Optimal design of experiments via linear programming

Katarína Burclová¹ and Andrej Pázman²
Comenius University Bratislava³

April 24, 2015

Abstract

We investigate the possibility of extending some results of Pázman and Pronzato (2014) to a larger set of optimality criteria. Namely, in a linear regression model the problem of computing $D_r$, $A_r$, $E_k$-optimal designs, of combining these optimality criteria, and the “criterion robust” problem of Harman (2004) are reformulated here as “infinite-dimensional” linear programming problems. Approximate optimum designs can then be computed by a modified cutting-plane method, and this is checked on examples. Finally, the expressions for these criteria are reformulated in terms of the response function of an even nonlinear model.

Keywords: Regression models, optimality criteria, concave maximization, cutting-plane method, criterion-robust design.

1 Introduction

We consider a regression model

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

where $y_i$ are observed variables, $\varepsilon_i$ are observation errors, which satisfy $E(\varepsilon_i) = 0$, and $V\sigmar(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, $\sigma^2$ is not supposed to be known. The value of $\theta$ is a priori restricted to a parameter space $\Theta$. In a vector notation the model is

$$y = \eta_X(\theta) + \varepsilon,$$
$$E(\varepsilon) = 0, V\sigmar(\varepsilon) = \sigma^2 I.$$

Here $X = (x_1, \ldots, x_N)$ is the exact design with points $x_i \in \mathcal{X}$, $y = (y_1, \ldots, y_N)^\top$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^\top$, $\eta_X(\theta) = (\eta(x_1, \theta), \ldots, \eta(x_N, \theta))^\top$. The design space $\mathcal{X}$ is supposed here to be finite. Instead of the exact design $X$ we can consider equivalently for any $x \in \mathcal{X}$ the value $\xi_X(x)$ of the relative frequency of $x$ within the design $X$. By a standard

¹katarina.burclova@fmph.uniba.sk
²pazman@fmph.uniba.sk
³Faculty of Mathematics, Physics and Informatics, Mlynská dolina, 842 48 Bratislava, Slovak Republic
approximation procedure, we consider the set $\Xi$ of all probability measures defined on $\mathcal{X}$, as the set of all approximate designs allowed in the experiment.

In the main part of the present paper we suppose the linearity of the response function $\eta (x_i, \theta) = f^\top (x_i) \theta$, and we suppose $\Theta = \mathbb{R}^p$. In a standard way, to any $\xi \in \Xi$ is associated its information matrix

$$M (\xi) = \sum_{x \in \mathcal{X}} f (x) f^\top (x) \xi (x),$$

with $f(x) = (f_1(x), \ldots, f_p(x))^\top$. According to the aim of the experiment, we may choose an optimality criterion $\phi (\xi)$, and a design $\mu$ is considered $\phi$-optimal when $\phi (\mu) = \max_{\xi \in \Xi} \phi (\xi)$. Standard criteria $\phi (\cdot)$ are concave functions on $\Xi$ having a statistical interpretation.

In Pronzato and Pázmán (2013) the criteria of $E$-, $c$-, and $G$-optimality were considered, and the corresponding criterion functions have been rewritten in a form

$$\phi (\xi) = \min_{u \in \mathbb{R}^p} \sum_{x \in \mathcal{X}} T (u, x) \xi (x)$$

with given $T (u, x)$. This, together with the standard restrictions on $\xi$, defines an “infinite-dimensional” linear programming (LP) problem: to choose the values of $\xi (x); x \in \mathcal{X}$ and of $t \in \mathbb{R}$ so to maximize $t$ under infinitely many linear restrictions:

$$\sum_{x \in \mathcal{X}} T (u, x) \xi (x) \geq t \quad \text{for any } u \in \mathbb{R}^p,$$

$$\xi (x) \geq 0 \quad \text{for any } x \in \mathcal{X}, \text{ and } \sum_{x \in \mathcal{X}} \xi (x) = 1.$$

In particular, for $E$-optimality, with $\phi_E (\xi)$ equal to the minimum eigenvalue of $M (\xi)$, we have

$$\phi_E (\xi) = \min_{u \in \mathbb{R}^p} \frac{u^\top M (\xi) u}{u^\top u} = \min_{u \in \mathbb{R}^p} \sum_{x \in \mathcal{X}} \frac{|f^\top (x) u|^2}{u^\top u} \xi (x).$$

The main idea of Pázmán and Pronzato (2014) was to substitute the nonlinear response function $\eta (x, \theta)$ instead of $f^\top (x) \theta$ and so to obtain new criteria for nonlinear models, with the aim to detect the lack of identifiability under the design $\xi$. However, a second aim of Pázmán and Pronzato (2014) was to point attention to the fact that for those expressions for criteria an LP method could be used to obtain nearly optimum designs in linear models.

In the present paper we follow this second aim, but for $D$-, $A$-, and $E_k$-optimality criteria and also for the computationally not easy task to find the “criterion robust” optimum design in linear models, or to find a $D$-optimum design under the condition that the $A$-optimality criterion exceeds a given value. The difficulties to achieve also the first aim for $D$-, $A$-, and $E_k$-criteria are discussed in Appendix.

We notice that LP method has been used to compute $c$-optimal design in Harman and Jurík (2008) but under a quite different set-up.

2 Reformulation of the optimality criteria

The $D$-optimal design maximizes $\det (M (\xi))$, hence minimizes the generalized variance of $\hat{\theta}$, the BLUE of $\theta$. The $A$-optimal design minimizes the sum of the variances of $\hat{\theta}_1, \ldots, \hat{\theta}_p$. 

2
The $E_k$-optimal design maximizes the sum of the smallest $k$ eigenvalues of $M(\xi)$. There are many forms of expressing the corresponding criteria function $s$. All forms of $s$ representing the same criterion maintain the ordering of the designs but differ by the scaling of this ordering, say $\phi(\xi) = \ln \det [M(\xi)]$ and $\phi(\xi) = \det^{1/p} [M(\xi)]$ for $D$-optimality, and similarly for the other criteria. Here we prefer criteria functions which are not only concave, but also positively homogeneous, $\phi(\alpha\xi) = \alpha\phi(\xi)$ for $\alpha > 0$ (see Pukelsheim (1993) for a justification). So for the $D$-optimality $\phi_D(\xi) = \det^{1/p} [M(\xi)]$, for the $A$-optimality $\phi_A(\xi) = 1/tr [M^{-1}(\xi)]$ when $M(\xi)$ is nonsingular, and for the $E_k$-optimality $\phi_{E_k}(\xi) = \lambda_1(\xi) + \ldots + \lambda_k(\xi)$ where $\lambda_1(\xi) \leq \lambda_2(\xi) \leq \ldots \leq \lambda_p(\xi)$ is the ordering of eigenvalues of $M(\xi)$ respecting their multiplicity. Denote $u_1(\xi), \ldots, u_p(\xi)$ the corresponding orthonormal eigenvectors of $M(\xi)$. Denote also $\Xi^+ = \{ \mu \in \Xi : M(\mu) \text{ is nonsingular} \}$. $D$- and $A$-optimal designs are evidently localized on $\Xi^+$, what need not to be true for the $E_k$-optimality.

