Optimum design accounting for the global nonlinear behavior of the model
Andrej Pazman, Luc Pronzato

To cite this version:

HAL Id: hal-00875302
https://hal.archives-ouvertes.fr/hal-00875302v2
Submitted on 4 Jun 2014
OPTIMUM DESIGN ACCOUNTING FOR THE GLOBAL NONLINEAR BEHAVIOR OF THE MODEL

BY ANDREJ PÁZMAN *
Comenius University, Bratislava, Slovakia

AND

BY LUC PRONZATO
Laboratoire I3S, University of Nice-Sophia Antipolis/CNRS, France

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 are 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 and necessary and sufficient conditions for optimality (Equivalence Theorems) can be formulated. They are not differentiable, but when the design space is finite and the set $\Theta$ of admissible $\theta$ is discretized, optimal design forms a linear programming problem which can be solved directly or via relaxation when $\Theta$ is just compact. Several examples are presented.

1. Introduction. We consider a nonlinear regression model with observations

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

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

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

*The work of Andrej Pázman was supported by the VEGA grant No. 1/0163/13
MSC 2010 subject classifications: Primary 62K05; secondary 62J02
Keywords and phrases: optimal design, nonlinear least-squares, estimability, curvature
where $\eta_X(\theta) = (\eta(x_1, \theta), \ldots, \eta(x_N, \theta))^T$, $y = (y_1, \ldots, y_N)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$, and $X$ denotes the $N$-point exact design $(x_1, \ldots, x_N)$. The more general non-stationary (heteroscedastic) case where $\text{var}(\varepsilon_i) = \sigma_i^2(x_i)$ can easily be transformed into the model (1) with $\sigma_i^2 = 1$ via the division of $y_i$ and $\eta(x_i, \theta)$ by $\sigma(x_i)$. We suppose that $\eta(x, \theta)$ is twice continuously differentiable with respect to $\theta \in \text{int}(\Theta)$ for any $x \in \mathcal{X}$, a compact subset of $\mathbb{R}^d$. The model is assumed to be identifiable over $\mathcal{X}$; that is, we suppose that

\[(2) \quad \eta(x, \theta') = \eta(x, \theta) \quad \text{for all } x \in \mathcal{X} \implies \theta' = \theta.\]

We shall denote by $\Xi$ the set of design measures $\xi$, i.e., of probability measures on $\mathcal{X}$. The information matrix (for $\sigma_i^2 = 1$) for the design $X$ at $\theta$ is

\[M(X, \theta) = \sum_{i=1}^{N} \frac{\partial \eta(x_i, \theta)}{\partial \theta} \frac{\partial \eta(x_i, \theta)}{\partial \theta^T}\]

and, for any $\xi \in \Xi$, we shall write

\[M(\xi, \theta) = \int_{\mathcal{X}} \left[\partial \eta(x, \theta)/\partial \theta\right] \left[\partial \eta(x, \theta)/\partial \theta^T\right] \xi(dx).\]

Denoting $\xi_N = (1/N) \sum_{i=1}^{N} \delta_{x_i}$ the empirical design measure associated with $X$, with $\delta_x$ the delta measure at $x$, we have $M(X, \theta) = N M(\xi_N, \theta)$. Note that (2) implies the existence of a $\xi \in \Xi$ satisfying the Least-Squares (LS) estimability condition

\[(3) \quad \eta(x, \theta') = \eta(x, \theta) \text{-almost everywhere} \implies \theta' = \theta.\]

Given an exact $N$-point design $X$, the set of all hypothetical means of the observed vectors $y$ in the sample space $\mathbb{R}^N$ forms the expectation surface $S_\eta = \{\eta_X(\theta) : \theta \in \Theta\}$. Since $\eta_X(\theta)$ is supposed to have continuous first and second-order derivatives in $\text{int}(\Theta)$, $S_\eta$ is a smooth surface in $\mathbb{R}^N$ with a (local) dimension given by $r = \text{rank}[\partial \eta_X(\theta)/\partial \theta^T]$. If $r = p$ (which means full rank), the model (1) is said regular. In regular models with no overlapping of $S_\eta$, i.e. when $\eta_X(\theta) = \eta_X(\theta')$ implies $\theta = \theta'$, the LS estimator

\[(4) \quad \hat{\theta}_{LS} = \hat{\theta}_{LS}^N = \arg \min_{\theta \in \Theta} ||y - \eta_X(\theta)||^2\]

is uniquely defined with probability one (w.p.1). Indeed, when the distributions of errors $\varepsilon_i$ have probability densities (in the standard sense) it can be proven that $\eta[\hat{\theta}_{LS}(y)]$ is unique w.p.1, see Pázmán (1984) and Pázmán (1993, p. 107). However, there is still a positive probability that the function...
\[ \theta \rightarrow \| y - \eta_X(\theta) \|^2 \] 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 \( S_\eta \) 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 \( \theta \) and \( \theta' \) in \( \Theta \) such that \( \| \theta' - \theta \| \) is large but \( \| \eta_X(\theta') - \eta_X(\theta) \| \) 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. Classically, those issues are ignored at the design stage and the experiment is chosen on the basis of asymptotic local properties of the estimator. Even when the design relies on small-sample properties of the estimator, like in (Pázman and Pronzato, 1992; Gauchi and Pázmán, 2006), a non-overlapping assumption is used (see Pázmán (1993, pp. 66 and 157)) which permits to avoid the aforementioned difficulties. Note that putting restrictions on curvature measures is not enough: consider the case \( \text{dim}(\theta) = 1 \) with the overlapping \( S_\eta \) formed by a circle of arbitrarily large radius and thus arbitrarily small curvature (see the example in Sect. 2 below).

Important and precise results are available concerning the construction of subsets of \( \Theta \) where such difficulties 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 \( \eta_X(\theta) \) 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. Other design criteria using a second-order development of the model response, or an approximation of the density of \( \hat{\theta}_{LS} \) (Hamilton and Watts, 1985; Pronzato and Pázmán, 1994), are also inadequate.

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 \( E_\varepsilon \), \( c \) - or \( G \)-optimality (in fact, they coincide with those criteria when the regression model is linear). Classical optimality criteria focus on efficiency, i.e., aim at ensuring a precise estimation of \( \theta \), asymptotically, provided that the model is locally identifiable at \( \theta \). On the other hand, the new extended criteria account for the global behavior of the model and enforce identifiability.

An elementary example is given in the next section and illustrates the motivation of our work. The criterion of extended \( E \)-optimality is considered in
Sect. 3; its main properties are detailed and algorithms for the construction of optimal designs are presented. Sections 4 and 5 are respectively devoted to the criteria of extended $c$-optimality and extended $G$-optimality. Several illustrative examples are presented in Sect. 6. Section 7 suggests some extensions and further developments and Sect. 8 concludes.

2. An elementary motivating example.

Example 1. Suppose that $\theta \in \Theta = [0, 1]$ and that, for any design point $x = (t, u) \in X' = \{0, \pi/2\} \times [0, u_{\text{max}}]$, we have

$$\eta(x, \theta) = r \cos(t - u \theta),$$

with $r$ a known positive constant. We take $u_{\text{max}} = 7 \pi/4$; the difficulties mentioned below are even more pronounced for values of $u_{\text{max}}$ closer to $2\pi$.

We shall consider exclusively two-point designs $X = (x_1, x_2)$ of the form

$$x_1 = (0, u)^\top, \quad x_2 = (\pi/2, u)^\top$$

and denote $\nu_u$ the associated design measure, $\nu_u = (1/2)[\delta_{x_1} + \delta_{x_2}]$. We shall look for an optimal design, that is, an optimal choice of $u \in [0, u_{\text{max}}]$, where optimality is considered in terms of information.

It is easy to see that for any design $\nu_u$ we have

$$\eta_{X}(\theta) = \left( \begin{array}{c} \eta(x_1, \theta) \\ \eta(x_2, \theta) \end{array} \right) = \left( \begin{array}{c} r \cos(u \theta) \\ r \sin(u \theta) \end{array} \right).$$

The expectation surface is then an arc of a circle, with central angle $u$, see Figure 1 for the case $u = u_{\text{max}} = 7\pi/4$. The model is nonlinear but parametrically linear since the information matrix $M(X, \theta)$ for $\sigma^2 = 1$ (here scalar since $\theta$ is scalar) equals $r^2 u^2$ and does not depend on $\theta$. Also, the intrinsic curvature (see Sect. 6) is constant and equals $1/r$, and the model is also almost intrinsically linear if $r$ gets large.

Any classical optimality criterion ($A$-, $D$-, $E$-) indicates that one should observe at $u = u_{\text{max}}$, and setting a constraint on the intrinsic curvature is not possible here. However, if the true value of $\theta$ is $\bar{\theta} = 0$ and $\sigma^2$ is large enough, there is a chance that the LS estimator will be $\hat{\theta}_{\text{LS}} = 1$, and thus very far from $\bar{\theta}$, see Figure 1. The situation gets even worse if $u_{\text{max}}$ gets closer to $2\pi$, since $S_\eta$ then almost overlaps.

Now, consider $H_E(\nu_u, \theta) = (1/2) ||\eta_{X}(\theta) - \eta_{X}(\theta^0)||^2/|\theta - \theta^0|^2$, see (5), with $\theta^0 = 0$. For all $u \in [0, u_{\text{max}}]$, the minimum of $H_E(\nu_u, \theta)$ with respect to $\theta \in \Theta$ is obtained at $\theta = 1$, $H_E(\nu_u, 1) = r^2 |1 - \cos(u)|$ is then maximum in $[0, u_{\text{max}}]$ for $u = u_s = \pi$. This choice $u = u_s$ seems preferable to $u = u_{\text{max}}$ since the
Fig 1. Expectation surface $S_{\eta}$ for $\theta \in \Theta = [0, 1]$, $r = 1$ and $u = u_{\text{max}} = 7\pi/4$.

The rest of the paper will formalize these ideas and show how to implement them for general nonlinear models through the definition of suitable design criteria that can be easily optimized.


3.1. Definition of $\phi_{\text{E}}(\cdot)$. Take a fixed point $\theta^0$ in $\Theta$ and denote

\[
H_E(\xi, \theta) = H_E(\xi, \theta; \theta^0) = \frac{\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_2^2}{\|\theta - \theta^0\|^2},
\]

expectation surface $S_{\eta}$ is then a half-circle, so that $\eta_X(0)$ and $\eta_X(1)$ are as far away as possible. On the other hand, as shown in Sect. 3, $\min_{\theta \in \Theta} H_E(\nu_u, \theta)$ possesses most of the attractive properties of classical optimality criteria and even coincides with one of them in linear models.

