, \ldots, z_p\}$. Generalisations for prediction with a singular information matrix $M(\xi, \theta^0)$ will be discussed briefly in § 5.

In connecting D-/G-, A- and E-optimality through the Φ_k-optimality criterion, Kiefer (1974) uses different weighted sums of the eigenvalues of the inverse Fisher information matrix $M^{-1}(\xi, \theta^0)$. Analogously, we weight the variance function in the following definition with an exponent L and a weight measure μ.

DEFINITION 1. For $L \in (0, \infty)$ a design ξ_L^* with nonsingular information matrix $M(\xi_L^*, \theta^0)$ is said to be locally I_L-optimal if it minimises the function

$$\psi_L(\xi) = \left\{ \int_{\mathscr{Z}} d^L(z, \xi, \theta^0) \, d\mu(z) \right\}^{1/L} \tag{7}$$

over the space of all approximate designs; here μ is a probability measure and $\mathscr{Z}$ is the corresponding integration space, usually, but not always, taken to coincide with the design space, $\mathscr{X}$; see Example 5 below.

Remark 1. Note that this definition is easily extended to allow for the cases $L = 0$ and $L = \infty$ by defining the criterion functions in these cases to be

$$\psi_0(\xi) = \exp\left\{ \int_{\mathscr{Z}} \log d(z, \xi, \theta^0) \, d\mu(z) \right\}, \quad \psi_\infty(\xi) = \sup_{z \in \mathscr{Z}} d(z, \xi, \theta^0).$$

When the design and integration spaces coincide, that is $\mathscr{X} = \mathscr{Z}$, and $d(x, \xi, \theta^0)$ is bounded on $\mathscr{X}$, I_∞-optimality is equivalent to D-/G-optimality and can be obtained from I_L-optimality as $L \to \infty$. Moreover, I_1-optimality corresponds to ordinary I-optimality, which is in fact a special case of A-optimality.

Some special cases of our I_L-optimality criterion have already been considered; see e.g. Studden (1977) or Wong (1992). There is also a close relationship to optimal designs for extrapolation and interpolation; see e.g. Kiefer & Wolfowitz (1964), Levine (1966), Studden (1971), Herzberg & Cox (1972) and Spruill (1987, 1990a, b). One motivation for considering various powers in Definition 1 is to allow for different averages of the variance function over the space $\mathscr{Z}$. Another is to connect different design criteria such as D-/G- and I-optimality, which focus on the predicted variance, as Kiefer (1974) connected the A-, D- and E-optimality criterion by (4). Furthermore, the measure μ in Definition 1 can be used to weight specific regions of the design space differently; see also § 6 below, Studden (1977), Cook & Nachtsheim (1982) and Haines (1987).

In comparing Kiefer's Φ_k class and our I_L class, the following theorem provides an important preference for the latter class.

THEOREM 1. Let the design ξ_L^* be locally I_L-optimal for a nonlinear model $\eta(x, \theta)$ which depends on the p-vector θ, and let the p-vector $\phi = \phi(\theta)$ be any linear or nonlinear reparameterisation of η such that the $p \times p$ Jacobian $T = \partial \theta / \partial \phi$ is nonsingular. Then ξ_L^* is locally I_L-optimal for $\eta(x, \phi)$. That is, I_L-optimality is invariant with respect to reparameterisations.

Proof. Let V_θ be the $r \times p$ Jacobian of η with ith row equal to $\partial\theta(x_i, \theta)/\partial\theta^{\mathrm{T}}$ and V_ϕ be the $r \times p$ Jacobian of η with ith row equal to $\partial\eta(x_i, \phi)/\partial\phi^{\mathrm{T}}$. Then, by the chain rule, $V_\phi = V_\theta T$, so that $M(\xi, \phi) = V_\phi^{\mathrm{T}}\Omega V_\phi = T^{\mathrm{T}} M(\xi, \theta)T$ and

$$d(x, \xi, \phi) = \frac{\partial\eta(x, \phi)}{\partial\phi^{\mathrm{T}}} M^{-1}(\xi, \phi) \frac{\partial\eta(x, \phi)}{\partial\phi} = d(x, \xi, \theta).$$

Therefore the variance function and corresponding optimality criteria are invariant with respect to reparameterisations, and as a consequence the I_L-optimal designs are also invariant. □

Remark 2. As a consequence, when the regression is linear, i.e.

$$\eta(x, \theta) = \theta_1 f_1(x) + \ldots + \theta_p f_p(x),$$

the I_L-optimal designs do not depend on the specific choice of a basis for the vector space of the regression functions. In particular, if the assumed model function is a kth-degree polynomial model but we are entertaining several design spaces of the form $[a, b]$, we need only obtain the I_L-optimal design for this model using the design space $[0, 1]$, since this design could then be suitably modified for any design space $[a, b]$; see Example 3 below.

Remark 3. In regression settings, invariance of an optimal design to reparameterisation of the model function is typically an absolute necessity, and, since D-optimality is the only criterion of the Kiefer Φ_k class which possesses this property, the I_L class is therefore favoured in this regard. An advantage of using Φ_k-optimal designs is that their optimality can be verified by using the general equivalence theorem of Kiefer & Wolfowitz (1960); in § 3 we show that this same advantage applies also to I_L-optimal designs.