**Theorem 1.** We can write

\[
\phi_D(\xi) = \min_{\mu \in \Xi^+} \sum_{x \in X} H_D(\mu, x)\xi(x)
\]

\[
= \min_{\mu \in \Xi^+} \sum_{x \in X} \left\{ \frac{\det^{1/p} [M(\mu)]}{p} \right\} f^\top(x) M^{-1}(\mu) f(x) \xi(x),
\]

(1)

\[
\phi_A(\xi) = \min_{\mu \in \Xi^+} \sum_{x \in X} H_A(\mu, x)\xi(x)
\]

\[
= \min_{\mu \in \Xi^+} \sum_{x \in X} \left\{ \frac{\|M^{-1}(\mu) f(x)\|^2}{\|\tr(M^{-1}(\mu))\|^2} \right\} \xi(x)
\]

(2)

for any $\xi \in \Xi^+$, and

\[
\phi_{E_k}(\xi) = \min_{\mu \in \Xi^+} \sum_{x \in X} H_{E_k}(\mu, x)\xi(x) = \min_{\mu \in \Xi^+} \sum_{x \in X} \|P^{(k)}(\mu) f(x)\|^2 \xi(x)
\]

(3)

for any $\xi \in \Xi$. Here $P^{(k)}(\mu)$ is the $k$-dimensional orthogonal projector $P^{(k)}(\mu) = \sum_{i=1}^k u_i(\mu) u_i^\top(\mu)$, and $\|\cdot\|$ denotes the Euclidean norm.

**Proof.** In the proof we shall often use that $\tr[AB] = \tr[BA]$ for any matrices $A = A_{l \times s}, B = B_{s \times l}$ (Harville, 2000). By the known inequality between the geometric and arithmetic means of positive numbers (cf. (Steele, 2004, Chap. 2)), we obtain

\[
\left\{ \det S^\top M(\xi) S \right\}^{1/p} = \{\Pi_{i=1}^p \alpha_i\}^{1/p} \leq \frac{1}{p} \sum_{i=1}^p \alpha_i = \frac{1}{p} \tr [S^\top M(\xi) S]
\]

for any nonsingular $p \times p$ matrix $S$. Here $\alpha_1, \ldots, \alpha_p$ are the eigenvalues of $S^\top M(\xi) S$. So $\det^{1/p} [M(\xi)] \leq \frac{1}{p} \det^{1/p} [S^\top S] \tr S^\top S f(x) \xi(x)$, and we have just to put $S = M^{-1/2}(\mu)$ to obtain the expression in (1). If $S = M^{-1/2}(\xi)$, then $\alpha_i = 1; i = 1 \ldots p$, and the geometric mean is equal to the arithmetic mean, so the minimum is attained.

For any nonsingular $p \times p$ matrix $S$ we obtain from the Schwarz inequality

\[
\left[\tr(S)\right]^2 = \left\{ \tr [M^{-1/2}(\xi) S M^{1/2}(\xi)] \right\}^2 \leq \tr [M^{-1}(\xi)] \tr [M^{1/2}(\xi) S^\top S M^{1/2}(\xi)] = \tr [M^{-1}(\xi)] \tr [S M(\xi) S^\top]
\]
since in general \( tr \left[A^\top B \right] \) is a scalar product of matrices \( A, B, \) and since \( M^{-1/2} (\xi) \) and \( M^{1/2} (\xi) \) are symmetric matrices. So \( \{ M^{-1} (\xi) \}^{-1} = tr \left[ SM (\xi) S^\top \right] / [tr (S)]^2 = \sum_x \| S f (x) \|^2 \xi (x) / [tr (S)]^2, \) and we have just to put \( S = M^{-1} (\mu) \) to obtain the expression in (2). When \( S = M^{-1} (\xi), \) we obtain evidently an equality in the Schwarz inequality.

Denote \( P = P^{(k)} (\mu). \) By the definition of \( P^{(k)} (\mu) \) we have \( PP = P \) and \( P = P^\top. \) So

\[
\sum_{x \in X} \| Pf (x) \|^2 \xi (x) = tr \left[ PM (\xi) P \right].
\]

On the other hand, denote \( U = (u_1 (\xi), \ldots, u_p (\xi)), \lambda = \text{diag} \{ \lambda_1 (\xi), \ldots, \lambda_p (\xi) \}, \) and use that \( M (\xi) = U \Lambda U^\top \) to obtain

\[
tr \left[ PM (\xi) P \right] = tr \left[ P U \Lambda U^\top P \right] = tr \left[ \Lambda (PU)^\top (PU) \right] = \sum_{i=1}^p \lambda_i (\xi) \left\{ (PU)^\top (PU) \right\}_{ii} = \sum_{i=1}^p \lambda_i (\xi) \left\{ P u_i (\xi) \right\}^2 = \sum_{i=1}^p \lambda_i (\xi) w_i,
\]

where we denoted \( w_i = \left\{ (PU)^\top (PU) \right\}_{ii} = \| P u_i (\xi) \|^2. \) Since \( UU^\top = U^\top U = I, \) we have

\[
k = tr [P] = tr [P^\top P] = tr [P^\top PU U^\top] = \sum_{i=1}^p \left\{ (PU)^\top (PU) \right\}_{ii} = \sum_{i=1}^p w_i.
\]

Further \( w_i \in [0, 1], \) since \( 0 \leq \| P u_i (\xi) \|^2 \leq \| u_i (\xi) \|^2 = 1. \) So, using that \( \lambda_1 (\xi) \leq \ldots \leq \lambda_p (\xi) \) we obtain that \( \sum_{i=1}^p \lambda_i (\xi) w_i \) is minimized exactly when the weights \( w_i \) have maximum value (= 1) at the smallest \( k \) values of \( \lambda_i (\xi). \)

Summarizing we obtain

\[
\sum_{x \in X} \| Pf (x) \|^2 \xi (x) = tr \left[ PM (\xi) P \right] = \sum_{i=1}^p \lambda_i (\xi) w_i \geq \sum_{i=1}^k \lambda_i (\xi) = \phi_{E_k} (\xi). \tag{4}
\]

In the particular case that \( P = P^{(k)} (\xi) = \sum_{j=1}^k u_j (\xi) u_j^\top (\xi) \) we have \( w_i = \left\| P^{(k)} (\xi) u_i (\xi) \right\|^2 = \left\| u_i (\xi) \right\|^2 = 1 \) if \( i \leq k, \) \( \left\| P^{(k)} (\xi) u_i (\xi) \right\|^2 = 0 \) if \( i > k, \) hence \( \sum_{x \in X} \left\| P^{(k)} (\xi) f (x) \right\|^2 \xi (x) = \sum_{i=1}^k \lambda_i (\xi) = \phi_{E_k} (\xi), \) which together with (4) yields an expression in (3).