Figure 2-left shows $H_E(\nu_u, \theta)$ as a function of $\theta$ for three values of $u$ and illustrates the fact that the minimum of $H_E(\nu_u, \theta)$ with respect to $\theta \in \Theta$ is maximized for $u = u^\star$. Figure 2-right shows that the design with $u = u_{\text{max}}$ (dashed line) is optimal locally at $\theta = \theta^0$, in the sense that it yields the fastest increase of $\|\eta_X(\theta) - \eta_X(\theta^0)\|$ as $\theta$ slightly deviates from $\theta^0$. On the other hand, $u = \pi$ maximizes $\min_{\theta \in \Theta} \|\eta_X(\theta) - \eta_X(\theta^0)\|/|\theta - \theta^0|$ (solid line) and realizes a better protection against the folding effect of $S_{\eta}$, at the price of a slightly less informative experiment for $\theta$ close to $\theta^0$. Smaller values of $u$ (dotted line) are worse than $u^\star$, both locally for $\theta$ close to $\theta^0$ and globally in terms of the folding of $S_{\eta}$.

The rest of the paper will formalize these ideas and show how to implement them for general nonlinear models through the definition of suitable design criteria that can be easily optimized.
where $\|\cdot\|_\xi$ denotes the norm in $L^2(\xi)$; that is, $\|l\|_\xi = \left( \int_X l^2(x) \xi(dx) \right)^{1/2}$ for any $l \in L^2(\xi)$. When $\xi$ is a discrete measure, like in the examples considered in the paper, then $\|l\|^2_\xi$ is simply the sum $\sum_{x : \xi(\{x\}) > 0} \xi(\{x\}) l^2(x)$.

The extended $E$-optimality criterion is defined by

$$
\phi_{eE}(\xi) = \phi_{eE}(\xi; \theta^0) = \min_{\theta \in \Theta} H_E(\xi, \theta),
$$

to be maximized with respect to the design measure $\xi$.

In a nonlinear regression model $\phi_{eE}(\cdot)$ depends on the value chosen for $\theta^0$ 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 $\theta^0$. This (limited) locality can be removed by considering $\phi_{MeE}(\xi) = \min_{\theta^0 \in \Theta} \phi_{eE}(\xi; \theta^0)$ instead of (6), but only the case of $\phi_{eE}(\cdot)$ will be detailed in the paper, the developments being similar for $\phi_{MeE}(\cdot)$, see Sect. 7.2.

For a linear regression model with $\eta(x, \theta) = f^T(x)\theta + v(x)$ and $\Theta = \mathbb{R}^p$, for any $\theta^0$ and any $\xi \in \Xi$ we have $\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|^2_\xi = (\theta - \theta^0)^T M(\xi)(\theta - \theta^0)$, so that

$$
\phi_{eE}(\xi) = \min_{\theta - \theta^0 \in \mathbb{R}^p} \frac{(\theta - \theta^0)^T M(\xi)(\theta - \theta^0)}{\|\theta - \theta^0\|^2} = \lambda_{\text{min}}[M(\xi)],
$$

the minimum eigenvalue of $M(\xi)$, and corresponds to the $E$-optimality criterion.

For a nonlinear model with $\Theta = \mathcal{B}(\theta^0, \rho)$, the ball with centre $\theta^0$ and radius $\rho$, direct calculation shows that

$$
\lim_{\rho \to 0} \phi_{eE}(\xi; \theta^0) = \lambda_{\text{min}}[M(\xi, \theta^0)].
$$
In a nonlinear regression model with larger \( \Theta \), the determination of an optimum design \( \xi^*_{eE} \) maximizing \( \phi_{eE}(\xi) \) ensures some protection against \( \|y(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi \) being small for some \( \theta \) far from \( \theta^0 \). In particular, when \( \theta^0 \in \text{int}(\Theta) \) then \( \phi_{eE}(\xi; \theta^0) = 0 \) if either \( M(\xi, \theta^0) \) is singular or \( \|y(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi = 0 \) for some \( \theta \neq \theta^0 \). Therefore, under the condition (2), \( \xi^*_{eE} \) satisfies the estimability condition (3) at \( \theta = \theta^0 \) and is necessarily non-degenerate, i.e., \( M(\xi^*_{eE}, \theta^0) \) is nonsingular, when \( \theta^0 \in \text{int}(\Theta) \) (provided that there exists a non-degenerate design in \( \Xi \)). Notice that (7) implies that \( \phi_{eE}(\xi; \theta^0) \leq \lambda_{\min}[M(\xi, \theta^0)] \) when \( \Theta \) contains some open neighborhood of \( \theta^0 \). In contrast with the \( E \)-optimality criterion, maximizing \( \phi_{eE}(\xi; \theta^0) \) in nonlinear models does not require computation of the derivatives of \( y(x, \theta) \) with respect to \( \theta \) at \( \theta^0 \), see the algorithms proposed in Sects. 3.3 and 3.4. Also note that the influence of points that are very far from \( \theta^0 \) can be suppressed by modification of the denominator of (5) without changing the relation with \( E \)-optimality, see Sect. 7.1.

Before investigating properties of \( \phi_{eE}(\cdot) \) as a criterion function for optimum design in the next section, we state a property relating \( \phi_{eE}(\xi) \) to the localization of the LS estimator \( \hat{\theta}_{LS} \).

**Theorem 1** For any given \( \theta \in \Theta \), the LS estimator \( \hat{\theta}_{LS} \) given by (4) in the model (1) satisfies

\[
\hat{\theta}_{LS} \in \Theta \cap \mathcal{B} \left( \theta, \frac{2 \|y - \eta_X(\theta)\|}{\sqrt{N} \sqrt{\phi_{eE}(\xi_N; \theta)}} \right),
\]

with \( \xi_N \) the empirical measure associated with the design \( X \) used to observe \( y \).

**Proof.** The result follows from the following chain of inequalities

\[
\|\hat{\theta}_{LS} - \theta\| \leq \frac{\|y(\cdot, \hat{\theta}_{LS}) - \eta(\cdot, \theta)\|_{\xi_N}}{\sqrt{\phi_{eE}(\xi_N; \theta)}} = \frac{\|\eta_X(\hat{\theta}_{LS}) - \eta_X(\theta)\|}{\sqrt{N} \sqrt{\phi_{eE}(\xi_N; \theta)}} \leq \frac{2 \|y - \eta_X(\theta)\|}{\sqrt{N} \sqrt{\phi_{eE}(\xi_N; \theta)}}.
\]

Note that although the bound (8) is tight in general nonlinear situations (due to the possibility that \( S_0 \) overlaps), it is often pessimistic. In particular, in the linear regression model \( y(x, \theta) = f^\top(x)\theta + v(x) \), direct calculation gives

\[
\|\hat{\theta}_{LS} - \theta\| \leq \sqrt{\lambda_{\max}((\mathbf{F}^\top \mathbf{F})^{-1})} \|y - \eta_X(\theta)\| = \frac{\|y - \eta_X(\theta)\|}{\sqrt{N} \sqrt{\phi_{eE}(\xi_N)}},
\]
where \( \mathbf{F} \) is the \( N \times p \) matrix with \( i \)th line equal to \( \mathbf{f}^\top(x_i) \). We also have \( \| \hat{\theta}_{LS} - \theta \| \leq \| y - \eta_X(\theta) \| / \sqrt{N} \sqrt{\varphi_{eE}(\xi_N, \theta)} \) in intrinsically linear models (with a flat expectation surface \( S_\eta \)) since then \( \| \eta_X(\hat{\theta}_{LS}) - \eta_X(\theta) \| \leq \| y - \eta_X(\theta) \| \).

In the following we shall omit the dependence in \( \theta^0 \) and simply write \( \varphi_{eE}(\xi) \) for \( \varphi_{eE}(\xi; \theta^0) \) when there is no ambiguity.

3.2. Properties of \( \varphi_{eE}(\cdot) \). As the minimum of linear functions of \( \xi \), \( \varphi_{eE}(\cdot) \) is concave: for all \( \xi, \nu \in \Xi \) and all \( \alpha \in [0, 1] \), \( \varphi_{eE}((1 - \alpha)\xi + \alpha\nu) \geq (1 - \alpha)\varphi_{eE}(\xi) + \alpha\varphi_{eE}(\nu) \). It is also positively homogeneous: \( \varphi_{eE}(a\xi) = a \varphi_{eE}(\xi) \) for all \( \xi \in \Xi \) and \( a > 0 \), see, e.g., Pukelsheim (1993, Chap. 5). The criterion of \( eE \)-efficiency can then be defined as

\[
E_{eE}(\xi) = \frac{\varphi_{eE}(\xi)}{\varphi_{eE}(\xi^*_eE)}, \quad \xi \in \Xi,
\]

where \( \xi^*_eE \) maximizes \( \varphi_{eE}(\xi) \).

The concavity of \( \varphi_{eE}(\cdot) \) implies the existence of directional derivatives and, due to the linearity in \( \xi \) of \( H_E(\xi, \theta) \), we have the following, see, e.g., Dem'yanov and Malozemov (1974).

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

\[
F_{\varphi_{eE}}(\xi; \nu) = \min_{\theta \in \Theta_{E}(\xi)} H_E(\nu, \theta) - \varphi_{eE}(\xi),
\]

where \( \Theta_{E}(\xi) = \{ \theta \in \Theta : H_E(\xi, \theta) = \varphi_{eE}(\xi) \} \).

Note that we can write \( F_{\varphi_{eE}}(\xi; \nu) = \min_{\theta \in \Theta_{E}(\xi)} \int_{\mathcal{X}} \Psi_{eE}(x, \theta, \xi) \nu(dx) \), where

\[
(9) \quad \Psi_{eE}(x, \theta, \xi) = \frac{[\eta(x, \theta) - \eta(x, \theta^0)]^2 - \| \eta(\cdot, \theta) - \eta(\cdot, \theta^0) \|^2_{\xi}}{\| \theta - \theta^0 \|^2}.
\]

Due to the concavity of \( \varphi_{eE}(\cdot) \), a necessary and sufficient condition for the optimality of a design measure \( \xi^*_eE \) is that

\[
(10) \quad \sup_{\nu \in \Xi} F_{\varphi_{eE}}(\xi^*_eE; \nu) = 0,
\]

a condition often called “Equivalence Theorem” in optimal design theory; see, e.g., Fedorov (1972); Silvey (1980). An equivalent condition is as follows.
Theorem 3 A design $\xi^*_E \in \Xi$ is optimal for $\phi_{eE}(\cdot)$ if and only if
\[
\max_{x \in \mathcal{X}} \int_{\Theta_E(\xi^*_E)} \Psi_E(x, \theta, \xi) \mu^*(d\theta) = 0 \text{ for some measure } \mu^* \in \mathcal{M}[\Theta_E(\xi^*_E)],
\]
the set of probability measures on $\Theta_E(\xi^*_E)$.