Example 2. The range of designs produced using the I_L-criterion is illustrated with the intermediate product model function in (5) for $\mathcal{Z} = \mathcal{X} = [0, 20]$, $\theta_1^0 = 0.70$, $\theta_2^0 = 0.20$ and
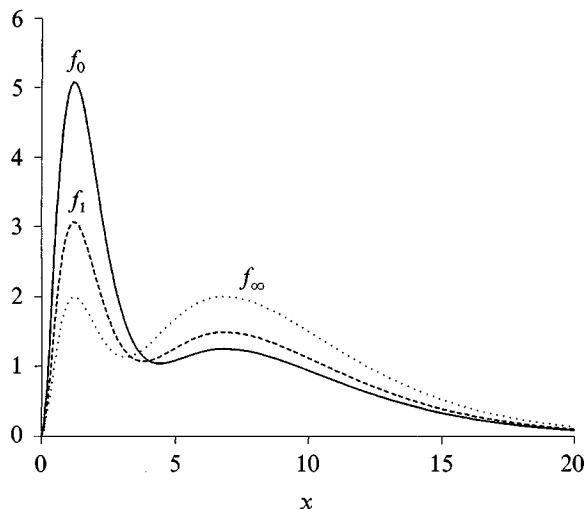

Fig. 2: Example 2. The variance function $f_L(x) = d(x, \xi_L^*)$ for various I_L-optimal designs for the intermediate product model ($\theta_1^0 = 0.7$, $\theta_2^0 = 0.2$). Solid line, $L = 0$; dashed line, $L = 1$; dotted line, $L = \infty$.

$d\mu(z) = dz$ by examining the locally I_L-optimal designs for $L = 0$, 1 and ∞,

$$\xi_0^* = \begin{Bmatrix} 1 \cdot 380 & 6 \cdot 693 \\ 0 \cdot 200 & 0 \cdot 800 \end{Bmatrix}, \quad \xi_1^* = \begin{Bmatrix} 1 \cdot 311 & 6 \cdot 768 \\ 0 \cdot 328 & 0 \cdot 672 \end{Bmatrix}, \quad \xi_\infty^* = \begin{Bmatrix} 1 \cdot 229 & 6 \cdot 858 \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}.$$

The corresponding variance functions are graphed in Fig. 2, indicating how the I_L-optimal designs weight the regions of the design space differently; I_L-optimal designs with increasing L yield a smaller variance function in the region $[4, 20]$, while for $x \in [0, 4]$ the I_∞-optimal design produces the smallest variance function. To illustrate Theorem 1 we note that the above designs are also I_L-optimal for the model

$$\eta(x, a, b) = \frac{a}{e^{ax}} \frac{e^{bx} - 1}{b},$$

which is obtained by reparameterisation from the intermediate product model ($\theta_1 = a$, $\theta_1 - \theta_2 = b$).

Example 3. In the spirit of Cook & Nachtsheim (1982), we also examine the three-parameter quadratic model function

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

The I_L-optimal designs for $L = 0$, 1 and ∞ for this model over $\mathscr{X} = \mathscr{X} = [0, 1]$ and with $d\mu(z) = dz$ are

$$\xi_0^* = \begin{Bmatrix} 0 & \frac{1}{2} & 1 \\ 0 \cdot 2285 & 0 \cdot 5430 & 0 \cdot 2285 \end{Bmatrix}, \quad \xi_1^* = \begin{Bmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{Bmatrix}, \quad \xi_\infty^* = \begin{Bmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}.$$

Results on symmetry and admissibility of optimal designs dictate that the support points of these designs must be $x = 0$, $\frac{1}{2}$ and 1, and the corresponding weights must be of the form $\omega = \alpha$, $1 - 2\alpha$ and α for $\alpha \in [0, \frac{1}{2}]$; see Pukelsheim (1993, p. 253). As with the previous example, the corresponding plots of the variance functions, given in Fig. 3, show how I_L-optimal designs with smaller values of L focus on regions of $\mathscr{X}$ where the variance
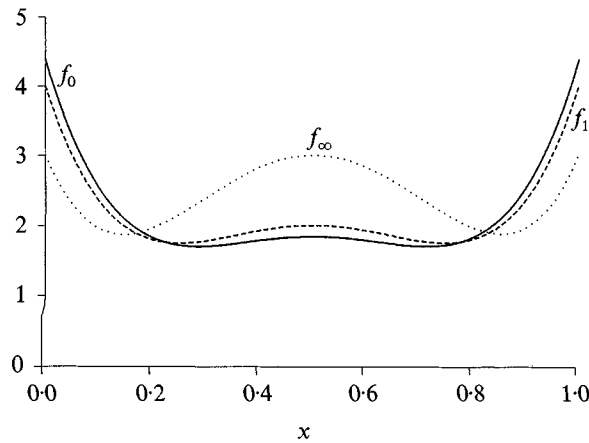

Fig. 3: Example 3. The variance function $f_L(x) = d(x, \xi_L^*)$ for various I_L-optimal designs in the quadratic regression model. Solid line, $L = 0$; dashed line, $L = 1$; dotted line, $L = \infty$.

function takes on lower values and vice versa. For model functions of this type, we recommend I_L-optimal designs with smaller values of L, such as $L = 1$, rather than those with larger values since the variance function of the I_1-optimal design lies below the variance function of the D-/I_∞-optimal design for roughly two-thirds of the integration space; see Cook & Nachtsheim (1982). Incidentally, to illustrate the use of Remark 2, note that the I_1-optimal design for this model function over the interval $[a, b]$ is

$$\xi_1^* = \left\{ \begin{matrix} a & \frac{1}{2}(a + b) & b \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix} \right\}.$$

The I_0- and I_∞-optimal designs are obtained in the same way.