\( \square \)

**Remark 1.** We could write in (2) \( \phi_A (\xi) = \min_{B \in \mathcal{B}} \sum_{x \in X} \left\{ \frac{\| B f (x) \|^2}{\| tr (B) \|^2} \right\} \xi (x), \) where \( \mathcal{B} \) is any set of nonsingular matrices containing \( M^{-1} (\xi). \) When this formula should hold for all \( \xi \in \Xi, \) then the set \( \mathcal{B} = \{ M^{-1} (\mu) : \mu \in \Xi^+ \} \) is the smallest of such sets. A similar modification could be done for \( D-\text{optimality in (1).} \) In (3) we could minimize over any set of \( k \)-dimensional projectors containing \( P^{(k)} (\xi). \)

**Remark 2.** As follows from (Pronzato and Pázmán, 2013, Chap. 9.5) we could obtain similar results as in Theorem 1 by considering gradients or subgradients of \( \phi (\xi). \) However, the presented direct proofs, without using a not very common notion of subgradients, can be more attractive for people in applications.
3 The iterative computation by LP; the algorithms and examples

3.1 Algorithm for \(D-, A-,\) and \(E_k\)-optimality

Let us write \(H(\mu, x)\) instead of \(H_D(\mu, x), H_A(\mu, x),\) or \(H_{E_k}(\mu, x)\) from Theorem 1. For the maximization of \(\phi\) we apply a modification of the cutting-plane method Kelley (1960) as presented in Pronzato and Pázman (2013) and Pázman and Pronzato (2014):

0. Take any vector \(\xi^{(0)}\) such that \(\sum_{x \in \mathcal{X}} \xi^{(0)}(x) = 1\) and \(\xi^{(0)}(x) \geq 0 \forall x \in \mathcal{X}\), choose \(\epsilon > 0\), set \(\Xi^{(0)} = \emptyset\) and \(n = 0\).

1. Set \(\Xi^{(n+1)} = \Xi^{(n)} \cup \{\xi^{(n)}\}\).

2. Use the LP solver to find \((\xi^{(n+1)}, t^{(n+1)})\) so to maximize \(t\) satisfying the constraints:
   - \(t > 0, \xi(x) \geq 0 \forall x \in \mathcal{X}, \sum_{x \in \mathcal{X}} \xi(x) = 1\),
   - \(\sum_{x \in \mathcal{X}} H(\mu, x)\xi(x) \geq t \forall \mu \in \Xi^{(n+1)}\).

3. Set \(\Delta^{(n+1)} = t^{(n+1)} - \phi(\xi^{(n+1)})\), if \(\Delta^{(n+1)} < \epsilon\) take \(\xi^{(n+1)}\) as an \(\epsilon\)-optimal design and stop, or else \(n \leftarrow n + 1\) and continue by step 1.

Notice that \(\min_{\mu \in \Xi^{(n+1)}} \sum_{x \in \mathcal{X}} H(\mu, x)\xi(x)\) is an upper piecewise linear approximation of \(\phi(\xi)\). Increasing \(n\), the set \(\Xi^{(n+1)} \subseteq \Xi\) becomes larger and the approximation is better.

On the other hand, when \(n\) is small, the information matrix \(M(\xi^{(n)})\) could be ill-conditioned or even singular. In order to avoid the difficulty with inverse matrices in \(D-\) and \(A\)-optimality, it is possible to use any symmetric positive definite matrix as a substitute for \(M(\xi^{(n)})\) as justified in Remark 1. Alternatively, Pronzato and Pázman (2013) recommend the regularization \(M(\xi^{(n)}) + \gamma I\), where \(\gamma\) is a small positive number and \(I\) is the identity matrix. Note that it is also possible to take \(\Xi^{(0)}\) as a nonempty set containing \(s \geq 1\) initial designs. If \(s\) or \(n\) is large, the probability of ill-conditioned or singular information matrix \(M(\xi^{(n)})\) is less.

The problem of singular information matrix does not appear in \(E_k\)-optimality criteria.

The stopping rule used in the above algorithm follows from the upper and lower bounds for \(\max_{\xi \in \Xi} \phi(\xi)\):

\[
\phi(\xi^{(n+1)}) \leq \max_{\xi \in \Xi} \phi(\xi) \leq t^{(n+1)}.
\]

The first inequality is obvious. Note that \(t^{(n+1)} = \max_{\xi \in \Xi} \min_{\mu \in \Xi^{(n+1)}} \sum_{x \in \mathcal{X}} H(\mu, x)\xi(x)\), while \(\max_{\xi \in \Xi} \phi(\xi) = \max_{\xi \in \Xi} \min_{\mu \in \Xi} \sum_{x \in \mathcal{X}} H(\mu, x)\xi(x), \text{ and } \Xi \supseteq \Xi^{(n+1)}\). This yields the second inequality.

There are also available stopping rules based on the equivalence theorem (Kiefer, 1974; Kiefer and Wolfowitz, 1959), which are considered as standard. Let \(\epsilon_{\text{stop}}\) be a chosen small nonnegative number. An iterative algorithm will stop if \(\phi(\xi^{(n)}) < \epsilon_{\text{stop}}\), where for \(D\)-optimality \(d(\xi^{(n)}) = \max_{\xi \in \mathcal{X}} f^T(x)M^{-1}(\xi^{(n)})f(x) - p\) and for the criterion of \(A\)-optimality \(d(\xi^{(n)}) = \max_{\xi \in \mathcal{X}} f^T(x)M^{-2}(\xi^{(n)})f(x) - tr[M^{-1}(\xi^{(n)})]\) as seen e.g. in Kiefer (1974, 1975). According to Harman (2004) the stopping rule for \(E_k\)-optimality criteria is \(d(\xi^{(n)}) = \phi_{E_k}(\xi^{(n)}) - \max_{\xi \in \mathcal{X}} \sum_{i=1}^k [f^T(x)u_i(\xi^{(n)})]^2\), which can be used only if \(\lambda_k(\xi^{(n)}) < \lambda_{k+1}(\xi^{(n)})\).


