Metrika (2004) 60: 137–153
DOI 10.1007/s001840300301

Metrika

Minimal efficiency of designs under the class of orthogonally invariant information criteria

Radoslav Harman

Department of Probability and Statistics, Faculty of Mathematics, Physics and Informatics
Comenius Univ., Mlynská dolina, 842 48 Bratislava, Slovakia
E-mail: harman@fmph.uniba.sk

Abstract. Consider the linear regression model with uncorrelated errors and an experimental design ξ. In the article, we address the problem of calculating the minimal efficiency of ξ with respect to the class $\mathbb{O}$ of orthogonally invariant information criteria, containing all Kiefer's criteria of ϕ_p-optimality, among others. We show that the $\mathbb{O}$-minimal efficiency of ξ is equal to the minimal efficiency of ξ with respect to a finite class of criteria which generalize the criterion of E-optimality. We also formulate conditions under which a design is maximin efficient, i.e. the most efficiency-stable for criteria from $\mathbb{O}$. To illustrate the results, we calculated the $\mathbb{O}$-minimal efficiency of ϕ_p (in particular D, A and E) optimal designs for polynomial regression on $[-1, 1]$ up to degree 4. Moreover, for the quadratic model we explicitly constructed the $\mathbb{O}$-maximin efficient design.

Key words: Optimal design; Efficiency of designs; E-optimality; Maximin design; Polynomial regression.

MSC 2000: 62K05

1 Introduction and model assumptions

A common situation in design of experiments is that general, statistically motivated considerations lead to a broad, often infinite class of reasonable criteria of optimality. In the same time, usually no design is simultaneously optimal for the entire class. One possibility how to approach this problem is

Supported by: VEGA grant of the Slovak Agency No. 1/0264/03

to select a design which is practically efficient enough, or even the most efficiency-stable with respect to all criteria in consideration.

The aim of this article is to analyze the problem outlined above for the class $\mathbb{O}$ of orthogonally invariant information criteria. For an information criterion, the property of orthogonal invariance corresponds to the geometrical assumption, that the quality of a design should depend only on the shape of the resulting confidence ellipsoid for the vector of parameters, and not on its orthogonal rotation or a shift.

In Section 2, we characterize the class $\mathbb{O}$ in detail, and specify the notion of design efficiency. In Section 3, we focus on a class of criteria generalizing E-optimality, which we call criteria of E_k-optimality. For these criteria we prove an equivalence theorem to facilitate identification of E_k-optimal designs. Next, in Section 4 we show that for any design, the minimal efficiency with respect to the uncountable class $\mathbb{O}$ is equal to the efficiency with respect to the finite class of criteria of E_k-optimality. In Section 5, we define a criterion of minimal efficiency with respect to $\mathbb{O}$, the maximization of which can lead to designs with efficiency satisfactory under all orthogonally invariant information criteria. Finally, in Section 6 we illustrate the obtained results on polynomial regression on the interval $[-1, 1]$; for degrees up to 4 we will calculate the minimal efficiency of Kiefer's ϕ_p-optimal designs. For the quadratic case we also construct the design which is maximin efficient with respect to $\mathbb{O}$. We note that all necessary proofs are deferred to the Appendix.

In the paper, we will consider the usual linear regression model on a compact experimental domain $\mathfrak{X} \subseteq \mathfrak{R}^s$. For each design point $x \in \mathfrak{X}$, we can observe a random variable $Y = f^T(x)\beta + \varepsilon$, where $f : \mathfrak{X} \to \mathfrak{R}^m$ is a continuous vector of known regression functions, $\beta \in \mathfrak{R}^m$ is an unknown vector of parameters, and ε is an unobservable random error. For different observations, the errors are assumed to be uncorrelated, with zero mean and the same variance which is assumed to be 1 without loss of generality.

By an experimental design we understand a probability measure ξ finitely supported on $\mathfrak{X}$. The set of all designs on $\mathfrak{X}$ is denoted by Ξ. The performance of a design $\xi \in \Xi$ is based on the information matrix associated with ξ, which is a positive semidefinite matrix defined by the formula

$$\mathbf{M}(\xi) = \sum_{x;\xi(x)>0} \xi(x) f(x) f^T(x)$$

On the set of information matrices we define a real-valued optimality criterion Φ which measures the largeness of an information matrix, i.e. the quality of the corresponding design. A design ξ^* is Φ-optimal iff $\Phi(\mathbf{M}(\xi^*)) = \sup\{\Phi(\mathbf{M}(\xi)); \xi \in \Xi\}$. In this case, $\mathbf{M}(\xi^*)$ is called a Φ-optimal information matrix. We shall also presume that the set of all information matrices $\mathscr{M} = \{\mathbf{M}(\xi); \xi \in \Xi\}$ contains a regular matrix.

In the article, the symbols $\mathscr{S}^m$, $\mathscr{S}^m_+$, and $\mathscr{S}^m_{++}$ denote the sets of all symmetric, positively semidefinite, resp. positively definite matrices of type $m \times m$. On $\mathscr{S}^m$ we define a function λ such that $\lambda(\mathbf{A}) = (\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}), \ldots, \lambda_m(\mathbf{A}))^T$ is the vector of all (not necessarily distinct) eigenvalues of $\mathbf{A}$ in nondecreasing order; $\lambda_1(\mathbf{A}) \leq \lambda_2(\mathbf{A}) \leq \cdots \leq \lambda_m(\mathbf{A})$. Clearly, $\lambda(\mathbf{A})$ has nonnegative (positive) components for all $\mathbf{A} \in \mathscr{S}^m_+$ (resp. $\mathbf{A} \in \mathscr{S}^m_{++}$). For a vector $a \in \mathfrak{R}^m$, the symbol $diag(a)$ denotes the diagonal matrix of type $m \times m$ with diagonal entries equal to the components of a. On $\mathscr{S}^m$ we use the

Loewner partial ordering $\leq$, which is defined as $\mathbf{A} \leq \mathbf{B}$ iff $\mathbf{B} - \mathbf{A} \in \mathscr{S}^m_+$. For vectors $a, b \in \Re^m$, by $a \leq b$ we mean $a_i \leq b_i$ for all $i = 1, \ldots, m$.

2 The class of orthogonally invariant information criteria

In the article, we will consider the class $\mathbb{O}$ of all orthogonally invariant information criteria, i.e. the set of functions $\Phi : \mathscr{S}^m_+ \to [0, \infty)$, which are not identically zero, and which satisfy all of the following properties:

(I) *isotonicity:* $\mathbf{C} \leq \mathbf{D} \Rightarrow \Phi(\mathbf{C}) \leq \Phi(\mathbf{D})$ for all $\mathbf{C}, \mathbf{D} \in \mathscr{S}^m_+$
(C) *concavity:* $\Phi(\alpha\mathbf{C} + (1-\alpha)\mathbf{D}) \geq \alpha\Phi(\mathbf{C}) + (1-\alpha)\Phi(\mathbf{D})$ for all $\mathbf{C}, \mathbf{D} \in \mathscr{S}^m_+$ and $\alpha \in [0, 1]$
(H) *positive homogeneity:* $\Phi(\alpha\mathbf{C}) = \alpha\Phi(\mathbf{C})$ for all $\mathbf{C} \in \mathscr{S}^m_+$ and $\alpha \geq 0$
(S) *upper semicontinuity:* The level sets $\{\mathbf{C} \in \mathscr{S}^m_+ ; \Phi(\mathbf{C}) \geq c\}$ are closed for all $c \in \Re$
(O) *orthogonal invariance:* $\Phi(\mathbf{U}\mathbf{C}\mathbf{U}^T) = \Phi(\mathbf{C})$ for all $\mathbf{C} \in \mathscr{S}^m_+$ and orthogonal $m \times m$ matrices $\mathbf{U}$

We remark that (I), (C), (H), and (S) are well established properties based on natural demands for a statistically reasonable measure of information gained by the experiment (see [8], Chapter 5). In particular, positive homogeneity is essential for definition of relative efficiency of $\xi \in \Xi$ with respect to $\zeta \in \Xi$ and absolute efficiency of ξ in the form (cf. [8] p. 115, 132)

$$\mathrm{eff}_\Phi(\xi|\zeta) = \frac{\Phi(\mathbf{M}(\xi))}{\Phi(\mathbf{M}(\zeta))}, \text{ resp. } \mathrm{eff}_\Phi(\xi) = \frac{\Phi(\mathbf{M}(\xi))}{\sup_{\zeta \in \Xi} \Phi(\mathbf{M}(\zeta))}$$

For simplicity we will not define the relative efficiency for a design ξ with singular information matrix, but we do include the situation of $\mathbf{M}(\zeta)$ being singular (we assume that $c/0 = +\infty$ for $c > 0$). As Φ is positive on $\mathscr{S}^m_{++}$ (see [8] p. 117) and we assume that $\mathscr{M} \cap \mathscr{S}^m_{++} \neq \emptyset$, the Φ-optimal value $\sup_{\zeta \in \Xi} \Phi(\mathbf{M}(\zeta))$ is positive, which implies that $\mathrm{eff}_\Phi(\xi)$ is a well defined number between 0 and 1.

