

Stochastic Design Criteria in Linear Models

Alexander Zaigraev
N. Copernicus University, Toruń, Poland

Abstract: Within the framework of classical linear regression model stochastic optimal design criteria are considered. As examples a line fit model and a k -way linear fit model are taken. If the optimal design does not exist, an approach consisting in choosing the efficient design is suggested.

Keywords: Linear Regression Model, Stochastic Criteria, Efficient Design.

1 Introduction

A literature on optimal design criteria is very extensive. For references see Pukelsheim (1993) and Liski et al. (2002), for example. Among the criteria there are the so-called classical criteria like A-, D- or E-optimality as well as the relatively new ones like *stochastic optimality* criteria. The stochastic criteria have gained a momentum only recently, though the most known criterion of such type was put forward more than thirty years ago (see Sinha, 1970).

In the paper we consider the classical linear regression model

$$Y \sim N_n(X\beta, \sigma^2 I_n) \quad (1)$$

where the $n \times 1$ response vector $Y = (Y_1, Y_2, \dots, Y_n)^T$ follows a multivariate normal distribution, $X = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T$ is the $n \times k$ design matrix of the full rank $k \leq n$, $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ is the $k \times 1$ parameter vector, $E(Y) = X\beta$ is the expectation vector of Y and $\sigma^2 I_n$ is the covariance matrix of Y , where I_n is the $n \times n$ identity matrix and $\sigma > 0$ is unknown.

Let $\hat{\beta}$ be the least squares estimator of β being at the same time the best linear unbiased estimator. As it is well-known,

$$\hat{\beta} = (X^T X)^{-1} X^T Y \sim N_k(\beta, \sigma^2 (X^T X)^{-1}).$$

In the sequel, we deal with the so-called *continuous* or *approximate* design. Each approximate design ξ is a discrete probability measure taking values $p_i > 0$ at vectors $x^{(i)}$ (support vectors), $i = 1, 2, \dots, m$, that is

$$\xi = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}; p_1, p_2, \dots, p_m\}, \quad \sum_{i=1}^m p_i = 1$$

where $x^{(i)} \in \mathcal{X}$, $i = 1, 2, \dots, m$. The set \mathcal{X} is called the *experimental domain*. We denote by Ξ the set of all approximate designs on \mathcal{X} .

With a design ξ we associate its $k \times k$ moment matrix $M(\xi) = \sum_{i=1}^m p_i x^{(i)} x^{(i)T}$. If $p_i = n_i/n$, $i = 1, 2, \dots, m$, $m \leq n$, where n_i are integers and $\sum_{i=1}^m n_i = n$, then the covariance matrix of $\hat{\beta}$ is $(\sigma^2/n)M^{-1}(\xi)$. Throughout the paper, we write $\hat{\beta} = \hat{\beta}(\xi)$ or $\hat{\beta} = \hat{\beta}(M)$ to emphasize the dependence of $\hat{\beta}$ from the design ξ or from the moment matrix M , respectively.

In the paper we refer to a line fit model when we have $n \geq 2$ uncorrelated responses

$$Y_{ij} = \beta_1 + \beta_2 z_i + e_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n_i \quad (2)$$

with expectations $E(Y_{ij}) = \beta_1 + \beta_2 z_i$ and variances $V(Y_{ij}) = \sigma^2$. In this case an approximate design ξ specifies distinct values z_1, \dots, z_m chosen from a given experimental domain (usually an interval $[a, b]$) and assigns to them weights $p_1 > 0, \dots, p_m > 0$, respectively. Of course, these weights satisfy $\sum_{i=1}^m p_i = 1$. Here,

$$M(\xi) = \begin{pmatrix} 1 & \sum_{i=1}^m p_i z_i \\ \sum_{i=1}^m p_i z_i & \sum_{i=1}^m p_i z_i^2 \end{pmatrix}.$$