As mentioned in (Pronzato and Pázman, 2013, Chap. 9.5), the cutting-plane method can have bad convergence properties (referenced to Bonnans et al. (2006); Nesterov (2004)), one can then use the level method (see Nesterov (2004) or Pronzato and Pázman (2013)), which adds the quadratic programming step in the method of cutting planes.

In the examples below we compare the known optimal designs with results of our algorithm. The computations were performed in Matlab on a bi-processor PC (3.10 Ghz) equipped with 6GB of RAM and with 64 bits Windows 8.1. LP problems were solved with interior point method.

**Example 1.** Consider the nonlinear regression model of Atkinson et al. (1993).

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

We use the algorithm of Sec. 3.1 to compute local \(D\)- and \(E_1\)-optimal designs for the nominal value of the parameter \(\theta^0 = (21.8, 0.05884, 4.298)^T\), so we shall write \(\partial \eta(x, \theta)/\partial \theta\vert_{\theta_0}\) instead of \(f(x)\) everywhere. We take a finite design space containing 24,000 points \(X = \{0.001, 0.002, \ldots, 23.999, 24.000\}, \quad \epsilon = 10^{-10}\) with \(\xi(0)(x) = 1/3\) if \(x \in [0.2, 1, 23]\) and \(\xi(0)(x) = 0\) otherwise. The computed designs are given in Table 1. Notice that the computed results correspond to those in Atkinson et al. (1993).

| \(\phi\) | \(\xi^*\) | \(\phi^*\) | iter. | time | \(d(\xi^*)\) |
|---|---|---|---|---|---|
| \(D\) | 0.229 | 1.389 | 18.417 | \(\phi_D^* = 11.739\) | 64 | 16m 9s | \(1.5 \cdot 10^{-5}\) |
| & 0.3333 | 0.3333 | 0.3333 | | | |
| \(E_1\) | 0.169 | 1.394 | 23.402 | 23.403 | \(\phi_{E_1}^* = 0.3163\) | 49 | 5m 53s | \(3.89 \cdot 10^{-6}\) |
| & 0.1993 | 0.6623 | 0.0415 | 0.0969 | | |

Table 1: Example 1: the locally optimal designs are \(\xi_D^*\) and \(\xi_{E_1}^*\) (column 2); \(\phi_D^* = \phi_D(\xi_D^*)\) and \(\phi_{E_1}^* = \phi_{E_1}(\xi_{E_1}^*)\) (column 3); the number of iterations (column 4) and the computational time (column 5) required until the algorithm stopped; the corresponding value of \(d(\xi^*)\) based on the equivalence theorem (column 6).

### 3.2 Algorithm for computing criterion robust designs

The criteria of \(E_k\)-optimality play a special role in experimental design. We say that the design \(\xi\) is not worse than the design \(\mu\) with respect to the Schur ordering of designs if \(\phi_{E_k}(\xi) \geq \phi_{E_k}(\mu)\) for all \(k = 1, \ldots, p\). Then also \(\phi(\xi) \geq \phi(\mu)\) for many other optimality criteria. However, the Schur ordering is a partial ordering of designs, and a Schur-optimal design exists only in some very particular cases. On the other hand, if we denote by \(\mathcal{O}\) the set of all criteria functions \(\phi(\xi)\), which are concave and positive homogeneous, and moreover are orthogonally invariant in the sense that \(\phi(\xi) = \Phi[M(\xi)]\) with \(\Phi[M(\xi)]=\Phi[U^TM(\xi)U]\) for any orthogonal matrix \(U\), it makes sense to look for a design \(\xi_{ef}\) which is maximin efficient with respect to such criteria, i.e.

\[
\xi_{ef} = \arg \max_{\xi \in \mathcal{O}} \min_{\phi \in \mathcal{O}} \left[ \frac{\phi(\xi)}{\max_{\zeta \in \mathcal{O}} \phi(\zeta)} \right].
\]

Here the ratio \(\frac{\phi(\xi)}{\max_{\zeta \in \mathcal{O}} \phi(\zeta)}\) is called the \(\phi\)-efficiency of the design \(\xi\). This maximin efficiency problem can be simplified (cf. Harman (2004)), the solution \(\xi_{ef}\) coincides with the solution of

\[
\xi_{ef} = \arg \max_{\xi \in \mathcal{O}} \min_{1 \leq k \leq p} \left[ \frac{\phi_{E_k}(\xi)}{\max_{\zeta \in \mathcal{O}} \phi_{E_k}(\zeta)} \right],
\]
i.e. with the design which is maximin efficient in the (finite) class of all $E_k$-optimality criteria. Such a design is called also “criterion robust” in Harman (2004). But even this problem is computationally difficult, mainly because the $E_k$-optimality criteria are not differentiable. For us it is important that we can approach the solution of this problem by the LP programming technique. First, using Theorem 1 we compute $E_k(opt) = \max_{\xi \in \Xi} \phi_{E_k}(\xi)$ for all $k$ (see Sec. 3.1), and then we can formulate another “infinite-dimensional” LP problem: to choose the values of $\xi(x)$; $x \in \mathcal{X}$ and of $t \in \mathbb{R}$ so to maximize $t$ under linear constraints:

$$\sum_{x \in \mathcal{X}} \frac{H_{E_k}(\mu, x)}{E_k(opt)} \xi(x) \geq t \text{ for any } \mu \in \Xi^+ \text{ and for every } k \in \{1, \ldots, p\},$$

$$\xi(x) \geq 0 \text{ for any } x \in \mathcal{X}, \text{ and } \sum_{x \in \mathcal{X}} \xi(x) = 1.$$

In order to compute the maximin efficient design, the algorithm of Sec. 3.1 needs to be modified in step 2. Actually, the constraints in the LP problem will be:

- $t > 0$, $\xi(x) \geq 0 \forall x \in \mathcal{X}$, $\sum_{x \in \mathcal{X}} \xi(x) = 1,$
- $\sum_{x \in \mathcal{X}} \frac{H_{E_k}(\mu, x)}{E_k(opt)} \xi(x) \geq t \forall \mu \in \Xi^{(n+1)}$ and $\forall k = 1, \ldots, p,$

where $E_k(opt) = \max_{\xi \in \Xi} \phi_{E_k}(\xi)$ is computed using the unmodified algorithm of Sec. 3.1 for all $k = 1, \ldots, p$.