Next, the property of orthogonal invariance is equivalent to the assumption that $\Phi(\mathbf{C})$ depends only on the eigenvalues of $\mathbf{C}$, i.e. if $\lambda(\mathbf{C}) = \lambda(\mathbf{D})$ for some $\mathbf{C}, \mathbf{D} \in \mathscr{S}^m_+$, then $\Phi(\mathbf{C}) = \Phi(\mathbf{D})$. This is an immediate consequence of the fact that $\Phi(\mathbf{C}) = \Phi(\mathrm{diag}\lambda(\mathbf{C}))$ for any $\mathbf{C} \in \mathscr{S}^m_+$ and an orthogonally invariant Φ. (For more properties of orthogonally invariant matrix functions, see e.g. [1] , [13], [4] or [2] p. 104-108.) Notice also, that the assumption of orthogonal invariance corresponds to the assumption of invariance under orthogonal reparametrization of the model.

It turns out that the properties (I) and (O) could be substituted by a single assumption of "spectral monotonicity"; more precisely:

Proposition 1. *Let* $\Phi : \mathscr{S}^m_+ \to [0, \infty)$. *Then the following two statements are equivalent*: (i) Φ *is isotonic and orthogonally invariant*. (ii) *For all* $\mathbf{C}, \mathbf{D} \in \mathscr{S}^m_+$: *If* $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ *then* $\Phi(\mathbf{C}) \leq \Phi(\mathbf{D})$.

The previous proposition can be geometrically formulated in terms of confidence ellipsoids, as we will specify in the sequel. For $\widehat{\beta} \in \Re^m$, $c > 0$, and $\mathbf{C} \in \mathscr{S}^m_+$ let

$$\mathfrak{E}_{\widehat{\beta},c}(\mathbf{C}) = \left\{ b \in \mathfrak{R}^m; \left(b - \widehat{\beta}\right)^T \mathbf{C}\left(b - \widehat{\beta}\right) \leq c \right\}$$

If the errors are normally distributed and $D(\widehat{\beta}) \in \mathscr{S}^m_{++}$ is the covariance matrix of the least squares estimator $\widehat{\beta}$, then $\mathbf{C} = \left(D(\widehat{\beta})\right)^{-1}$ is the information matrix and $\mathfrak{E}_{\widehat{\beta},c}(\mathbf{C})$ is a confidence ellipsoid covering the true parameter β with probability $P\left[\chi^2_m \leq c\right]$ (For details, see [7] p. 79,80). In the following proposition, a rigid-motion transformation is the composition of an orthogonal transformation and a shift by a vector.

Proposition 2. *Let $\Phi : \mathscr{S}^m_+ \to [0, \infty)$. Then the following two statements are equivalent: (i) Φ is isotonous and orthogonally invariant. (ii) For all $\mathbf{C}, \mathbf{D} \in \mathscr{S}^m_+$: If for any $\widehat{\beta}_1, \widehat{\beta}_2 \in \mathfrak{R}^m$ and $c > 0$ there exists a rigid-motion transformation $\rho : \mathfrak{R}^m \to \mathfrak{R}^m$ such that $\rho(\mathfrak{E}_{\widehat{\beta}_1,c}(\mathbf{C})) \supseteq \mathfrak{E}_{\widehat{\beta}_2,c}(\mathbf{D})$, then $\Phi(\mathbf{C}) \leq \Phi(\mathbf{D})$.*

Among orthogonally invariant information criteria we find all criteria of ϕ_p-optimality for $p \in [-\infty, 1]$ (we use the parametrization as defined in [8] p. 139,140, cf. also [7] p. 94). The class $\mathbb{O}$ also contains the criteria of Ch_k-optimality (see [12]) for all $k = 1, \ldots, m$ defined in their homogeneous and concave version:

$$\Phi_{Ch_k}(\mathbf{C}) = \left(\binom{m}{k}^{-1} \sum_{1 \leq i_1 < \ldots < i_k \leq m} \lambda^{-1}_{i_1}(\mathbf{C}) \ldots \lambda^{-1}_{i_k}(\mathbf{C}) \right)^{-1/k} \quad \text{for} \quad \mathbf{C} \in \mathscr{S}^m_{++}$$

and $\Phi_{Ch_k}(\mathbf{C}) = 0$ for $\mathbf{C} \in \mathscr{S}^m_+ \setminus \mathscr{S}^m_{++}$.

As special cases we have D-optimality ($\approx \phi_0$, or Ch_m): $\Phi_D(\mathbf{C}) = (\det(\mathbf{C}))^{1/m}$; A-optimality ($\approx \phi_{-1}$, or Ch_1): $\Phi_A(\mathbf{C}) = m\left(\mathrm{tr}(\mathbf{C}^{-1})\right)^{-1}$ for $\mathbf{C} \in \mathscr{S}^m_{++}$ and $\Phi_A(\mathbf{C}) = 0$ for $\mathbf{C} \in \mathscr{S}^m_+ \setminus \mathscr{S}^m_{++}$; E-optimality ($\approx \phi_{-\infty}$): $\Phi_E(\mathbf{C}) = \lambda_1(\mathbf{C})$; The trace optimality: ($\approx \phi_1$) $\Phi_T(\mathbf{C}) = m^{-1}\mathrm{tr}(\mathbf{C})$.

Notice also, that any convex combination, or a minimum of a finite set of criteria from $\mathbb{O}$ is again an orthogonally invariant information criterion (cmp. [8], p.124-125), so the set $\mathbb{O}$ is very rich.

3 Criteria of E_k-optimality

For $k \in \{1, \ldots, m\}$, let $\widetilde{\Phi}_{E_k}(\mathbf{A})$ be the sum of the k smallest eigenvalues of $\mathbf{A} \in \mathscr{S}^m$, i.e.

$$\widetilde{\Phi}_{E_k} : \mathscr{S}^m \to (-\infty, \infty), \widetilde{\Phi}_{E_k}(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i(\mathbf{A})$$

Let $\Phi_{E_k} : \mathscr{S}^m_+ \to [0, \infty)$ be the restriction of the function $\widetilde{\Phi}_{E_k}$ onto $\mathscr{S}^m_+$. We will call Φ_{E_k} the criterion of E_k-optimality. Obviously $\Phi_{E_1} = \Phi_E$ and $\Phi_{E_m} = m.\Phi_T$, therefore E_k-optimality can be considered as a generalization of both E and the trace optimality. Clearly, the functions Φ_{E_k} are isotonic, positively homogeneous, upper semicontinuous and orthogonally invariant. Moreover, for any $\mathbf{C} \in \mathscr{S}^m_+$ we have $\Phi_{E_k}(\mathbf{C}) = \min_{\mathbf{U} \in \mathscr{U}_{m,k}} \mathrm{tr}\mathbf{U}^T\mathbf{C}\mathbf{U}$, where $\mathscr{U}_{m,k}$ is the set of all matrices $\mathbf{U}$ of type $m \times k$, such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ (the theorem of Ky Fan; see e.g. [3] p.191). It follows that Φ_{E_k} is a minimum of linear functions, which entails concavity of Φ_{E_k} (see also [4], [6]). Consequently Φ_{E_k} are orthogonally invariant information criteria for all $k = 1, \ldots, m$.

The functions $\widetilde{\Phi}_{E_k}$ are not differentiable everywhere on $\mathscr{S}^m$ (with the exception of the linear function $\widetilde{\Phi}_{E_m}$), but it is possible to find the subdifferential $\partial\widetilde{\Phi}_{E_k}(\mathbf{A})$ of the function $\widetilde{\Phi}_{E_k}$ in any fixed $\mathbf{A} \in \mathscr{S}^m$, i.e. the set of all subgradients of $\widetilde{\Phi}_{E_k}$ in $\mathbf{A}$ (see [11] p. 308 or [10]; notice that the trace of a product of two matrices corresponds to the scalar product):

$$\partial\widetilde{\Phi}_{E_k}(\mathbf{A}) = \left\{\mathbf{Y} \in \mathscr{S}^m; \widetilde{\Phi}_{E_k}(\mathbf{B}) \leq \widetilde{\Phi}_{E_k}(\mathbf{A}) + \mathrm{tr}((\mathbf{B} - \mathbf{A})\mathbf{Y}) \text{ for all } \mathbf{B} \in \mathscr{S}^m\right\}.$$

The following proposition is in all important aspects equal to a claim published in [4] (see also [6]).

Proposition 3. *Let* $\mathbf{A} \in \mathscr{S}^m$, *and let* $\lambda(\mathbf{A}) = \lambda = (\lambda_1, \ldots, \lambda_m)^T$. *Then* $\partial\widetilde{\Phi}_{E_k}(\mathbf{A})$ *is the set of all matrices* $\mathbf{U}\mathrm{diag}(\gamma)\mathbf{U}^T$, *where the orthogonal matrix* $\mathbf{U}$ *satisfies* $\mathbf{A} = \mathbf{U}\mathrm{diag}(\lambda)\mathbf{U}^T$, *and the vector* $\gamma = (\gamma_1, \ldots, \gamma_m)^T$ *satisfies:* $\gamma_i \in [0, 1]$ *for all* $i = 1, \ldots, m$, $\gamma_i = 1$ *if* $\lambda_i < \lambda_k$, $\gamma_i = 0$ *if* $\lambda_i > \lambda_k$, and $\sum_{i=1}^m \gamma_i = k$.