Proof. This is a classical result for maximin design problems, see, e.g., Fedorov and Hackl (1997, Sect. 2.6). We have
\[
0 \leq \sup_{\nu \in \Xi} F_{\phi_{eE}}(\xi; \nu) = \sup_{\nu \in \Xi} \min_{\theta \in \Theta_E(\xi)} \int_{\mathcal{X}} \int_{\Theta_E(\xi)} \Psi_E(x, \theta, \xi) \nu(dx) \mu(d\theta)
\]
\[
= \sup_{\nu \in \Xi} \min_{\mu \in \mathcal{M}[\Theta_E(\xi)]} \int_{\mathcal{X}} \int_{\Theta_E(\xi)} \Psi_E(x, \theta, \xi) \mu(d\theta) \nu(dx)
\]
\[
= \min_{\mu \in \mathcal{M}[\Theta_E(\xi)]} \sup_{\nu \in \Xi} \int_{\mathcal{X}} \int_{\Theta_E(\xi)} \Psi_E(x, \theta, \xi) \mu(d\theta) \nu(dx)
\]
\[
(12) = \min_{\mu \in \mathcal{M}[\Theta_E(\xi)]} \max_{x \in \mathcal{X}} \int_{\Theta_E(\xi)} \Psi_E(x, \theta, \xi) \mu(d\theta).
\]

Therefore, the necessary and sufficient condition (10) can be written as (11).

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 $\phi_{eE}(\cdot)$. Indeed, the minimax problem (12) has generally several solutions $x^{(i)}$ for $x, i = 1, \ldots, s$, and the optimal $\nu^*$ is then a linear combination $\sum_{i=1}^s w_i \delta_{x^{(i)}}$, with $w_i \geq 0$ and $\sum_{i=1}^s w_i = 1$; see Pronzato, Huang and Walter (1991) for developments on a similar difficulty in $T$-optimum design for model discrimination. This property, due to the fact that $\phi_{eE}(\cdot)$ is not differentiable, has the important consequence that the determination of a maximin-optimal design cannot be obtained via standard design algorithms used for differentiable criteria.

To avoid that difficulty, a regularized version $\phi_{eE,\lambda}(\cdot)$ of $\phi_{eE}(\cdot)$ is considered in (Pronzato and Pázman, 2013, Sects. 7.7.3 and 8.3.2), with the property that $\lim_{\lambda \to \infty} \phi_{eE,\lambda}(\xi) = \phi_{eE}(\xi)$ for any $\xi \in \Xi$ (the convergence being uniform when $\Theta$ is a finite set), $\phi_{eE}(\cdot)$ is concave and such 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 $\xi$). However, although $\phi_{eE,\lambda}(\cdot)$ is smooth for any finite $\lambda$, its maximization tends to be badly conditioned for large $\lambda$.

In the next section we show that optimal design for $\phi_{eE}(\cdot)$ reduces to linear programming when $\Theta$ and $\mathcal{X}$ are finite. This is an important property. An
algorithm based on a relaxation of the maximin problem is then considered in Sect. 3.4 for the case where \( \Theta \) is compact.

3.3. Optimal design via linear-programming (\( \Theta \) is finite). To simplify the construction of an optimal design, one may take \( \Theta \) as a finite set, \( \Theta = \Theta^{(m)} = \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}\} \); \( \phi_{eE}(\xi) \) can then be written as \( \phi_{eE}(\xi) = \min_{j=1, \ldots, m} H_E(\xi, \theta^{(j)}) \), with \( H_E(\xi, \theta) \) given by (5). If the design space \( \mathcal{X} \) is also finite, with \( \mathcal{X} = \{x^{(1)}, x^{(2)}, \ldots, x^{(\ell)}\} \), then the determination of an optimal design measure for \( \phi_{eE}(\cdot) \) amounts to the determination of a scalar \( t \) and of a vector of weights \( w = (w_1, w_2, \ldots, w_\ell)\) satifying the constraints

\[
\begin{align*}
\sum_{i=1}^\ell w_i &= 1, \\
w_i &\geq 0, \quad i = 1, \ldots, \ell, \\
\sum_{i=1}^\ell w_i h_i(\theta^{(j)}) &\geq t, \quad j = 1, \ldots, m, \tag{13}
\end{align*}
\]

where we denoted

\[
h_i(\theta) = \frac{[\eta(x^{(i)}, \theta) - \eta(x^{(i)}, \theta_0)]^2}{\|\theta - \theta_0\|^2}. \tag{14}
\]

This is a linear programming (LP) problem, which can easily be solved using standard methods (the simplex algorithm for instance), even for large \( m \) and \( \ell \). We shall denote by \((\hat{w}, \hat{t}) = LP_{eE}(\mathcal{X}, \Theta^{(m)})\) the solution of this problem.

We show below how a compact subset \( \Theta \) of \( \mathbb{R}^p \) with nonempty interior can be replaced by a suitable discretized version \( \Theta^{(m)} \) that can be enlarged iteratively.

3.4. Optimal design via relaxation and the cutting-plane method (\( \Theta \) is a compact subset of \( \mathbb{R}^p \)). Suppose now that \( \mathcal{X} \) is finite and that \( \Theta \) is a compact subset of \( \mathbb{R}^p \) with nonempty interior. In the LP formulation above, \((w, t)\) must satisfy an infinite number of constraints: \( \sum_{i=1}^{\ell} w_i h_i(\theta) \geq t \) for all \( \theta \in \Theta \), see (13). One may then use the method of Shimizu and Aiyoshi (1980) and consider the solution of a series of relaxed LP problems, using at step \( k \) a finite set of constraints only, i.e., consider \( \tilde{\Theta} \in \Theta^{(k)} \) finite. Once a solution \((w^k, t^k) = LP_{eE}(\mathcal{X}, \Theta^{(k)})\) of this problem is obtained, using a standard LP solver, the set \( \Theta^{(k)} \) is enlarged to \( \Theta^{(k+1)} = \Theta^{(k)} \cup \{\theta^{(k+1)}\} \) with
\[ \theta^{(k+1)} = \arg \min_{\theta \in \Theta} H_E(\mathbf{w}^k, \theta), \]

where, with a slight abuse of notation, we write \( H_E(\mathbf{w}, \theta) = H_E(\xi, \theta) \), see (5), when \( \xi \) allocates mass \( w_i \) at the support point \( x^{(i)} \in \mathcal{X} \) for all \( i \). This yields the following algorithm for the maximization of \( \phi_e E(\cdot) \).

0) Take any vector \( \mathbf{w}^0 \) of nonnegative weights summing to one, choose \( \epsilon > 0 \), set \( \Theta^{(0)} = \emptyset \) and \( k = 0 \).

1) Compute \( \theta^{(k+1)} \) given by (15), set \( \Theta^{(k+1)} = \Theta^{(k)} \cup \{ \theta^{(k+1)} \} \).

2) Use a LP solver to determine \( (\mathbf{w}^{k+1}, t^{k+1}) = \text{LP}_{\epsilon E}(\mathcal{X}, \Theta^{(k+1)}) \)

3) If \( \Delta^{k+1} = t^{k+1} - \phi_e E(\mathbf{w}^{k+1}) < \epsilon \), take \( \mathbf{w}^{k+1} \) as an \( \epsilon \)-optimal solution and stop; otherwise \( k \leftarrow k + 1 \), return to step 1.

The optimal value \( \phi_e E^* = \max_{\xi \in \Xi} \phi_e E(\xi) \) satisfies

\[ \phi_e E(\mathbf{w}^{k+1}) \leq \phi_e E^* \leq t^{k+1} \]

at every iteration, so that \( \Delta^{k+1} \) of step 3 gives an upper bound on the distance to the optimum in terms of criterion value.

The algorithm can be interpreted in terms of the cutting-plane method. Indeed, from (5) and (14) we have \( H_E(\mathbf{w}, \theta^{(j+1)}) = \sum_{i=1}^\ell w_i h_i(\theta^{(j+1)}) \) for any vector of weights \( \mathbf{w} \). From the definition of \( \theta^{(j+1)} \) in (15) we obtain

\[ \phi_e E(\mathbf{w}) \leq H_E(\mathbf{w}, \theta^{(j+1)}) = H_E(\mathbf{w}^j, \theta^{(j+1)}) + \sum_{i=1}^\ell h_i(\theta^{(j+1)})\{\mathbf{w} - \mathbf{w}^j\}_i \]

so that the vector with components \( h_i(\theta^{(j+1)}) \), \( i = 1, \ldots, \ell \), forms a subgradient of \( \phi_e E(\cdot) \) at \( \mathbf{w}^j \), which we denote \( \nabla \phi_e E(\mathbf{w}^j) \) below (it is sometimes called supergradient since \( \phi_e E(\cdot) \) is concave). Each of the constraints

\[ \sum_{i=1}^\ell w_i h_i(\theta^{(j+1)}) \geq t, \]

used in the LP problem of step 2, with \( j = 0, \ldots, k \), can be written as

\[ \nabla^T \phi_e E(\mathbf{w}^j) \mathbf{w} = \phi_e E(\mathbf{w}^j) + \nabla^T \phi_e E(\mathbf{w}^j)(\mathbf{w} - \mathbf{w}^j) \geq t. \]
Therefore, $w^{k+1}$ determined at step 2 maximizes the piecewise-linear approximation
\[
\min_{j=0,\ldots,k} \{ \phi_{E}(w^j) + \nabla^\top \phi_{E}(w^j)(w - w^j) \}
\]
of $\phi_{E}(w)$ with respect to the vector of weights $w$, and the algorithm corresponds to the cutting-plane method of Kelley (1960).