3. A GENERAL EQUIVALENCE THEOREM

Obtaining and verifying efficient I_L-optimal designs is greatly facilitated by the following theorems which are proved in the Appendix.

THEOREM 2. *For any* $L \in [0, \infty)$, *I_L-optimal designs may equivalently be characterised by either of the following:*
  (i) ξ_L^* *minimises* $\psi_L(\xi) = \{ \int_{\mathscr{X}} d^L(z, \xi, \theta^0) \, d\mu(z) \}^{1/L}$ *over the set of all approximate designs;*
  (ii) *for all* $x \in \mathscr{X}$,

$$\int_{\mathscr{X}} d^{L-1}(z, \xi_L^*, \theta^0) d^2(x, z, \xi_L^*, \theta^0) \, d\mu(z) \leqslant \int_{\mathscr{X}} d^L(z, \xi_L^*, \theta^0) \, d\mu(z); \tag{8}$$

here the covariance function $d(x, z, \xi, \theta)$ *is given by*

$$\frac{\partial \eta(x, \theta)}{\partial \theta^{\mathrm{T}}} M^{-1}(\xi, \theta) \frac{\partial \eta(z, \theta)}{\partial \theta}$$

and the variance function $d(z, \xi, \theta)$ *is defined in* (3). *Moreover, equality is achieved in* (8) *at the support points of* ξ_L^*.

Remark 4. The second condition in this theorem is equivalent to $\phi_L(x, \xi_L^*) \leqslant 1$ for all $x \in \mathscr{X}$, where

$$\phi_L(x, \xi_L^*) = \frac{\int_{\mathscr{X}} d^{L-1}(z, \xi_L^*, \theta^0) d^2(x, z, \xi_L^*, \theta^0) \, d\mu(z)}{\int_{\mathscr{X}} d^L(z, \xi_L^*, \theta^0) \, d\mu(z)}. \tag{9}$$

It follows from Theorem 2 that we can verify that a given design is indeed I_L-optimal by plotting the graph of $\phi_L(x, \xi_L^*)$ versus x to make sure that this curve first does not go above the line $y = 1$, and secondly touches the line $y = 1$ at the support points of ξ_L^*. Moreover, the following result shows that, if

$$\mathrm{eff}_L(\xi) = \frac{\psi_L(\xi_L^*)}{\psi_L(\xi)} \tag{10}$$

denotes the local I_L-efficiency of a design ξ, then the maximum excess of $\phi_L(x, \xi)$ over the line $y = 1$ provides a lower bound for $\mathrm{eff}_L(\xi)$.

THEOREM 3. *If* ξ *is a design with nonsingular information matrix* $M(\xi, \theta_0)$, *then*

$$\mathrm{eff}_L(\xi) \geqslant \frac{1}{\sup_{x \in \mathscr{X}} \phi_L(x, \xi)}.$$
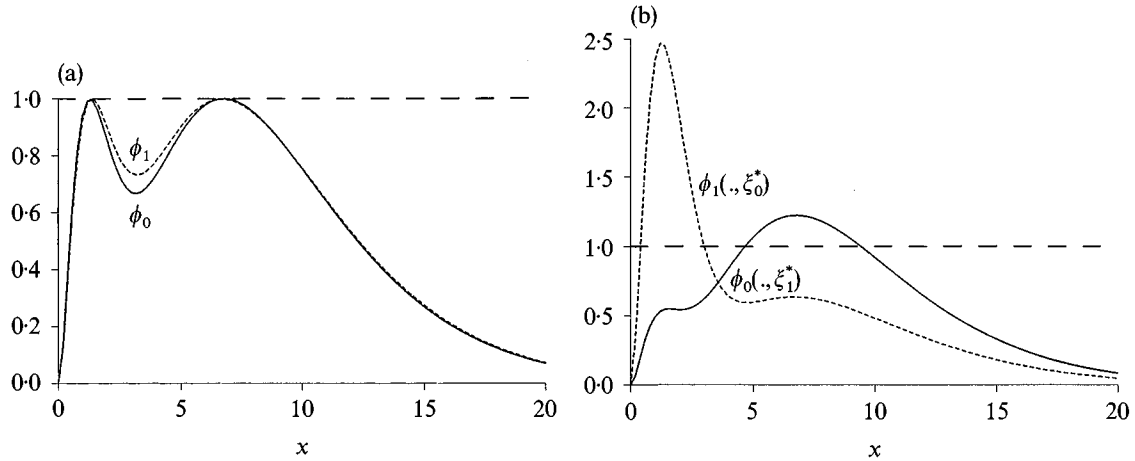
Fig. 4. (a) The function $\phi_L(x, \xi_L^*)$ for two I_L-optimal designs ($\mathcal{X} = \mathcal{Z} = [0, 20]$) in the intermediate product model ($\theta_1^0 = 0.7$, $\theta_2^0 = 0.2$). Solid line, $L = 0$; dashed line, $L = 1$. (b) The functions $\phi_0(x, \xi_1^*)$ (solid line) and $\phi_1(x, \xi_0^*)$ (dashed line) in the intermediate product model ($\mathcal{X} = \mathcal{Z} = [0, 20]$, $\theta_1^0 = 0.7$, $\theta_2^0 = 0.2$).