In the case of $\lambda_k(\mathbf{A})$ being *strictly* less than $\lambda_{k+1}(\mathbf{A})$, the subdifferential contains elements $\sum_{i=1}^k u_i u_i^T$, where $\mathbf{A}u_i = \lambda_i(\mathbf{A})u_i$, $\langle u_i, u_j\rangle = \delta_{ij}$ (Kronecker delta) for $i, j = 1, \ldots, k$. It is easy to see that for any choice of vectors u_i, the sum $\sum_{i=1}^k u_i u_i^T$ is a unique matrix: the orthogonal projector on the linear space generated by the eigenvectors of $\mathbf{A}$ corresponding to the k smallest eigenvalues. That is, in this case the function $\widetilde{\Phi}_{E_k}$ is differentiable in $\mathbf{A}$, and the gradient is the matrix of orthogonal projection.

Using $\partial\widetilde{\Phi}_{E_k}$, we can formulate a characterization of Φ_{E_k}-optimal designs, i.e. an "equivalence theorem" for E_k-optimality. (For a simple proof see the Appendix. Cf. also [10].)

Theorem 4. *Let* $\xi \in \Xi$. *Then the following three statements are equivalent:* (*i*) ξ *is* Φ_{E_k}*-optimal.* (*ii*) *There exists* $\mathbf{Y} \in \partial\widetilde{\Phi}_{E_k}(\mathbf{M}(\xi))$, *such that* $\Phi_{E_k}(\mathbf{M}(\xi)) = \max_{x \in \mathfrak{X}} f^T(x)\mathbf{Y}f(x)$. (*iii*) *There exists* $\mathbf{Y} \in \mathscr{S}^m_+$, $\mathrm{tr}(\mathbf{Y}) = k$, $\mathbf{Y} \leq \mathbf{I}$, *such that* $\Phi_{E_k}(\mathbf{M}(\xi)) = \max_{x \in \mathfrak{X}} f^T(x)\mathbf{Y}f(x)$.

Notice that the previous theorem gives us a generalization of the equivalence theorem for E-optimality in the form of [8], p. 182. The equivalence theorem for the trace optimality [8], p. 240 is a direct consequence as well.

According to the discussion above, Theorem 4 provides an easy test of E_k-optimality for those designs ξ, that $\lambda_k(\mathbf{M}(\xi)) < \lambda_{k+1}(\mathbf{M}(\xi))$, i.e. if $\widetilde{\Phi}_{E_k}$ is differentiable in $\mathbf{M}(\xi)$. In such a case we only need to check that

$$\Phi_{E_k}(\mathbf{M}(\xi)) = \max_{x \in \mathfrak{X}} \sum_{i=1}^k \left(f^T(x)u_i\right)^2$$

where $u_1, \ldots, u_k$ are (arbitrary) orthonormal eigenvectors corresponding to $\lambda_1(\mathbf{M}(\xi)), \ldots, \lambda_k(\mathbf{M}(\xi))$.

Notice also, that criteria Φ_{E_k} are not strictly concave (i.e. we can have more than one Φ_{E_k}-optimal information matrix) and, with the exception of Φ_{E_1}, they can be positive or even formally optimal for a singular information matrix (as is the case in the polynomial regression model analyzed in Section 6).

4 Minimal efficiency under the class of orthogonally invariant information criteria

The theorems of this section are central to the article.

Theorem 5. *Let* $\xi, \zeta \in \Xi$, $\mathbf{M}(\xi) \in \mathcal{S}^m_{++}$. *Then*

$$\inf_{\Phi \in \mathbb{O}} \mathrm{eff}_\Phi(\xi|\zeta) = \min_{k=1,\ldots,m} \mathrm{eff}_{\Phi_{E_k}}(\xi|\zeta)$$

A proof of this theorem, as given in the Appendix, relies on the theory of vector majorization. Firstly, it can be proved that the constant $\delta = \min_{k=1,\ldots,m} \mathrm{eff}_{\Phi_{E_k}}(\xi|\zeta)$ is sufficiently small to ensure that $\lambda(\mathbf{M}(\xi))$ majorizes a vector which is equal or greater than $\delta\lambda(\mathbf{M}(\zeta))$ in componentwise comparison. Secondly, it can be shown that this type of dominance of eigenvalues is enough to guarantee that $\Phi(\mathbf{M}(\xi)) \geq \delta\Phi(\mathbf{M}(\zeta))$, i.e. eff $_\Phi(\xi|\zeta) \geq \delta$ for any orthogonally invariant criterion Φ. Therefore $\inf_{\Phi \in \mathbb{O}} \mathrm{eff}_\Phi(\xi|\zeta) \geq \delta$. The converse inequality follows from the fact that $\Phi_{E_k} \in \mathbb{O}$ for all k.

An immediate consequence of the previous theorem is that a design ξ is optimal for all orthogonally invariant information criteria (i.e. ξ is "universally" optimal for the class $\mathbb{O}$) if and only if ξ is Φ_{E_k}-optimal for all $k = 1, \ldots, m$. Notice, that this is the same condition characterizing universally optimal designs as the one obtained by Bondar for a somewhat different class of criteria (see [1] for details). It turns out that designs optimal with respect to all orthogonally invariant criteria do exist in some special models. An example is the equispaced support design in the trigonometric regression on the full circle (compare with [7], p. 185 or [8], p. 241) as we state in the following proposition.

Proposition 6. *Let* $d \in \mathbb{N}$, $f_{2j-1}(x) = \cos(jx)$, $f_{2j}(x) = \sin(jx)$ for $j = 1, \ldots, d$ *and* $f_{2d+1}(x) = 1$. *Let* $\mathfrak{X} = [0, 2\pi]$ *and let* ξ^* *be a uniform design on* $\{x_1, \ldots, x_n\} \subseteq \mathfrak{X}$, *where* $n \geq 2d + 1$ and $x_{i+1} - x_i = 2\pi/n$ *for all* $i = 1, \ldots, n-1$. *Then* ξ^* *is optimal with respect to all* $\Phi \in \mathbb{O}$.

For the purpose of this article, a more important consequence of Theorem 5 is that it gives us a method how to compute the minimal (absolute) efficiency with respect to $\mathbb{O}$ for *any* regular design, even in the case of models which do not permit a universally optimal design. From the previous theorem we immediately obtain:

Theorem 7. *Let* $\xi \in \Xi$. *Then*

$$\inf_{\Phi \in \mathbb{O}} \mathrm{eff}_\Phi(\xi) = \min_{k=1,\ldots,m} \mathrm{eff}_{\Phi_{E_k}}(\xi)$$

Hence, the minimal efficiency with respect to the uncountable set of all orthogonally invariant criteria is simply the minimal efficiency with respect to the set of criteria Φ_{E_k}, which numbers only m elements. In other words, once we computed the model-specific values $v_k = \sup_{\zeta \in \Xi} \Phi_{E_k}(\mathbf{M}(\zeta))$, $k = 1, \ldots, m$, we can directly calculate the minimal efficiency for any design ξ as

$\inf_{\Phi \in \mathbb{O}} \mathrm{eff}_{\Phi}(\xi) = \min_{k=1,\ldots,m} \left\{ v_k^{-1} \Phi_{E_k}(\mathbf{M}(\xi)) \right\}$. (We will call the value $\inf_{\Phi \in \mathbb{O}} \mathrm{eff}_{\Phi}(\xi)$ also the $\mathbb{O}$-minimal efficiency of ξ.)

To demonstrate this method, we will find the values v_k for the polynomial regression models on $[-1, 1]$ up to degree 4. This will allow us to compute the $\mathbb{O}$-minimal efficiency for ϕ_p-optimal designs (see Section 6).

5 Maximin efficient designs with respect to the class $\mathbb{O}$

The results of the previous section entail a natural question: If we are able to compute the $\mathbb{O}$-minimal efficiency of a given design, is there a method how to find the design which maximizes this value?

Consider the matrix function

$$\widetilde{\Phi}_v : \mathscr{S}^m \to (-\infty, \infty); \widetilde{\Phi}_v(\mathbf{A}) = \min_{k=1,\ldots,m} v_k^{-1} \widetilde{\Phi}_{E_k}(\mathbf{A}),$$

where $v = (v_1, \ldots, v_m)^T$ is defined as $v_k = \sup_{\zeta \in \Xi} \Phi_{E_k}(\mathbf{M}(\zeta)) > 0$, $k = 1, \ldots, m$. Clearly, the function Φ_v, defined as the restriction of $\widetilde{\Phi}_v$ onto $\mathscr{S}^m_+$, is an orthogonally invariant information criterion and, in accord with Theorem 7, we can consider it to be the criterion of minimal efficiency with respect to $\mathbb{O}$. Therefore, the Φ_v-optimal design can be called the maximin efficient design with respect to $\mathbb{O}$ (shortly $\mathbb{O}$-maximin efficient design).