**Example 2.** Consider the quadratic regression model on a $q$-dimensional cube:

$$y = \beta_0 + \sum_{i=1}^{q} \beta_i x_i^2 + \sum_{i=1}^{q} \beta(i) x_i + \sum_{i<j} \beta_{ij} x_i x_j + \varepsilon, \quad x = (x_1, \ldots, x_q)^\top \in [-1, 1]^q \quad (5)$$

with a parameter $\beta = (\beta_0, \beta_1, \ldots, \beta_q, \beta(1), \ldots, \beta(q), \beta_{12}, \ldots, \beta_{q-1,q})^\top$ of dimension $p = 1 + 3/2q + q^2/2$. The criterion robust design in the model (5) was analytically studied for $q = 1$ in Harman (2004) and for $q = 2$ in Filová and Harman (2013). The case of $q = 3$ was numerically solved in Filová and Harman (2013).

Consider the set $C_i = \{x \in \{-1, 0, 1\}^q : \sum_{j=1}^{q} |x_j| = i\}$ for $i = 0, 1, \ldots, q$. Thus, $C_0 = \{(0, \ldots, 0)^\top\}$ and $C_q$ is the set of all vertices of the $q$-dimensional cube. We shall denote $C = \bigcup_{i=0}^{q} C_i$ and $\xi(C_i) = \sum_{x \in C_i} \xi(x)$. As mentioned in Filová and Harman (2013), for every $\phi \in \mathcal{C}$ there exists a $\phi$-optimal design $\xi^*$ with support on $C$, such that for all $i = 0, 1, \ldots, q$ the measure $\xi^*(C_i)$ is uniformly distributed over points $x \in C_i$ (see also Gaffke (1987); Heiligers (1992)).

Before computing the criterion robust designs, we needed to evaluate $E_k(opt)$ for $k = 1, \ldots, p$. The algorithm of Sec. 3.1 initialized with the uniform measure on $\mathcal{X} = C$ and with $\varepsilon = 10^{-10}$ gave the optimal values $E_k(opt)$ summarized in Table 2 for $q = 1, 2, 3, 4$. We observed the same optimal designs as calculated in Harman (2004) for $q = 1$ and in Filová and Harman (2013) for $q = 2, 3, 4$.

Then using the algorithm of Sec. 3.2 we computed criterion robust designs on $\mathcal{X} = C$ for $q = 1, 2, 3, 4$ obtaining the same results (except $q = 4$) as in Harman (2004); Filová and Harman (2013), and the optimal mass concentrated on $C_i$ is listed in Table 3. Note, that for $q = 3$ and $q = 4$ the optimal design $\xi^*$ computed by algorithm of Sec. 3.2 does not put mass uniformly among $x \in C_i$ with $i = 0, \ldots, q$. By redistributing the mass $\xi^*(C_i)$
Table 2: Example 2: the optimal values of the $E_k$-optimality criteria on a $q$-dimensional cube for the model (5) and the total computational time required until the optimal values for all $k = 1, \ldots, p$ together were evaluated.

| $q$ | $E_k(\text{opt})$ | $\xi$ | $\Psi$ | iter. | time   |
|-----|-------------------|-------|--------|-------|--------|
| 1   | $\xi^*(C_0)$      | 0.3532|        | 0.7646| 1s     |
|     | $\xi^*(C_1)$      | 0.6468|        |       |        |
| 2   | $\xi^*(C_0)$      | 0.1775| 0.2924 | 0.5304|       |
|     | $\xi^*(C_1)$      |       |        | 0.7060| 108    |
|     | $\xi^*(C_2)$      |       |        |       | 1m 20s |
| 3   | $\xi^*(C_0)$      | 0.0884| 0.2343 | 0.2306| 0.4467 |
|     | $\xi^*(C_1)$      |       | 0.1437 |       | 0.6642 | 464   |
|     | $\xi^*(C_2)$      |       | 0.3062 |       |        | 17m 9s|
|     | $\xi^*(C_3)$      |       | 0.3845 |       |        |        |
| 4   | $\xi^*(C_0)$      | 0.1097| 0.0559 | 0.1437| 0.3062 | 0.3845|
|     | $\xi^*(C_1)$      |       |        |       | 0.6526 | 1453  |
|     | $\xi^*(C_2)$      |       |        |       |        | 10h 20m7s |

Table 3: Example 2: criterion robust designs $\xi^*$ (column 2) and the $O$-minimal efficiency of $\xi^*$, i.e. $\Psi^* = \max_{\xi \in \Xi} \min_k \frac{\phi_{E_k(\xi)}}{E_k(\text{opt})}$ (column 3) on a $q$-dimensional cube for the model (5); the number of iterations (column 4) and the computational time (column 5) required until the algorithm stopped.

uniformly over $x \in C_i$ for $i = 0, \ldots, q$, we obtained new design $\xi^{**}$ of identical $O$-minimal efficiency as achieved in $\xi^*$. Thus, $\xi^{**}$ is another criterion robust design with required uniform measure on $C_i$ for any $i = 0, \ldots, q$.

Alternatively, we computed the criterion robust design for $q = 2$ (thus $p = 6$) on a modified design space $\mathcal{X} = \{-1, -0.95, \ldots, 0.9, 0.95, 1\}^2$ (i.e. $\mathcal{X}$ is a grid consisting of 1,681 two-dimensional points including the set $C$). Assuming that the values $E_k(\text{opt})$ are known or previously computed for all $k = 1, \ldots, p$, the algorithm of Sec. 3.2 initialized with uniform measure on $\mathcal{X}$ and $\epsilon = 10^{-10}$ converged after 102 iterations in 36m and 5s with the same results as given in Table 3.

### 3.3 Algorithm for $D$-optimality conditioned by prescribed $A$-optimality

It is not difficult to see that in the considered LP problems we can easily add some supplementary constraints linear in $\xi$, say a cost constraint $\sum_{x \in \mathcal{X}} c(x) \xi(x) = c$, where $c(x)$ is the cost of an observation at $x$ and $c$ is proportional to the total cost allowed for the whole experiment. What is less evident is that we can combine optimality criteria. Say, when we want to obtain a $D$-optimal design under the condition that the $A$-optimality criterion attains a prescribed value $a$, we have to solve the “infinite-dimensional” LP
problem: to choose the values of $\xi(x)$; $x \in \mathcal{X}$ and of $t \in \mathbb{R}$ so to maximize $t$ under linear constraints:

$$\sum_{x \in \mathcal{X}} H_D(\mu, x)\xi(x) \geq t \text{ for any } \mu \in \Xi^+,$$

$$\sum_{x \in \mathcal{X}} H_A(\mu, x)\xi(x) \geq a \text{ for any } \mu \in \Xi^+,$$

$$\xi(x) \geq 0 \text{ for any } x \in \mathcal{X}, \text{ and } \sum_{x \in \mathcal{X}} \xi(x) = 1.$$

This problem can be solved by the algorithm of Sec. 3.1 with a modification in constraints of the LP problem and in the stopping rule.