In the paper we also consider a k -way line fit model with or without an intercept. In the first case we have $n \geq k + 1$ uncorrelated responses

$$Y_{ij} = \beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_k x_k^{(i)} + e_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n_i \quad (3)$$

with unknown parameters $(\beta_0, \beta_1, \dots, \beta_k)$ and experimental conditions $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})^T$, $i = 1, 2, \dots, m$. In the second case we have $n \geq k$ uncorrelated responses

$$Y_{ij} = \beta_1 x_1^{(i)} + \dots + \beta_k x_k^{(i)} + e_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n_i \quad (4)$$

with unknown parameters $(\beta_1, \beta_2, \dots, \beta_k)$ and experimental conditions $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})^T$, $i = 1, 2, \dots, m$. In both cases the assumptions on e_{ij} are the same as in (2).

The paper is organized as follows. Classical and stochastic optimality criteria are discussed in Section 2. There the problem of establishing the corresponding optimal designs in the above-mentioned models is also considered. Section 3 is devoted to the concept of efficient designs in the situations where the optimal designs do not exist. The proofs of theorems can be found in Appendix.

2 Optimality Criteria and Optimal Designs

2.1 Optimality Criteria

An optimality criterion F is a function from the closed cone of nonnegative definite matrices into \mathbb{R}_+^1 . We say that the design ξ^* is F -optimal if and only if

$$\xi^* \in \text{Arg} \min_{\xi \in \Xi} F(M(\xi)).$$

Recall the definitions of the classical optimality criteria.

- The criterion $F(M) = \det(M^{-1})$ is called D-criterion.
- The criterion $F(M) = \text{tr}(M^{-1})$ is called A-criterion.
- The criterion $F(M) = \lambda_{\max}(M^{-1})$ is called E-criterion.

Here, $\lambda_{\max}(M^{-1})$ denotes the maximal eigenvalue of the matrix M^{-1} . It is well-known that $\det(M^{-1})$, called the generalized variance, determines the volume of the ellipsoid of concentration for $\widehat{\beta}$. Its minimization leads to the ellipsoid of concentration with the smallest volume. On the other hand, minimization of $\text{tr}(M^{-1})$, called the average variance, is the minimization of the sum of variances of $\widehat{\beta}_i$, $i = 1, 2, \dots, k$. At last, minimization of $\lambda_{\max}(M^{-1})$ leads to the ellipsoid of concentration having the smallest length of the maximal axis. In all the three cases we set $F(M) = \infty$ if the matrix M is degenerate. This means that the designs with degenerate moment matrices can be excluded from the consideration. Therefore, we can assume that the support size of designs satisfies the inequality $m \geq k$ (or $m \geq k + 1$ when we deal with a k -way line fit model with an intercept).

Saying 'stochastic optimality criteria' we mean functions depending on the moment matrices through a probability. A typical example is the family of criteria

$$F_A(M) = -\ln P([\widehat{\beta}(M) - \beta] \in A), \quad A \in \mathcal{A} \quad (5)$$

where \mathcal{A} is a given class of bounded subsets of \mathbb{R}^k containing the origin. The class \mathcal{A} can be interpreted as a collection of sets determining a system of neighbourhoods of the origin. Here, we would like to choose a design which guarantees the maximal probability for the estimator $\widehat{\beta}$ of being 'close' to β . Of course, the terminology 'stochastic' is rather relative. It is due to Sinha (1970) who introduced the concept of the *distance stochastic* (DS) criterion in certain treatment design settings. Liski et al. (1999) studied the properties of this criterion under the classical linear regression model (1). In case of a degenerate matrix M , it is natural to set $P([\widehat{\beta}(M) - \beta] \in A) = 0$.