The only difficult step in the algorithm corresponds to the determination of $\theta^{(k+1)}$ in (15) when $\Theta$ is a compact set. We found that the following simple procedure is rather efficient. Construct a finite grid, or a space-filling design, $G^0$ in $\Theta$. Then, for $k = 0, 1, 2, \ldots$

\begin{equation}
\begin{cases}
(i) & \text{compute } \hat{\theta}^{k+1} = \arg \min_{\theta' \in G^k} H_E(w^k, \theta'); \\
(ii) & \text{perform a local minimization of } H_E(w^k, \theta) \\
& \text{with respect to } \theta \in \Theta, \text{ initialized at } \hat{\theta}^{k+1}; \\
(iii) & \text{let } \theta^{(k+1)} \text{ denote the solution;}
\end{cases}
\end{equation}

The optimal value $\phi_{E}(\xi^*_E)$ can then be approximated by $H_E(w^{k+1}, \theta^{(k+2)})$ when the algorithm stops (step 3).

The method of cutting planes is known to have sometimes rather poor convergence properties, see, e.g., Bonnans et al. (2006, Chap. 9), Nesterov (2004, Sect. 3.3.2). A significant improvement consists in restricting the search for $w^{k+1}$ at step 2 to some neighborhood of the best solution obtained so far, which forms the central idea of bundle methods, see Lemaréchal, Nemirovski and Nesterov (1995), Bonnans et al. (2006, Chaps. 9-10). In particular, the level method of Nesterov (2004, Sect. 3.3.3) adds a quadratic-programming step to each iteration of the cutting planes algorithm presented above; one may refer for instance to Pronzato and Pázman (2013, Sect. 9.5.3) for an application of the level method to design problems. Notice that any linear constraint on $w$ can easily be taken into account in addition to those in (13), so that the method directly applies to optimal design with linear cost-constraints, see, e.g., Fedorov and Leonov (2014, Sect. 4.2).

4. Extended (globalized) $c$-optimality.

4.1. Definition and properties. Consider the case where one wants to estimate a scalar function of $\theta$, denoted by $g(\theta)$, possibly nonlinear. We assume that
\[
c = c(\theta) = \frac{\partial g(\theta)}{\partial \theta} \bigg|_{\theta = \theta^0} \neq 0.
\]
Denote

\[ H_c(\xi, \theta) = H_c(\xi, \theta; \theta^0) = \frac{||\eta(\cdot, \theta) - \eta(\cdot, \theta^0)||_2^2}{|g(\theta) - g(\theta^0)|^2} \]

and consider the design criterion defined by

\[ \phi_{ec}(\xi) = \min_{\theta \in \Theta} H_c(\xi, \theta), \]

to be maximized with respect to the design measure \( \xi \).

When \( \eta(x, \theta) \) and the scalar function \( g(\theta) \) are both linear in \( \theta \), with \( g(\theta) = c^T \theta \), we get

\[ \phi_{ec}(\xi) = \min_{\theta \in \Theta} \left( \theta - \theta^0 \right)^T M(\xi)(\theta - \theta^0) \left( \theta - \theta^0 \right)^T \]

and therefore \( \phi_{ec}(\xi) = [c^T M^{-1}(\xi) c]^{-1} \), using the well-known formula \( c^T M^{-1} c = \max_{\alpha \neq 0} (c^T \alpha)^2 / (\alpha^T M \alpha) \), c.f. Harville (1997, eq. 10.4). Also, for a nonlinear model with \( \Theta = B(\theta^0, \rho) \) and a design \( \xi \) such that \( M(\xi, \theta^0) \) has full rank, one has

\[ \lim_{\rho \to 0} \phi_{ec}(\xi) = [c^T M^{-1}(\xi, \theta^0) c]^{-1}, \]

which justifies that we consider \( \phi_{ec}(\xi) \) as an extended \( c \)-optimality criterion.

At the same time, in a nonlinear situation with larger \( \Theta \) the determination of an optimal design \( \xi_{ec}^* \) maximizing \( \phi_{ec}(\xi) \) ensures some protection against \( ||\eta(x, \theta) - \eta(x, \theta^0)||_2^2 \) being small for some \( \theta \) such that \( g(\theta) \) is significantly different from \( g(\theta^0) \). The condition (2) guarantees the existence of a \( \xi \in \Xi \) such that \( \phi_{ec}(\xi) > 0 \) and thus the LS estimability of \( g(\theta) \) at \( \theta^0 \) for \( \xi_{ec}^* \), that is,

\[ \eta(x, \theta) = \eta(x, \theta^0) \xi_{ec}^*-\text{almost everywhere} \implies g(\theta) = g(\theta^0), \]

see Pronzato and Pázman (2013, Sect. 7.4.4). When \( \Theta \) contains an open neighborhood of \( \theta^0 \), then \( \phi_{ec}(\xi) \leq [c^T M^{-1}(\xi, \theta^0) c]^{-1} \).

Similarly to \( \phi_{eE}(\cdot) \), the criterion \( \phi_{ec}(\cdot) \) is concave and positively homogeneous; its concavity implies the existence of directional derivatives.

**Theorem 4** For any \( \xi, \nu \in \Xi \), the directional derivative of the criterion \( \phi_{ec}(\cdot) \) at \( \xi \) in the direction \( \nu \) is given by

\[ F_{\phi_{ec}}(\xi; \nu) = \min_{\theta \in \Theta_c(\xi)} H_c(\nu, \theta) - \phi_{ec}(\xi), \]

where \( \Theta_c(\xi) = \{ \theta \in \Theta : H_c(\xi, \theta) = \phi_{ec}(\xi) \} \).

A necessary and sufficient condition for the optimality of $\xi^*$ maximizing $\phi_{\text{ec}}(\cdot)$ is that $\sup_{\nu \in \Xi} F_{\phi_{\text{ec}}}(\xi^*; \nu) = 0$, which yields an Equivalence Theorem similar to Th. 3.

When both $\Theta$ and $\mathcal{X}$ are finite, an optimal design for $\phi_{\text{ec}}(\cdot)$ is obtained by solving a LP problem. Compared with Sect. 3.3, we simply need to substitute $H_e$ for $H_E$ and use $h_i(\theta) = [\eta(x^{(i)}, \theta) - \eta(x^{(i)}, \theta^0)]^2 / |g(\theta) - g(\theta^0)|^2$, $i = 1, \ldots, \ell$, instead of (14). Also, a relaxation method similar to that in Sect. 3.4 can be used when $\Theta$ is a compact subset of $\mathbb{R}^p$.

5. Extended (globalized) $G$-optimality. Following the same lines as above, we can also define an extended $G$-optimality criterion by

$$
\phi_{\text{ec}}(\xi) = \min_{\theta \in \Theta} \frac{\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_2^2}{\max_{x \in \mathcal{X}} [\eta(x, \theta) - \eta(x, \theta^0)]^2}.
$$

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 $\eta(x, \theta) = f^\top(x)\theta + v(x)$ we have

$$
\left\{ \sup_{x \in \mathcal{X}} \frac{N}{\sigma^2} \var \left[ f^\top(x) \hat{\theta}_{\text{LS}} \right] \right\}^{-1} = \inf_{x \in \mathcal{X}} \left[ f^\top(x) M^{-1}(\xi_N) f(x) \right]^{-1}
$$

$$
= \inf_{x \in \mathcal{X}} \inf_{u \in \mathbb{R}^p} \frac{u^\top M(\xi_N) u}{\| f^\top(x) u \|^2}
$$

$$
= \inf_{u \in \mathbb{R}^p} \max_{x \in \mathcal{X}} \frac{\| f^\top(x) u \|^2}{\| u \|^2},
$$

where $\xi_N$ denotes the empirical design measure corresponding to $X$, assumed to be nonsingular, and the second equality follows from Harville (1997, eq. 10.4). The equivalence theorem of Kiefer and Wolfowitz (1960) indicates that $D$- and $G$-optimal designs coincide; therefore, $D$-optimal designs are optimal for $\phi_{\text{ec}}(\cdot)$ in linear models. Moreover, the optimum (maximum) value of $\phi_{\text{ec}}(\xi)$ equals $1/p$ with $p = \dim(\theta)$.

In a nonlinear model, a design $\xi^*_G$ maximizing $\phi_{\text{ec}}(\xi)$ satisfies the estimability condition (3) at $\theta = \theta^0$. Indeed, $\max_{x \in \mathcal{X}} [\eta(x, \theta) - \eta(x, \theta^0)]^2 > 0$ for any $\theta \neq \theta^0$ from (2), so that there exists some $\xi \in \Xi$ such that $\phi_{\text{ec}}(\xi) > 0$. Therefore, $\phi_{\text{ec}}(\xi^*_G) > 0$, and $\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_{\xi_{G}}^2 = 0$ implies that $\eta(x, \theta) = \eta(x, \theta^0)$ for all $x \in \mathcal{X}$, that is, $\theta = \theta^0$ from (2). Notice that when $\Theta$ contains an open neighborhood of $\theta^0$, then $\phi_{\text{ec}}(\xi) \leq 1/p$ for all $\xi \in \Xi$.

Again, directional derivatives can easily be computed and an optimal design can be obtained by linear programming when $\Theta$ and $\mathcal{X}$ are both finite, or with the algorithm of Sect. 3.4 when $\mathcal{X}$ is finite but $\Theta$ has nonempty
interior. Note that there are now \( m \times \ell \) inequality constraints in (13), given by
\[
\sum_{i=1}^{\ell} w_i h_i(\theta^{(j)}, x^{(k)}) \geq t, \text{ } j = 1, \ldots, m, \text{ } k = 1, \ldots, \ell,
\]
where now
\[
h_i(\theta, x) = \left[ \frac{\eta(x^{(i)}, \theta) - \eta(x^{(i)}, \theta^0)}{\eta(x^{(i)}, \theta) - \eta(x, \theta^0)} \right]^2.
\]
Also note that in the algorithm of Sect. 3.4 we need to construct two sequences of sets, \( \Theta^{(k)} \) and \( X^{(k)} \), with \( \Theta^{(k+1)} = \Theta^{(k)} \cup \{ \theta^{(k+1)} \} \) and \( X^{(k+1)} = X^{(k)} \cup \{ \hat{x}^{(k+1)} \} \) at step 2, and (15) replaced by
\[
\{ \theta^{(k+1)}, \hat{x}^{(k+1)} \} = \arg \min_{\{\theta, x\} \in \Theta \times X} \| \eta(\cdot, \theta) - \eta(\cdot, \theta^0) \|_{\xi_k}^2,
\]
with \( \xi_k \) the design measure corresponding to the weights \( w^k \).

