Optimum design accounting for the global nonlinear behavior of the model

Andrej Pazman, Luc Pronzato

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 θ^0 (local optimal design), account for the behavior of the model response $\eta(\theta)$ for θ far from θ^0. In particular, they ensure some protection against close-to-overlapping situations where $\|\eta(\theta) - \eta(\theta^0)\|$ is small for some θ far from θ^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 Θ of admissible θ is discretized, optimal design forms a linear programming problem which can be solved directly or via relaxation when Θ is just compact. Several examples are presented.

1. Introduction. We consider a nonlinear regression model with observations
$$y_i = y(x_i) = \eta(x_i, \bar{\theta}) + \varepsilon_i, \; i = 1, \ldots, N,$$
where the errors ε_i satisfy $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{var}(\varepsilon_i) = \sigma^2$ and $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, $i, j = 1, \ldots, N$, and the true value $\bar{\theta}$ of the vector of model parameter θ belongs to Θ, a compact subset of $\mathbb{R}^p$ such that $\Theta \subset \overline{\mathrm{int}(\Theta)}$, the closure of the interior of Θ. In a vector notation, we write

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

where $\eta_X(\theta) = (\eta(x_1, \theta), \ldots, \eta(x_N, \theta))^\top$, $\mathbf{y} = (y_1, \ldots, y_N)^\top$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^\top$, and X denotes the N-point exact design $(x_1, \ldots, x_N)$. The more general non-stationary (heteroscedastic) case where $\mathrm{var}(\varepsilon_i) = \sigma^2(x_i)$ can easily be transformed into the model (1) with $\sigma^2 = 1$ 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 $\theta \in \mathrm{int}(\Theta)$ for any $x \in \mathscr{X}$, a compact subset of $\mathbb{R}^d$. The model is assumed to be identifiable over $\mathscr{X}$; that is, we suppose that

$$(2) \qquad \eta(x, \theta') = \eta(x, \theta) \ \text{ for all } x \in \mathscr{X} \implies \theta' = \theta \,.$$

We shall denote by Ξ the set of design measures ξ, i.e., of probability measures on $\mathscr{X}$. The information matrix (for $\sigma^2 = 1$) for the design X at θ is

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

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

$$\mathbf{M}(\xi, \theta) = \int_{\mathscr{X}} [\partial \eta(x, \theta)/\partial \theta] \, [\partial \eta(x, \theta)/\partial \theta^\top] \, \xi(\mathrm{d}x) \,.$$

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

$$(3) \qquad \eta(x, \theta') = \eta(x, \theta) \ \xi\text{-almost everywhere} \implies \theta' = \theta \,.$$

Given an exact N-point design X, the set of all hypothetical means of the observed vectors $\mathbf{y}$ in the sample space $\mathbb{R}^N$ forms the expectation surface $\mathbb{S}_\eta = \{\eta_X(\theta) : \theta \in \Theta\}$. Since $\eta_X(\theta)$ is supposed to have continuous first and second-order derivatives in $\mathrm{int}(\Theta)$, $\mathbb{S}_\eta$ is a smooth surface in $\mathbb{R}^N$ with a (local) dimension given by $r = \mathrm{rank}[\partial \eta_X(\theta)/\partial \theta^\top]$. 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 $\eta_X(\theta) = \eta_X(\theta')$ implies $\theta = \theta'$, the LS estimator

$$(4) \qquad \hat{\theta}_{LS} = \hat{\theta}_{LS}^N = \arg \min_{\theta \in \Theta} \|\mathbf{y} - \eta_X(\theta)\|^2$$

is uniquely defined with probability one (w.p.1). Indeed, when the distributions of errors ε_i have probability densities (in the standard sense) it can be proven that $\eta[\hat{\theta}_{LS}(\mathbf{y})]$ is unique w.p.1, see Pázman (1984) and Pázman (1993, p. 107). However, there is still a positive probability that the function

$\theta \longrightarrow \|\mathbf{y} - \eta_X(\theta)\|^2$ 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 $\mathbb{S}_\eta$ 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 θ and θ' in Θ 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 Pázman, 2006), a non-overlapping assumption is used (see Pázman (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 $\dim(\theta) = 1$ with the overlapping $\mathbb{S}_\eta$ 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 Θ 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 θ; 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 Pázman, 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-, 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 θ, asymptotically, provided that the model is locally identifiable at θ. 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 $x = (t, \ u)^\top \in \mathscr{X} = \{0, \pi/2\} \times [0, u_{\max}]$, we have

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

with r a known positive constant. We take $u_{\max} = 7 \, \pi/4$; the difficulties mentioned below are even more pronounced for values of $u_{\max}$ closer to 2π. We shall consider exclusively two-point designs $X = (x_1, x_2)$ of the form

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

and denote ν_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 $u \in [0, u_{\max}]$, where optimality is considered in terms of information.

It is easy to see that for any design ν_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 $u = u_{\max} = 7\pi/4$. The model is nonlinear but parametrically linear since the information matrix $M(X, \theta)$ for $\sigma^2 = 1$ (here scalar since θ is scalar) equals $r^2 \, u^2$ and does not depend on θ. 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 $u = u_{\max}$, and setting a constraint on the intrinsic curvature is not possible here. However, if the true value of θ is $\bar{\theta} = 0$ and σ^2 is large enough, there is a chance that the LS estimator will be $\hat{\theta}_{LS} = 1$, and thus very far from $\bar{\theta}$, see Figure 1. The situation gets even worse if $u_{\max}$ gets closer to $2 \, \pi$, since $\mathbb{S}_\eta$ 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 $u \in [0, u_{\max}]$, 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 $[0, u_{\max}]$ for $u = u_* = \pi$. This choice $u = u_*$ seems preferable to $u = u_{\max}$ since the
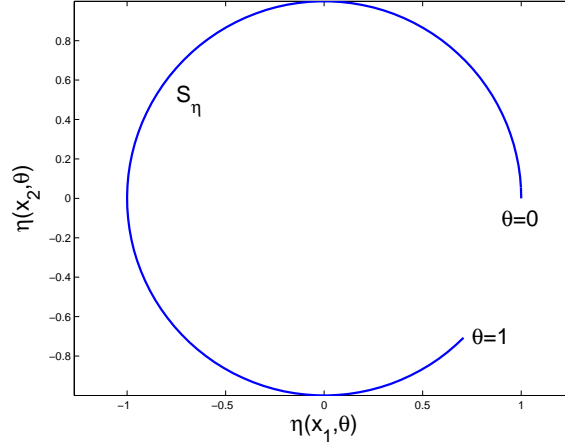
FIG 1. *Expectation surface $\mathbb{S}_\eta$ for $\theta \in \Theta = [0,1]$, $r = 1$ and $u = u_{\max} = 7\pi/4$.*

expectation surface $\mathbb{S}_\eta$ 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 θ 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 $u = u_*$. Figure 2-right shows that the design with $u = u_{\max}$ (*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 θ slightly deviates from θ^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 $\mathbb{S}_\eta$, at the price of a slightly less informative experiment for θ close to θ^0. Smaller values of u (*dotted line*) are worse than u_*, both locally for θ close to θ^0 and globally in terms of the folding of $\mathbb{S}_\eta$.

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.

