



# Extended optimality criteria for optimum design in nonlinear regression

Andrej Pazman, Luc Pronzato

## ► To cite this version:

Andrej Pazman, Luc Pronzato. Extended optimality criteria for optimum design in nonlinear regression. mODa 10, Jun 2013, Lagow Lubuski, Poland. pp.195-202, 10.1007/978-3-319-00218-7\_23 . hal-00934877

**HAL Id: hal-00934877**

**<https://hal.science/hal-00934877>**

Submitted on 22 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Extended optimality criteria for optimum design in nonlinear regression

Andrej Pázman<sup>1</sup> and Luc Pronzato<sup>2</sup>

**Abstract** Among the major difficulties that one may encounter when estimating parameters in a nonlinear regression model are the non-uniqueness of the estimator, its instability with respect to small perturbations of the observations and the presence of local optimizers of the estimation criterion.

We show that these estimability issues can be taken into account at the design stage, through the definition of suitable design criteria. Extensions of  $E$ ,  $c$  and  $G$ -optimality criteria will be considered, which, when evaluated at a given  $\theta^0$  (local optimal design), account for the behavior of the model response  $\eta(\theta)$  for  $\theta$  far from  $\theta^0$ . In particular, they ensure some protection against close-to-overlapping situations where  $\|\eta(\theta) - \eta(\theta^0)\|$  is small for some  $\theta$  far from  $\theta^0$ . These extended criteria are concave, their directional derivative can be computed and necessary and sufficient conditions for optimality (Equivalence Theorems) can be formulated. They are not differentiable, but a relaxation based on maximum-entropy regularization is proposed to obtain concave and differentiable alternatives. When the design space is finite and the set of admissible  $\theta$  is discretized, their optimization forms a linear programming problem.

## 1 Introduction

We consider a nonlinear regression model with observations

$$y_i = y(x_i) = \eta(x_i, \bar{\theta}) + \varepsilon_i, \quad i = 1, \dots, N,$$

---

<sup>1</sup>Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia [pazman@fmph.uniba.sk](mailto:pazman@fmph.uniba.sk)

<sup>2</sup>Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, France [pronzato@i3s.unice.fr](mailto:pronzato@i3s.unice.fr)

where the errors  $\varepsilon_i$  satisfy  $\mathbb{E}(\varepsilon_i) = 0$ ,  $\text{var}(\varepsilon_i) = \sigma^2$  and  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  for  $i \neq j$ ,  $i, j = 1, \dots, N$ , and  $\bar{\theta} \in \Theta$ , a compact subset of  $\mathbb{R}^p$  such that  $\Theta \subset \text{int}(\bar{\Theta})$ , the closure of  $\text{int}(\Theta)$ . In a vector notation, we write

$$\mathbf{y} = \eta(\bar{\theta}) + \varepsilon, \quad \text{with } \mathbb{E}(\varepsilon) = \mathbf{0}, \quad \text{Var}(\varepsilon) = \sigma^2 \mathbf{I}_N, \quad (1)$$

where  $\eta(\theta) = \eta_X(\theta) = (\eta(x_1, \theta), \dots, \eta(x_N, \theta))^\top$ ,  $\mathbf{y} = (y_1, \dots, y_N)^\top$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^\top$  and  $X$  is the exact design  $(x_1, \dots, x_N)$ . We suppose that  $\eta(x, \theta)$  is twice continuously differentiable with respect to  $\theta \in \text{int}(\Theta)$  for any  $x \in \mathcal{X}$  compact. We shall consider design measures  $\xi$ , which correspond to probability measures on  $\mathcal{X}$ . The information matrix for the design  $X$  at  $\theta$  is

$$\mathbf{M}_\theta(X) = \sum_{i=1}^N \frac{\partial \eta(x_i, \theta)}{\partial \theta} \frac{\partial \eta(x_i, \theta)}{\partial \theta^\top}$$

and we have  $\mathbf{M}_\theta(\xi) = \int_{\mathcal{X}} [\partial \eta(x, \theta) / \partial \theta] [\partial \eta(x, \theta) / \partial \theta^\top] \xi(dx)$ . Denoting  $\xi_N$  the empirical design measure associated with  $X$ ,  $\xi_N = (1/N) \sum_{i=1}^N \delta_{x_i}$  with  $\delta_x$  the delta measure at  $x$ , we have  $\mathbf{M}_\theta(X) = N \mathbf{M}_\theta(\xi_N)$ .