Example 4. To illustrate Theorem 2, the functions $\phi_1(x, \xi_1^*)$ and $\phi_0(x, \xi_0^*)$ have been plotted against x in Fig. 4(a) for the intermediate product model function in (5) and the locally I_0- and I_1-optimal designs,

$$\xi_0^* = \begin{cases} 1.380 & 6.693 \\ 0.200 & 0.800 \end{cases}, \quad \xi_1^* = \begin{cases} 1.311 & 6.768 \\ 0.328 & 0.672 \end{cases}.$$

The I_L-optimality of ξ_L^* is verified by noting that $\phi_L(x, \xi_L^*) \leqslant 1$ for all $x \in [0, 20]$; $L = 0, 1$. In Fig. 4(b) we illustrate the application of Theorem 3 and plot the functions $\phi_1(x, \xi_0^*)$ and $\phi_0(x, \xi_1^*)$. By Theorem 2 the design ξ_0^* is not I_1-optimal and ξ_1^* not I_0-optimal. On the other hand, Theorem 3 shows that the I_1-efficiency of ξ_0^* is at least 40% while the I_0-efficiency of ξ_1^* is at least 81·7%. In order to demonstrate the accuracy of these bounds we also calculated the exact efficiencies, which are $\text{eff}_1(\xi_0^*) \approx 40.5\%$ and $\text{eff}_0(\xi_1^*) \approx 90.45\%$.

Remark 5. The above general equivalence theorem can be extended to the case $L = \infty$ in the following manner; see e.g. Wong & Cook (1992). Define

$$\Pi(\xi) = \left\{ z \in \mathcal{Z} \mid d(z, \xi, \theta^0) = \sup_{z \in \mathcal{Z}} d(z, \xi, \theta^0) \right\}.$$

Then the following are equivalent:
(i) ξ_∞^* minimises $\psi_\infty(\xi) = \sup_{z \in \mathcal{Z}} d(z, \xi, \theta^0)$, that is ξ_∞^* is I_∞-optimal;
(ii) there exists a probability measure μ^* on $\Pi(\xi_\infty^*)$ such that, for all $x \in \mathcal{X}$,

$$\int_{\mathcal{X}} d^2(x, z, \xi_\infty^*, \theta^0) \, d\mu^*(z) \leqslant \sup_{z \in \mathcal{Z}} d(z, \xi_\infty^*, \theta^0).$$

Note that for $\mathcal{X} \subset \mathcal{Z}$ we have $\Psi_\infty(\xi) \geqslant p$, which implies a sufficient condition for I_∞-optimality: if $\mathcal{Z} \supset \mathcal{X}$ and $\Psi_\infty(\xi^*) = p$, then ξ^* is I_∞-optimal. Moreover, if $\mathcal{Z} = \mathcal{X}$, then the converse is also true. This follows because, in the case $\mathcal{Z} = \mathcal{X}$, I_∞-optimality is equivalent to D/G-optimality.

Example 5. As pointed out above, it is not necessary that the design space $\mathscr{X}$ and the integration space $\mathscr{Z}$ coincide. One instance involves prediction or extrapolation in which the design space is a proper subset of the integration space; for example, design support points may only be chosen in the interval $[a, b]$, but we wish to take account of the predicted variance outside this region; see Spruill (1987; 1990a, b). Our general equivalence theorem applies also in this situation. By way of illustration, consider again the problem of determining the I_1-optimal design for the quadratic model of Example 2. When we take $\mathscr{X} = \mathscr{Z} = [0, 1]$, we obtain ξ_1^* of Example 3.

In contrast, if we are interested in predicting out to $z = 2$, so that $\mathscr{Z} = [0, 2]$, but can only take design points in the interval $\mathscr{X} = [0, 1]$, the I_1-optimal design is

$$\xi_1^{**} = \left\{ \begin{matrix} 0 & \frac{1}{2} & 1 \\ 0.165 & 0.452 & 0.383 \end{matrix} \right\}.$$

The result of shifting the prediction space from $[0, 1]$ to $[0, 2]$ is thus an increase in the weight on the upper support point.

The application of the general equivalence theorem shows that the function $\phi_1(x, \xi_1^{**})$ does not cross over the line $y = 1$ on the design space $\mathscr{X} = [0, 1]$. The inverse maximal excess of the corresponding function for the design ξ_1^* over the line $y = 1$ gives a lower bound for the I_1-efficiency, if ξ_1^* is used to extrapolate on $\mathscr{Z} = [0, 2]$, that is $\text{eff}_1(\xi^*) \geqslant 56.49\%$. We also note that the exact I_1-efficiency of ξ_1^* for extrapolation on $\mathscr{Z} = [0, 2]$ is given by 80.5%.

A second instance in which the design and prediction spaces differ is interpolation, in which $\mathscr{Z}$ is a proper subset of $\mathscr{X}$; some practical examples are provided in Cook & Nachtsheim (1982). For the quadratic model with $\mathscr{X} = [0, 1]$ but with focus on the sub-space $\mathscr{Z} = [\frac{1}{4}, \frac{3}{4}]$, the I_1-optimal design is now

$$\xi_1^{***} = \left\{ \begin{matrix} 0 & \frac{1}{2} & 1 \\ 0.126 & 0.748 & 0.126 \end{matrix} \right\}$$

and, from Theorem 3, the efficiency of ξ_1^* for interpolation in $\mathscr{Z} = [\frac{1}{4}, \frac{3}{4}]$ is at least 55.66%. The exact efficiency is 80.23%.