3. Extended (globalized) *E*-optimality.

3.1. *Definition of $\phi_{eE}(\cdot)$.* Take a fixed point θ^0 in Θ and denote

$$(5) \qquad H_E(\xi, \theta) = H_E(\xi, \theta; \theta^0) = \frac{\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2}{\|\theta - \theta^0\|^2},$$
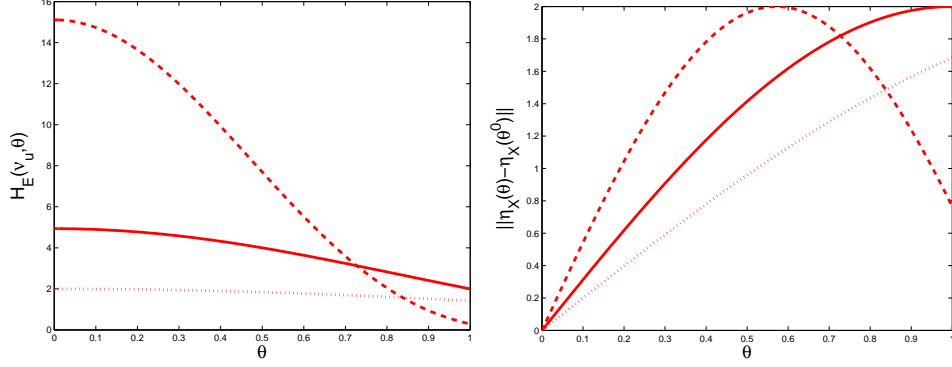
FIG 2. $H_E(\nu_u, \theta)$ (left) and $\|\eta_X(\theta) - \eta_X(\theta^0)\|$ (right) as functions of $\theta \in \Theta = [0,1]$ for $r = 1$, $u = 2$ (dotted line), $u = u_{\max} = 7\pi/4$ (dashed line) and $u = u_* = \pi$ (solid line).

where $\|\cdot\|_\xi$ denotes the norm in $\mathscr{L}_2(\xi)$; that is, $\|l\|_\xi = \left[\int_{\mathscr{X}} l^2(x)\,\xi(\mathrm{d}x)\right]^{1/2}$ for any $l \in \mathscr{L}_2(\xi)$. When ξ is a discrete measure, like in the examples considered in the paper, then $\|l\|_\xi^2$ is simply the sum $\sum_{x:\xi(\{x\})>0} \xi(\{x\})\,l^2(x)$.

The extended E-optimality criterion is defined by

$$(6) \qquad \phi_{eE}(\xi) = \phi_{eE}(\xi; \theta^0) = \min_{\theta \in \Theta} H_E(\xi, \theta),$$

to be maximized with respect to the design measure ξ.

In a nonlinear regression model $\phi_{eE}(\cdot)$ depends on the value chosen for θ^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 θ far from θ^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 $\eta(x, \theta) = \mathbf{f}^\top(x)\theta + v(x)$ and $\Theta = \mathbb{R}^p$, for any θ^0 and any $\xi \in \Xi$ we have $\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 = (\theta - \theta^0)^\top \mathbf{M}(\xi)(\theta - \theta^0)$, so that

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

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

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

$$(7) \qquad \lim_{\rho \to 0} \phi_{eE}(\xi; \theta^0) = \lambda_{\min}[\mathbf{M}(\xi, \theta^0)].$$

In a nonlinear regression model with larger Θ, the determination of an optimum design ξ_{eE}^* maximizing $\phi_{eE}(\xi)$ ensures some protection against $\|\eta(\cdot,\theta) - \eta(\cdot,\theta^0)\|_\xi$ being small for some θ far from θ^0. In particular, when $\theta^0 \in \mathrm{int}(\Theta)$ then $\phi_{eE}(\xi;\theta^0) = 0$ if either $\mathbf{M}(\xi,\theta^0)$ is singular or $\|\eta(\cdot,\theta) - \eta(\cdot,\theta^0)\|_\xi = 0$ for some $\theta \neq \theta^0$. Therefore, under the condition (2), ξ_{eE}^* satisfies the estimability condition (3) at $\theta = \theta^0$ and is necessarily non-degenerate, i.e., $\mathbf{M}(\xi_{eE}^*,\theta^0)$ is nonsingular, when $\theta^0 \in \mathrm{int}(\Theta)$ (provided that there exists a non-degenerate design in Ξ). Notice that (7) implies that $\phi_{eE}(\xi;\theta^0) \leq \lambda_{\min}[\mathbf{M}(\xi,\theta^0)]$ when Θ contains some open neighborhood of θ^0. In contrast with the E-optimality criterion, maximizing $\phi_{eE}(\xi;\theta^0)$ in non-linear models does not require computation of the derivatives of $\eta(x,\theta)$ with respect to θ at θ^0, see the algorithms proposed in Sects. 3.3 and 3.4. Also note that the influence of points that are very far from θ^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*

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

with ξ_N the empirical measure associated with the design X used to observe $\mathbf{y}$.

Proof. The result follows from the following chain of inequalities

$$
\begin{aligned}
\|\hat{\theta}_{LS} - \theta\| &\leq \frac{\|\eta(\cdot,\hat{\theta}_{LS}) - \eta(\cdot,\theta)\|_{\xi_N}}{\sqrt{\phi_{eE}(\xi_N;\theta)}} = \frac{\|\eta_X(\hat{\theta}_{LS}) - \eta_X(\theta)\|}{\sqrt{N}\,\sqrt{\phi_{eE}(\xi_N;\theta)}} \\
&\leq \frac{\|\mathbf{y} - \eta_X(\hat{\theta}_{LS})\| + \|\mathbf{y} - \eta_X(\theta)\|}{\sqrt{N}\,\sqrt{\phi_{eE}(\xi_N;\theta)}} \leq \frac{2\,\|\mathbf{y} - \eta_X(\theta)\|}{\sqrt{N}\,\sqrt{\phi_{eE}(\xi_N;\theta)}}\,.
\end{aligned}
$$

(8)

$\blacksquare$

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

$$\|\hat{\theta}_{LS} - \theta\| \leq \sqrt{\lambda_{\max}[(\mathbf{F}^\top\mathbf{F})^{-1}]}\,\|\mathbf{y} - \eta_X(\theta)\| = \frac{\|\mathbf{y} - \eta_X(\theta)\|}{\sqrt{N}\,\sqrt{\phi_{eE}(\xi_N)}}\,,$$

where $\mathbf{F}$ is the $N \times p$ matrix with ith line equal to $\mathbf{f}^\top(x_i)$. We also have $\|\hat{\theta}_{LS} - \theta\| \le \|\mathbf{y} - \eta_X(\theta)\|/[\sqrt{N}\sqrt{\phi_{eE}(\xi_N, \theta)}]$ in intrinsically linear models (with a flat expectation surface $\mathbb{S}_\eta$) since then $\|\eta_X(\hat{\theta}_{LS}) - \eta_X(\theta)\| \le \|\mathbf{y} - \eta_X(\theta)\|$.

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

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

$$\mathscr{E}_{eE}(\xi) = \frac{\phi_{eE}(\xi)}{\phi_{eE}(\xi^*_{eE})}, \ \xi \in \Xi,$$

where ξ^*_{eE} maximizes $\phi_{eE}(\xi)$.

The concavity of $\phi_{eE}(\cdot)$ implies the existence of directional derivatives and, due to the linearity in ξ 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 $\phi_{eE}(\cdot)$ at ξ in the direction ν is given by*

$$F_{\phi_{eE}}(\xi; \nu) = \min_{\theta \in \Theta_E(\xi)} H_E(\nu, \theta) - \phi_{eE}(\xi)\,,$$

where $\Theta_E(\xi) = \{\theta \in \Theta : H_E(\xi, \theta) = \phi_{eE}(\xi)\}$.