We obtain the DS-criterion taking \mathcal{A} to be the class of all k -dimensional balls centered at the origin:

$$\mathcal{A} = \{\varepsilon A, \varepsilon > 0\}, \quad A = \{x \in \mathbb{R}^k : \|x\| \leq 1\}$$

where $\|\cdot\|$ stands for the usual Euclidean norm in \mathbb{R}^k . Here, we assume the system of neighbourhoods to be the balls, one of the most natural choice. Observe that the DS-criterion, in fact, is a family of DS(ε)-criteria indexed by $\varepsilon > 0$. We say that the design is optimal with respect to the family of criteria if and only if it is optimal with respect to the each criterion from this family. One can remember the families of criteria popular in the literature: the ϕ_p -criterion (see Pukelsheim, 1993, Chapter 6) or the characteristic criterion (see Rodrigues-Diaz and López-Fidalgo, 2003). It should be emphasized, however, that the DS(ε)-optimal design, i.e. the design which is optimal with respect to the DS(ε)-criterion for given $\varepsilon > 0$, is not of great interest itself since usually it depends on unknown σ .

Liski et al. (1999, Theorem 5.1) studied the behavior of the DS(ε)-criterion when ε approaches 0 and ∞ . These limiting cases have an interesting relationship with the classical D- and E-optimality criteria. It turns out that the DS(ε)-criterion is equivalent to the D-criterion as $\varepsilon \rightarrow 0$ and to the E-criterion as $\varepsilon \rightarrow \infty$. Moreover, minimization of the probability $P(\|\widehat{\beta}(\xi) - \beta\| > \varepsilon)$ simultaneously for all $\varepsilon > 0$ is equivalent to minimization of $Eg(\|\widehat{\beta}(\xi) - \beta\|)$ for all increasing functions g such that the expectation exists (see Marshall and Olkin, 1979, Chapter 17.A). In particular, one can take $g(x) = x$. Therefore, if a design is DS-optimal then it is also A-optimal.

Zaigraev (2002) suggested a natural extension of the DS-criterion called the *shape stochastic* (SS_ρ) criterion. Here,

$$\mathcal{A} = \{\varepsilon A_\rho, \varepsilon > 0\}, \quad A_\rho = \{x \in \mathbb{R}^k : \|x\| \leq \rho(x/\|x\|)\} \cup \{0\}$$

and ρ is a positive continuous function defined on the unit sphere \mathbb{S}^{k-1} in \mathbb{R}^k . In particular, the SS_ρ -criterion is simply the DS-criterion if $\rho(\cdot) \equiv 1$. Again one can note that the $SS_\rho(\varepsilon)$ -optimal design, in general, depends on unknown σ .

In the sequel, we confine ourselves to the case where A_ρ is a convex and symmetric (with respect to the origin) set. This restriction has the following sense. We say that a design ξ_1 *dominates* ξ_2 in the Loewner ordering sense if $M_1 - M_2$ is a nonnegative definite matrix (in such a case we write $M_1 - M_2 \geq 0$ or $M_1 \geq M_2$), where M_1 and M_2 are the moment matrices of the designs ξ_1 and ξ_2 , respectively. Thus the Loewner partial ordering among moment matrices induces a partial ordering among associated designs. Observe that D-, A-, E- and DS-criterion are *antitonic* relative to Loewner ordering, that is for any two moment matrices M_1 and M_2 the inequality $M_1 \geq M_2$ implies $F(M_1) \leq F(M_2)$. Such a property is a desirable if we deal with an optimality criterion. As it follows from Theorem 2 of Liski and Zaigraev (2001), the SS_ρ -criterion is antitonic relative to Loewner ordering if and only if the set A_ρ is convex and symmetric with respect to the origin.

Zaigraev (2002, Theorems 1 and 2) established the limit behavior of the SS_ρ -criterion when ε approaches 0 and ∞ . It turns out that the $SS_\rho(\varepsilon)$ -criterion is equivalent to the D-criterion as $\varepsilon \rightarrow 0$ and, under the mild regularity conditions on ρ , to the minimax criterion

$$F(M) = \max_{x \in \partial A_\rho} (x^T M x)^{-1}$$

as $\varepsilon \rightarrow \infty$. Here, ∂A_ρ is the boundary of the set A_ρ .