4. Computational aspects

The I_L-optimal designs may be obtained using software packages such as Gauss, Matlab, SAS or S-Plus. For example, all I_L-optimal designs given in this paper were obtained using Gauss computer programs in which the objective function

$$\psi_L(\xi) = \left\{ \int_{\mathscr{X}} d^L(x, \xi, \theta^0) \, d\mu(x) \right\}^{1/L}$$

is specified for the given model function and initial parameter estimate θ^0. The Broyden–Fletcher–Goldfarb–Shanno minimisation routine, which calculates a numerical gradient and Hessian, is used to obtain the corresponding locally I_L-optimal design.

An alternative strategy would be to use the following algorithm which has been adapted from Fedorov (1972).

ALGORITHM
1. *Start with an arbitrary admissible p-point design* ξ_0 *and let* $K = 0$.

2. *Obtain* $\phi_L(x, \xi_K)$ *and* $x_K^* = \arg\sup_{x \in \mathscr{X}} \phi_L(x, \xi_K)$.
3. *For*

$$\xi_\alpha = \alpha\xi_K + (1 - \alpha)\begin{Bmatrix} x_K^* \\ 1 \end{Bmatrix},$$

find $\alpha_K^* \in [0, 1]$ *to minimise* $\phi_L(x_K^*, \xi_\alpha)$.
4. *Let*

$$\xi_{K+1} = \alpha_K^* \xi_K + (1 - \alpha_K^*)\begin{Bmatrix} x_K^* \\ 1 \end{Bmatrix}.$$

5. *Repeat the loop and go to step 2 until* $\{\sup_{x \in \mathscr{X}} \phi_L(x, \xi_K)\}^{-1} \geqslant 1 - \varepsilon$ *for some pre-specified* ε.

Under fairly general conditions, the sequence $\{\xi_K\}_{K \in N}$ obtained using this algorithm converges to the I_L-optimal design; see Fedorov (1972). Gauss computer programs to obtain I_L-optimal designs are available from T. E. O'Brien.

5. SOME EXTENSIONS

5·1. *Bayesian designs*

In nonlinear models the locally optimal design criterion ψ_L is useful if a best guess for the unknown parameter is available; see e.g. Chernoff (1953). If less precise knowledge of the unknown parameter is available a minimax approach (Fedorov & Atkinson, 1988) or a Bayesian (Chaloner & Verdinelli, 1995) or robust (Pronzato & Walter, 1985) version of the I_L-optimality criterion could be used. For the sake of brevity we concentrate on the last named criterion. More precisely, if π denotes a prior distribution for the unknown parameter $\theta \in \Theta$, we define a Bayesian I_L-optimal design with respect to the prior π as a design which minimises

$$\psi_L^\pi(\xi) := \left[\int_{\mathscr{X}} \left\{ \int_{\Theta} d(z, \xi, \theta)\, d\pi(\theta) \right\}^L d\mu(z) \right]^{1/L}.$$

Note that the case $L = 1$ gives the Bayesian A-optimality criterion considered by Chaloner & Larntz (1989) while the case $L = \infty$, $\mathscr{Z} = \mathscr{X}$ corresponds to the Bayesian D-optimality criterion (Pronzato & Walter, 1985). For an interpretation of the integrated variance $\int_{\Theta} d(z, \xi, \theta)\pi(d\theta)$ in terms of approximate expected utility we refer to Chaloner & Verdinelli (1995, p. 286). The criterion ψ_L^π is based on a posterior normal approximation which only depends on the data through the maximum likelihood estimator $\hat{\theta}$. An alternative criterion can be obtained by using a normal approximation of the joint posterior distribution of θ; see Chaloner & Verdinelli (1995, p. 286). This is based on the matrix $M(\xi, \theta) + n^{-1}R$, where R is the matrix of second derivatives of the logarithm of the prior density function or the precision matrix of the prior distribution.

All results of the previous sections can be easily transferred to the Bayesian criterion ψ_L^π. As a simple demonstration we consider the bound for the efficiency of Theorem 3. For the Bayesian ψ_L^π-optimality criterion this bound is given by

$$\mathrm{eff}_L^\pi(\xi) := \frac{\inf_\eta \psi_L^\pi(\eta)}{\psi_L^\pi(\xi)} \geqslant \frac{1}{\sup_{x \in \mathscr{X}} \phi_L^\pi(x, \xi)},$$

where

$$\phi_L^\pi(x, \xi) = \frac{\int_{\mathcal{Z}} \{\int_\Theta d(z, \theta, \xi) \, d\pi(\theta)\}^{L-1} \{\int_\Theta d^2(x, z, \xi, \theta) \, d\pi(\theta)\} \, d\mu(z)}{\int_{\mathcal{Z}} \{\int_\Theta d(z, \xi, \theta) \, d\pi(\theta)\}^L \, d\mu(z)}$$

if $0 \leqslant L < \infty$, and

$$\phi_\infty^\pi(x, \xi) = \frac{1}{p} \int_\Theta d(x, \xi, \theta) \, d\pi(\theta)$$

if $L = \infty$, $\mathcal{Z} = \mathcal{X}$. Similarly, an analogue of Theorem 2 shows that ξ_L^π minimises ψ_L^π if and only if $\phi_L^\pi(x, \xi_L^\pi) \leqslant 1$ for all $x \in \mathcal{X}$.