Note that we can write $F_{\phi_{eE}}(\xi; \nu) = \min_{\theta \in \Theta_E(\xi)} \int_{\mathscr{X}} \Psi_{eE}(x, \theta, \xi)\,\nu(\mathrm{d}x)$, where

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

Due to the concavity of $\phi_{eE}(\cdot)$, a necessary and sufficient condition for the optimality of a design measure ξ^*_{eE} is that

$$(10) \qquad\qquad \sup_{\nu \in \Xi} F_{\phi_{eE}}(\xi^*_{eE}; \nu) = 0\,,$$

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 $\xi_{eE}^* \in \Xi$ is optimal for $\phi_{eE}(\cdot)$ if and only if*
(11)
$$\max_{x \in \mathscr{X}} \int_{\Theta_E(\xi_{eE}^*)} \Psi_{eE}(x, \theta, \xi) \, \mu^*(\mathrm{d}\theta) = 0 \text{ for some measure } \mu^* \in \mathscr{M}[\Theta_E(\xi_{eE}^*)],$$

the set of probability measures on $\Theta_E(\xi_{eE}^*)$.

Proof. This is a classical result for maximin design problems, see, e.g., Fedorov and Hackl (1997, Sect. 2.6). We have

$$
\begin{aligned}
0 \leq \sup_{\nu \in \Xi} F_{\phi_{eE}}(\xi; \nu) &= \sup_{\nu \in \Xi} \min_{\theta \in \Theta_E(\xi)} \int_{\mathscr{X}} \Psi_{eE}(x, \theta, \xi) \, \nu(\mathrm{d}x) \\
&= \sup_{\nu \in \Xi} \min_{\mu \in \mathscr{M}[\Theta_E(\xi)]} \int_{\mathscr{X}} \int_{\Theta_E(\xi)} \Psi_{eE}(x, \theta, \xi) \, \mu(\mathrm{d}\theta) \, \nu(\mathrm{d}x) \\
&= \min_{\mu \in \mathscr{M}[\Theta_E(\xi)]} \sup_{\nu \in \Xi} \int_{\mathscr{X}} \int_{\Theta_E(\xi)} \Psi_{eE}(x, \theta, \xi) \, \mu(\mathrm{d}\theta) \, \nu(\mathrm{d}x) \\
(12) \qquad &= \min_{\mu \in \mathscr{M}[\Theta_E(\xi)]} \max_{x \in \mathscr{X}} \int_{\Theta_E(\xi)} \Psi_{eE}(x, \theta, \xi) \, \mu(\mathrm{d}\theta).
\end{aligned}
$$

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 ν 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 ν^* 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 Θ 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 ν is the delta measure δ_{x^*} at some $x^* \in \mathscr{X}$ (depending on ξ). However, although $\phi_{eE,\lambda}(\cdot)$ is smooth for any finite λ, its maximization tends to be badly conditioned for large λ.

In the next section we show that optimal design for $\phi_{eE}(\cdot)$ reduces to linear programming when Θ and $\mathscr{X}$ 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 Θ is compact.

3.3. *Optimal design via linear-programming (Θ is finite).* To simplify the construction of an optimal design, one may take Θ 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 $\mathscr{X}$ is also finite, with $\mathscr{X} = \{x^{(1)}, x^{(2)}, \ldots, x^{(\ell)}\}$, 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 $\mathbf{w} = (w_1, w_2, \ldots, w_\ell)^\top$, w_i being allocated at $x^{(i)}$ for each $i = 1, \ldots, \ell$, such that $\mathbf{c}^\top [\mathbf{w}^\top, \, t]^\top$ is maximized, with $\mathbf{c} = (0, 0, \ldots, 0, 1)^\top$ and $\mathbf{w}$ and t satisfying the constraints

$$\sum_{i=1}^{\ell} w_i = 1 \,,$$
$$w_i \geq 0 \,, \ i = 1, \ldots, \ell \,,$$
$$(13) \qquad \sum_{i=1}^{\ell} w_i \, h_i(\theta^{(j)}) \geq t \,, \ j = 1, \ldots, m \,,$$

where we denoted

$$(14) \qquad h_i(\theta) = \frac{[\eta(x^{(i)}, \theta) - \eta(x^{(i)}, \theta^0)]^2}{\|\theta - \theta^0\|^2} \,.$$

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 ℓ. We shall denote by $(\hat{\mathbf{w}}, \hat{t}) = LP_{eE}(\mathscr{X}, \Theta^{(m)})$ the solution of this problem.

We show below how a compact subset Θ 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 (Θ is a compact subset of $\mathbb{R}^p$).* Suppose now that $\mathscr{X}$ is finite and that Θ is a compact subset of $\mathbb{R}^p$ with nonempty interior. In the LP formulation above, $(\mathbf{w}, t)$ 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 $\theta \in \Theta^{(k)}$ finite. Once a solution $(\mathbf{w}^k, t^k) = LP_{eE}(\mathscr{X}, \Theta^{(k)})$ 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

$\theta^{(k+1)}$ given by the constraint (13) most violated by $\mathbf{w}^k$, i.e.,

$$(15) \qquad \theta^{(k+1)} = \arg\min_{\theta \in \Theta} H_E(\mathbf{w}^k, \theta) \,,$$

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

0) Take any vector $\mathbf{w}^0$ of nonnegative weights summing to one, choose $\epsilon > 0$, set $\Theta^{(0)} = \varnothing$ 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 $(\mathbf{w}^{k+1}, t^{k+1}) = LP_{eE}(\mathscr{X}, \Theta^{(k+1)})$

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

The optimal value $\phi_{eE}^* = \max_{\xi \in \Xi} \phi_{eE}(\xi)$ satisfies

$$\phi_{eE}(\mathbf{w}^{k+1}) \leq \phi_{eE}^* \leq t^{k+1}$$

at every iteration, so that Δ_{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

$$
\begin{aligned}
\phi_{eE}(\mathbf{w}) \leq H_E(\mathbf{w}, \theta^{(j+1)}) \;&=\; H_E(\mathbf{w}^j, \theta^{(j+1)}) + \sum_{i=1}^{\ell} h_i(\theta^{(j+1)})\{\mathbf{w} - \mathbf{w}^j\}_i \\
&=\; \phi_{eE}(\mathbf{w}^j) + \sum_{i=1}^{\ell} h_i(\theta^{(j+1)})\{\mathbf{w} - \mathbf{w}^j\}_i \,,
\end{aligned}
$$

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

$$\nabla^\top \phi_{eE}(\mathbf{w}^j)\mathbf{w} = \phi_{eE}(\mathbf{w}^j) + \nabla^\top \phi_{eE}(\mathbf{w}^j)(\mathbf{w} - \mathbf{w}^j) \geq t \,.$$

Therefore, $\mathbf{w}^{k+1}$ determined at step 2 maximizes the piecewise-linear approximation

$$\min_{j=0,\ldots,k}\{\phi_{eE}(\mathbf{w}^j) + \nabla^\top \phi_{eE}(\mathbf{w}^j)(\mathbf{w} - \mathbf{w}^j)\}$$

of $\phi_{eE}(\mathbf{w})$ with respect to the vector of weights $\mathbf{w}$, 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 Θ is a compact set. We found that the following simple procedure is rather efficient. Construct a finite grid, or a space-filling design, $\mathscr{G}^0$ in Θ. Then, for $k = 0, 1, 2, \ldots$

$$(16) \quad \begin{cases} (i) & \text{compute } \hat{\theta}^{k+1} = \arg\min_{\theta' \in \mathscr{G}^k} H_E(\mathbf{w}^k, \theta'); \\ (ii) & \text{perform a local minimization of } H_E(\mathbf{w}^k, \theta) \\ & \text{with respect to } \theta \in \Theta, \text{ initialized at } \hat{\theta}^{k+1}; \\ & \text{let } \theta^{(k+1)} \text{ denote the solution}; \\ (iii) & \text{set } \mathscr{G}^{k+1} = \mathscr{G}^k \cup \{\theta^{(k+1)}\}. \end{cases}$$

The optimal value $\phi_{eE}(\xi_{eE}^*)$ can then be approximated by $H_E(\mathbf{w}^{k+1}, \theta^{(k+2)})$ 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 $\mathbf{w}^{k+1}$ at step 2 to some neighborhood of the best solution obtained so far, which forms the central idea of bundle methods, see Lemaréchal, Nemirovskii 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 $\mathbf{w}$ 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 θ, denoted by $g(\theta)$, possibly nonlinear. We assume that

$$\mathbf{c} = \mathbf{c}(\theta) = \left.\frac{\partial g(\theta)}{\partial \theta}\right|_{\theta = \theta^0} \neq \mathbf{0}.$$

Denote

$$(17) \qquad H_c(\xi, \theta) = H_c(\xi, \theta; \theta^0) = \frac{\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2}{|g(\theta) - g(\theta^0)|^2}$$

and consider the design criterion defined by

$$(18) \qquad \phi_{ec}(\xi) = \min_{\theta \in \Theta} H_c(\xi, \theta) \,,$$

to be maximized with respect to the design measure ξ.