2.2 Optimal Designs for a Line Fit Model

Consider model (2) with $z_i \in [a, b]$, $i = 1, 2, \dots, m$. Searching for optimal designs, in this situation it is enough to consider only two-point designs of the form $\xi_p = \{a, b; p, 1-p\}$, $0 < p < 1$ (see e.g. de la Garza, 1954, Liski and Zaigraev, 2001, Lemma 1). That is, the support consists of the extreme points of the experimental domain. The moment matrix of such a design has the form

$$M(\xi_p) = \begin{pmatrix} 1 & ap + b(1-p) \\ ap + b(1-p) & a^2p + b^2(1-p) \end{pmatrix}.$$

It is not difficult to calculate the optimal designs with respect to the classical optimality criteria. They are given as follows:

- D-optimal design $\xi^D = \xi_{0.5}$;
- A-optimal design $\xi^A = \xi_{p^*}$, $p^* = \sqrt{1+b^2}/(\sqrt{1+b^2} + \sqrt{1+a^2})$;
- E-optimal design $\xi^E = \xi_{\tilde{p}}$,

$$\tilde{p} = \begin{cases} b/(b-a), & \text{if } b+a=0 \text{ or } ab \leq -1, \\ (b(b+a)+2)/((b+a)^2+4), & \text{otherwise.} \end{cases}$$

Observe that if $a = -b$ (symmetric experimental domain), then the design $\{-b, b; 0.5, 0.5\}$ is D-, A- and E-optimal. As to DS-optimality, the following general result holds.

Theorem 1. *The design $\{-b, b; 0.5, 0.5\}$ is optimal for model (2) on the symmetric experimental domain $[-b, b]$ with respect to any stochastic criterion of the form (5) with \mathcal{A} to be a class of convex and symmetric (with respect to the axes) sets in \mathbb{R}^2 .*

Theorem 1 is a direct extension of Lemma 2 of Liski and Zaigraev (2001). Its proof, in fact, repeats the proof of that lemma modifying it slightly accordingly with the fact that in the situation considered we have

$$M(\xi_p) = \begin{pmatrix} 1 & b(1-2p) \\ b(1-2p) & b^2 \end{pmatrix} = DM(\xi'_p)D$$

where $D = \text{diag}(1, b)$ while

$$M(\xi'_p) = \begin{pmatrix} 1 & 1-2p \\ 1-2p & 1 \end{pmatrix}$$

is the moment matrix of the design $\xi'_p = \{-1, 1; p, 1-p\}$ for model (2) on the experimental domain $[-1, 1]$.

Observe that if $a \neq -b$ (asymmetric experimental domain), then D-, A- and E-optimal designs are different and, therefore, DS-optimal design does not exist. However, it is of interest to note that sometimes in such a situation SS_ρ -optimal designs exist (see Zaigraev, 2002, Section 3).

Dealing with the SS_ρ -criterion in model (2), we confine ourselves to the following classes of sets centered at the origin:

- squares with $A_\rho = \{(x_1, x_2)^T \in \mathbb{R}^2 : |x_1| \leq 1, |x_2| \leq 1\}$;
- rectangulars with $A_\rho = \{(x_1, x_2)^T \in \mathbb{R}^2 : 2|x_1| \leq 1, |x_2| \leq 1\}$;
- ellipses with $A_\rho = \{(x_1, x_2)^T \in \mathbb{R}^2 : x_1^2 + 2x_1x_2 + x_2^2 \leq 1\}$.

The choice between those cases can be made basing on the problems to be solved or our personal preferences. In the first two cases, by Theorem 1, the SS_ρ -optimal design on the experimental domain $[-b, b]$ exists; it is $\{-b, b; 0.5, 0.5\}$. In the last case, however, the SS_ρ -optimal design does not exist due to lack of symmetry with respect to the axes.