As $\widetilde{\Phi}_v$ is a finite minimum of concave functions, it is not difficult to find the subdifferential of $\widetilde{\Phi}_v$, and consequently formulate an equivalence theorem for the $\mathbb{O}$-maximin efficient design. The following proposition is a special case of a known and more general formula which can be found e.g. in [2] p.47. (Cf. also [11] p.223.)

Proposition 8. *Let* $\mathbf{A} \in \mathscr{S}^m$ *and let* $I = \left\{ k \in \{1, \ldots, m\}; v_k^{-1} \widetilde{\Phi}_{E_k}(\mathbf{A}) = \widetilde{\Phi}_v(\mathbf{A}) \right\}$. *Then subdifferential of* $\widetilde{\Phi}_v$ in $\mathbf{A}$ *is the set*

$$\partial \widetilde{\Phi}_v(\mathbf{A}) = \bigcup \left\{ \sum_{k \in I} \alpha_k v_k^{-1} \partial \widetilde{\Phi}_{E_k}(\mathbf{A}); \sum_{k \in I} \alpha_k = 1, \text{ and } \alpha_k \geq 0 \text{ for } k \in I \right\}$$

Theorem 9. *A design* ξ *is maximin efficient with respect to* $\mathbb{O}$ *if and only if there exists* $\mathbf{Y} \in \partial \widetilde{\Phi}_v(\mathbf{M}(\xi))$, *such that* $\Phi_v(\mathbf{M}(\xi)) = \max_{x \in \mathfrak{X}} f^T(x) \mathbf{Y} f(x)$.

The proof of the previous theorem is similar to the proof of (i)⇔(ii) in Theorem 4 and is therefore omitted. Notice that if $\mathbf{M}(\xi)$ is a point of non-differentiability of $\widetilde{\Phi}_v$, then $\partial \widetilde{\Phi}_v(\mathbf{M}(\xi))$ contains a full continuum of subgradients, and the previous theorem does not give us a method how to choose the right one for the equivalence determining optimality. In fact, the search for the appropriate subgradient can be a difficult optimization problem itself. We also remark that the previous theorem can be formulated in terms of directional derivatives of $\widetilde{\Phi}_v$, similarly as e.g. in [5]. Nevertheless, both approaches are essentially same, and they suffer analogous disadvantages.

To illustrate the concept of $\mathbb{O}$-maximin efficiency on an example, we will exhibit the maximin efficient design for the quadratic regression on $[-1, 1]$ in the next section.

6 Example: polynomial regression on $[-1, 1]$

As a special case of the regression model defined in the introduction, consider the polynomial regression of degree d on the experimental domain $\mathfrak{X} = [-1, 1]$ given by the model equation

$$y = \beta_1 + \beta_2 x + \ldots + \beta_{d+1} x^d + \varepsilon$$

In the notation of this article, the vector of unknown parameters of interest $\beta = (\beta_1, \beta_2, \ldots, \beta_{d+1})^T$ is $m = d + 1$ dimensional and the vector of regression functions is $f(x) = (1, x, \ldots, x^d)^T$.

We will analyze this model for degrees $d \in \{1, 2, 3, 4\}$ with the aim to evaluate the minimal efficiency of ϕ_p-optimal designs, $p \in [-\infty, 1]$, with respect to the class $\mathbb{O}$ of all orthogonally invariant information criteria. As special cases, we will obtain the $\mathbb{O}$-minimal efficiency for the D, A, and E optimal designs. Moreover, for the quadratic regression ($d = 2$), we will identify the design which is maximin efficient with respect to $\mathbb{O}$.

Firstly, we need to find the optimal values for the criteria of E_k-optimality. It turns out that the Φ_{E_k}-optimal designs for degrees $d = 1, 2, 3, 4$, and for $k = 1, \ldots, d + 1$ can be found explicitly. For $k = 1$, that is for the ordinary E-optimality, the optimal designs are known (see [9] or [8] p. 232-237), and we denote them by $\xi_1^{(d)}$. In particular, the simplest design $\xi_1^{(1)}$ assigns the weight $\frac{1}{2}$ to -1 and 1. Next, by $\xi_3^{(4)}$ we denote the design which assigns the weight $1/6$ to -1,1, and the weight $2/3$ to 0. Using Theorem 4, it is simple to verify that for any considered combination of d and k, some of the five designs described (i.e. $\xi_1^{(1)}, \ldots, \xi_1^{(4)}$ or $\xi_3^{(4)}$) is Φ_{E_k} -optimal. The Φ_{E_k}-optimal designs $\xi_k^{(d)}$ and the corresponding optimal values $v_k^{(d)} = \Phi_{E_k}(\mathbf{M}(\xi_k^{(d)}))$ are summarized in Table 1.

Table 1.

$\xi_k^{(d)}; v_k^{(d)}$	d = 1	d = 2	d = 3	d = 4
k = 1	$\xi_1^{(1)}; 1$	$\xi_1^{(2)}; 1/5$	$\xi_1^{(3)}; 1/25$	$\xi_1^{(4)}; 1/129$
k = 2	$\xi_1^{(1)}; 2$	$\xi_1^{(1)}; 1$	$\xi_1^{(2)}; 1/5$	$\xi_1^{(3)}; 1/25$
k = 3	–	$\xi_1^{(1)}; 3$	$\xi_1^{(1)}; 2$	$\xi_3^{(4)}; 1/3$
k = 4	–	–	$\xi_1^{(1)}; 4$	$\xi_1^{(1)}; 2$
k = 5	–	–	–	$\xi_1^{(1)}; 5$

Notice that in the case of line regression ($d = 1$) the design $\xi_1^{(1)}$ is Φ_{E_k}-optimal for both $k = 1, 2$. This means that $\xi_1^{(1)}$ is optimal for all orthogonally invariant criteria.

For $d = 2, 3, 4$, the values $v_k^{(d)}$ allow us compute the $\mathbb{O}$-minimal efficiency for the ϕ_p-optimal designs, $p \in [-\infty, 1]$. For polynomial regression on $[-1, 1]$, the $E \approx \phi_{-\infty}$, $A \approx \phi_{-1}$, and $D \approx \phi_0$ optimal designs are known (see e.g. [8] Chapter 9). For these designs, Tables 2,3 and 4 give the E_k-efficiencies.

Table 2. (d = 2)

eff	E_1	E_2	E_3	$\mathbb{O}$-minimal
D	0.730745	0.812816	0.777778	0.730745
A	0.954915	0.690983	0.666667	0.666667
E	1.000000	0.600000	0.600000	0.600000

Table 3. (d = 3)

eff	E_1	E_2	E_3	E_4	$\mathbb{O}$-minimal
D	0.744733	0.717848	0.608890	0.656000	0.608890
A	0.967451	0.721081	0.432425	0.523166	0.432425
E	1.000000	0.638861	0.396386	0.501250	0.396386

Table 4. (d = 4)

eff	E_1	E_2	E_3	E_4	E_5	$\mathbb{O}$-minimal
D	0.738857	0.714339	0.683360	0.603928	0.577976	0.577976
A	0.969005	0.709550	0.637469	0.443494	0.448069	0.443494
E	1.000000	0.627669	0.603550	0.433000	0.441860	0.433000

According to Theorem 7, the minimum of these $d+1$ efficiencies equals to the $\mathbb{O}$-minimal efficiency.

For an arbitrary $p \in [-\infty, 1]$, the ϕ_p-optimal design can be computed using general iterative methods (see [7] Chapter V). The Graphs 1,2,3 in the Appendix plot the numerically computed E_k-efficiencies, and the $\mathbb{O}$-minimal efficiency of ϕ_p-optimal designs. The parameter r, which corresponds to the horizontal axis, relates to the parameter p via the function $p(r) = \frac{2r}{1+r}$ for $r \in (-1, 1]$ and $p(r) = -\infty$ for $r = -1$. Hence, E, A, D and the trace optimality correspond to $r = -1, -1/3, 0$, resp. 1. Notice also that the function $p(\cdot)$ is chosen such that $p(r) + p(-r) = p(r)p(-r)$, which means that $p(r)$ and $p(-r)$ are conjugate numbers.

From the graphs of minimal efficiency with respect to $\mathbb{O}$ we see that the D-optimal design performs well, yet it does not maximize the $\mathbb{O}$-minimal efficiency even within the class of ϕ_p-optimal designs.