We illustrate these results for the intermediate product model (5) using independent uniform priors π_1 and π_2 on the intervals $[0\cdot3, 1\cdot1]$ and $[0\cdot15, 0\cdot25]$ for the parameters θ_1 and θ_2, respectively. By numerical calculation the Bayesian ψ_∞^π-optimal design with respect to the prior $\pi = \pi_1 \times \pi_2$, with $\mathcal{Z} = \mathcal{X} = [0, 20]$, is found as

$$\xi_\infty^\pi = \begin{Bmatrix} 1\cdot236 & 6\cdot15 \\ 0\cdot5 & 0\cdot5 \end{Bmatrix}.$$

The function $\phi_\infty^\pi(x, \xi_\infty^\pi)$ does not cross the line $y = 1$, showing that ξ_∞^π is Bayesian ψ_∞^π-optimal. We also investigated the locally I_∞-optimal design ($\theta_1^0 = 0\cdot7$, $\theta_2^0 = 0\cdot2$) given in Example 2. The inverse maximal excess over the line $y = 1$ gives the lower bound for the efficiency of ξ_∞^* with respect to the Bayesian I_∞-optimal design, that is $\mathrm{eff}_\infty^\pi(\xi) \geqslant 94\cdot76\%$.

5·2. *Model-robust designs*

Since in practice the regression model is not known, Läuter (1974, 1976) proposed a model-robust optimality criterion which averages various design criteria. Cook & Nachtsheim (1982) applied such a criterion to design for polynomial regression when the degree of the polynomial is unknown. In a similar way we can extend the I_L-optimality criterion to a design criterion for a class of models

$$\mathcal{F} = \{\eta_1(x, \theta_1), \dots, \eta_k(x, \theta_k)\},$$

$$\psi_L^{\mathcal{F}}(\xi) = \sum_{j=1}^k \alpha_j \left\{ \int_{\mathcal{X}} d_j^L(z, \xi, \theta_j) d\mu(z) \right\}^{1/L},$$

where $d_j(x, \xi, \theta_j)$ denotes the variance function for the model $\eta_j(x, \theta_j)$ ($j = 1, \dots, k$) and $\alpha_j \geqslant 0$ reflects the experimenter's belief about the adequacy of the model $\eta_j(x, \theta_j)$ ($j = 1, \dots, k$). A design is called I_L-optimal for the class of models $\mathcal{F}$ with respect to the prior $\alpha = (\alpha_1, \dots, \alpha_k)$ if it minimises the function $\psi_L^{\mathcal{F}}$ over the set of all approximate designs. The cases $L = \infty$ and $L = 1$ give similar optimality criteria ($\mathcal{Z} = \mathcal{X}$) to those considered by Läuter (1974) and Cook & Nachtsheim (1982). The corresponding function $\phi_L^{\mathcal{F}}$ is given by

$$\phi_L^{\mathcal{F}}(x, \xi) = \frac{\sum_{j=1}^k \alpha_j \int_{\mathcal{X}} d_j^{L-1}(z, \xi, \theta_j) d_j^2(x, z, \xi, \theta_j) \, d\mu(z)}{\sum_{j=1}^k \alpha_j \int_{\mathcal{X}} d_j^L(z, \xi, \theta_j) \, d\mu(z)} \qquad (0 \leqslant L < \infty)$$

and a design is $I_L^{\mathcal{F}}$-optimal for the class $\mathcal{F}$ if and only if $\phi_L^{\mathcal{F}}(x, \xi) \leqslant 1$ for all $x \in \mathcal{X}$. Similarly, a lower bound for the $I_L^{\mathcal{F}}$-efficiency of ξ is given by $\{\sup_{x \in \mathcal{X}} \phi_L^{\mathcal{F}}(x, \xi)\}^{-1}$.

As a simple example consider the class of quadratic polynomials $\mathscr{F}_2 = \{a_1 + b_1 x, a_2 + b_2 x + c_2 x^2\}$ and a uniform prior for the linear and quadratic models, that is, $\alpha_1 = \alpha_2 = 0\cdot5$. The $I_1^{\mathscr{F}_2}$-optimal design, for $\mathscr{Z} = \mathscr{X} = [0, 1]$, was found by Cook & Nachtsheim (1982) to be

$$\zeta_1^{\mathscr{F}_2} = \left\{ \begin{matrix} 0 & \frac{1}{2} & 1 \\ 0\cdot293 & 0\cdot414 & 0\cdot293 \end{matrix} \right\},$$

and the $I_\infty^{\mathscr{F}_2}$-optimal design, determined by the methods described in § 4, is given by

$$\zeta_\infty^{\mathscr{F}_2} = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

5·3. *Singular Fisher information*

We can often extend the analysis to the case of singular information matrices. Typically, singular optimal designs occur for prediction at fewer than p points in the design space, that is $\#\mathscr{Z} < p$, $\mathscr{Z} \subset \mathscr{X}$ (Kiefer & Wolfowitz, 1964; Herzberg & Cox, 1972; Spruill, 1987, 1990a, b).