0. Take any vector $\xi^{(0)}$ such that $\sum_{x \in \mathcal{X}} \xi^{(0)}(x) = 1$ and $\xi^{(0)}(x) \geq 0 \forall x \in \mathcal{X}$, choose $\epsilon_D > 0$, $\delta_A \approx 0$ , set $\Xi^{(0)} = \emptyset$ and $n = 0$.

1. Set $\Xi^{(n+1)} = \Xi^{(n)} \cup \{\xi^{(n)}\}$.

2. Use the LP solver to find $(\xi^{(n+1)}, t^{(n+1)})$ so to maximize $t$ satisfying the constraints:

- $t > 0$, $\xi(x) \geq 0 \forall x \in \mathcal{X}$, $\sum_{x \in \mathcal{X}} \xi(x) = 1$,
- $\sum_{x \in \mathcal{X}} H_D(\mu, x)\xi(x) \geq t \forall \mu \in \Xi^{(n+1)}$,
- $\sum_{x \in \mathcal{X}} H_A(\mu, x)\xi(x) \geq a \forall \mu \in \Xi^{(n+1)}$.

3. Set $\Delta_D^{(n+1)} = t^{(n+1)} - \phi_D(\xi^{(n+1)})$ and $\Delta_A^{(n+1)} = \phi_A(\xi^{(n+1)}) - a$. If $\Delta_D^{(n+1)} < \epsilon_D$ and $\Delta_A^{(n+1)} > \delta_A$ take $\xi^{(n+1)}$ as an $(\epsilon_D, \delta_A)$-optimal design and stop, or else $n \leftarrow n + 1$ and continue by step 1.

The constant $\delta_A$ is chosen at the beginning of the algorithm. The preferred value is $\delta_A = 0$, however choosing $\delta_A < 0$ but small, we can reduce the strictness of the condition on $A$-optimality.

Now consider the set $\mathcal{A}^{(n)} = \{\xi \in \Xi : \sum_{x \in \mathcal{X}} H_A(\mu, x)\xi(x) \geq a \forall \mu \in \Xi^{(n)}\}$, then $\mathcal{A}^{(n)} \supset \mathcal{A}^{(n+1)} \supset \mathcal{A} = \{\xi \in \Xi : \phi_A(\xi) \geq a\}$. So the exact solution of our problem would be $\xi^* = \max_{\xi \in \mathcal{A}} \phi_D(\xi)$. We can write:

$$t^{(n+1)} = \max_{\xi \in \mathcal{A}^{(n+1)}} \min_{\mu \in \Xi^{(n+1)}} \sum_{x \in \mathcal{X}} H_D(\mu, x)\xi(x)$$

$$\geq \max_{\xi \in \mathcal{A}^{(n+1)}} \min_{\mu \in \Xi^{(n+1)}} \sum_{x \in \mathcal{X}} H_D(\mu, x)\xi(x) = \max_{\xi \in \mathcal{A}^{(n+1)}} \phi_D(\xi),$$

and then

$$\max_{\xi \in \mathcal{A}} \phi_D(\xi) \leq \max_{\xi \in \mathcal{A}^{(n+1)}} \phi_D(\xi) \leq t^{(n+1)}, \quad (6)$$

$$\phi_D(\xi^{(n+1)}) \leq \max_{\xi \in \mathcal{A}^{(n+1)}} \phi_D(\xi) \leq t^{(n+1)}, \quad (7)$$

where

$$\xi^{(n+1)} = \arg \max_{\xi \in \mathcal{A}^{(n+1)}} \min_{\mu \in \Xi^{(n+1)}} \sum_{x \in \mathcal{X}} H_D(\mu, x)\xi(x).$$

Assume that $\delta_A = 0$ and the algorithm stopped, i.e. $t^{(n+1)} - \phi_D(\xi^{(n+1)}) < \epsilon_D$ and $\phi_A(\xi^{(n+1)}) \geq a$. According to (6) and (7) there are only two possibilities: first, if
max_{ξ ∈ A} φ_D(ξ) ≤ φ_D(ξ^{(n+1)}) ≤ t^{(n+1)}, then ξ^{(n+1)} ∈ A^{(n+1)} is even “better” design than we expected; second, φ_D(ξ^{(n+1)}) ≤ max_{ξ ∈ A} φ_D(ξ) ≤ t^{(n+1)}, and the stopping rule implies that max_{ξ ∈ A} φ_D(ξ) − φ_D(ξ^{(n+1)}) < ε_D, thus ξ^{(n+1)} is an ε_D-optimal design in both cases.

**Example 3.** Consider the polynomial regression model of degree d:

\[ y = θ_0 + θ_1 x + θ_2 x^2 + \ldots + θ_d x^d + ε, \quad x ∈ [-1, 1], \quad θ = (θ_0, θ_1, \ldots, θ_d)^Γ. \]

Denote by ξ_{D|a} the D-optimal design under the condition that the A-criterion exceeds a value a. Set \( X = \{-1.00, -0.99, -0.98, \ldots, 0.99, 1.00\} \) as the design space, suppose that the initial design ξ(0) allocates the unit mass uniformly to each \( x ∈ X \), \( ε_D = 10^{-10} \), and \( δ_A = 0 \). In Table 4 are given optimal designs \( ξ_{D|a}^∗ \) for some particular values of a and for \( d = 4 \) computed by the algorithm of Sec. 3.3 with abovementioned setting. Notice that the D- and A-optimal (maximum) values are \( φ_D^∗ = 0.1339 \) and \( φ_A^∗ = 0.0053 \) respectively (see the optimal designs in polynomial regression in (Atkinson and Donev, 1992, Chap. 11) and Pukelsheim and Torsney (1991)). When a is small, the algorithm of Sec. 3.3 will compute the D-optimal design. The initial knowledge of \( φ_A^∗ \) is necessary because if a exceeds \( φ_A^∗ \), the algorithm does not work. Figure 1 displays \( φ_D \) and \( φ_A \) efficiencies of \( ξ_{D|a}^∗ \) as a function of a, i.e. \( eff_D(a) = φ_D(ξ_{D|a}^∗)/φ_D^∗ \) and \( eff_A(a) = φ_A(ξ_{D|a}^∗)/φ_A^∗ \).