In general, it turns out to be difficult to find the $\mathbb{O}$-maximin efficient design without resorting to numerical procedures for maximization of a nondifferentiable function (see e.g. [14] for such a numerical algorithm in the context of optimal experimental design). Nevertheless, for the case of the quadratic regression, we can specify the maximin efficient design explicitly:

Proposition 10. *Let ξ be the design which assigns the weight $w = \frac{46}{251} + \frac{15}{502}\sqrt{22} \doteq 0.32342$ to the points -1,1 and the weight $1 - 2w \doteq 0.35316$ to the point* 0. *Then ξ is maximin efficient with respect to $\mathbb{O}$ for the quadratic regression on $[-1, 1]$. The $\mathbb{O}$-minimal efficiency of ξ is $\frac{145}{251} + \frac{10}{251}\sqrt{22} \doteq 0.76456$.*

The $\phi_{p(r)}$-efficiencies of D, A, E-optimal designs and the design ξ from the previous proposition are depicted in Graph 4 in the Appendix (the graph of $r \to eff_{\phi_{p(r)}}(\xi)$ is denoted by M). Notice, that the efficiencies of ξ and the D-optimal design are close to each other, although ξ is more efficient for the criterion of E-optimality and slightly less efficient for the criterion of trace optimality.

It can be shown that in the case of quadratic regression, the $\mathbb{O}$-maximin efficient design ξ must be ϕ_{p_ξ}-optimal for some $p_\xi \in [-\infty, 1]$. (More generally, if $d = 2$ then *any* design which is optimal with respect to some $\Phi \in \mathbb{O}$ is also ϕ_p-optimal for some p; see the considerations in [8] p. 334. On the other hand, our preliminary numerical computations suggest that for $d > 2$ the $\mathbb{O}$-maximin efficient design does *not* belong to the class of ϕ_p-optimal designs). One can calculate that $p_\xi \doteq -0.0648$ (i.e. $r_\xi \doteq -0.0314$; cf. with Graph 4).

7 Appendix

7.1 Proof of Proposition 1

Let $\Phi : \mathscr{S}_+^m \to [0,\infty)$ be isotonic and orthogonally invariant, and let $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ for some $\mathbf{C},\mathbf{D} \in \mathscr{S}_+^m$. Then $\mathrm{diag}(\lambda(\mathbf{C})) \leq \mathrm{diag}(\lambda(\mathbf{D}))$ in Loewner ordering, thus $\Phi(\mathbf{C}) = \Phi(\mathrm{diag}(\lambda(\mathbf{C}))) \leq \Phi(\mathrm{diag}(\lambda(\mathbf{D}))) = \Phi(\mathbf{D})$. This proves the "(i)⇒(ii)" part of the proposition.

To prove the converse, suppose that for any $\mathbf{C},\mathbf{D} \in \mathscr{S}_+^m$ such that $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$, we have $\Phi(\mathbf{C}) \leq \Phi(\mathbf{D})$. If $\mathbf{C} \leq \mathbf{D}$, then from 7.7.4 (c) in [3], p. 471 we have $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ which implies $\Phi(\mathbf{C}) \leq \Phi(\mathbf{D})$ by the assumption . (See also [1], p. 327.) This proves isotonicity of Φ. Moreover, if $\lambda(\mathbf{C}) = \lambda(\mathbf{D})$ then we have both $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ and $\lambda(\mathbf{D}) \leq \lambda(\mathbf{C})$, which means that $\Phi(\mathbf{C}) \leq \Phi(\mathbf{D}) \leq \Phi(\mathbf{C})$, i.e. $\Phi(\mathbf{C}) = \Phi(\mathbf{D})$. Hence, $\Phi(\mathbf{C})$ depends only on the eigenvalues of $\mathbf{C}$, which means that Φ is orthogonally invariant. ∎

7.2 Proof of Proposition 2

Proposition 2 follows from Proposition 1 once we prove that for all $\mathbf{C},\mathbf{D} \in \mathscr{S}_+^m$ these two statements are equivalent: (iii) For any $\widehat{\beta}_1, \widehat{\beta}_2 \in \mathfrak{R}^m$ and $c > 0$ there exists a rigid-motion transformation $\rho : \mathfrak{R}^m \to \mathfrak{R}^m$ such that $\rho(\mathfrak{E}_{\widehat{\beta}_1,c}(\mathbf{C})) \supseteq \mathfrak{E}_{\widehat{\beta}_2,c}(\mathbf{D})$. (iv) $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ componentwise.

It is simple to show that for any $\mathbf{H} \in \mathscr{S}_+^m$, $\widehat{\beta}, s \in \mathfrak{R}^m$, $c > 0$ and a regular matrix $\mathbf{A}$ of type $m \times m$ we have $\mathbf{A}\cdot\mathfrak{E}_{\widehat{\beta},c}(\mathbf{H}) + s = \mathfrak{E}_{\mathbf{A}\widehat{\beta}+s,c}\left(\left(\mathbf{A}^{-1}\right)^T \mathbf{H}\mathbf{A}^{-1}\right)$.

"(iii)⇒(iv)" If (iii) holds then choosing $\widehat{\beta}_1 = \widehat{\beta}_2 = 0$ and $c = 1$ we see that there must exist $\delta \in \mathfrak{R}^m$ and an orthogonal matrix $\mathbf{U}$ such that for $\mathbf{Q} = \mathbf{U}\mathbf{C}\mathbf{U}^T$ we have $\mathfrak{E}_{\delta,1}(\mathbf{Q}) \supseteq \mathfrak{E}_{0,1}(\mathbf{D})$. Let $a \in \mathfrak{R}^m$, $d > 0$ and $a^T\mathbf{D}a \leq d$. Then the vectors $\frac{1}{\sqrt{d}}a$ and $-\frac{1}{\sqrt{d}}a$ are members of $\mathfrak{E}_{0,1}(\mathbf{D})$, hence they both belong to $\mathfrak{E}_{\delta,1}(\mathbf{Q})$. It follows that $\left(\frac{1}{\sqrt{d}}a - \delta\right)^T \mathbf{Q}\left(\frac{1}{\sqrt{d}}a - \delta\right) \leq 1$ as well as $\left(-\frac{1}{\sqrt{d}}a - \delta\right)^T \mathbf{Q}\left(-\frac{1}{\sqrt{d}}a - \delta\right) \leq 1$. Summing up these two inequalities we obtain $\frac{1}{d}a^T\mathbf{Q}a + \delta^T\mathbf{Q}\delta \leq 1$. As $\mathbf{Q} \in \mathscr{S}_+^m$ we have $\delta^T\mathbf{Q}\delta \geq 0$, thence $a^T\mathbf{Q}a \leq d$. This clearly implies that for any $a \in \mathfrak{R}^m$ we have $a^T\mathbf{Q}a \leq a^T\mathbf{D}a$ and consequently $\mathbf{Q} \leq \mathbf{D}$ in Loewner ordering. Therefore $\lambda(\mathbf{C}) = \lambda(\mathbf{Q}) \leq \lambda(\mathbf{D})$, where the inequality follows from 7.7.4 (c) in [3].

"(iii)⇐(iv)" Let (iv) hold, $\mathbf{D} = \mathbf{V}\mathrm{diag}(\lambda(\mathbf{D}))\mathbf{V}^T$ and $\mathbf{C} = \mathbf{R}\mathrm{diag}(\lambda(\mathbf{C}))\mathbf{R}^T$ for some orthogonal matrices $\mathbf{V}$, $\mathbf{R}$. Let $\widehat{\beta}_1, \widehat{\beta}_2 \in \mathfrak{R}^m$ and $c > 0$ be arbitrary. Choose the following rigid-motion transformation:

$$\rho(\cdot) = \mathbf{U}\times(\cdot) + r, \quad \text{where } \mathbf{U} = \mathbf{V}\mathbf{R}^T, r = \left(\widehat{\beta}_2 - \mathbf{U}\widehat{\beta}_1\right).$$

Evidently, $\lambda(\mathbf{C}) \leq \lambda(\mathbf{D})$ implies that for all $\widehat{\beta} \in \Re^m$: $\mathfrak{E}_{\widehat{\beta},c}(\mathrm{diag}(\lambda(\mathbf{C}))) \supseteq \mathfrak{E}_{\widehat{\beta},c}(\mathrm{diag}(\lambda(\mathbf{D})))$. Therefore

$$\begin{aligned}\rho(\mathfrak{E}_{\widehat{\beta}_1,c}(\mathbf{C})) &= \mathfrak{E}_{\mathbf{U}\widehat{\beta}_1+r,c}(\mathbf{U}\mathbf{C}\mathbf{U}^T) = \mathfrak{E}_{\widehat{\beta}_2,c}(\mathbf{V}\mathrm{diag}(\lambda(\mathbf{C}))\mathbf{V}^T)\\ &= \mathbf{V}\cdot\mathfrak{E}_{\mathbf{V}^T\widehat{\beta}_2,c}(\mathrm{diag}(\lambda(\mathbf{C}))) \supseteq \mathbf{V}\cdot\mathfrak{E}_{\mathbf{V}^T\widehat{\beta}_2,c}(\mathrm{diag}(\lambda(\mathbf{D}))) = \mathfrak{E}_{\widehat{\beta}_2,c}(\mathbf{D})\end{aligned}$$