When $\eta(x, \theta)$ and the scalar function $g(\theta)$ are both linear in θ, with $g(\theta) = \mathbf{c}^\top \theta$, we get

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

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

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

which justifies that we consider $\phi_{ec}(\xi)$ as an *extended c-optimality criterion*. At the same time, in a nonlinear situation with larger Θ the determination of an optimal design ξ_{ec}^* maximizing $\phi_{ec}(\xi)$ ensures some protection against $\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2$ being small for some θ 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 θ^0 for ξ_{ec}^*, that is,

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

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

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 ξ in the direction ν 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 ξ^* maximizing $\phi_{ec}(\cdot)$ is that $\sup_{\nu \in \Xi} F_{\phi_{ec}}(\xi^*; \nu) = 0$, which yields an Equivalence Theorem similar to Th. 3.

When both Θ and $\mathscr{X}$ are finite, an optimal design for $\phi_{ec}(\cdot)$ is obtained by solving a LP problem. Compared with Sect. 3.3, we simply need to substitute H_c 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 Θ 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

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

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

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

where ξ_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 $\phi_{eG}(\cdot)$ in linear models. Moreover, the optimum (maximum) value of $\phi_{eG}(\xi)$ equals $1/p$ with $p = \dim(\theta)$.

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

Again, directional derivatives can easily be computed and an optimal design can be obtained by linear programming when Θ and $\mathscr{X}$ are both finite, or with the algorithm of Sect. 3.4 when $\mathscr{X}$ is finite but Θ has nonempty

interior. Note that there are now $m \times \ell$ inequality constraints in (13), given by

$$\sum_{i=1}^{\ell} w_i \, h_i(\theta^{(j)}, x^{(k)}) \geq t \,, \ j = 1, \ldots, m \,, \ k = 1, \ldots, \ell \,,$$

where now

$$h_i(\theta, x) = \frac{[\eta(x^{(i)}, \theta) - \eta(x^{(i)}, \theta^0)]^2}{[\eta(x, \theta) - \eta(x, \theta^0)]^2} \,.$$

Also note that in the algorithm of Sect. 3.4 we need to construct two sequences of sets, $\Theta^{(k)}$ and $\mathscr{X}^{(k)}$, with $\Theta^{(k+1)} = \Theta^{(k)} \cup \{\theta^{(k+1)}\}$ and $\mathscr{X}^{(k+1)} = \mathscr{X}^{(k)} \cup \{\hat{x}^{(k+1)}\}$ at step 2, and (15) replaced by

$$\{\theta^{(k+1)}, \hat{x}^{(k+1)}\} = \arg \min_{\{\theta, x\} \in \Theta \times \mathscr{X}} \frac{\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_{\xi_k}^2}{[\eta(x, \theta) - \eta(x, \theta^0)]^2}$$

with ξ_k the design measure corresponding to the weights $\mathbf{w}^k$.

6. Examples.   We shall use the common notation

$$\xi = \left\{ \begin{array}{ccc} x_1 & \cdots & x_m \\ w_1 & \cdots & w_m \end{array} \right\}$$

for a discrete design measure with m support points x_i and such that $\xi(\{x_i\}) = w_i$, $i = 1, \ldots, m$. In the three examples considered we indicate the values of the parametric, intrinsic and total measure of curvatures at θ^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 θ^0), so that a small curvature does not mean good performance in terms of extended optimality. They are given by

$$
\begin{aligned}
C_{int}(\xi, \theta) &= \sup_{\mathbf{u} \in \mathbb{R}^p - \{\mathbf{0}\}} \frac{\|[I - P_\theta] \sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta)/\partial\theta_i \partial\theta_j] u_j\|_\xi}{\mathbf{u}^\top \mathbf{M}(\xi, \theta) \mathbf{u}} \,, \\
C_{par}(\xi, \theta) &= \sup_{\mathbf{u} \in \mathbb{R}^p - \{\mathbf{0}\}} \frac{\|P_\theta \sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta)/\partial\theta_i \partial\theta_j] u_j\|_\xi}{\mathbf{u}^\top \mathbf{M}(\xi, \theta) \mathbf{u}} \,, \\
C_{tot}(\xi, \theta) &= \sup_{\mathbf{u} \in \mathbb{R}^p - \{\mathbf{0}\}} \frac{\|\sum_{i,j=1}^{p} u_i [\partial^2 \eta(\cdot, \theta)/\partial\theta_i \partial\theta_j] u_j\|_\xi}{\mathbf{u}^\top \mathbf{M}(\xi, \theta) \mathbf{u}} \,, \\
&\leq C_{int}(\xi, \theta) + C_{par}(\xi, \theta) \,,
\end{aligned}
$$

with P_θ the projector

$$(P_\theta f)(x') = \frac{\partial \eta(x', \theta)}{\partial \theta^\top} \, \mathbf{M}^{-1}(\xi, \theta) \int_{\mathscr{X}} \frac{\partial \eta(x, \theta)}{\partial \theta} \, f(x) \, \xi(\mathrm{d}x) \,,$$

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 ξ instead of an exact design $(x_1, \ldots, x_N)$. The connection with the curvature arrays of Bates and Watts (1980) is presented in (Pázman, 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 Pázman (2013, Sect. 9.5.3); LP problems are solved with simplex algorithm; we use sequential quadratic programming for the local minimization of $H(\mathbf{w}^k, \theta)$ 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

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

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

$$\xi_{D,\theta^0}^* \simeq \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 0.4134 \quad 0.3184 \quad 0.2682 \right\}, \ \xi_{E,\theta^0}^* \simeq \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0.5113 \quad 0.4887 \right\}.$$

When only the design points $\mathbf{x}_1 = (0 \ 1)^\top$ and $\mathbf{x}_2 = (1 \ 0)^\top$ are used, the parameters are only locally estimable. Indeed, the equations in θ'

$$\begin{aligned} \eta(\mathbf{x}_1, \theta') &= \eta(\mathbf{x}_1, \theta) \\ \eta(\mathbf{x}_2, \theta') &= \eta(\mathbf{x}_2, \theta) \end{aligned}$$

give not only the trivial solutions $\theta_1' = \theta_1$ and $\theta_2' = \theta_2$ but also θ_1' and θ_2' as roots of two univariate polynomials of the fifth degree (with coefficients depending on θ). Since these polynomials always admit at least one real root, at least one solution exists for θ' that is different from θ. In particular, the vector $\theta^{0\prime} = (-0.9760, 1.0567)^\top$ gives approximately the same values as θ^0 for the responses at $\mathbf{x}_1$ and $\mathbf{x}_2$.