| a         | ξ_{D|a} | φ_{D|a} | φ_A(ξ_{D|a}^∗) | iter. time |
|------------|---------|---------|----------------|------------|
| 0.0052     | -1      | -0.68   | 0.68           | 1          |
|            |         | 0.136   | 0.2338         | 0.2604     |
|            |         | 0.2338  | 0.2338         | 0.136      |
| 0.005      | -1      | -0.68   | 0.68           | 1          |
|            |         | 0.1623  | 0.2194         | 0.2366     |
|            |         | 0.2194  | 0.2194         | 0.1623     |
| 0.002      | -1      | -0.66   | -0.65          | 0.65       |
|            |         | 0.0847  | 0.1152         | 0.2        |
|            |         | 0.1151  | 0.0850         | 0.2        |
|            |         | 0.1338  | 0.0044         | 153        |

Table 4: Example 3: the optimal designs \( ξ_{D|a}^∗ \) (column 2) with different choices of a (column 1); \( φ_{D|a}^∗ = φ_D(ξ_{D|a}^∗) \)- the value of the D-optimality criterion (column 3); \( φ_A(ξ_{D|a}^∗) \)- the value of the A-optimality criterion (column 4); the number of iterations (column 5) and the computational time (column 6) required until the algorithm stopped.

### 4 Reformulation of AVE criteria in nonlinear experiments

In general, the information matrix in nonlinear regression model \( y = η_X(θ) + ε \) is a function of the parameter \( θ \). Similarly as in Theorem 1, we rewrite (local) D-, A-, and \( E_k \)-optimality criteria in nonlinear regression model to a form:

\[
φ(ξ, θ) = \min_{μ ∈ Ξ} \sum_{x ∈ X} H(μ, x, θ)ξ(x).
\]  

Here \( Ξ^∗ \) can be replaced by \( Ξ \) or \( Ξ^+ \) depending on the considered criterion like in Theorem 1. The reformulation of expressions in Theorem 1 in terms of average (AVE) optimality criteria \( \int_θ φ(x, θ)dπ(θ) \), where \( π(θ) \) is supposed to be known prior distribution, is also possible and is given in Theorem 2.
Figure 1: The graph of $\phi_A$-efficiency (dashed line) and of $\phi_D$-efficiency (solid line) of $\xi^*_D|_a$ as a function of prescribed value $a$ in Example 3.

**Theorem 2.** We can write

$$\int_\Theta \phi(\xi, \theta) d\pi(\theta) = \min_{\mu \in \Xi} \sum_{x \in X} K(\mu, x, \theta) \xi(x),$$

where $K(\mu, x, \theta) = \int_\Theta H(\mu, x, \theta) d\pi(\theta)$.

**Proof.** The design space $X$ is assumed to be finite, hence the summation and the integration are interchangeable. From (8) we have $\phi(\xi, \theta) \leq \sum_{x \in X} H(\mu, x, \theta) \xi(x)$ for any $\mu \in \Xi^*$ and for all $\theta \in \Theta$. We can write

$$\int_\Theta \phi(\xi, \theta) d\pi(\theta) \leq \sum_{x \in X} \left[ \int_\Theta H(\mu, x, \theta) d\pi(\theta) \right] \xi(x). \quad (9)$$

Since the inequality (9) holds for every $\mu \in \Xi^*$, evidently:

$$\int_\Theta \phi(\xi, \theta) d\pi(\theta) \leq \min_{\mu \in \Xi^*} \sum_{x \in X} \left[ \int_\Theta H(\mu, x, \theta) d\pi(\theta) \right] \xi(x). \quad (10)$$

Theorem 1 implies that minimum is in (8) attained at $\mu = \xi$ for any $\theta \in \Theta$, so we obtain an equality in (9) for $\mu = \xi$, which together with (10) proofs the theorem. \qed

**Appendix: Reformulation of criteria in terms of nonlinear models**

Using the notation $\eta(x, \theta) = f^\top(x) \theta$ we can rewrite the expressions from Theorem 1 to a form, which formally allows an extension of criteria to a nonlinear model

$$y_x = \eta(x, \theta) + \varepsilon_x,$$

$$\theta \in \Theta \subset \mathbb{R}^p.$$
Theorem 3. Let \( \theta^{(0)} \in \Theta \) be a given vector. Denote

\[ \mathcal{V}_{\theta^{(0)}} = \left\{ (\theta^{(1)}, \ldots, \theta^{(p)}) : \forall i, \theta^{(i)} \in \Theta, (\theta^{(i)} - \theta^{(0)}) \neq 0, (\theta^{(i)} - \theta^{(0)})^\top (\theta^{(i)} - \theta^{(0)}) = 0, i \neq j \right\}. \]

Further denote by \( \|\theta^{(i)} - \theta^{(0)}\| \) the Euclidean norm of \( \theta^{(i)} - \theta^{(0)} \), and

\[ \|\eta (\cdot, \theta^{(i)}) - \eta (\cdot, \theta^{(0)})\|_\xi^2 = \sum_{x \in \mathcal{X}} \left[ \eta (x, \theta^{(i)}) - \eta (x, \theta^{(0)}) \right]^2 \xi (x). \]

The “extended” criteria defined as:

\[ \phi_{cD} (\xi) = \min_{(\theta^{(1)}, \ldots, \theta^{(p)}) \in \mathcal{V}_{\theta^{(0)}}} \frac{(1/p) \sum_{i=1}^p \|\eta (\cdot, \theta^{(i)}) - \eta (\cdot, \theta^{(0)})\|_\xi^2}{\left[ \prod_{j=1}^p \|\theta^{(j)} - \theta^{(0)}\| \right]^{1/p}}, \]

\[ \phi_{cA} (\xi) = \min_{(\theta^{(1)}, \ldots, \theta^{(p)}) \in \mathcal{V}_{\theta^{(0)}}} \frac{\sum_{i=1}^p \|\theta^{(i)} - \theta^{(0)}\| \|\eta (\cdot, \theta^{(i)}) - \eta (\cdot, \theta^{(0)})\|_\xi^2}{\left[ \sum_{j=1}^p \|\theta^{(j)} - \theta^{(0)}\| \right]^2}, \]

\[ \phi_{cE_k} (\xi) = \min_{(\theta^{(1)}, \ldots, \theta^{(p)}) \in \mathcal{V}_{\theta^{(0)}}} \frac{\sum_{i=1}^k \|\eta (\cdot, \theta^{(i)}) - \eta (\cdot, \theta^{(0)})\|_\xi^2}{\|\theta^{(i)} - \theta^{(0)}\|^2}, \]

coincide with those in Theorem 1 in case that the model is linear.