■

7.3 Proof of Theorem 4

A well known theorem from convex analysis (see e.g. [11] or [10]) implies that $\mathbf{M} = \mathbf{M}(\xi)$ maximizes $\widetilde{\Phi}_{E_k}$ (i.e. also Φ_{E_k}) on $\mathscr{M}$ if and only if there exists $\mathbf{Y} \in \partial\widetilde{\Phi}_{E_k}(\mathbf{M})$, such that $\mathrm{tr}(\mathbf{N}\mathbf{Y}) \leq \mathrm{tr}(\mathbf{M}\mathbf{Y})$ for all $\mathbf{N} \in \mathscr{M}$, resp. iff $\sup_{\mathbf{N}\in\mathscr{M}} \mathrm{tr}(\mathbf{N}\mathbf{Y}) = \mathrm{tr}(\mathbf{M}\mathbf{Y})$. But

$$\sup_{\mathbf{N}\in\mathscr{M}} \mathrm{tr}(\mathbf{N}\mathbf{Y}) = \sup_{\zeta\in\Xi} \mathrm{tr}(\sum_{\zeta(x)>0} \zeta(x)f(x)f^T(x)\mathbf{Y}) =$$

$$\sup_{\zeta\in\Xi} \sum_{\zeta(x)>0} \zeta(x)f^T(x)\mathbf{Y}f(x) = \max_{x\in\mathfrak{X}} f^T(x)\mathbf{Y}f(x)$$

Also, for any choice $\mathbf{Y} = \mathbf{U}\mathrm{diag}(\gamma)\mathbf{U}^T$ where $\mathbf{U}$, γ are given in Proposition 3, we have $\mathrm{tr}(\mathbf{M}\mathbf{Y}) = \mathrm{tr}(\mathrm{diag}(\lambda(\mathbf{M}))\mathrm{diag}(\gamma)) = \Phi_{E_k}(\mathbf{M})$. This proves (i)⇔(ii).

Next, the implication (ii)⇒(iii) follows from Proposition 3, as for any $\mathbf{Y} \in \partial\widetilde{\Phi}_{E_k}(\mathbf{M})$ we have $\mathrm{tr}(\mathbf{Y}) = k$, and $\lambda(\mathbf{Y}) \in [0,1]^m$ i.e. $\mathbf{Y} \leq \mathbf{I}$.

We will prove (iii)⇒(ii) by simply showing that the matrix $\mathbf{Y}$ from (iii) is a subgradient of $\widetilde{\Phi}_{E_k}$ in $\mathbf{M} = \mathbf{M}(\xi)$. Let (iii) hold. Obviously

$$\mathrm{tr}(\mathbf{M}\mathbf{Y}) = \mathrm{tr}(\sum_{\xi(x)>0} \xi(x)f(x)f^T(x)\mathbf{Y}) =$$

$$\sum_{\xi(x)>0} \xi(x)f^T(x)\mathbf{Y}f(x) \leq \max_{x\in\mathfrak{X}} f^T(x)\mathbf{Y}f(x) = \Phi_{E_k}(\mathbf{M}).$$

For any $\mathbf{B} \in \mathscr{S}^m$ we thus have $\widetilde{\Phi}_{E_k}(\mathbf{M}) + \mathrm{tr}((\mathbf{B}-\mathbf{M})\mathbf{Y}) \geq \mathrm{tr}(\mathbf{B}\mathbf{Y})$. Moreover from Theorem 3.4 in [6] it follows that:

$$\widetilde{\Phi}_{E_k}(\mathbf{B}) = \min\left\{\mathrm{tr}(\mathbf{B}\mathbf{H}); \mathbf{H} \in \mathscr{S}^m_+, \mathbf{H} \leq \mathbf{I}, \mathrm{tr}(\mathbf{H}) = \mathrm{k}\right\}.$$

Hence $\mathrm{tr}(\mathbf{B}\mathbf{Y}) \geq \widetilde{\Phi}_{E_k}(\mathbf{B})$, which entails $\widetilde{\Phi}_{E_k}(\mathbf{M}) + \mathrm{tr}((\mathbf{B}-\mathbf{M})\mathbf{Y}) \geq \widetilde{\Phi}_{E_k}(\mathbf{B})$. By the definition of subgradient, this means that $\mathbf{Y} \in \partial\widetilde{\Phi}_{E_k}(\mathbf{M})$.

7.4 Proof of Theorem 5

Recall that an $m \times m$ matrix $\mathbf{S}$ is said to be doubly stochastic, if it has non-negative entries and each column and row is summing to one. An $m \times m$ matrix $\mathbf{P}$ is a permutation matrix, if each row and column contains exactly one element 1, and $m-1$ elements 0. Evidently, there are $m!$ such matrices.

Let $\mathfrak{R}_+^m$, and $\mathfrak{R}_\leq^m$ denote the set of all m -dimensional vectors with nonnegative components (the nonnegative orthant), resp. the set of all vectors with components in a nondecreasing order. Let $x = (x_1, \ldots, x_m)^T$, $y = (y_1, \ldots, y_m)^T \in \mathfrak{R}_\leq^m$. If $\sum_{i=1}^k x_i \geq \sum_{i=1}^k y_i$ for $k = 1, \ldots, m-1$, and $\sum_{i=1}^m x_i = \sum_{i=1}^m y_i$, then we will say that x majorizes y, and denote this fact by $x \succeq y$. (We use this notion as defined e.g. in [3] p.192, cf. also with [8], p.144-5.)

In the proof, we will use the following theorems:

1. (Birkhoff theorem, see e.g. [3] p. 527) An $m \times m$ matrix $\mathbf{S}$ is doubly stochastic if and only if $\mathbf{S}$ is a convex combination of permutation matrices.
2. (Hardy-Littlewood-Pólya theorem; see e.g. [3] p. 197) If $x, y \in \mathfrak{R}_+^m \cap \mathfrak{R}_\leq^m$, then $x \succeq y$ if and only if $x = \mathbf{S}y$ for some doubly stochastic matrix $\mathbf{S}$.

In Theorem 5, the inequality $\inf_{\Phi \in \mathbb{O}} \text{eff}_\Phi(\xi|\zeta) \leq \min_{k=1,\ldots,m} \text{eff}_{\Phi_{E_k}}(\xi|\zeta)$ is clear because $\Phi_{E_k} \in \mathbb{O}$. We will prove the converse inequality.

If $\mathbf{M}(\zeta)$ is the zero matrix, then $\text{eff}_\Phi(\xi|\zeta) = \Phi(\mathbf{M}(\xi))/\Phi(\mathbf{M}(\zeta)) = \infty$ for any $\Phi \in \mathbb{O}$ and the inequality is trivial. Let $\mathbf{M}(\zeta) \neq \mathbf{0}$. Denote

$$\delta = \min_{k=1,\ldots,m} \text{eff}_{\Phi_{E_k}}(\xi|\zeta) = \min_{k=1,\ldots,m} \left(\sum_{i=1}^k l_i \right) \left(\sum_{i=1}^k g_i \right)^{-1},$$

where $l = \lambda(\mathbf{M}(\xi))$, $g = \lambda(\mathbf{M}(\zeta))$; $l, g \in \mathfrak{R}_+^m \cap \mathfrak{R}_\leq^m$

Clearly, δ is a positive number less than ∞, because it is a finite minimum of positive values at least one of which is not infinite. Let

$$g_m^* = \delta^{-1} \left(\sum_{i=1}^m l_i \right) - \sum_{i=1}^{m-1} g_i, \text{ and } g^* = \left(g_1, \ldots, g_{m-1}, g_m^* \right)^T$$

As $\delta \leq \left(\sum_{i=1}^m l_i \right) \left(\sum_{i=1}^m g_i \right)^{-1}$, we have $g_m^* \geq g_m$, hence $g^* \geq g$ and $g^* \in \mathfrak{R}_+^m \cap \mathfrak{R}_\leq^m$. Moreover, $\sum_{i=1}^k l_i \geq \sum_{i=1}^k \delta g_i$ for $k = 1, \ldots, m-1$, and $\sum_{i=1}^m l_i = \left(\sum_{i=1}^{m-1} \delta g_i \right) + \delta g_m^*$, which implies that l majorizes δg^*. We thus have

$$l \succeq \delta g^* \geq \delta g$$

Hence, the Hardy-Littlewood-Pólya theorem entails that there exists a doubly stochastic matrix $\mathbf{S}$, such that $l = \mathbf{S}(\delta g^*)$. In the same time, as the elements of $\mathbf{S}$, as well as the coordinates of g and g^* are nonnegative, the inequality $g^* \geq g$ implies $\mathbf{S}(\delta g^*) \geq \mathbf{S}(\delta g)$, therefore $l \geq \mathbf{S}(\delta g)$. Moreover, from the Birkhoff theorem we know that $\mathbf{S} = \sum_{j=1}^{m!} \alpha_j \mathbf{P}_j$, where $\mathbf{P}_j$ is the j -th permutation matrix, $\alpha_j \in [0, 1]$ for $j = 1, \ldots, m!$, and $\sum_{j=1}^{m!} \alpha_j = 1$. Consequently

$$l \geq \sum_{j=1}^{m!} \alpha_j \mathbf{P}_j(\delta g)$$

Let $\Phi \in \mathbb{O}$. Notice, that for any permutation matrix $\mathbf{P}_j$ it holds that $\Phi\big(\text{diag}\mathbf{P}_j(\delta g)\big) = \Phi(\text{diag}(\delta g))$, since Φ is orthogonally invariant. Subsequently using properties (O), (I), (C), (O), (O) and (H), we finally obtain:

$$\Phi(\mathbf{M}(\xi)) = \Phi(\mathrm{diag} l) \geq \Phi\left(\mathrm{diag}\left(\sum_{j=1}^{m!} \alpha_j \mathbf{P}_j(\delta g)\right)\right)$$

$$\geq \sum_{j=1}^{m!} \alpha_j \big(\Phi\big(\mathrm{diag}\mathbf{P}_j(\delta g)\big)\big) = \Phi(\mathrm{diag}(\delta g)) = \Phi(\delta \mathbf{M}(\zeta)) = \delta \Phi(\mathbf{M}(\zeta))$$

Therefore $\mathrm{eff}_\Phi(\xi|\zeta) = \Phi(\mathbf{M}(\xi))/\Phi(\mathbf{M}(\zeta)) \geq \delta$ which concludes the proof.