Direct calculations indicate that, for any θ, the maximum of $\|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2$ 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 $\mathscr{X} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. We take $\Theta = [-3, 4] \times [-2, 2]$ and use the algorithm of Sect. 3.4, with the grid $\mathscr{G}^0$ of (16)-(iii) given by a random Latin hypercube design with $10\,000$ points in $[0, 1]^2$ renormalized to Θ (see, e.g., Tang (1993)), to determine optimal designs for $\phi_{eE}(\cdot)$ and $\phi_{eG}(\cdot)$. When initialized with the uniform measure on the four points of $\mathscr{X}$, 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

$$
\xi^*_{eE,\theta^0} \simeq \left\{ \begin{array}{ccc} \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 0.32 & 0.197 & 0.483 \end{array} \right\},
$$

$$
\xi^*_{eG,\theta^0} \simeq \left\{ \begin{array}{cccc} \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ 0.258 & 0.258 & 0.258 & 0.226 \end{array} \right\}.
$$

The performances of the designs ξ^*_D, ξ^*_E, ξ^*_{eE} and ξ^*_{eG} are given in Table 1. The values $\phi_{eE}(\xi^*_E) = \phi_{eG}(\xi^*_E) = 0$ indicate that E-optimal design is not suitable here, the model being only locally identifiable for ξ^*_E. The parametric, intrinsic and total measures of curvature at θ^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 θ^0 do not reveal any particular difficulty concerning ξ^*_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 θ' in a neighborhood of θ, 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 $\phi_{eE}(\cdot)$ and $\phi_{eG}(\cdot)$ would be zero for Θ 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),

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

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

ξ	$\det^{1/3}$	$\lambda_{\min}$	ϕ_{eE}	ϕ_{eG}	C_{par}	C_{int}	C_{tot}
ξ_D^*	**0.652**	0.273	$3.16 \cdot 10^{-3}$	0.108	1.10	0.541	1.22
ξ_E^*	0.625	**0.367**	0	0	1.19	0	1.19
ξ_{eE}^*	0.453	$8.45 \cdot 10^{-2}$	**$8.78 \cdot 10^{-3}$**	$9.74 \cdot 10^{-2}$	3.33	2.69	4.28
ξ_{eG}^*	0.540	0.195	$5.68 \cdot 10^{-3}$	**0.340**	1.33	1.26	1.83

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

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

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

see Atkinson et al. (1993).

We take Θ 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 $\mathscr{G}^0$ of (16)-(*iii*) is taken as a random Latin hypercube design with 10 000 points in $[0, 1]^3$ renormalized to Θ. The number of iterations and computational time depend on ℓ, the number of elements of $\mathscr{X}$. For instance, when $\mathscr{X}$ is the finite set $\{0.2, 0.4, 0.6, \ldots, 24\}$ with $\ell = 120$, and the required precision ϵ 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 $\mathscr{X}$ 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 $\phi_{ec}(\cdot)$ and in Example 4 for $\phi_{eE}(\cdot)$. The performances of the designs ξ_D^*, ξ_E^* and ξ_{eE}^* are indicated in Table 2. One may notice that the design ξ_{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 ξ_D^*, ξ_E^* and ξ_{eE}^* (since they all have $3 = \dim(\theta)$ support points) and the parametric curvatures at θ^0 are rather small (the smallest one is for ξ_{eE}^*). This explains that, the domain Θ being

not too large, the values of $\phi_{eE}(\xi)$ do not differ very much from those of $\phi_E(\xi) = \lambda_{\min}[\mathbf{M}(\xi, \theta^0)]$.

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

$$g_1(\theta) = \int_0^\infty \eta(x, \theta)\, \mathrm{d}x = \theta_1\, (1/\theta_2 - 1/\theta_3)\,;$$

$g_2(\theta)$ 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 $\mathbf{c}_i = \mathbf{c}_i(\theta^0) = \partial g_i(\theta)/\partial\theta|_{\theta^0}$ and denote ξ_{c_i, θ^0} the (locally) optimal design for $g_i(\theta)$ which maximizes $\phi_{c_i}(\xi; \theta^0) = [\mathbf{c}_i^\top \mathbf{M}^-(\xi, \theta^0)\mathbf{c}_i]^{-1}$, for $i = 1, 2, 3$. The ξ_{c_i, θ^0}^* are singular and are approximately given by

$$\xi_{c_1, \theta^0}^* \simeq \left\{ \begin{array}{cc} 0.2327 & 17.63 \\ 0.0135 & 0.9865 \end{array} \right\}\,,$$

$$\xi_{c_2, \theta^0}^* \simeq \left\{ \begin{array}{cc} 0.1793 & 3.5671 \\ 0.6062 & 0.3938 \end{array} \right\}\,,$$

$$\xi_{c_3, \theta^0}^* \simeq \left\{ \begin{array}{c} 1.0122 \\ 1 \end{array} \right\}\,,$$

see Atkinson et al. (1993).

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

$$\xi_{ec_1, \theta^0}^* \simeq \left\{ \begin{array}{ccc} 0.2327 & 1.389 & 23.36 \\ 9 \cdot 10^{-4} & 1.2 \cdot 10^{-2} & 0.9871 \end{array} \right\}\,,$$

$$\xi_{ec_2, \theta^0}^* \simeq \left\{ \begin{array}{cccc} 0.1793 & 0.229 & 3.5671 & 18.42 \\ 5.11 \cdot 10^{-2} & 0.5375 & 0.3158 & 9.56 \cdot 10^{-2} \end{array} \right\}\,,$$

$$\xi_{ec_3, \theta^0}^* \simeq \left\{ \begin{array}{cccc} 0.229 & 1.0122 & 1.389 & 18.42 \\ 8.42 \cdot 10^{-2} & 0.4867 & 0.4089 & 2.02 \cdot 10^{-2} \end{array} \right\}\,.$$

The performances of $\xi_{c_i}^*$ and $\xi_{ec_i}^*$, $i = 1, \ldots, 3$, are indicated in Table 2, together with the curvature measures at θ^0 for $\xi_{ec_i}^*$ (which are nonsingular). For each function g_i of interest, the design $\xi_{ec_i}^*$ performs slightly worse than $\xi_{c_i}^*$ in terms of c-optimality, but, contrarily to $\xi_{c_i}^*$, it allows us to estimate the three parameters θ and guarantees good estimability properties for $g_i(\theta)$ for all $\theta \in \Theta$. Notice that, apart from the c-optimality criteria $\phi_{c_i}(\cdot)$, all criteria considered take the value 0 at the c-optimal designs $\xi_{c_i}^*$. 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 θ^0 into account. Consider a prior distribution $\pi(\cdot)$ on the two components of θ that intervene nonlinearly in $\eta(x, \theta)$, and let $\mathbb{E}_\pi\{\cdot\}$ denote the expectation for $\pi(\cdot)$. Atkinson et al. (1993) indicate that when π equals π_A uniform on $[\theta_2^0 - 0.01, \theta_2^0 + 0.01] \times [\theta_3^0 - 1, \theta_3^0 + 1]$, the design that maximizes $\mathbb{E}_\pi\{\log \det[\mathbf{M}(\xi, \theta)]\}$ 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 $\mathbb{E}_\pi\{\mathbf{c}_i^\top(\theta)\mathbf{M}^-(\xi, \theta)\mathbf{c}_i(\theta)\}$, $i = 1, 2, 3$, are

$$\xi_{Ac_1-A}^* \simeq \left\{ \begin{array}{ccc} 0.2449 & 1.4950 & 18.4903 \\ 0.0129 & 0.0387 & 0.9484 \end{array} \right\},$$

$$\xi_{Ac_2-A}^* \simeq \left\{ \begin{array}{ccc} 0.1829 & 2.4639 & 8.8542 \\ 0.6023 & 0.2979 & 0.0998 \end{array} \right\},$$

$$\xi_{Ac_3-A}^* \simeq \left\{ \begin{array}{ccc} 0.3608 & 1.1446 & 20.9218 \\ 0.0730 & 0.9094 & 0.0176 \end{array} \right\}.$$