The set of all hypothetical means of the observed vectors  $\mathbf{y}$  in the sample space  $\mathbb{R}^N$  forms the expectation surface  $\mathbb{S}_\eta = \{\eta(\theta) : \theta \in \Theta\}$ . Since  $\eta(\theta)$  is supposed to have continuous first and second-order derivatives in  $\text{int}(\Theta)$ ,  $\mathbb{S}_\eta$  is a smooth surface in  $\mathbb{R}^N$  with a (local) dimension given by  $r = \text{rank}[\partial \eta(\theta) / \partial \theta^\top]$ . If  $r = p$  (which means full rank), the model (1) is said regular. In regular models with no overlapping of  $\mathbb{S}_\eta$ , i.e. when  $\eta(\theta) = \eta(\theta')$  implies  $\theta = \theta'$ , the LS estimator

$$\hat{\theta} = \hat{\theta}_{LS}^N = \arg \min_{\theta \in \Theta} \|\mathbf{y} - \eta(\theta)\|^2$$

is uniquely defined, since as soon as the distributions of errors  $\varepsilon_i$  have probability densities (in the standard sense) it can be proven that  $\eta[\hat{\theta}_{LS}^N(\mathbf{y})]$  is unique with probability one, see Pázman (1984) and Pázman (1993, p.107). However, there is still a positive probability that the function  $\theta \longrightarrow \|\mathbf{y} - \eta(\theta)\|^2$  has a local minimizer different from the global one when the regression model is intrinsically curved in the sense of Bates and Watts (1980), i.e., when  $\mathbb{S}_\eta$  is a curved surface in  $\mathbb{R}^N$ , see Demidenko (1989, 2000). Moreover, a curved surface can “almost overlap”; that is, there may exist points  $\theta, \theta'$  such that  $\|\theta' - \theta\|$  is large but  $\|\eta(\theta') - \eta(\theta)\|$  is small (or even equals zero in case of strict overlapping). This phenomenon can cause serious difficulties in parameter estimation, leading to instabilities of the estimator, and one should thus attempt to reduce its effects by choosing an adequate experimental design. Note that putting restrictions on curvature measures is not enough: consider the case  $\dim(\theta) = 1$  with the overlapping  $\mathbb{S}_\eta$  formed by a circle of arbitrarily large radius and thus arbitrarily small curvature.

Important and precise results are available concerning the construction of subsets of  $\Theta$  where such effects are guaranteed not to occur, see, e.g.,

Chavent (1983, 1990, 1991); however, their exploitation for choosing adequate designs is far from straightforward. Also, the construction of designs with restricted curvatures, as proposed by Clyde and Chaloner (2002), is based on the curvature measures of Bates and Watts (1980) and uses derivatives of  $\eta(\theta)$  at a certain  $\theta$ ; this local approach is unable to catch the problem of overlapping for two points that are distant in the parameter space.

The aim of this paper is to present new optimality criteria for optimum design in nonlinear regression models that may reduce such effects, especially overlapping, and are at the same time closely related to classical optimality criteria like  $E$ ,  $c$  or  $G$ -optimality (in fact, they coincide with those criteria when the regression model is linear).

## 2 Extended (globalized) $E$ -optimality

Consider the design criterion defined by

$$\phi_{eE}(\xi) = \phi_{eE}(\xi; \theta) = \min_{\theta' \in \Theta} \left\{ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 (K + \|\theta' - \theta\|^{-2}) \right\},$$

to be maximized with respect to the design measure  $\xi$ , where  $K$  is some positive tuning constant (to be chosen in advance) and  $\|\cdot\|_{\xi}$  denotes the norm in  $\mathcal{L}_2(\xi)$ ; that is,  $\|\phi\|_{\xi} = [\int_{\mathcal{X}} \phi^2(x) \xi(dx)]^{1/2}$  for any  $\phi \in \mathcal{L}_2(\xi)$ .