7.5 Proof of Proposition 6

Similarly as in the proof of Proposition VI.9 in [7] or Claim 9.16. in [8], we can show that $\mathbf{M}^* = \mathbf{M}(\xi^*) = \mathrm{diag}(1/2, \ldots, 1/2, 1)$. By Theorem 5, to prove the Φ-optimality of ξ^* for all $\Phi \in \mathbb{O}$ we only need to check that ξ^* is Φ_{E_k}-optimal for all $k = 1, \ldots, m$, where $m = 2d + 1$ is the number of parameters. The Φ_m-optimality is simple to prove (see e.g. [8], Claim 9.16.), so we only need to analyze the case of a fixed $k < m$. Proposition 3 implies that $\mathrm{diag}(\gamma) \in \partial\Phi_{E_k}(\mathbf{M}^*)$ for any choice of γ which satisfies $\gamma_i \in [0, 1]$ for $i = 1, \ldots, m-1$, $\gamma_m = 0$, and $\sum_{i=1}^m \gamma_i = k$. If k is even, let $\gamma_1 = \ldots = \gamma_k = 1$ and $\gamma_{k+1} = \ldots = \gamma_m = 0$. If k is odd, then $k \leq m-2$, hence we can choose $\gamma_1 = \ldots = \gamma_{k-1} = 1$, $\gamma_k = \gamma_{k+1} = 1/2$, $\gamma_{k+2} = \ldots = \gamma_m = 0$. In both cases we have for any $x \in \mathfrak{X}$:

$$f^T(x)\mathrm{diag}(\gamma)f(x) = \sum_{j=1}^{\frac{m-1}{2}} \big(\gamma_{2j-1}\cos^2(jx) + \gamma_{2j}\sin^2(jx)\big) = k/2 = \Phi_{E_k}(\mathbf{M}^*).$$

By Theorem 4, this equality proves Φ_{E_k}-optimality of ξ^*.

7.6 Proof of Proposition 10

It is simple to verify that the eigenvalues of $\mathbf{M} = \mathbf{M}(\xi)$ (cf. [8] p.333) are

$$\lambda_1(\mathbf{M}) = \frac{29}{251} + \frac{2}{251}\sqrt{22} \doteq 0.15291, \lambda_2(\mathbf{M}) = \frac{92}{251} + \frac{15}{251}\sqrt{22} \doteq 0.64684,$$

$$\lambda_3(\mathbf{M}) = \frac{314}{251} + \frac{13}{251}\sqrt{22} \doteq 1.49393$$

We already know (see Table 1) that the optimal values for Φ_{E_k}, $k = 1, 2, 3$ are $v_1 = 1/5$, $v_2 = 1$, and $v_3 = 3$, therefore

$$v_1^{-1}\Phi_{E_1}(\mathbf{M}) = 5\lambda_1(\mathbf{M}) = \frac{145}{251} + \frac{10}{251}\sqrt{22} \doteq 0.76456$$

$$v_2^{-1}\Phi_{E_2}(\mathbf{M}) = \lambda_1(\mathbf{M}) + \lambda_2(\mathbf{M}) = \frac{121}{251} + \frac{17}{251}\sqrt{22} \doteq 0.79975$$

$$v_3^{-1}\Phi_{E_3}(\mathbf{M}) = \frac{1}{3}\{\lambda_1(\mathbf{M}) + \lambda_2(\mathbf{M}) + \lambda_3(\mathbf{M})\} = \frac{145}{251} + \frac{10}{251}\sqrt{22} \doteq 0.76456$$

Thence the $\mathbb{O}$-minimal efficiency of ξ is (cmp. also with Table 2)

$$\Phi_v(\mathbf{M}) = \min_{k=1,2,3} v_k^{-1}\widetilde{\Phi}_{E_k}(\mathbf{M}) = \frac{145}{251} + \frac{10}{251}\sqrt{22} \doteq 0.76456$$

The eigenvalues of $\mathbf{M}$ are mutually distinct, hence by the discussion following Proposition 3 we know that the functions $\widetilde{\Phi}_{E_k}$ are differentiable in $\mathbf{M}$, i.e. there exist unique gradients of $\widetilde{\Phi}_{E_k}$ in $\mathbf{M}$ for all k. Moreover, as the active set used in Proposition 8 is $I = \{1, 3\}$, we see that the subgradients of $\widetilde{\Phi}_v$ in $\mathbf{M}$ are of the form

$$\mathbf{Y} = \alpha v_1^{-1}\nabla\widetilde{\Phi}_{E_1}(\mathbf{M}) + (1-\alpha)v_3^{-1}\nabla\widetilde{\Phi}_{E_3}(\mathbf{M}), \alpha \in [0, 1].$$

In the sequel, we shall use the subgradient $\mathbf{Y}$ which corresponds to $\alpha = \frac{11}{251} + \frac{285}{5522}\sqrt{22} \doteq 0.28591$. From Proposition 3, one can calculate that

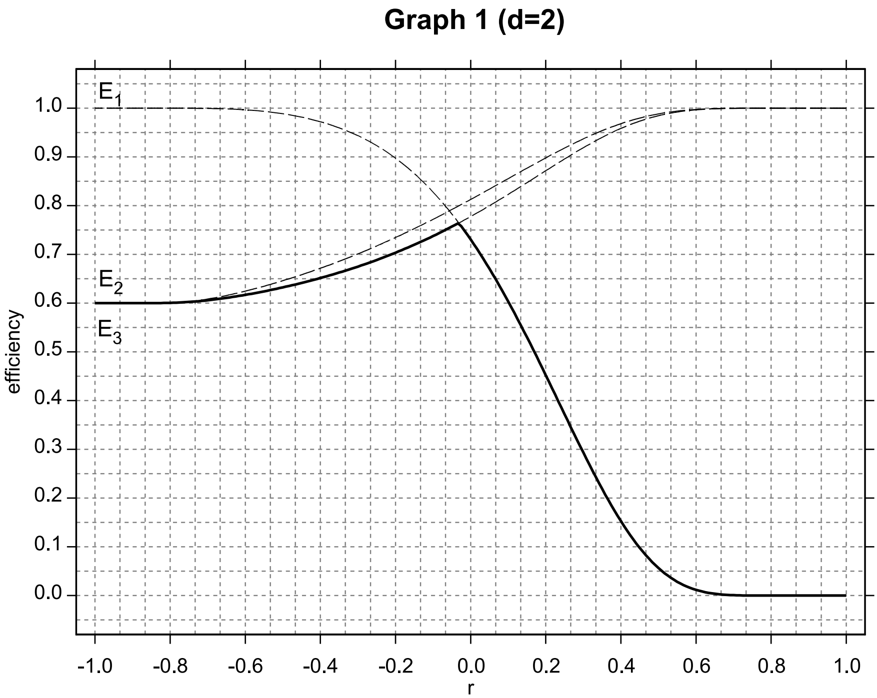

Fig. 1. The E_1, E_2, E_3 and the $\mathbb{O}$-minimal efficiency of $\phi_{p(r)}$-optimal designs for the quadratic polynomial regression model