Now, consider the case of an asymmetric experimental domain. For definiteness, we take $[a, b] = [0, 1]$. In accordance with the above-mentioned calculations, $\xi^D = \{0, 1; 0.5, 0.5\}$; $\xi^A = \{0, 1; 2-\sqrt{2}, \sqrt{2}-1\}$; $\xi^E = \{0, 1; 0.6, 0.4\}$.

The DS-optimal design does not exist; for given $\varepsilon > 0$ the $\text{DS}(\varepsilon)$ -optimal design depends on ε . Denote by C the class of all $\text{DS}(\varepsilon)$ -optimal designs and add to C two limiting designs when $\varepsilon \rightarrow 0$ and $\varepsilon \rightarrow \infty$. The same notation concerns the SS_ρ -criterion as well.

As we have mentioned earlier, for model (2) on the experimental domain $[0, 1]$ it is enough to consider only two-point designs of the form $\{0, 1; p, 1-p\}$, $0 < p < 1$. Below we give numerical results of calculating C in four cases: for the DS-criterion and for the SS_ρ -criterion (squares, rectangulars and ellipses). For given $\varepsilon > 0$ the weight p

for the corresponding optimal design depends on ε . However, it is more comfortable to use $\delta = \sqrt{n}\varepsilon/\sigma$ instead of ε to express this dependence. Observe that limiting optimal designs when $\varepsilon \rightarrow \infty$ coincide with those calculated theoretically (see Zaigraev, 2002, Section 3).

Table 1: DS-criterion, $C = \{0, 1; p, 1 - p\}, 0.5 \leq p \leq 0.6\}$

δ	0.05	0.10	0.25	0.50	0.75	1.00	1.25	1.50
p	0.5001	0.5004	0.5019	0.5075	0.5159	0.5261	0.5371	0.5480
δ	1.75	2.00	3.00	4.00	5.00	6.00	7.00	8.00
p	0.5572	0.5665	0.5871	0.5938	0.5963	0.5975	0.5683	0.5986

Table 2: SS_ρ -criterion, squares, $C = \{0, 1; p, 1 - p\}, 0.5 \leq p \leq 0.54(54)\}$

δ	0.05	0.10	0.25	0.50	0.75	1.00	1.25	1.50
p	0.5001	0.5004	0.5025	0.5094	0.5189	0.5286	0.5368	0.5424
δ	1.75	2.00	3.00	4.00	5.00	6.00	7.00	8.00
p	0.5437	0.5451	0.5313	0.5152	0.5056	0.5015	0.5012	0.5009

Table 3: SS_ρ -criterion, rectangles, $C = \{0, 1; p, 1 - p\}, 0.5 \leq p \leq 0.75\}$

δ	0.05	0.10	0.25	0.50	0.75	1.00	1.25	1.50
p	0.5001	0.5004	0.5026	0.5100	0.5216	0.5361	0.5524	0.5693
δ	1.75	2.00	3.00	4.00	5.00	6.00	8.00	10.00
p	0.5860	0.6018	0.6518	0.6821	0.7005	0.7123	0.7262	0.7338

Table 4: SS_ρ -criterion, ellipses, $C = \{0, 1; p, 1 - p\}, 0.4 \leq p \leq 0.5\}$

δ	0.05	0.10	0.25	0.50	0.75	1.00	1.25	1.50
p	0.4999	0.4997	0.4981	0.4925	0.4841	0.4739	0.4629	0.4520
δ	1.75	2.00	3.00	4.00	5.00	6.00	7.00	8.00
p	0.4427	0.4335	0.4129	0.4062	0.4037	0.4025	0.4018	0.4013

2.3 Optimal Designs for a k -way Line Fit Model

Consider model (3), that is a k -way line fit model with an intercept. Here, $x^{(i)} = (1, z^{(i)\top})^\top$, $z^{(i)} \in \mathbb{R}^k$, $i = 1, 2, \dots, m$. As an experimental domain, let us take $\mathcal{Z}_r = \{z \in \mathbb{R}^k : \|z\| \leq r\}$ (see Pukelsheim, 1993, Section 15.5, Liski et al., 1999, Liski and Zaigraev, 2001). Recall that for model (3) the smallest possible support size of a feasible design is $m = k + 1$.