6. Examples. We shall use the common notation
\[
\xi = \left\{ \begin{array}{c} x_1 \cdots x_m \\ w_1 \cdots w_m \end{array} \right\}
\]
for a discrete design measure with \( m \) support points \( x_i \) and such that \( \xi(\{x_i\}) = w_i, \text{ } i = 1, \ldots, m. \) In the three examples considered we indicate the values of the parametric, intrinsic and total measure of curvatures at \( \theta^0 \) (for \( \sigma = 1 \)), see Tables 1, 2 and 3. They are not used for the construction of optimal designs, and the examples illustrate the fact that they provide information on the local behavior only (at \( \theta^0 \)), so that a small curvature does not mean good performance in terms of extended optimality. They are given by
\[
C_{\text{int}}(\xi, \theta) = \sup_{u \in \mathbb{R}^p \setminus \{0\}} \frac{\| [(I - P_\theta) \sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta) / \partial \theta_i \partial \theta_j] u_j \|_\xi}{u^\top M(\xi, \theta) u},
\]
\[
C_{\text{par}}(\xi, \theta) = \sup_{u \in \mathbb{R}^p \setminus \{0\}} \frac{\| P_\theta \sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta) / \partial \theta_i \partial \theta_j] u_j \|_\xi}{u^\top M(\xi, \theta) u},
\]
\[
C_{\text{tot}}(\xi, \theta) = \sup_{u \in \mathbb{R}^p \setminus \{0\}} \frac{\| \sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta) / \partial \theta_i \partial \theta_j] u_j \|_\xi}{u^\top M(\xi, \theta) u} \leq C_{\text{int}}(\xi, \theta) + C_{\text{par}}(\xi, \theta),
\]
with \( P_\theta \) the projector
\[
(P_\theta f)(x') = \frac{\partial \eta(x', \theta)}{\partial \theta} M^{-1}(\xi, \theta) \int_{\mathcal{X}} \frac{\partial \eta(x, \theta)}{\partial \theta} f(x) \xi(dx),
\]
and correspond to the original measures of nonlinearity of Bates and Watts (1980) for $\sigma = 1$, with an adaptation to the use of a design measure $\xi$ instead of an exact design $(x_1, \ldots, x_N)$. The connection with the curvature arrays of Bates and Watts (1980) is presented in (Pázmán, 1993, Sect. 5.5); a procedure for their numerical computation is given in (Bates and Watts, 1980; Ratkowsky, 1983).

All computations are performed in Matlab on a bi-processor PC (2.5 Ghz) with 64 bits, equipped with 32 Go RAM. Classical optimal designs ($D$-, $E$- and $c$-optimality) are computed with the cutting-plane method, see Pronzato and Pázmán (2013, Sect. 9.5.3); LP problems are solved with simplex algorithm; we use sequential quadratic programming for the local minimization of $H(w^k, \theta)$ that yields $\theta^{(k+1)}$ in (16)-(ii).

Example 2. This example is artificial and constructed to illustrate the possible pitfall of using a local approach (here $E$-optimal design) for designing an experiment. The model response is given by

$$\eta(x, \theta) = \theta_1 x_1 + \theta_3 (1 - x_1) + \theta_2 x_2 + \theta_2^2 (1 - x_2), \quad \theta = (\theta_1, \theta_2)^	op,$$

with $x \in \mathcal{X} = [0, 1]^2$ and $\{x\}_i$ denoting the $i$-th component of $x$. We consider local designs for $\theta^0 = (1/8, 1/8)^\top$. One may notice that the set $\{\partial \eta(x, \theta)/\partial \theta |_{\theta^0} : x \in \mathcal{X}\}$ is the rectangle $[3/64, 1] \times [1/4, 1]$, so that optimal designs for any isotonic criterion function of the information matrix $M(\xi)$ are supported on the vertices $(0, 1), (1, 0)$ and $(1, 1)$ of $\mathcal{X}$. The classical $D$- and $E$-optimal designs are supported on three and two points respectively,

$$\xi^*_{D, \theta^0} \simeq \left\{ \begin{array}{ccc} 0 & 1 & 1 \\ 0.4134 & 0.3184 & 0.2682 \end{array} \right\}, \quad \xi^*_{E, \theta^0} \simeq \left\{ \begin{array}{ccc} 0 & 1 \\ 0.5113 & 0.4887 \end{array} \right\}.$$

When only the design points $x_1 = (0 1)^\top$ and $x_2 = (1 0)^\top$ are used, the parameters are only locally estimable. Indeed, the equations in $\theta'$

$$\eta(x_1, \theta') = \eta(x_1, \theta), \quad \eta(x_2, \theta') = \eta(x_2, \theta)$$

give not only the trivial solutions $\theta'_1 = \theta_1$ and $\theta'_2 = \theta_2$ but also $\theta'_3$ and $\theta'_2$ as roots of two univariate polynomials of the fifth degree (with coefficients depending on $\theta$). Since these polynomials always admit at least one real root, at least one solution exists for $\theta'$ that is different from $\theta$. In particular, the vector $\theta'^0 = (-0.9760, 1.0567)^\top$ gives approximately the same values as $\theta^0$ for the responses at $x_1$ and $x_2$. 
Direct calculations indicate that, for any $\theta$, the maximum of $||\eta(\cdot, \theta) - \eta(\cdot, \theta^0)||^2_\xi$ with respect to $\xi \in \Xi$ is reached for a measure supported on $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. Also, the maximum of $[\eta(x, \theta) - \eta(x, \theta^0)]^2$ with respect to $x$ is attained on the same points. We can thus restrict our attention to the design space $\mathcal{X} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. We take $\Theta = [-3, 4] \times [-2, 2]$ and use the algorithm of Sect. 3.4, with the grid $G_0$ of (16)-(iii) given by a random Latin hypercube design with 10,000 points in $[0, 1]^2$ renormalized to $\Theta$ (see, e.g., Tang (1993)), to determine optimal designs for $\varphi_{eE}(\cdot)$ and $\varphi_{eG}(\cdot)$. When initialized with the uniform measure on the four points of $\mathcal{X}$, and with $\epsilon = 10^{-10}$, the algorithm stops after 46 and 15 iterations, respectively requiring 0.67 s and 0.28 s in total, and gives the designs

$$
\xi_{eE, \theta^0}^* \simeq \left\{ \begin{array}{ccc}
(0) & (0) & (1) \\
0 & 1 & 1 \\
0.32 & 0.197 & 0.483
\end{array} \right\},
\xi_{eG, \theta^0}^* \simeq \left\{ \begin{array}{ccc}
(0) & (0) & (1) \\
0 & 1 & 1 \\
0.258 & 0.258 & 0.258 & 0.226
\end{array} \right\}.
$$

The performances of the designs $\xi_D^*$, $\xi_E^*$, $\xi_{eE}^*$ and $\xi_{eG}^*$ are given in Table 1. The values $\varphi_{eE}(\xi_E^*) = \varphi_{eG}(\xi_E^*) = 0$ indicate that $E$-optimal design is not suitable here, the model being only locally identifiable for $\xi_E^*$. The parametric, intrinsic and total measures of curvature at $\theta^0$ (for $\sigma^2 = 1$) are also indicated in Table 1, see Pronzato and Pázman (2013, p. 223). Notice that the values of these curvature at $\theta^0$ do not reveal any particular difficulty concerning $\xi_E^*$, but that the lack of identifiability for this design is pointed out by the extended optimality criteria.

This example is very particular and situations where the model is locally, but not globally, identifiable are much more common: in that case, (2) is only satisfied locally, for $\theta'$ in a neighborhood of $\theta$, and one may refer, e.g., to Walter (1987); Walter and Pronzato (1995) for a precise definition and examples. The lack of global identifiability would then not be detected by classical optimal design, but the maximum of $\varphi_{eE}(\cdot)$ and $\varphi_{eG}(\cdot)$ would be zero for $\Theta$ large enough, showing that the model is not globally identifiable.

**Example 3.** Consider the regression model (one-compartment with first-order absorption input) used in (Atkinson et al., 1993),

$$
\eta(x, \theta) = \theta_1[\exp(-\theta_2 x) - \exp(-\theta_3 x)], \ \theta = (\theta_1, \theta_2, \theta_3)^\top, \ x \in \mathbb{R}^+,
$$

\[19\]
A. PÁZMAN AND L. PRONZATO

Table 1

Performances of designs $\xi^*_D$, $\xi^*_E$, $\xi^*_eE$ and $\xi^*_eG$ and curvature measures at $\theta^0$ in Example 2; $\det^{1/3} = \phi_D(\xi) = \{\det[M(\xi, \theta^0)]\}^{1/3}$, $\lambda_{\min} = \phi_E(\xi) = \lambda_{\min}[M(\xi, \theta^0)]$. The optimal (maximum) values of the criteria are indicated in boldface.

<table>
<thead>
<tr>
<th>$\xi$</th>
<th>$\det^{1/3}$</th>
<th>$\lambda_{\min}$</th>
<th>$\phi_eE$</th>
<th>$\phi_eG$</th>
<th>$C_{par}$</th>
<th>$C_{int}$</th>
<th>$C_{tot}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi^*_D$</td>
<td>0.652</td>
<td>0.273</td>
<td>3.16 · 10^{-3}</td>
<td>0.108</td>
<td>1.10</td>
<td>0.541</td>
<td>1.22</td>
</tr>
<tr>
<td>$\xi^*_E$</td>
<td>0.625</td>
<td>0.367</td>
<td>0</td>
<td>0</td>
<td>1.19</td>
<td>0</td>
<td>1.19</td>
</tr>
<tr>
<td>$\xi^*_eE$</td>
<td>0.453</td>
<td>8.45 · 10^{-2}</td>
<td>8.78 · 10^{-3}</td>
<td>9.74 · 10^{-2}</td>
<td>3.33</td>
<td>2.69</td>
<td>4.28</td>
</tr>
<tr>
<td>$\xi^*_eG$</td>
<td>0.540</td>
<td>0.195</td>
<td>5.68 · 10^{-3}</td>
<td>0.340</td>
<td>1.33</td>
<td>1.26</td>
<td>1.83</td>
</tr>
</tbody>
</table>

with nominal parameters $\theta^0 = (21.80, 0.05884, 4.298)^T$. The $D$- and $E$-optimal designs for $\theta^0$ are respectively given by

$$
\xi^*_{D, \theta^0} \simeq \left\{ \begin{array}{ccc}
0.229 & 1.389 & 18.42 \\
1/3 & 1/3 & 1/3 \\
\end{array} \right\},
$$

$$
\xi^*_{E, \theta^0} \simeq \left\{ \begin{array}{ccc}
0.170 & 1.398 & 23.36 \\
0.199 & 0.662 & 0.139 \\
\end{array} \right\},
$$

see Atkinson et al. (1993).