More precisely, assume that ξ is a design such that $\eta(x, \theta^0)$ is estimable for all $z \in \text{supp}(\mu)$; that is the range inclusion (6) holds for all $z \in \text{supp}(\mu)$. The I_L-optimality can be exactly defined as in § 2, where the inverse in the variance function (3) has to be replaced by a generalised inverse of $M(\xi, \theta^0)$. The results of Theorems 2 and 3 can be rewritten as follows. If $0 \leqslant L < \infty$, then ξ_L^* is I_L-optimal if and only if there exists a generalised inverse, G, say, of $M(\xi, \theta^0)$ such that the inequality

$$\phi_L(x, \xi_L^*, G) := \frac{\displaystyle\int_{\mathscr{X}} d^{L-1}(z, \xi_L^*, \theta^0) \left\{ \left.\frac{\partial \eta(x, \theta)}{\partial \theta^T}\right|_{\theta=\theta_0} G \left.\frac{\partial \eta(z, \theta)}{\partial \theta}\right|_{\theta=\theta_0} \right\}^2 d\mu(z)}{\displaystyle\int_{\mathscr{X}} d^L(z, \xi_L^*, \theta^0)\, d\mu(z)} \leqslant 1$$

holds for all $x \in \mathscr{X}$. Moreover, the I_L-efficiency of a design ξ is bounded from below by

$$\text{eff}_L(\xi) \geqslant \frac{1}{\inf_G \sup_{x \in \mathscr{X}} \phi_L(x, \xi_L^*, G)},$$

where the infimum is taken over the set of all generalised inverses of $M(\xi, \theta^0)$. An analogous statement holds for the case $L = \infty$ and is omitted for the sake of brevity. These results can be derived from a general bound for design efficiencies given by Dette (1996).

ACKNOWLEDGEMENT

APPENDIX

Proofs of Theorems 2 and 3

In definition of ψ_L in Definition 1 put $L = -p$ where $p \in [-\infty, 0]$. We obtain

$$\psi_L^{-1}(\xi) = \left[\int_{\mathscr{X}} \{d(z, \xi, \theta_0)\}^{-p}\, d\mu(z) \right]^{1/p}.$$

It follows from results of Pukelsheim (1993, p. 77) that $\psi_L^{-1}(\xi)$ is a concave function on the set of designs with nonsingular information matrices. For fixed designs ξ, η with nonsingular information matrices define $\xi_\alpha = (1 - \alpha)\xi + \alpha\eta$, $g(\alpha) = \psi_L^{-1}(\xi_\alpha) = \psi_L^{-1}\{(1 - \alpha)\xi + \alpha\eta\}$ and assume that the order of differentiation and integration can be interchanged in the I_L-optimality criterion, i.e.

$$\frac{d}{d\alpha} \int_{\mathscr{X}} [\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)B(z)\}]^L \, d\mu(z) = \int_{\mathscr{X}} \frac{d}{d\alpha} [\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)B(z)\}]^L \, d\mu(z),$$

where

$$B(z) = \left(\frac{\partial \eta(z, \theta)}{\partial \theta}\right)\left(\frac{\partial \eta(z, \theta)}{\partial \theta}\right)^{\mathrm{T}}\bigg|_{\theta = \theta_0}.$$

Well-known results (Fedorov, 1972, p. 21) show that

$$\frac{d}{d\alpha} g(\alpha) = \{\psi_L(\xi_\alpha)\}^{-L-1} \int_{\mathscr{X}} [\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)B(z)\}]^{L-1}$$
$$\times \operatorname{tr}[M^{-1}(\xi_\alpha, \theta_0)\{M(\eta, \theta_0) - M(\xi, \theta_0)\}M^{-1}(\xi, \theta_0)B(z)] \, d\mu(z) \qquad \text{(A1)}$$

$$= \psi_L^{-L-1}(\xi_\alpha)\left(\int_{\mathscr{X}} \int_{\mathscr{X}} [\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)B(z)\}]^{L-1}\right.$$
$$\times [\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)B(x)M^{-1}(\xi_\alpha, \theta_0)B(z)\}$$
$$\left. - \operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)M(\xi, \theta_0)M^{-1}(\xi_\alpha, \theta_0)B(z)\}] \, d\mu(z) \, d\eta(x)\right) \qquad \text{(A2)}$$

$$\leqslant \psi_L^{-L-1}(\xi_\alpha)\left[\sup_{x \in \mathscr{X}} \int_{\mathscr{X}} d^{L-1}(z, \xi_\alpha, \theta_0)d^2(x, z, \xi_\alpha, \theta_0) \, d\mu(z)\right.$$
$$\left. - \int_{\mathscr{X}} d^{L-1}(z, \xi_\alpha, \theta_0) \operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)M(\xi, \theta_0)M^{-1}(\xi_\alpha, \theta_0)B(z)\} \, d\mu(z)\right].$$

Observing the definition of ϕ_L in (9) we obtain

$$\frac{d}{d\alpha} g(\alpha) \leqslant \psi_L^{-1}(\xi_\alpha) \sup_{x \in \mathscr{X}} \phi_L(x, \xi_\alpha) - \psi_L^{-L-1}(\xi_\alpha)$$
$$\times \int_{\mathscr{X}} d^{L-1}(z, \xi_\alpha, \theta_0)\operatorname{tr}\{M^{-1}(\xi_\alpha, \theta_0)M(\xi, \theta_0)M^{-1}(\xi_\alpha, \theta_0)B(z)\} \, d\mu(z). \qquad \text{(A3)}$$