When π equals π_B uniform on $[\theta_2^0 - 0.04, \theta_2^0 + 0.04] \times [\theta_3^0 - 4, \theta_3^0 + 4]$, the

average-optimal designs are

$$\xi^*_{AD-B} \simeq \left\{ \begin{array}{ccccc} 0.2034 & 1.1967 & 2.8323 & 7.8229 & 20.1899 \\ 0.2870 & 0.2327 & 0.1004 & 0.0678 & 0.3120 \end{array} \right\},$$

$$\xi^*_{Ac_1-B} \simeq \left\{ \begin{array}{cccc} 0.2909 & 1.7269 & 13.0961 & 39.58 \\ 0.0089 & 0.0365 & 0.2570 & 0.6976 \end{array} \right\},$$

$$\xi^*_{Ac_2-B} \simeq \left\{ \begin{array}{ccccc} 0.2513 & 0.9383 & 2.7558 & 8.8381 & 26.6564 \\ 0.2914 & 0.2854 & 0.1468 & 0.2174 & 0.0590 \end{array} \right\},$$

$$\xi^*_{Ac_3-B} \simeq \left\{ \begin{array}{ccccc} 0.3696 & 1.1383 & 2.4370 & 6.0691 & 24.0831 \\ 0.0971 & 0.3584 & 0.3169 & 0.1634 & 0.0641 \end{array} \right\}.$$

Their performances are indicated in the bottom part of Table 2. Notice that the average-optimal designs for the vague prior π_B are supported on more than three points and thus allow model checking; this is the case too for the two designs $\xi^*_{ec_2}$ and $\xi^*_{ec_3}$. 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 Θ: 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 θ^0, although larger than those for ξ^*_{E,θ^0} and ξ^*_{eE,θ^0}. The performances of ξ^*_{AD-A} and ξ^*_{AD-B} are close to those of ξ^*_{D,θ^0}, and the most interesting features concern designs for estimation of functions of interest $g_i(\theta)$. The designs $\xi^*_{c_i,\theta^0}$ cannot be used if $\theta \neq \theta^0$ and are thus useless in practice. The average-optimal designs $\xi^*_{Ac_i-B}$ perform significantly worse than $\xi^*_{ec_i,\theta^0}$ in terms of $\phi_{ec_i}(\cdot)$ for $i = 1, 2$ and 3 and in terms of $\phi_{c_i}(\cdot)$ for $i = 2$ and 3. On the other hand, the designs $\xi^*_{Ac_i-A}$, constructed for the precise prior π_A, perform significantly better than $\xi^*_{ec_i,\theta^0}$ in terms of $\phi_{c_i}(\cdot)$ for all i. Figure 3 presents $\phi_{c_3}(\xi; \theta)$ as a function of θ, for the three designs $\xi^*_{Ac_3-A}$ (*dashed line*), $\xi^*_{Ac_3-B}$ (*dash-dotted line*) and $\xi^*_{ec_3,\theta^0}$ (*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 θ of the set Θ used for extended c-optimality is intermediate between the supports of π_A and π_B. Although average-optimal designs $\xi^*_{Ac_i-A,B}$ and extended-optimal designs $\xi^*_{ec_i,\theta^0}$ 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.

TABLE 2. *Performances of different designs and curvature measures at θ^0 for the model* (19) *with* $\theta^0 = (21.80,\ 0.05884,\ 4.298)^\top$ *and* $\Theta = [16, 27] \times [0.03, 0.08] \times [3, 6]$; $\det^{1/3} = \phi_D(\xi) = \{\det[\mathbf{M}(\xi, \theta^0)]\}^{1/3}$, $\lambda_{\min} = \phi_E(\xi) = \lambda_{\min}[\mathbf{M}(\xi, \theta^0)]$. *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., 1993).*

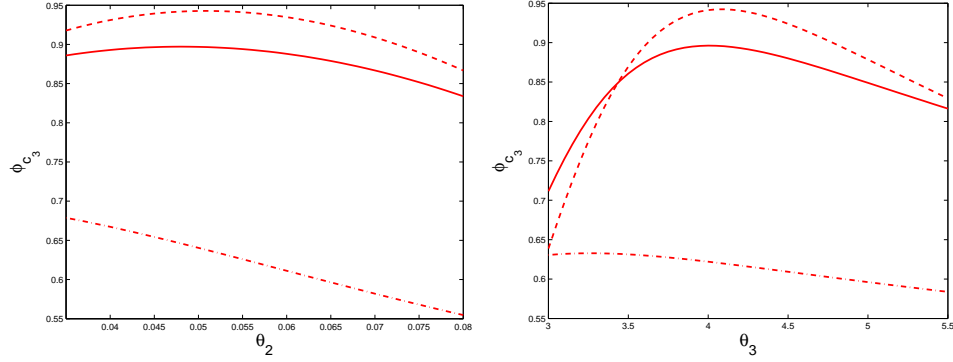| ξ	$\det^{1/3}$	$\lambda_{\min}$	ϕ_{eE}	ϕ_{c_1}	ϕ_{ec_1}	ϕ_{c_2}	ϕ_{ec_2}	ϕ_{c_3}	ϕ_{ec_3}	C_{par}	C_{int}	C_{tot}
ξ_D^*	**11.74**	0.191	0.178	$1.56 \cdot 10^{-4}$	$6.68 \cdot 10^{-5}$	23.43	18.31	0.361	0.356	0.526	0	0.526
ξ_E^*	8.82	**0.316**	0.274	$6.07 \cdot 10^{-5}$	$3.08 \cdot 10^{-5}$	15.89	10.35	0.675	0.667	0.370	0	0.370
ξ_{eE}^*	9.05	0.311	**0.281**	$6.45 \cdot 10^{-5}$	$3.01 \cdot 10^{-5}$	16.62	11.03	0.656	0.644	0.358	0	0.358
$\xi_{c_1}^*$	0	0	0	$\mathbf{4.56 \cdot 10^{-4}}$	0	0	0	0	0			
$\xi_{ec_1}^*$	0.757	$2.70 \cdot 10^{-3}$	$1.92 \cdot 10^{-3}$	$2.26 \cdot 10^{-4}$	$\mathbf{2.17 \cdot 10^{-4}}$	$8.55 \cdot 10^{-2}$	$6.12 \cdot 10^{-2}$	$1.12 \cdot 10^{-2}$	$1.09 \cdot 10^{-2}$	6.51	0	6.51
$\xi_{c_2}^*$	0	0	0	0	0	**35.55**	0	0	0			
$\xi_{ec_2}^*$	7.86	$7.20 \cdot 10^{-2}$	$5.99 \cdot 10^{-2}$	$4.55 \cdot 10^{-5}$	$1.81 \cdot 10^{-5}$	28.82	**27.20**	0.157	0.145	1.12	0.028	1.12
$\xi_{c_3}^*$	0	0	0	0	0	0	0	**1**	0			
$\xi_{ec_3}^*$	4.06	0.162	0.137	$9.70 \cdot 10^{-6}$	$4.19 \cdot 10^{-6}$	6.77	4.36	0.890	**0.865**	1.11	0.263	1.14
ξ_{AD-A}^*	11.74	0.191	0.177	$1.56 \cdot 10^{-4}$	$6.68 \cdot 10^{-5}$	23.53	18.43	0.360	0.355	0.522	0	0.522
$\xi_{Ac_1-A}^*$	2.74	$2.07 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$4.36 \cdot 10^{-4}$	$1.50 \cdot 10^{-4}$	1.12	0.864	$4.06 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$	1.82	0	1.82
$\xi_{Ac_2-A}^*$	6.71	$7.22 \cdot 10^{-2}$	$6.66 \cdot 10^{-2}$	$2.21 \cdot 10^{-5}$	$6.67 \cdot 10^{-6}$	35.16	20.31	0.175	0.175	0.909	0	0.909
$\xi_{Ac_3-A}^*$	3.31	0.118	$8.23 \cdot 10^{-2}$	$8.06 \cdot 10^{-6}$	$3.86 \cdot 10^{-6}$	4.37	3.17	0.937	0.838	1.82	0	1.82
ξ_{AD-B}^*	11.08	0.179	0.159	$1.62 \cdot 10^{-4}$	$6.99 \cdot 10^{-5}$	21.15	15.93	0.338	0.335	0.505	0.056	0.507
$\xi_{Ac_1-B}^*$	2.18	$2.23 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$	$2.34 \cdot 10^{-4}$	$1.56 \cdot 10^{-4}$	0.791	0.644	$5.37 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$	2.12	0.133	2.13
$\xi_{Ac_2-B}^*$	9.45	0.162	0.134	$8.07 \cdot 10^{-5}$	$3.03 \cdot 10^{-5}$	20.05	16.28	0.385	0.358	0.753	0.118	0.761
$\xi_{Ac_3-B}^*$	6.16	0.149	$9.92 \cdot 10^{-2}$	$5.07 \cdot 10^{-5}$	$2.06 \cdot 10^{-5}$	6.60	6.13	0.615	0.587	1.22	0.256	1.25