Notice that in a nonlinear regression model  $\phi_{eE}(\cdot)$  depends on the value chosen for  $\theta$  and can thus be considered as a local optimality criterion. On the other hand, the criterion is global in the sense that it depends on the behavior of  $\eta(\cdot, \theta')$  for  $\theta'$  far from  $\theta$ . We could remove this (limited) locality by considering  $\phi_{MeE}(\xi) = \min_{\theta \in \Theta} \phi_{eE}(\xi)$ , but this will not be considered in what follows.

For a linear regression model with  $\eta(\theta) = \mathbf{F}(X)\theta + \mathbf{v}(X)$  and  $\Theta = \mathbb{R}^p$ , we have  $\min_{\theta' \in \Theta, \|\theta' - \theta\|^2 = \delta} \|\eta(\theta') - \eta(\theta)\|^2 = \min_{\theta' \in \Theta, \|\theta' - \theta\|^2 = \delta} (\theta' - \theta)^\top [N\mathbf{M}(\xi_N)](\theta' - \theta) = N\delta \lambda_{\min}[\mathbf{M}(\xi_N)]$ , so that  $\phi_{eE}(\xi) = \lambda_{\min}[\mathbf{M}(\xi)]$  for any  $K \geq 0$ , which corresponds to the  $E$ -optimality criterion. For a nonlinear regression model with no overlapping  $\phi_{eE}(\xi; \theta)$  can be made arbitrarily close to  $\lambda_{\min}[\mathbf{M}_{\theta}(\xi)]$  by choosing  $K$  large enough;  $\phi_{eE}(\cdot)$  can thus be considered as an *extended  $E$ -optimality criterion*. At the same time, choosing  $K$  not too large ensures some protection against  $\|\eta(\theta') - \eta(\theta)\|$  being small for some  $\theta'$  far from  $\theta$  for a  $\phi_{eE}$ -optimum design  $\xi_{eE}^*$ . Note that  $\xi_{eE}^*$  is necessarily non-degenerate, i.e.,  $\mathbf{M}(\xi_{eE}^*)$  is nonsingular.

### *Properties of $\phi_{eE}(\cdot)$*

$\phi_{eE}(\cdot)$  is the minimum of linear functions of  $\xi$  and is thus *concave*: for all  $\xi, \nu \in \Xi$ , the set of design measures on  $\mathcal{X}$ , for all  $\alpha \in [0, 1]$ , for all  $\theta \in \Theta$ ,

$\phi_{eE}[(1-\alpha)\xi + \alpha\nu] \geq (1-\alpha)\phi_{eE}(\xi) + \alpha\phi_{eE}(\nu)$ . It is also *positively homogeneous*:  $\phi_{eE}(a\xi) = a\phi_{eE}(\xi)$  for all  $\xi \in \Xi$ ,  $\theta \in \Theta$  and  $a > 0$ . Its concavity implies the existence of directional derivatives and we have the following, see, e.g., Dem'yanov and Malozemov (1974).

**Theorem 1.** *For any  $\xi, \nu \in \Xi$ , the directional derivative of the criterion  $\phi_{eE}(\cdot)$  at  $\xi$  in the direction  $\nu$  is given by*

$$F_{\phi_{eE}}(\xi; \nu) = \min_{\theta' \in \Theta_\theta(\xi)} \left\{ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\nu^2 (K + \|\theta' - \theta\|^{-2}) \right\} - \phi_{eE}(\xi),$$

where  $\Theta_\theta(\xi) = \left\{ \theta' \in \Theta : \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2 (K + \|\theta' - \theta\|^{-2}) = \phi_{eE}(\xi) \right\}$ .