Proof. Consider first the expression for \( \phi_{cD} (\xi) \) in Theorem 1. Using the notation from Sec. 2 for every \( \mu \in \Xi^+ \) we can write \( M^{-1} (\mu) = \sum_{i=1}^p \nu_i (\mu) \nu_i^\top (\mu) \) with \( \nu_i (\mu) = u_i (\mu) / \sqrt{\lambda_i (\mu)} \) (the normed eigenvector divided by the square root of the eigenvalue), and \( \|\nu_i (\mu)\|^2 = \lambda_i^{-1} (\mu) \). It follows that

\[ \frac{\det^{1/p} [M (\mu)] f^\top (x) M^{-1} (\mu) f (x)}{p} = \frac{(1/p) \sum_{i=1}^p \left[ f^\top (x) \nu_i (\mu) \right]^2}{\left[ \prod_{j=1}^p \|\nu_j (\mu)\|^2 \right]^{1/p}}. \]

Denote \( \theta^{(i)} (\mu) = \theta^{(0)} + \nu_i (\mu) \). In the linear model \( f^\top (x) \nu_i (\mu) = \eta (x, \theta^{(i)} (\mu)) - \eta (x, \theta^{(0)} (\mu)) \). So from Theorem 1 it follows that

\[ \phi_{cD} (\xi) = \min_{\mu \in \Xi^+} \frac{(1/p) \sum_{i=1}^p \|\eta (\cdot, \theta^{(i)} (\mu)) - \eta (\cdot, \theta^{(0)} (\mu))\|_\xi^2}{\left[ \prod_{j=1}^p \|\theta^{(j)} (\mu) - \theta^{(0)} (\mu)\| \right]^{1/p}}. \] (11)

Evidently \( (\theta^{(1)} (\mu), \ldots, \theta^{(p)} (\mu)) \in \mathcal{V}_{\theta^{(0)}} \). On the other hand, for any \( (\theta^{(1)}, \ldots, \theta^{(p)}) \in \mathcal{V}_{\theta^{(0)}} \) we define \( B = \left[ \sum_{i=1}^p (\theta^{(i)} - \theta^{(0)}) (\theta^{(i)} - \theta^{(0)})^\top \right]^{-1} \). From Remark 1 of Theorem 1 it follows that we can take the minimum in (11) with respect to all \( (\theta^{(1)}, \ldots, \theta^{(p)}) \in \mathcal{V}_{\theta^{(0)}} \) and not with respect to all \( \mu \in \Xi^+ \).

We proceed similarly for \( A \)-optimality. We have \( M^{-2} (\mu) = \sum_{i=1}^p \|\nu_i (\mu)\|^2 \nu_i (\mu) \nu_i^\top (\mu) \) and \( \text{tr} [M^{-1} (\mu)] = \sum_{i=1}^p \lambda_i^{-1} (\mu) \), so

\[ \frac{\|M^{-1} (\mu) f (x)\|^2}{\{\text{tr} [M^{-1} (\mu)]\}^2} = \frac{\sum_{i=1}^p \|\nu_i (\mu)\|^2 \left[ f^\top (x) \nu_i (\mu) \right]^2}{\sum_{j=1}^p \|\nu_j (\mu)\|^2} = \frac{\sum_{i=1}^p \|\theta^{(i)} (\mu) - \theta^{(0)} \|^2 \left[ \eta (x, \theta^{(i)} (\mu)) - \eta (x, \theta^{(0)} (\mu)) \right]^2}{\sum_{j=1}^p \|\theta^{(j)} (\mu) - \theta^{(0)} \|^2}. \]
For the $E_k$-optimality criterion we write $P^{(k)}(\mu) = \sum_{i=1}^{k} \|\nu_i(\mu)\|^{-2} \nu_i(\mu) \nu_i^\top(\mu)$, hence

$$\left\| P^{(k)}(\mu) f(x) \right\|^2 = \sum_{i=1}^{k} \|\nu_i(\mu)\|^{-2} \left[ f^\top(x) \nu_i(\mu) \right]^2$$

$$= \sum_{i=1}^{k} \left[ \eta(x, \theta^{(i)}(\mu)) - \eta(x, \theta^{(0)}(\mu)) \right]^2 \left\| \theta^{(i)}(\mu) - \theta^{(0)} \right\|^2.$$

\[\square\]

**Remark 3.** The expressions in Theorem 3 are evidently linear in $\xi$, so maximization of $\phi_D(\xi)$, $\phi_A(\xi)$, and $\phi_{E_k}(\xi)$ with respect to $\xi$ corresponds to an “infinite-dimensional” LP problem even in a nonlinear model. However this problem is too complex to be used for experimental design. Moreover, in contrast to the criteria considered in Pázman and Pronzato (2014), a clear statistical interpretation is still missing.

**Acknowledgements.** We would like to thank Luc Pronzato for helpful advises. The paper was supported by the Slovak VEGA-Grant No. 1/0163/13.

**References**

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimal experimental designs for properties of a compartmental model. *Biometrics*, 49:325–327.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford University Press, Oxford.

Bonnans, J., Gilbert, J., Lemaréchal, C., and Sagastizábal, C. (2006). *Numerical Optimization. Theoretical and Practical Aspects*. Springer, Heidelberg, 2nd edition.

Filová, L. and Harman, R. (2013). Criterion-robust experimental designs for the quadratic regression on a square and a cube. *Communications in Statistics - Theory and Methods*, 42(11):2044–2055.

Gaffke, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. *Annals of Statistics*, 115(3):942–957.

Harman, R. (2004). Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika*, 60:137–153.

Harman, R. and Jurík, T. (2008). Computing $c$-optimal experimental designs using the simplex method of linear programming. *Computational Statistics and Data Analysis*, 59:247–254.

Harville, D. A. (2000). *Matrix Algebra From a Statistician’s Perspective*. Springer, New York.

Heiligers, B. (1992). Admissible experimental designs in multiple polynomial regression. *Journal of Statistical Planning and Inference*, 31(2):219–233.

Kelley, J. (1960). The cutting plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712.
Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2(5):849–879.

Kiefer, J. (1975). Optimal design: Variation in structure and performance under change of criterion. *Biometrika*, 62(2):277–288.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Annals of Mathematical Statistics*, 30:271–294.

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

Pázman, A. and Pronzato, L. (2014). Optimum design accounting for the global nonlinear behavior of the model. *Annals of Statistics*, 42(4):194–219.

Pronzato, L. and Pázman, A. (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Lecture Notes in Statistics, Vol. 212. Springer, New York, Heidelberg.

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

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Annals of Statistics*, 19(3):1614–1625.

Steele, J. M. (2004). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, New York.