We take $\Theta$ as the rectangular region $[16, 27] \times [0.03, 0.08] \times [3, 6]$ and use the algorithm of Sect. 3.4 to compute an optimal design for $\phi_eE(\cdot)$; the grid $\mathcal{G}$ of (16)-(iii) is taken as a random Latin hypercube design with 10000 points in $[0, 1]^3$ renormalized to $\Theta$. The number of iterations and computational time depend on $\ell$, the number of elements of $\mathcal{X}$. For instance, when $\mathcal{X}$ is the finite set $\{0.2, 0.4, 0.6, \ldots, 24\}$ with $\ell = 120$, and the required precision $\epsilon$ equals $10^{-10}$, the algorithm initialized at the uniform measure on the three points 0.2, 1 and 23 converges after 42 iterations in about 26 s. By refining $\mathcal{X}$ iteratively around the support points of the current optimal design, after a few steps we obtain

$$
\xi^*_{eE, \theta^0} \simeq \left\{ \begin{array}{ccc}
0.1785 & 1.520 & 20.95 \\
0.20 & 0.66 & 0.14 \\
\end{array} \right\}.
$$

A similar approach is used below for the construction of optimal designs for $\phi_eE(\cdot)$ and in Example 4 for $\phi_eE(\cdot)$. The performances of the designs $\xi^*_D$, $\xi^*_E$ and $\xi^*_eE$ are indicated in Table 2. One may notice that the design $\xi^*_eE$ is best or second best for $\phi_D(\cdot)$, $\phi_E(\cdot)$ and $\phi_eE(\cdot)$ among all locally optimal designs considered.

The intrinsic curvature is zero for $\xi^*_D$, $\xi^*_E$ and $\xi^*_eE$ (since they all have $3 = \dim(\theta)$ support points) and the parametric curvatures at $\theta^0$ are rather small (the smallest one is for $\xi^*_eE$). This explains that, the domain $\Theta$ being...
not too large, the values of $\phi_{ee}(\xi)\,d\xi$ do not differ very much from those of $\phi_{E}(\xi) = \lambda_{\min}[M(\xi, \theta^0)]$.

Consider now the same three functions of interest as in (Atkinson et al., 1993): $g_1(\theta)$ is the area under the curve,

$$g_1(\theta) = \int_{0}^{\infty} \eta(x, \theta) \, dx = \theta_1 (1/\theta_2 - 1/\theta_3);$$

$g_2(\theta)$ is the time to maximum concentration,

$$g_2(\theta) = \frac{\log \theta_3 - \log \theta_2}{\theta_3 - \theta_2},$$

and $g_3(\theta)$ is the maximum concentration

$$g_3(\theta) = \eta[g_2(\theta), \theta].$$

We shall write $c_i = c_i(\theta^0) = \partial g_i(\theta)/\partial \theta|_{\theta^0}$ and denote $\xi_{c_i, \theta^0}$ the (locally) optimal design for $g_i(\theta)$ which maximizes $\phi_{c_i}(\xi; \theta^0) = [c_i^\top M^{-1}(\xi, \theta^0)c_i]^{-1}$, for $i = 1, 2, 3$. The $\xi_{c_i, \theta^0}$ are singular and are approximately given by

$$\begin{align*}
\xi_{c_1, \theta^0} &\approx \begin{bmatrix} 0.2327 \\ 0.0135 \end{bmatrix} \\
\xi_{c_2, \theta^0} &\approx \begin{bmatrix} 0.1793 \\ 0.6062 \end{bmatrix} \\
\xi_{c_3, \theta^0} &\approx \begin{bmatrix} 1.0122 \\ 1 \end{bmatrix}
\end{align*},$$

see Atkinson et al. (1993).

For each function $g_i$, we restrict the search of a design $\xi_{ec}$ optimal in the sense of the criterion $\phi_{ec}(-)$ to design measures supported on the union of the supports of $\xi_{D, \theta^0}$, $\xi_{E, \theta^0}$ and $\xi_{c_i, \theta^0}$. We then obtain the following designs

$$\begin{align*}
\xi_{ec_1, \theta^0} &\approx \begin{bmatrix} 0.2327 & 1.389 & 23.36 \\ 9 \cdot 10^{-4} & 1.2 \cdot 10^{-2} & 0.9871 \end{bmatrix} \\
\xi_{ec_2, \theta^0} &\approx \begin{bmatrix} 0.1793 & 0.229 & 3.5671 & 18.42 \\ 5.11 \cdot 10^{-2} & 0.5375 & 0.3158 & 9.56 \cdot 10^{-2} \end{bmatrix} \\
\xi_{ec_3, \theta^0} &\approx \begin{bmatrix} 0.229 & 1.0122 & 1.389 & 18.42 \\ 8.42 \cdot 10^{-2} & 0.4867 & 0.4089 & 2.02 \cdot 10^{-2} \end{bmatrix}
\end{align*}.$$
The performances of $\xi^*_c$ and $\xi^*_ec_i$, $i = 1, \ldots, 3$, are indicated in Table 2, together with the curvature measures at $\theta^0$ for $\xi^*_ec_i$ (which are nonsingular). For each function $g_i$ of interest, the design $\xi^*_ec_i$ performs slightly worse than $\xi^*_c$ in terms of $c$-optimality, but, contrarily to $\xi^*_c$, it allows us to estimate the three parameters $\theta$ and guarantees good estimability properties for $g_i(\theta)$ for all $\theta \in \Theta$. Notice that, apart from the $c$-optimality criteria $\phi_c(\cdot)$, all criteria considered take the value 0 at the $c$-optimal designs $\xi^*_c$. The construction of an optimal design for $\phi_{ec}(\cdot)$ thus forms an efficient method to circumvent the difficulties caused by singular $c$-optimal design in nonlinear models, see (Pronzato and Pázman, 2013, Chap. 3 & 5). One may also refer to (Pronzato, 2009) for alternative approaches for the regularization of singular $c$-optimal designs.

We conclude this example with a comparison with the average-optimal designs of (Atkinson et al., 1993), that aim at taking uncertainty on $\theta^0$ into account. Consider a prior distribution $\pi(\cdot)$ on the two components of $\theta$ that intervene nonlinearly in $\eta(x, \theta)$, and let $E_\pi\{\cdot\}$ denote the expectation for $\pi(\cdot)$. Atkinson et al. (1993) indicate that when $\pi$ equals $\pi_A$ uniform on $[\theta^0_2 - 0.01, \theta^0_2 + 0.01] \times [\theta^0_3 - 1, \theta^0_3 + 1]$, the design that maximizes $E_\pi\{\log \det[M(\xi, \theta)]\}$ is

$$\xi^*_{AD-A} \simeq \left\{ \begin{array}{ccc} 0.2288 & 1.4170 & 18.4513 \\ 1/3 & 1/3 & 1/3 \end{array} \right\},$$

and the designs that minimize $E_\pi\{c_i^T(\theta)M^{-1}(\xi, \theta)c_i(\theta)\}$, $i = 1, 2, 3$, are

$$\xi^*_{Ac1-A} \simeq \left\{ \begin{array}{ccc} 0.2449 & 1.4950 & 18.4903 \\ 0.0129 & 0.0387 & 0.9484 \end{array} \right\},$$

$$\xi^*_{Ac2-A} \simeq \left\{ \begin{array}{ccc} 0.1829 & 2.4639 & 8.8542 \\ 0.6023 & 0.2979 & 0.0998 \end{array} \right\},$$

$$\xi^*_{Ac3-A} \simeq \left\{ \begin{array}{ccc} 0.3608 & 1.1446 & 20.9218 \\ 0.0730 & 0.9094 & 0.0176 \end{array} \right\}.$$

When $\pi$ equals $\pi_B$ uniform on $[\theta^0_2 - 0.04, \theta^0_2 + 0.04] \times [\theta^0_3 - 4, \theta^0_3 + 4]$, the
average-optimal designs are

\[
\begin{align*}
\xi_{AD-B}^* & \approx \left\{ \begin{array}{cccc}
0.2034 & 1.1967 & 2.8323 & 7.8229 \\
0.2870 & 0.2327 & 0.1004 & 0.0678 \\
0.3120 & 0.2570 & 0.0697 & 0.0697 
\end{array} \right\}, \\
\xi_{Ac1-B}^* & \approx \left\{ \begin{array}{cccc}
0.2909 & 1.7269 & 13.0961 & 39.58 \\
0.0089 & 0.0365 & 0.2570 & 0.6976 
\end{array} \right\}, \\
\xi_{Ac2-B}^* & \approx \left\{ \begin{array}{cccc}
0.2513 & 0.9383 & 2.7558 & 8.8381 \\
0.2914 & 0.2854 & 0.1468 & 0.2174 \\
0.0050 & 0.0090 & 0.0050 & 0.0050 
\end{array} \right\}, \\
\xi_{Ac3-B}^* & \approx \left\{ \begin{array}{cccc}
0.3696 & 1.1383 & 2.4370 & 6.0691 \\
0.0971 & 0.3584 & 0.3169 & 0.1634 \\
0.0641 & 0.0641 & 0.0641 & 0.0641 
\end{array} \right\}.
\end{align*}
\]

Their performances are indicated in the bottom part of Table 2. Notice that the average-optimal designs for the vague prior \(\pi_B\) are supported on more than three points and thus allow model checking; this is the case too for the two designs \(\xi_{ec2}\) and \(\xi_{ec3}\). However, contrary to average-optimal design, the number of support points of optimal designs for extended optimality criteria does not seem to increase with uncertainty measured by the size of \(\Theta\); for instance, when \(\Theta = [\theta_1^0 - 5, \theta_1^0 + 5] \times [\theta_2^0 - 0.04, \theta_2^0 + 0.04] \times [\theta_3^0 - 4, \theta_3^0 + 4]\), the optimal design for \(\phi_{eE}(\cdot)\) is still supported on three points, approximately 0.1565, 1.552 and 19.73, receiving weights 0.268, 0.588 and 0.144 respectively.