We can write  $F_{\phi_{eE}}(\xi; \nu) = \min_{\theta' \in \Theta_\theta(\xi)} \int_{\mathcal{X}} \Psi_{eE}(x, \theta', \xi) \nu(dx)$ , where

$$\begin{aligned} \Psi_{eE}(x, \theta', \xi) &= (K + \|\theta' - \theta\|^{-2}) \\ &\times \left\{ [\eta(x, \theta') - \eta(x, \theta)]^2 - \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2 \right\}, \end{aligned} \quad (2)$$

and a necessary and sufficient condition for the optimality of a design measure  $\xi_{eE}^*$  for the criterion  $\phi_{eE}(\cdot)$  is that  $\sup_{\nu \in \Xi} F_{\phi_{eE}}(\xi^*; \nu) \leq 0$ . One should notice that  $\sup_{\nu \in \Xi} F_{\phi_{eE}}(\xi^*; \nu)$  is generally not obtained for  $\nu$  equal to a one-point (delta) measure, which prohibits the usage of classical vertex-direction algorithms for optimizing  $\phi_{eE}(\cdot)$ . This is why a regularized version  $\phi_{eE,\lambda}(\cdot)$  of  $\phi_{eE}(\cdot)$  is considered below, with the property that  $\sup_{\nu \in \Xi} F_{\phi_{eE,\lambda}}(\xi; \nu)$  is obtained when  $\nu$  is the delta measure  $\delta_{x^*}$  at some  $x^* \in \mathcal{X}$  (depending on  $\xi$ ).

*Maximum-entropy regularization of  $\phi_{eE}(\cdot)$*

The criterion  $\phi_{eE}(\cdot)$  can be equivalently defined by

$$\phi_{eE}(\xi) = \min_{\mu \in \mathcal{M}(\Theta)} \int_{\Theta} \left\{ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2 (K + \|\theta' - \theta\|^{-2}) \right\} \mu(d\theta'),$$

where  $\mathcal{M}(\Theta)$  denotes the set of probability measures on  $\Theta$ . We use the approach of Li and Fang (1997) and regularize  $\phi_{eE}(\xi)$  through a penalization of measures  $\mu$  having small (Shannon) entropy, with a penalty coefficient  $1/\lambda$  that sets the amount of regularization introduced. We then obtain the regularized criterion

$$\phi_{eE,\lambda}(\xi) = -\frac{1}{\lambda} \log \int_{\Theta} \exp \{-\lambda H_E(\xi, \theta')\} d\theta' \quad (3)$$

where

$$H_E(\xi, \theta') = \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2 (K + \|\theta' - \theta\|^{-2}). \quad (4)$$

It satisfies  $\lim_{\lambda \rightarrow \infty} \phi_{eE,\lambda}(\xi) = \phi_{eE}(\xi)$  for any  $\xi \in \Xi$  and the convergence is uniform when  $\Theta$  is a finite set. Moreover,  $\phi_{eE,\lambda}(\cdot)$  is concave, its directional derivative at  $\xi$  in the direction  $\nu$  is

$$F_{\phi_{eE},\lambda}(\xi; \nu) = \frac{\int_{\mathcal{X}} \int_{\Theta} \exp\{-\lambda H_E(\xi, \theta')\} \Psi_{eE}(x, \theta', \xi) d\theta' \nu(dx)}{\int_{\Theta} \exp\{-\lambda H_E(\xi, \theta')\} d\theta'}, \quad (5)$$

with  $\Psi_{eE}(x, \theta', \xi)$  given by (2). It is also differentiable (unlike  $\phi_{eE}(\cdot)$ ) and a necessary and sufficient condition for the optimality of  $\xi^*$  maximizing  $\phi_{eE,\lambda}(\cdot)$  is that  $\sup_{x \in \mathcal{X}} \int_{\Theta} \exp\{-\lambda H_E(\xi^*, \theta')\} \Psi_{eE}(x, \theta', \xi^*) d\theta' \leq 0$ . In order to facilitate computations, the integrals on  $\theta'$  in (3, 5) can be replaced by finite sums.