The mean value theorem, the concavity of g and the inequality (A2) imply for some $\alpha^* \in (0, 1)$ that

$$\frac{1}{\psi_L(\eta)} - \frac{1}{\psi_L(\xi)} = g(1) - g(0) = \frac{d}{d\alpha} g(\alpha)\bigg|_{\alpha = \alpha^*} \leqslant \frac{d}{d\alpha} g(\alpha)\bigg|_{\alpha = 0^+} \leqslant \frac{\sup_{x \in \mathscr{X}} \phi_L(x, \xi)}{\psi_L(\xi)} - \frac{1}{\psi_L(\xi)},$$

which is equivalent to

$$\frac{\psi_L(\eta)}{\psi_L(\xi)} \geqslant \frac{1}{\sup_{x \in \mathscr{X}} \phi_L(x, \xi)}. \qquad \text{(A4)}$$

The statement of Theorem 3 is now obtained by inserting $\eta = \xi_L^*$ in this inequality. Moreover, for $\eta = \xi = \xi_L^*$ we obtain from (A3) that $\sup_{x \in \mathscr{X}} \phi_L(x, \xi_L^*) \geqslant 1$. On the other hand ξ_L^* minimises $\psi_L(\xi)$ and an argument similar to that used in (A1) and (A2) shows, for a Dirac measure η_x at x, that

$$0 \geqslant \frac{d}{d\alpha} \psi_L^{-1}\{(1 - \alpha)\xi_L^* + \alpha\eta_x\}\bigg|_{\alpha = 0^+} = \psi_L^{-L-1}(\xi_L^*) \int_{\mathscr{X}} d^{L-1}(z, \xi_L^*, \theta_0)\{d^2(x, z, \xi_L^*, \theta_0) - d^1(z, \xi_L^*, \theta_0)\} \, d\mu(z)$$
$$= \frac{\phi_L(x, \xi_L) - 1}{\psi_L(\xi_L)},$$

which establishes the converse inequality. Consequently ξ_L^* is I_L-optimal if and only if

$$\sup_{x \in \mathscr{X}} \phi_L(x, \xi_L^*) = 1,$$

which proves inequality (8) of Theorem 2. The assertion regarding the support points follows by standard arguments of optimal design theory and the proof will be omitted. $\square$

REFERENCES

ATKINSON, A. C., CHALONER, K., HERZBERG, A. M. & JURITZ, J. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics* **49**, 325–37.

ATKINSON, A. C. & DONEV, A. N. (1992). *Optimum Experimental Designs*. Oxford: Clarendon Press.

BOX, G. E. P. & LUCAS, H. (1959). Design of experiments in non-linear situations. *Biometrika* **46**, 77–90.

CHALONER, K. & LARNTZ, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *J. Statist. Plan. Infer.* **21**, 191–208.

CHALONER, K. & VERDINELLI, I. (1995). Bayesian experimental design: a review. *Statist. Sci.* **10**, 273–304.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586–602.

COOK, R. D. & NACHTSHEIM, C. J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics* **22**, 315–24.

COOK, R. D. & NACHTSHEIM, C. J. (1982). Model robust, linear-optimal designs. *Technometrics* **24**, 49–54.

DETTE, H. (1996). A lower bound for efficiencies with applications. In *Festschrift in Honor of Madan L. Puri on the Occasion of his 65th Birthday*, Ed. E. Brunner and M. Denker, pp. 111–24. Utrecht: VSP.

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23**, 255–62.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.

FEDOROV, V. V. & ATKINSON, A. C. (1988). The optimum design of experiments in the presence of uncontrolled variability and prior information. In *Optimal Designs of Experiments*, Ed. Y. Dodge, V. Fedorov and H. Wynn, pp. 327–44. New York: North Holland.

HAINES, L. M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. *Technometrics* **29**, 439–47.

HERZBERG, A. M. & COX, D. R. (1972). Some optimal designs for interpolation and extrapolation. *Biometrika* **59**, 551–61.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–79.

KIEFER, J. & WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12**, 363–6.

KIEFER, J. & WOLFOWITZ, J. (1964). Optimum extrapolation and interpolation designs I. *Ann. Inst. Statist. Math.* **16**, 79–108.

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Oper. Statist.* **5**, 370–96.

LÄUTER, E. (1976). Optimal multipurpose designs for regression models. *Math. Oper. Statist.* **7**, 51–68.

LEVINE, A. (1966). A problem in minimax variance polynomial extrapolation. *Ann. Math. Statist.* **37**, 898–903.

PRONZATO, L. & WALTER, E. (1985). Robust experimental design via stochastic approximation. *Math. Biosci.* **75**, 103–20.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

PUKELSHEIM, F. & RIEDER, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79**, 763–70.

SPRUILL, M. C. (1987). Optimal design for interpolation. *J. Statist. Plan. Infer.* **16**, 219–29.

SPRUILL, M. C. (1990a). Good designs for polynomial extrapolation. *J. Statist. Plan. Infer.* **26**, 149–59.

SPRUILL, M. C. (1990b). Optimal designs for multivariate interpolation. *J. Mult. Anal.* **34**, 141–55.

STUDDEN, W. J. (1971). Optimal designs for multivariate polynomial extrapolation. *Ann. Math. Statist.* **42**, 828–32.

STUDDEN, W. J. (1977). Optimal designs for integrated variance in polynomial regression. In *Statistical Decision Theory and Related Topics II*, Ed. S. S. Gupta and D. S. Moore, pp. 411–20. New York: Academic Press.

WONG, W. K. (1992). A unified approach to the construction of minimax designs. *Biometrika* **79**, 611–9.

WONG, W. K. & COOK, R. D. (1992). Heteroscedastic G-optimal designs. *J. R. Statist. Soc.* B **55**, 871–80.