All average-optimal designs considered yield reasonably small curvatures at \(\theta^0\), although larger than those for \(\xi_{E,\theta^0}^*\) and \(\xi_{eE,\theta^0}^*\). The performances of \(\xi_{AD-A}^*\) and \(\xi_{AD-B}^*\) are close to those of \(\xi_{D,\theta^0}^*\), and the most interesting features concern designs for estimation of functions of interest \(g_i(\theta)\). The designs \(\xi_{ec,\theta^0}^*\) cannot be used if \(\theta \neq \theta^0\) and are thus useless in practice. The average-optimal designs \(\xi_{Ac1-B}^*\) perform significantly worse than \(\xi_{ec,\theta^0}^*\) in terms of \(\phi_{ec_i}(\cdot)\) for \(i = 1, 2, 3\) and in terms of \(\phi_{c_i}(\cdot)\) for \(i = 2, 3\). On the other hand, the designs \(\xi_{Ac1-A}^*\), constructed for the precise prior \(\pi_A\), perform significantly better than \(\xi_{ec,\theta^0}^*\) in terms of \(\phi_{c_i}(\cdot)\) for all \(i\). Figure 3 presents \(\phi_{c3}(\xi; \theta)\) as a function of \(\theta\), for the three designs \(\xi_{Ac3-A}^*\) (dashed line), \(\xi_{Ac3-B}^*\) (dash-dotted line) and \(\xi_{ec3,\theta^0}^*\) (solid line), when \(\theta_1 = \theta_1^0, \theta_3 = \theta_3^0\) (left) and \(\theta_1 = \theta_1^0, \theta_2 = \theta_2^0\) (right). Note that the projection on the last two components of \(\theta\) of the set \(\Theta\) used for extended c-optimality is intermediate between the supports of \(\pi_A\) and \(\pi_B\). Although average-optimal designs \(\xi_{Ac1-A,B}^*\) and extended-optimal designs \(\xi_{ec,\theta^0}^*\) pursue different objectives, the example indicates that they show some resemblance in terms of precision of estimation of \(g_i(\theta)\). The situation would be totally different in absence of global identifiability for \(g_i(\theta)\), a problem that would not be detected by average-optimal designs, see the discussion at the end of Example 2.

---

imsart-aos ver. 2013/03/06 file: Extended-optimality_AOS-REV.tex date: March 31, 2014
Table 2. Performances of different designs and curvature measures at $\theta^0$ for the model (19) with $\theta^0 = (21.80, 0.05884, 4.298)^T$ and $\Theta = [16, 27] \times [0.03, 0.08] \times [3, 6]$; $\det^{1/3} = \phi_D(\xi) = \{\det[M(\xi, \theta^0)]\}^{1/3}$, $\lambda_{\min} = \phi_E(\xi) = \lambda_{\min}[M(\xi, \theta^0)]$. The optimal (maximum) values of the criteria are on the main diagonal and indicated in boldface. The bottom part of the table corresponds to the average-optimal designs of (Atkinson et al., 1990).

| $\ell$  | $\det^{1/3}$ | $\lambda_{\min}$ | $\phi_D$ | $\phi_E$ | $\phi_{\mathcal{AC}}$ | $\phi_{\mathcal{C}}$ | $\phi_{\mathcal{CC}}$ | $\phi_{\mathcal{EC}}$ | $\phi_{\mathcal{EC}}$ | $\phi_{\mathcal{EC}}$ | $C_{\text{par}}$ | $C_{\text{tot}}$ | $C_{\text{tot}}$ |
|--------|--------------|-------------------|---------|---------|-----------------|----------------|----------------|^| | | | |
| $\xi_D$ | 11.74 | 0.191 | 0.178 | 1.56 $10^{-4}$ | 6.68 $10^{-5}$ | 23.43 | 18.31 | 0.361 | 0.356 | 0.356 | 0.526 | 0 | 0.526 |
| $\xi_E$ | 8.82 | 0.316 | 0.274 | 6.07 $10^{-5}$ | 3.08 $10^{-5}$ | 15.89 | 10.35 | 0.675 | 0.667 | 0.675 | 0.370 | 0 | 0.370 |
| $\xi_{\mathcal{E}}$ | 9.05 | 0.311 | 0.281 | 6.45 $10^{-5}$ | 3.01 $10^{-5}$ | 16.62 | 11.03 | 0.656 | 0.644 | 0.656 | 0.358 | 0 | 0.358 |
| $\xi_1$ | 0 | 0 | 0 | 4.56 $10^{-4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 6.51 | 0 | 6.51 |
| $\xi_{\mathcal{C}}$ | 0.757 | 2.70 $10^{-3}$ | 1.92 $10^{-3}$ | 2.26 $10^{-4}$ | 2.17 $10^{-4}$ | 8.55 $10^{-2}$ | 6.12 $10^{-2}$ | 1.12 $10^{-2}$ | 1.09 $10^{-2}$ | 1.12 $10^{-2}$ | 6.51 | 0 | 6.51 |
| $\xi_2$ | 0 | 0 | 0 | 0 | 0 | 35.55 | 0 | 0 | 0 | 0 | 1.12 | 0.028 | 1.12 |
| $\xi_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| $\xi_{\mathcal{CE}}$ | 4.06 | 0.162 | 0.137 | 9.70 $10^{-6}$ | 4.19 $10^{-6}$ | 6.77 | 4.36 | 0.890 | 0.865 | 0.865 | 1.11 | 0.263 | 1.14 |
| $\xi_{\mathcal{C-B}}$ | 11.74 | 0.191 | 0.177 | 1.56 $10^{-4}$ | 6.68 $10^{-5}$ | 23.53 | 18.43 | 0.360 | 0.355 | 0.355 | 0.522 | 0 | 0.522 |
| $\xi_{\mathcal{A}-C}$ | 2.74 | 2.07 $10^{-2}$ | 1.69 $10^{-2}$ | 4.36 $10^{-4}$ | 1.50 $10^{-4}$ | 1.12 | 0.864 | 4.06 $10^{-2}$ | 4.01 $10^{-2}$ | 4.01 $10^{-2}$ | 1.82 | 0 | 1.82 |
| $\xi_{\mathcal{A}-B}$ | 6.71 | 7.22 $10^{-2}$ | 6.66 $10^{-2}$ | 2.21 $10^{-2}$ | 6.67 $10^{-6}$ | 35.16 | 20.31 | 0.175 | 0.175 | 0.175 | 0.909 | 0 | 0.909 |
| $\xi_{\mathcal{A}-C}$ | 3.31 | 0.118 | 8.23 $10^{-2}$ | 8.06 $10^{-2}$ | 3.86 $10^{-6}$ | 4.37 | 3.17 | 0.937 | 0.838 | 0.838 | 1.82 | 0 | 1.82 |
| $\xi_{\mathcal{A}-B}$ | 11.08 | 0.179 | 0.159 | 1.62 $10^{-4}$ | 6.99 $10^{-5}$ | 21.15 | 15.93 | 0.338 | 0.335 | 0.335 | 0.505 | 0.056 | 0.507 |
| $\xi_{\mathcal{B}-C}$ | 2.18 | 2.23 $10^{-2}$ | 1.46 $10^{-2}$ | 2.34 $10^{-4}$ | 1.56 $10^{-4}$ | 0.791 | 0.644 | 5.37 $10^{-2}$ | 4.89 $10^{-2}$ | 4.89 $10^{-2}$ | 2.12 | 0.133 | 2.13 |
| $\xi_{\mathcal{B}-C}$ | 9.45 | 0.162 | 0.134 | 8.07 $10^{-5}$ | 3.03 $10^{-5}$ | 20.05 | 16.28 | 0.385 | 0.358 | 0.358 | 0.753 | 0.118 | 0.761 |
| $\xi_{\mathcal{B}-B}$ | 6.16 | 0.149 | 0.952 $10^{-2}$ | 5.07 $10^{-5}$ | 2.06 $10^{-5}$ | 6.60 | 6.13 | 0.615 | 0.587 | 0.587 | 1.22 | 0.256 | 1.25 |
Example 4. For the same regression model (19), we change the value of $\theta^0$ and the set $\Theta$ and take $\theta^0 = (0.773, 0.214, 2.09)^T$ and $\Theta = [0, 5] \times [0, 5] \times [0, 5]$, the values used by Kieffer and Walter (1998). With these values, from an investigation based on interval analysis, the authors report that for the 16-point design

$$
\xi_0 = \left\{ \frac{1}{16}, \frac{2}{16}, \cdots, \frac{16}{16} \right\},
$$

and the observations $y$ given in their Table 13.1, the LS criterion $\|y - \eta_X(\theta)\|^2$ has a global minimizer (the value we have taken here for $\theta^0$) and two other local minimizers in $\Theta$. The $D$- and $E$-optimal designs for $\theta^0$ are now given by

$$
\xi^*_{D,\theta^0} \simeq \left\{ 0.42, 1.82, 6.80 \right\},
$$

$$
\xi^*_{E,\theta^0} \simeq \left\{ 0.29, 1.83, 9.0 \right\}.
$$

Using the same approach as in Example 3, with the grid $\mathcal{G}^0$ of (16)-(iii) obtained from a random Latin hypercube design with 10000 points in $\Theta$, we obtain

$$
\xi^*_{E,\theta^0} \simeq \left\{ 0.38, 2.26, 7.91 \right\}.
$$

To compute an optimal design for $\phi_{cG}(\cdot)$, we consider the design space $\mathcal{X} = \{0, 0.1, 0.2, \ldots, 16\}$ (with 161 points) and use the algorithm of Sect. 3.4 with the grid $\mathcal{G}^0$ of (16)-(iii) taken as a random Latin hypercube design.

Fig. 3. $\psi_c(\xi; \theta)$ as a function of $\theta$, for $\theta_1 = \theta_1^0$ and $\theta_3 = \theta_3^0$ (left) and $\theta_1 = \theta_1^0$, $\theta_2 = \theta_2^0$ (right); $\xi = \xi_{cA-3}$ in dashed line, $\xi = \xi_{cA-3}$ in dash-dotted line and $\xi = \xi_{c3, \theta_0}$ in solid line.
with $10^5$ points. The same design space is used to evaluate $\phi_{eG}(\cdot)$ for the four designs above. For $\epsilon = 10^{-10}$, the algorithm initialized at the uniform measure on $\mathcal{X}$ converges after 34 iterations in about 52 s and gives

$$
\xi_{eG, \theta}^* \simeq \begin{cases} 
0.4 & 1.9 & 5.3 & 16 \\
0.278 & 0.258 & 0.244 & 0.22
\end{cases}.
$$

The performances and curvature measures at $\theta^0$ of $\xi_0$, $\xi^*_D$, $\xi^*_E$, $\xi^*_e$ and $\xi^*_G$ are given in Table 3. The large intrinsic curvature for $\xi_0$, associated with the small values of $\phi_{eE}(\xi^0)$ and $\phi_{eG}(\xi^0)$, explains the presence of local minimizers for the LS criterion and thus the possible difficulties for the estimation of $\theta$.