#### A solution via linear-programming

When  $\Theta$  is finite, i.e.,  $\Theta = \{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(m)}\}$ ,  $\phi_{eE}(\xi)$  can be written as  $\phi_{eE}(\xi) = \min_{j=1, \dots, m} H_E(\xi, \theta^{(j)})$ , with  $H_E(\xi, \theta')$  given by (4). If the design space  $\mathcal{X}$  is finite too,  $\mathcal{X} = \{x^{(1)}, x^{(2)}, \dots, x^{(q)}\}$ , then the determination of an optimal design measure for  $\phi_{eE}(\cdot)$  amounts to the determination of a scalar  $\gamma$  and of a vector of weights  $\mathbf{w} = (w_1, w_2, \dots, w_q)^\top$  such that  $\mathbf{c}^\top [\mathbf{w}^\top, \gamma]^\top$  is maximized, with  $\mathbf{c} = (0, 0, \dots, 0, 1)^\top$  and  $\mathbf{w}$  and  $\gamma$  satisfying the constraints

$$\begin{aligned} \sum_{i=1}^q w_i &= 1, \\ w_i &\geq 0, \quad i = 1, \dots, q, \\ \sum_{i=1}^q w_i [\eta(x^{(i)}, \theta^{(j)}) - \eta(x^{(i)}, \theta)]^2 (K + \|\theta^{(j)} - \theta\|^{-2}) &\geq \gamma, \quad j = 1, \dots, m. \end{aligned}$$

### 3 Extended (globalized) c-optimality

Define  $\phi_{ec}(\xi) = \min_{\theta' \in \Theta} \left\{ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2 (K + |h(\theta') - h(\theta)|^{-2}) \right\}$ , with  $K$  some positive constant. When  $\eta(x, \theta)$  and the scalar function  $h(\theta)$  are both linear in  $\theta$  we get

$$\phi_{ec}(\xi) = \min_{\theta' \in \Theta, \mathbf{c}^\top(\theta' - \theta) \neq 0} \frac{(\theta' - \theta)^\top \mathbf{M}(\xi) (\theta' - \theta)}{[\mathbf{c}^\top (\theta' - \theta)]^2}$$

and therefore  $\phi_{ec}(\xi) = [\mathbf{c}^\top \mathbf{M}^-(\xi) \mathbf{c}]^{-1}$ , which justifies that we consider  $\phi_{ec}(\xi)$  as an *extended c-optimality criterion*. Again,  $\phi_{ec}(\xi)$  can be approximated by  $[\mathbf{c}^\top \mathbf{M}^{-1}(\xi, \theta) \mathbf{c}]^{-1}$  for large  $K$ , whereas choosing  $K$  not too large ensures some protection against  $\|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_\xi^2$  being small for some  $\theta'$  such that  $h(\theta')$  is significantly different from  $h(\theta)$ . The criterion  $\phi_{ec}(\cdot)$  is concave and positively homogeneous; its concavity implies the existence of directional derivatives.

**Theorem 2.** *For any  $\xi, \nu \in \Xi$ , the directional derivative of the criterion  $\phi_{ec}(\cdot)$  at  $\xi$  in the direction  $\nu$  is given by*

$$F_{\phi_{ec}}(\xi; \nu) = \min_{\theta' \in \Theta_{\theta, c}(\xi)} \left\{ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 (K + |h(\theta') - h(\theta)|^{-2}) \right\} - \phi_{ec}(\xi),$$

where

$$\Theta_{\theta, c}(\xi) = \left\{ \theta' \in \Theta : \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 (K + |h(\theta') - h(\theta)|^{-2}) = \phi_{ec}(\xi) \right\}.$$

A necessary and sufficient condition for the optimality of  $\xi^*$  maximizing  $\phi_{ec}(\cdot)$  is that  $\sup_{\nu \in \Xi} F_{\phi_{ec}}(\xi^*; \nu) \leq 0$ . A regularized version of  $\phi_{ec}(\cdot)$  can be obtained through maximum-entropy regularization

$$\phi_{ec, \lambda}(\xi) = -\frac{1}{\lambda} \log \int_{\Theta} \exp \{-\lambda H_c(\xi, \theta')\} d\theta' \quad (6)$$

where  $H_c(\xi, \theta') = \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 (K + |h(\theta') - h(\theta)|^{-2})$ . The regularized criterion  $\phi_{ec, \lambda}(\cdot)$  is concave, differentiable with respect to  $\xi$ . Its directional derivative at  $\xi$  in the direction  $\nu$  is

$$F_{\phi_{ec, \lambda}}(\xi; \nu) = \frac{\int_{\mathcal{X}} \int_{\Theta} \exp \{-\lambda H_c(\xi, \theta')\} \Psi_{ec}(x, \theta', \xi) d\theta' \nu(dx)}{\int_{\Theta} \exp \{-\lambda H_c(\xi, \theta')\} d\theta'}, \quad (7)$$

where

$$\begin{aligned} \Psi_{ec}(x, \theta', \xi) &= (K + |h(\theta') - h(\theta)|^{-2}) \\ &\quad \times \left\{ [\eta(x, \theta') - \eta(x, \theta)]^2 - \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 \right\}. \end{aligned}$$