FIG 3. $\phi_{c_3}(\xi; \theta)$ as a function of θ, 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); $\xi = \xi_{Ac3-A}^*$ in dashed line, $\xi = \xi_{Ac3-B}^*$ in dash-dotted line and $\xi = \xi_{ec3,\theta^0}^*$ in solid line.

Example 4.   For the same regression model (19), we change the value of θ^0 and the set Θ and take $\theta^0 = (0.773,\ 0.214,\ 2.09)^\top$ 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

$$\xi_0 = \left\{ \begin{array}{cccc} 1 & 2 & \cdots & 16 \\ 1/16 & 1/16 & \cdots & 1/16 \end{array} \right\}$$

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

$$\begin{aligned} \xi_{D,\theta^0}^* &\simeq \left\{ \begin{array}{ccc} 0.42 & 1.82 & 6.80 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}, \\ \xi_{E,\theta^0}^* &\simeq \left\{ \begin{array}{ccc} 0.29 & 1.83 & 9.0 \\ 0.4424 & 0.3318 & 0.2258 \end{array} \right\}. \end{aligned}$$

Using the same approach as in Example 3, with the grid $\mathscr{G}^0$ of (16)-(*iii*) obtained from a random Latin hypercube design with 10 000 points in Θ, we obtain

$$\xi_{eE,\theta^0}^* \simeq \left\{ \begin{array}{ccc} 0.38 & 2.26 & 7.91 \\ 0.314 & 0.226 & 0.460 \end{array} \right\}.$$

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

Performances of different designs and curvature measures at θ^0 for the model (19) with $\theta^0 = (0.773,\ 0.214,\ 2.09)^\top$ and $\Theta = [0,5]^3$; $\det^{1/3} = \phi_D(\xi) = \{\det[\mathbf{M}(\xi,\theta^0)]\}^{1/3}$, $\lambda_{\min} = \phi_E(\xi) = \lambda_{\min}[\mathbf{M}(\xi,\theta^0)]$. The optimal (maximum) values of the criteria are indicated in boldface.

ξ	$\det^{1/3}$	$\lambda_{\min}$	ϕ_{eE}	ϕ_{eG}	C_{par}	C_{int}	C_{tot}
ξ_0	$1.85 \cdot 10^{-2}$	$1.92 \cdot 10^{-4}$	$2.28 \cdot 10^{-5}$	$5.66 \cdot 10^{-3}$	180.7	15.73	181.3
ξ_D^*	$\mathbf{5.19 \cdot 10^{-2}}$	$1.69 \cdot 10^{-3}$	$2.64 \cdot 10^{-4}$	$6.70 \cdot 10^{-2}$	58.0	0	58.0
ξ_E^*	$4.51 \cdot 10^{-2}$	$\mathbf{2.04 \cdot 10^{-3}}$	$1.32 \cdot 10^{-4}$	$7.95 \cdot 10^{-2}$	50.7	0	50.7
ξ_{eE}^*	$4.73 \cdot 10^{-2}$	$1.53 \cdot 10^{-3}$	$\mathbf{2.92 \cdot 10^{-4}}$	0.114	54.6	0	54.6
ξ_{eG}^*	$4.11 \cdot 10^{-2}$	$1.31 \cdot 10^{-3}$	$1.69 \cdot 10^{-4}$	$\mathbf{0.244}$	69.7	10.7	69.9

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 $\mathscr{X}$ converges after 34 iterations in about 52 s and gives

$$\xi_{eG,\theta^0}^* \simeq \left\{ \begin{array}{cccc} 0.4 & 1.9 & 5.3 & 16 \\ 0.278 & 0.258 & 0.244 & 0.22 \end{array} \right\}.$$

The performances and curvature measures at θ^0 of ξ_0, ξ_D^*, ξ_E^*, ξ_{eE}^* and ξ_{eG}^* are given in Table 3. The large intrinsic curvature for ξ_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 θ. The values of $\phi_{eE}(\cdot)$ and $\phi_{eG}(\cdot)$ reported in the table indicate that ξ_D^*, ξ_E^*, ξ_{eE}^* or ξ_{eG}^* 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
(20)
$$\phi_{eE}(\xi;\theta^0) = \max\{\alpha \in \mathbb{R} : \|\eta(\cdot,\theta) - \eta(\cdot,\theta^0)\|_\xi^2 \geq \alpha \|\theta - \theta^0\|^2, \text{ for all } \theta \in \Theta\}.$$

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

$$\phi_{eE|K}(\xi;\theta^0) = \max\left\{ \alpha \in \mathbb{R} : \|\eta(\cdot,\theta) - \eta(\cdot,\theta^0)\|_\xi^2 \geq \alpha \frac{\|\theta - \theta^0\|^2}{1 + K\|\theta - \theta^0\|^2}, \right.$$
(21)
$$\left. \text{for all } \theta \in \Theta \right\}$$

for some $K \geq 0$. Equivalently, $\phi_{eE|K}(\xi; \theta^0) = \min_{\theta \in \Theta} H_{E|K}(\xi, \theta)$, 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 $\phi_{eE|K}(\xi) = \lambda_{\min}[\mathbf{M}(\xi)]$ in a linear model and, for a nonlinear model with $\Theta = \mathscr{B}(\theta^0, \rho)$, $\lim_{\rho \to 0} \phi_{eE|K}(\xi; \theta^0) = \lambda_{\min}[\mathbf{M}(\xi, \theta^0)]$ for any $K \geq 0$. Moreover, in a nonlinear model with no overlapping $\phi_{eE|K}(\xi; \theta^0)$ can be made arbitrarily close to $\lambda_{\min}[\mathbf{M}(\xi, \theta^0)]$ 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 θ far from θ^0. Also, properties of $\phi_{eE}(\cdot)$ such as concavity, positive homogeneity, existence of directional derivatives, see Sect. 3.2, remain valid for $\phi_{eE|K}(\cdot)$, for any $K \geq 0$. The maximization of $\phi_{eE|K}(\cdot)$ forms a LP problem when both $\mathscr{X}$ and Θ are finite, see Sect. 3.3, and a relaxation procedure (cutting-plane method) can be used when Θ 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 $\phi_{ec|K}(\xi) = \min_{\theta' \in \Theta} H_{c|K}(\xi, \theta')$ with

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

and

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

for K a positive constant.