The values of $\phi_{eE}(\cdot)$ and $\phi_{eG}(\cdot)$ reported in the table indicate that $\xi^*_D$, $\xi^*_E$, $\xi^*_e$ or $\xi^*_G$ would have caused less difficulties.

7. Further extensions and developments.

7.1. An extra tuning parameter for a smooth transition to usual design criteria. The criterion $\phi_{eE}(\xi; \theta^0)$ can be written as

$$
\phi_{eE}(\xi; \theta^0) = \max \{ \alpha \in \mathbb{R} : \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|^2_\xi \geq \alpha \|\theta - \theta^0\|^2, \text{ for all } \theta \in \Theta \}.
$$

Instead of giving the same importance to all $\theta$ whatever their distance to $\theta^0$, one may wish to introduce a saturation and reduce the importance given to those $\theta$ very far from $\theta^0$, that is, consider

$$
\phi_{eE[K]}(\xi; \theta^0) = \max \left\{ \alpha \in \mathbb{R} : \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|^2_\xi \geq \alpha \frac{\|\theta - \theta^0\|^2}{1 + K \|\theta - \theta^0\|^2}, \text{ for all } \theta \in \Theta \right\}
$$

(21)
for some $K \geq 0$. Equivalently, $\phi_{cE|K}(\xi; \theta^0) = \min_{\theta \in \Theta} H_{E|K}(\xi, \theta)$, with

$$H_{E|K}(\xi, \theta) = \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 \left[ K + \frac{1}{\|\theta - \theta^0\|^2} \right].$$

As in Sect. 3.1, we obtain $\phi_{cE|K}(\xi) = \lambda_{\min}[M(\xi)]$ in a linear model and, for a nonlinear model with $\Theta = \mathcal{B}(\theta^0, \rho)$, $\lim_{\rho \to 0} \phi_{cE|K}(\xi; \theta^0) = \lambda_{\min}[M(\xi, \theta^0)]$ for any $K \geq 0$. Moreover, in a nonlinear model with no overlapping $\phi_{cE|K}(\xi; \theta^0)$ can be made arbitrarily close to $\lambda_{\min}[M(\xi, \theta^0)]$ by choosing $K$ large enough, whereas choosing $K$ not too large ensures some protection against $\|\eta_X(\theta) - \eta_X(\theta^0)\|$ being small for some $\theta$ far from $\theta^0$. Also, properties of $\phi_{cE}(\cdot)$ such as concavity, positive homogeneity, existence of directional derivatives, see Sect. 3.2, remain valid for $\phi_{cE|K}(\cdot)$, for any $K \geq 0$. The maximization of $\phi_{cE|K}(\cdot)$ forms a LP problem when both $\mathcal{X}$ and $\Theta$ are finite, see Sect. 3.3, and a relaxation procedure (cutting-plane method) can be used when $\Theta$ is a compact subset of $\mathbb{R}^p$, see Sect. 3.4.

A similar approach can be used with extended $c$- and $G$-optimality, which gives $\phi_{cG|K}(\xi) = \min_{\theta \in \Theta} H_{cG|K}(\xi, \theta^0)$ with

$$H_{cG|K}(\xi, \theta) = \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 \left[ K + \frac{1}{|g(\theta) - g(\theta^0)|^2} \right],$$

and

$$\phi_{cG|K}(\xi) = \min_{\theta \in \Theta} \left\{ \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 \left[ K + \frac{1}{\max_{x \in \mathcal{X}} |\eta(x, \theta) - \eta(x, \theta^0)|^2} \right] \right\},$$

for $K$ a positive constant.

7.2. Worst-case extended optimality criteria. The criterion defined by

$$\phi_{MeE}(\xi) = \min_{\theta^0 \in \Theta} \phi_{cE}(\xi; \theta^0) = \min_{(\theta, \theta^0) \in \Theta \times \Theta} H_E(\xi, \theta; \theta^0),$$

see (6), (5), accounts for the global behavior of $\eta(\cdot, \theta)$ for $\theta \in \Theta$ and obliterates the dependence on $\theta^0$ that is present in $\phi_{cE}(\xi; \theta^0)$. The situation is similar to that in Sect. 3, excepted that we consider now the minimum of $H_E$ with respect to two vectors $\theta$ and $\theta^0$ in $\Theta \times \Theta$. All the developments in Sect. 3 obviously remain valid (concavity, existence of directional derivative, etc.), including the algorithmic solutions of Sects. 3.3 and 3.4. The same is true for the worst-case versions of $\phi_{ec}(\cdot)$ and $\phi_{cG}(\cdot)$, respectively defined by $\phi_{MeE}(\xi) = \min_{(\theta, \theta^0) \in \Theta \times \Theta} H_E(\xi, \theta; \theta^0)$, see (17), and by $\phi_{MeG}(\xi) = \min_{(\theta, \theta^0) \in \Theta \times \Theta} \left\{ \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 / \max_{x \in \mathcal{X}} |\eta(x, \theta) - \eta(x, \theta^0)|^2 \right\}$, and for...
the worst-case versions of the extensions of previous section that include an additional tuning parameter $K$.

Note that the criterion $\phi_{MeE}(\cdot)$ may direct attention to a particularly pessimistic situation. Indeed, for $\Theta$ a compact set with nonempty interior and $\mu$ the Lebesgue measure on $\Theta$, one may have $\min_{\theta^0 \in \Theta} \phi_{eE}(\xi; \theta^0) = 0$ for all designs $\xi$ although $\mu\{\theta^0 \in \Theta : \phi_{eE}(\xi'; \theta^0) > 0\} = 1$ for some design $\xi'$. This corresponds to a situation where the model is structurally identifiable, in the sense that the property (2) is generic but is possibly false for $\theta$ in a subset of zero measure, see, e.g., Walter (1987)). Example 2 gives an illustration.

Example 2 (continued). When the three polynomial equations $\theta'_1 - \theta'_1^3 = \theta_1 - \theta_1^3$, $\theta'_2 - \theta'_2^2 = \theta_2 - \theta_2^2$, $\theta'_1 + \theta'_2^2 = \theta_1^3 + \theta_2^2$ are satisfied, then $\eta(x, \theta') = \eta(x, \theta)$ for all $x$. Since these equations have solutions $\theta' \neq \theta$ in $\Theta \times \Theta$, $\phi_{MeE}(\xi) = 0$ for all $\xi \in \Xi$. On the other hand, $\max_{\xi \in \Xi} \phi_{eE}(\xi; \theta^0) > 0$ w.p.1 when $\theta^0$ is randomly drawn with a probability measure having a density with respect to the Lebesgue measure on $\Theta$.

In a less pessimistic version of worst-case extended $E$-optimality we may thus consider a finite set $\Theta^0 \subset \Theta$ for $\theta^0$, obtained for instance by random sampling in $\Theta$, and maximize $\min_{\theta^0 \in \Theta^0} \phi_{eE}(\xi; \theta^0)$.

8. Conclusions. Two essential ideas have been presented. First, classical optimality criteria can be extended in a mathematically consistent way to criteria that preserve a nonlinear model against overlapping, and at the same time retain the main features of classical criteria, especially concavity. Moreover, they coincide with their classical counterpart for linear models. Second, designs that are nearly optimal for those extended criteria can be obtained by standard linear programming solvers, supposing that the approximation of the feasible parameter space $\Theta$ by a finite set is acceptable. A relaxation method, equivalent to the cutting-plane algorithm, can be used when $\Theta$ is a compact set with nonempty interior. Linear constraints on the design can easily be taken into account. As a by-product, this also provides simple algorithmic procedures for the determination of $E$-, $c$- or $G$-optimal designs in linear models with linear cost constraints.

As it is usually the case for optimal design in nonlinear models, the extended-optimality criteria are local and depend on a guessed value $\theta^0$ for the model parameters. However, the construction of a globalized, worst-case, version enjoying the same properties is straightforward (Sect. 7.2).

Finally, we recommend the following general procedure for optimal design in nonlinear regression. (i) Choose a parameter space $\Theta$ correspond-
ing to the domain of interest for $\theta$, select (e.g., randomly) a finite subset $\Theta^0$ in the interior of $\Theta$. (ii) for each $\theta^0$ in $\Theta^0$ compute an optimal design $\xi^*_{E,\theta^0}$ maximizing $\phi_{E}(\xi; \theta^0)$ and a $E$-optimal design $\xi^*_{E,\theta^0}$ maximizing $\phi_{E}(\xi; \theta^0) = \lambda_{\min}(M(\xi, \theta^0))$; (iii) if $\phi_{E}(\xi^*_{E,\theta^0}; \theta^0)$ is close enough to $\phi_{E}(\xi^*_{E,\theta^0}; \theta^0)$ for all $\theta^0$ in $\Theta^0$, one may consider that the risk of overlapping, or lack of identifiability in $\Theta$, is weak and classical optimal design that focuses on the precision of estimation can be used; otherwise, a design that maximizes $\min_{\theta^0 \in \Theta^0} \phi_{E}(\xi; \theta^0)$ should be preferred. When the extended $G$-optimality criterion $\phi_{G}(\xi; \theta^0)$ is substituted for $\phi_{E}(\xi; \theta^0)$, the comparison in (iii) should be between $\phi_{G}(\xi^*_{G,\theta^0}; \theta^0)$ and $1/\dim(\theta)$, see Sect. 5. Extended $c$-optimality can be used when one is interested in estimating a (nonlinear) function of $\theta$, the comparison in (iii) should then be with $c$-optimality.

Acknowledgements. The authors thank the referees for useful comments that helped to significantly improve the paper.

References.


---

**Department of Applied Mathematics and Statistics**
**Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia**
E-mail: pazman@fmph.uniba.sk

**Laboratoire I3S, Bât. Euclide, Les Algorithmes, BP 121 2000 route des Lucioles, 06903 Sophia Antipolis cedex, France**
E-mail: pronzato@i3s.unice.fr

---

imsart-aos ver. 2013/03/06 file: Extended-optimality_AOS-REV.tex date: March 31, 2014