A necessary and sufficient condition for the optimality of  $\xi^*$  maximizing  $\phi_{ec, \lambda}(\cdot)$  is that  $\sup_{x \in \mathcal{X}} \int_{\Theta} \exp \{-\lambda H_c(\xi^*, \theta')\} \Psi_{ec}(x, \theta', \xi^*) d\theta' \leq 0$ . Again, in order to facilitate computations, the integrals in (6, 7) can be replaced by finite sums. A linear programming solution can be obtained when both  $\Theta$  and  $\mathcal{X}$  are finite, following an approach similar to that in Sect. 2.

## 4 Extended (globalized) $G$ -optimality

Following the same lines as above, we can also define an extended  $G$ -optimality criterion by

$$\phi_{GG}(\xi) = \min_{\theta' \in \Theta} \left[ \|\eta(\cdot, \theta') - \eta(\cdot, \theta)\|_{\xi}^2 \left\{ K + \frac{1}{\max_{x \in \mathcal{X}} [\eta(x, \theta') - \eta(x, \theta)]^2} \right\} \right]$$

with  $K$  some positive constant. The fact that it corresponds to the  $G$ -optimality criterion for a linear model can easily be seen, noticing that in the model (1) with  $\eta(x, \theta) = \mathbf{f}^T(x)\theta + v(x)$  we have

$$\begin{aligned} \left\{ \sup_{x \in \mathcal{X}} \frac{N}{\sigma^2} \text{var} \left[ \mathbf{f}^\top(x) \hat{\theta}_{LS}^N \right] \right\}^{-1} &= \inf_{x \in \mathcal{X}} \inf_{\mathbf{u} \in \mathbb{R}^p, \mathbf{u}^\top \mathbf{f}(x) \neq 0} \frac{\mathbf{u}^\top \mathbf{M}(X) \mathbf{u}}{[\mathbf{f}^\top(x) \mathbf{u}]^2} \\ &= \inf_{\mathbf{u} \in \mathbb{R}^p, \mathbf{u}^\top \mathbf{f}(x) \neq 0} \left[ \mathbf{u}^\top \mathbf{M}(X) \mathbf{u} \left\{ K + \frac{1}{\max_{x \in \mathcal{X}} [\mathbf{f}^\top(x) \mathbf{u}]^2} \right\} \right]. \end{aligned}$$

Directional derivatives can be computed and a regularized version can be constructed similarly to the cases of extended  $E$  and  $c$ -optimality; an optimal design can be obtained by linear programming when  $\Theta$  and  $\mathcal{X}$  are both finite.

## 5 Example

The model response is given by

$$\eta(\mathbf{x}, \theta) = \theta_1 \{\mathbf{x}\}_1 + \theta_1^3 (1 - \{\mathbf{x}\}_1) + \theta_2 \{\mathbf{x}\}_2 + \theta_2^2 (1 - \{\mathbf{x}\}_2), \quad \theta = (\theta_1, \theta_2)^\top,$$

with  $\mathbf{x} \in \mathcal{X} = [0, 1]^2$  and  $\{\mathbf{x}\}_i$  denoting the  $i$ -th component of  $\mathbf{x}$ . We consider local designs for  $\theta^0 = (1/8, 1/8)^\top$ . The classical  $D$  and  $E$ -optimal designs are supported on three and two points respectively,

$$\xi_D^*(\theta^0) \simeq \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \quad \xi_E^*(\theta^0) \simeq \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}.$$