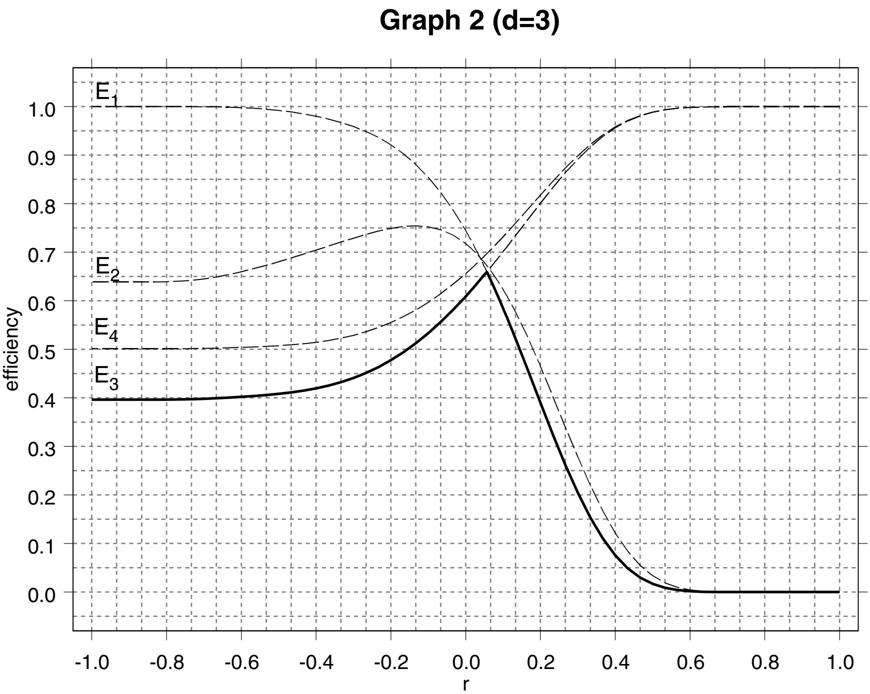

Fig. 2. The E_1, E_2, E_3, E_4 and the $\mathbb{O}$-minimal efficiency of $\phi_{p(r)}$-optimal designs for the cubic polynomial regression model

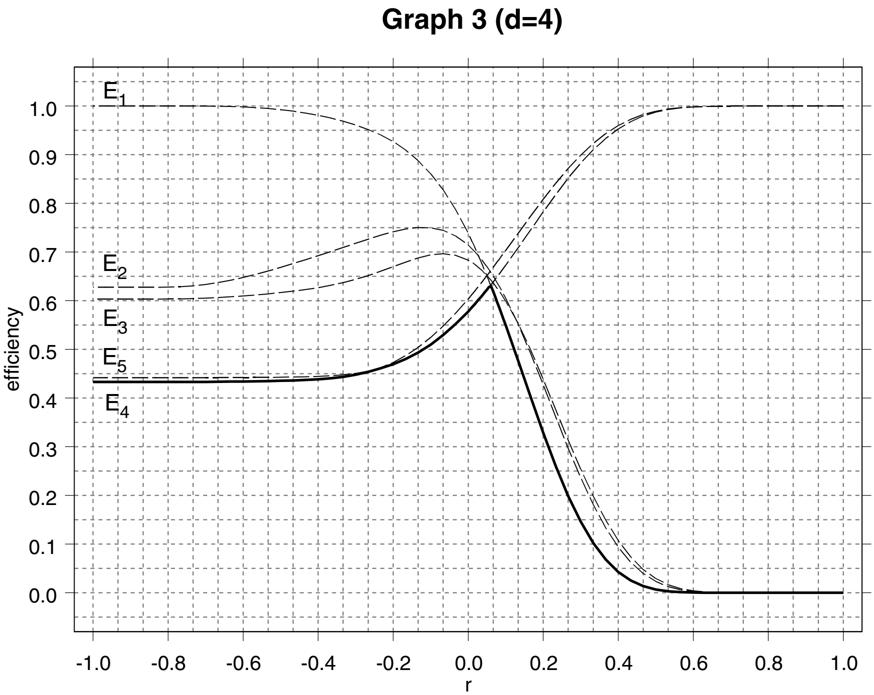

Fig. 3. The E_1, E_2, E_3, E_4, E_5 and the $\mathbb{O}$-minimal efficiency of the $\phi_{p(r)}$-optimal designs for the biquadratic polynomial regression model

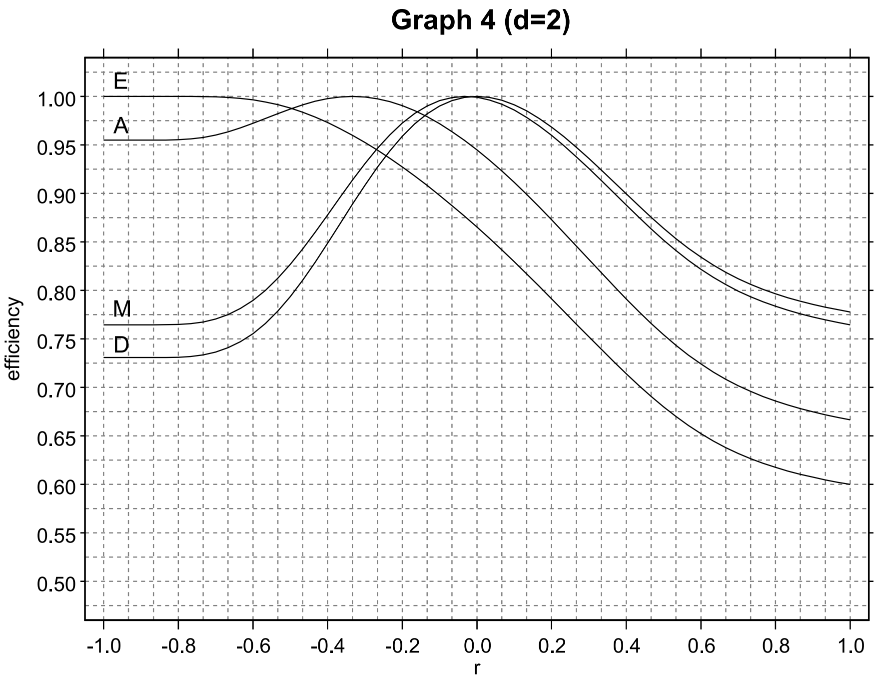

Fig. 4. The $\phi_{p(r)}$-efficiency of the D, A, E-optimal designs, and the $\mathbb{O}$-maximin efficient design (denoted by M) for the quadratic regression model

$$\nabla\widetilde{\Phi}_{E_1}(\mathbf{M}) = \begin{pmatrix} a & 0 & b \\ 0 & 0 & 0 \\ b & 0 & c \end{pmatrix}, \text{ where } a = \frac{59 + 12\sqrt{22}}{313},$$

$$b = \frac{-90 - 13\sqrt{22}}{313}, c = \frac{254 - 12\sqrt{22}}{313}.$$

Moreover, $\widetilde{\Phi}_{E_3}$ is simply the trace, therefore $\nabla\widetilde{\Phi}_{E_3}(\mathbf{M}) = \mathbf{I}$. Using algebraic simplifications and elementary calculus we finally obtain ■

$$f^T(x)\mathbf{Y}f(x) = \frac{220 + 145\sqrt{22}}{5522} \times \left(7x^4 - 7x^2 + \sqrt{22}\right)$$

$$\max_{x\in[0,1]} f^T(x)\mathbf{Y}f(x) = \frac{145}{251} + \frac{10\sqrt{22}}{251} = \Phi_v(\mathbf{M})$$

By Theorem 9, this proves that ξ is the $\mathbb{O}$-maximin efficient design.

Acknowledgments: The author would like to thank prof. Pázman as well as an anonymous referee for useful and inspiring comments on earlier versions of this article.

References

[1] Bondar JV (1983) Universal optimality of experimental designs: definitions and a criterion. The Canadian Journal of Statistics, 11(4):325–331
[2] Borwein JM, Lewis AS (2000) Convex analysis and nonlinear optimization, Theory and examples. CMS Books in Mathematics, Springer-Verlag, New York
[3] Horn RA, Johnson CR (1985) Matrix analysis, Cambridge University Press, Cambridge
[4] Lewis AS (1996) Convex analysis on the Hermitian matrices. SIAM Journal of Optimization 6:164–177
[5] Müller Ch, Pázman A (1998) Applications of necessary and sufficient conditions for maximin efficient designs. Metrika, 21(1):402–415
[6] Overton ML, Womersley RS (1993) Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. Mathematical Programming, Series B, 62:321–357
[7] Pázman A (1986) Foundations of optimum experimental design. D. Reidel, Dordrecht
[8] Pukelsheim F (1993) Optimal design of experiments. John Wiley &Sons, New York
[9] Pukelsheim F, Studden WJ (1993) E-optimal designs for polynomial regression. Annals of Statistics, 21(1)
[10] Pukelsheim F, Titterington DM (1983) General differential and Lagrangian theory for optimal experimental design. Annals of Statistics, 11(4):1060–1068
[11] Rockafellar RT (1970) Convex analysis, Princeton University Press, Princeton
[12] Rodríguez-Díaz JM, López-Fidalgo J (2000) Optimal characteristic designs for polynomial models. In: Atkinson, Zhigljavsky, Bogacka (eds.) Optimum Design 2000, Kluwer, Dordrecht, pp. 123–130
[13] Seeger A (1997) Convex analysis of spectrally defined matrix functions. Siam. J. Optim., 7(3):679–696
[14] Wong WK (1992) A unified approach to the construction of minimax designs, Biometrika, 79(3):611–619