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

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

see (6), (5), accounts for the global behavior of $\eta(\cdot, \theta)$ for $\theta \in \Theta$ and obliterates the dependence on θ^0 that is present in $\phi_{eE}(\xi; \theta^0)$. The situation is similar to that in Sect. 3, excepted that we consider now the minimum of H_E with respect to two vectors θ and θ^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 $\phi_{eG}(\cdot)$, respectively defined by $\phi_{Mec}(\xi) = \min_{(\theta, \theta^0) \in \Theta \times \Theta} H_c(\xi, \theta; \theta^0)$, see (17), and by $\phi_{MeG}(\xi) = \min_{(\theta, \theta^0) \in \Theta \times \Theta} \{ \|\eta(\cdot, \theta) - \eta(\cdot, \theta^0)\|_\xi^2 / \max_{x \in \mathscr{X}} \left[\eta(x, \theta) - \eta(x, \theta^0) \right]^2 \}$, 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 Θ a compact set with nonempty interior and μ the Lebesgue measure on Θ, one may have $\min_{\theta^0 \in \Theta} \phi_{eE}(\xi; \theta^0) = 0$ for all designs ξ although $\mu\{\theta^0 \in \Theta : \phi_{eE}(\xi'; \theta^0) > 0\} = 1$ for some design ξ'. 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 θ in a subset of zero measure, see, e.g., Walter (1987)). Example 2 gives an illustration.

Example 2 (continued). When the three polynomial equations $\theta_1' - {\theta_1'}^3 = \theta_1 - \theta_1^3$, $\theta_2' - {\theta_2'}^2 = \theta_2 - \theta_2^2$, ${\theta_1'}^3 + {\theta_2'}^2 = \theta_1^3 + \theta_2^2$ are satisfied, then $\eta(\mathbf{x}, \theta') = \eta(\mathbf{x}, \theta)$ for all $\mathbf{x}$. 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 θ^0 is randomly drawn with a probability measure having a density with respect to the Lebesgue measure on Θ.

In a less pessimistic version of worst-case extended E-optimality we may thus consider a finite set $\Theta^0 \subset \Theta$ for θ^0, obtained for instance by random sampling in Θ, 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 Θ by a finite set is acceptable. A relaxation method, equivalent to the cutting-plane algorithm, can be used when Θ 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 θ^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 Θ correspond-

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

References.

ATKINSON, A. C., CHALONER, K., HERZBERG, A. M. and JURITZ, J. (1993). Optimal experimental designs for properties of a compartmental model. *Biometrics* **49** 325–337.

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of Royal Statistical Society* **B42** 1–25.

BONNANS, J. F., GILBERT, J. C., LEMARÉCHAL, C. and SAGASTIZÁBAL, C. (2006). *Numerical Optimization. Theoretical and Practical Aspects.* Springer, Heidelberg. [2nd ed.].

CHAVENT, G. (1983). Local stability of the output least square parameter estimation technique. *Matematicada Applicada e Computacional* **2** 3-22.

CHAVENT, G. (1990). A new Sufficient Condition for the Wellposedness of Non-Linear Least-Square Problems arising in Identification and Control. In *Analysis and Optimization of Systems* (A. Bensoussan and J. L. Lions, eds.) 452-463. Springer.

CHAVENT, G. (1991). New Size × Curvature Conditions for Strict Quasiconvexity of Sets. *SIAM Journal on Control and Optimization* **29** 1348-1372.

CLYDE, M. and CHALONER, K. (2002). Constrained design strategies for improving normal approximations in nonlinear regression problems. *Journal of Statistical Planning and Inference* **104** 175–196.

DEMIDENKO, E. Z. (1989). *Optimization and Regression.* Nauka, Moscow. In Russian.

DEMIDENKO, E. Z. (2000). Is this the least squares estimate? *Biometrika* **87** 437–452.

DEM'YANOV, V. F. and MALOZEMOV, V. N. (1974). *Introduction to Minimax.* Dover, New York.

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

FEDOROV, V. V. and HACKL, P. (1997). *Model-Oriented Design of Experiments.* Springer, Berlin.

FEDOROV, V. V. and LEONOV, S. L. (2014). *Optimal Design for Nonlinear Response Models.* CRC Press, Boca Raton.

GAUCHI, J. P. and PÁZMAN, A. (2006). Designs in nonlinear regression by stochastic minimization of functionnals of the mean square error matrix. *Journal of Statistical Planning and Inference* **136** 1135–1152.

HAMILTON, D. C. and WATTS, D. G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **27** 241–250.

HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer, Heidelberg.

KELLEY, J. E. (1960). The cutting plane method for solving convex programs. *SIAM Journal* **8** 703–712.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12** 363–366.

KIEFFER, M. and WALTER, E. (1998). Interval analysis for guaranteed nonlinear parameter estimation. In *MODA5 – Advances in Model-Oriented Data Analysis and Experimental Design, Marseilles (France), June 1998* (A. C. Atkinson, L. Pronzato and H. P. Wynn, eds.) 115–125. Physica Verlag, Heidelberg.

LEMARÉCHAL, C., NEMIROVSKII, A. and NESTEROV, Y. (1995). New variants of bundle methods. *Math. Programming* **69** 111–147.

NESTEROV, Y. (2004). *Introductory Lectures to Convex Optimization: A Basic Course.* Kluwer, Dordrecht.

PÁZMAN, A. (1984). Nonlinear least squares – Uniqueness versus ambiguity. *Math. Operationsforschung u. Statist. Ser. Statistics* **15** 323–336.

PÁZMAN, A. (1993). *Nonlinear Statistical Models.* Kluwer, Dordrecht.

PÁZMAN, A. and PRONZATO, L. (1992). Nonlinear experimental design based on the distribution of estimators. *Journal of Statistical Planning and Inference* **33** 385–402.

PRONZATO, L. (2009). On the regularization of singular c-optimal designs. *Mathematica Slovaca* **59** 611–626.

PRONZATO, L., HUANG, C. Y. and WALTER, E. (1991). Nonsequential T-optimal design for model discrimination: new algorithms. In *Proc. PROBASTAT'91* (A. PÁZMAN and J. VOLAUFOVÁ, eds.) 130–136.

PRONZATO, L. and PÁZMAN, A. (1994). Second-order approximation of the entropy in nonlinear least-squares estimation. *Kybernetika* **30** 187–198. *Erratum* 32(1):104, 1996.

PRONZATO, L. and PÁZMAN, A. (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties.* Springer, LNS 212, New York, Heidelberg.

PUKELSHEIM, F. (1993). *Optimal Experimental Design.* Wiley, New York.

RATKOWSKY, D. A. (1983). *Nonlinear Regression Modelling.* Marcel Dekker, New York.

SHIMIZU, K. and AIYOSHI, E. (1980). Necessary conditions for min-max problems and algorithm by a relaxation procedure. *IEEE Transactions on Automatic Control* **25** 62–66.

SILVEY, S. D. (1980). *Optimal Design.* Chapman & Hall, London.

TANG, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association* **88** 1392–1397.

WALTER, E., ed. (1987). *Identifiability of Parametric Models.* Pergamon Press, Oxford.

WALTER, E. and PRONZATO, L. (1995). Identifiabilities and nonlinearities. In *Nonlinear Systems. Modeling and Estimation* (A. J. Fossard and D. Normand-Cyrot, eds.) 3, 111–143. Chapman & Hall, London.

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