We replace integrals by finite sums in (3, 5) and consider regular grids  $\mathcal{G}(\rho, M)$  formed of  $M$  points uniformly distributed on a circle centered at  $\theta^0$  with radius  $\rho$ . When  $\Theta = \cup_{i=1}^{20} \mathcal{G}(0.1 i, 100)$ ,  $K = 0.01$  and  $\lambda = 10^3$ , the opti-

mal design for  $\phi_{eE, \lambda}$  is  $\xi_{eE}^*(\theta^0) \simeq \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ . Fig. 1 presents

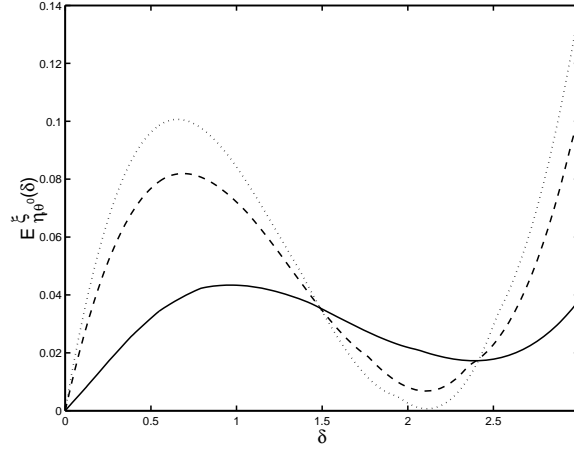
a plot of the function  $\delta \in \mathbb{R}^+ \longrightarrow E_{\eta, \theta^0}^\xi(\delta) = \min_{\theta' \in \Theta, \|\theta' - \theta^0\|^2 = \delta} \|\eta(\cdot, \theta') - \eta(\cdot, \theta^0)\|_\xi^2$  for the three designs  $\xi_D^*$ ,  $\xi_E^*$  and  $\xi_{eE}^*$ . The minimum of  $\|\eta(\cdot, \theta') - \eta(\cdot, \theta^0)\|_\xi$ , say for  $\|\theta' - \theta^0\| > 1$ , is 0.131 for  $\xi_{eE}^*$  and only 0.082 for  $\xi_D^*$ ; it is zero for  $\xi_E^*$  since the parameters are only locally estimable for this design.

**Acknowledgements** The work of Andrej Pázman was partly supported by the VEGA grant No. 2/0038/12.

## References

D.M. Bates and D.G. Watts. Relative curvature measures of nonlinearity. *Journal of Royal Statistical Society*, B42:1–25, 1980.





**Fig. 1** Plot of the function  $E_{\eta, \theta^0}^\xi(\cdot)$  for the three designs  $\xi_D^*$  (dashed line),  $\xi_E^*$  (dotted line) and  $\xi_{eE}^*$  (solid line).

- G. Chavent. Local stability of the output least square parameter estimation technique. *Matematicada Applicada e Computacional*, 2(1):3–22, 1983.
- G. Chavent. A new sufficient condition for the wellposedness of non-linear least-square problems arising in identification and control. In A. Bensoussan and J.L. Lions, editors, *Analysis and Optimization of Systems*, pages 452–463. Springer, 1990.
- G. Chavent. New size  $\times$  curvature conditions for strict quasiconvexity of sets. *SIAM Journal on Control and Optimization*, 29(6):1348–1372, 1991.
- M. Clyde and K. Chaloner. Constrained design strategies for improving normal approximations in nonlinear regression problems. *Journal of Statistical Planning and Inference*, 104:175–196, 2002.
- E.Z. Demidenko. *Optimization and Regression*. Nauka, Moscow, 1989. In Russian.
- E.Z. Demidenko. Is this the least squares estimate? *Biometrika*, 87(2):437–452, 2000.
- V.F. Dem'yanov and V.N. Malozemov. *Introduction to Minimax*. Dover, New York, 1974.
- X.-S. Li and S.-C. Fang. On the entropic regularization method for solving min-max problems with applications. *Mathematical Methods of Operations Research*, 46:119–130, 1997.
- A. Pázman. Nonlinear least squares – uniqueness versus ambiguity. *Math. Operationsforschung u. Statist. Ser. Statistics*, 15:323–336, 1984.
- A. Pázman. *Nonlinear Statistical Models*. Kluwer, Dordrecht, 1993.