As it is shown in Theorem 4.2 of Liski et al. (1999), if $r = \sqrt{k}$, then the DS-optimal design with the support size $m = k + 1$ on $\mathcal{Z}_{\sqrt{k}}$ exists. This is a so-called *regular simplex design* (see also Pukelsheim, 1993, Section 15.12). Such a design has the weights $1/(k + 1)$ and the support vectors $z^{(1)}, \dots, z^{(k+1)}$, which belong to the boundary of the experimental domain and form a regular simplex, that is $\|z^{(1)}\| = \sqrt{k}, \dots, \|z^{(k+1)}\| = \sqrt{k}$, and $z^{(i)\top} z^{(j)} = -1$ for all $i \neq j \leq k + 1$. Observe that a regular simplex design has the identity moment matrix and an orthogonal transformation of a regular simplex design is again a regular simplex design.

Now, we extend this result to the case of arbitrary $r > 0$.

Theorem 2. *A design ξ with the support size $m = k + 1$ for model (3) on the experimental domain \mathcal{Z}_r is DS-optimal if and only if it is a regular simplex design, that is the design having the weights $1/(k + 1)$ and the support vectors $z^{(1)}, \dots, z^{(k+1)}$ satisfying $\|z^{(1)}\| = r, \dots, \|z^{(k+1)}\| = r$, and $z^{(i)\top} z^{(j)} = -r^2/k$ for all $i \neq j \leq k + 1$. The moment matrix of such a design has the form $M = \text{diag}(1, r^2/k, \dots, r^2/k)$.*

It should be emphasized that the result of Theorem 2 extends also the corresponding result for model (2) on the experimental domain $[-b, b]$ (cf. Theorem 1).

Now, we consider model (4), that is a k -way line fit model without an intercept. As experimental domains, we take the sets $\mathcal{X}_1 = \{x \in \mathbb{R}^k : \|x\| \leq r\}$, $\mathcal{X}_2 = \{x \in \mathbb{R}_+^k : \sum_{i=1}^k x_i \leq r\}$, $\mathcal{X}_3 = \{x \in \mathbb{R}^k : \sum_{i=1}^k |x_i| \leq r\}$, $r > 0$.

Theorem 3. *DS-optimal designs with the smallest possible support size $m = k$ for model (4) on the experimental domains $\mathcal{X}_1, \mathcal{X}_2$ and \mathcal{X}_3 exist and coincide. This is a so-called orthonormal design, that is the design having the weights $1/k$ and the orthonormal support vectors $x^{(1)}, \dots, x^{(k)}$ satisfying $\|x^{(1)}\| = r, \dots, \|x^{(k)}\| = r$ and $x^{(i)\top} x^{(j)} = 0$ for all $i \neq j \leq k$. The moment matrix of such a design has the form $M = (r^2/k)I_k$.*

3 Efficient Designs

Consider model (2) and the asymmetric experimental domain $[a, b] = [0, 1]$. From Subsection 2.2 we know (see Table 1) that there is no DS-optimal design here and $C = \{0, 1; p, 1 - p\}$, $0.5 \leq p \leq 0.6$. What design should one choose then?

The natural approach consists in choosing an efficient design (see e.g. Li and Chan, 2002, Rodrigues-Diaz and López-Fidalgo, 2003), that is the design which is optimal for a certain $\text{DS}(\varepsilon^*)$ -criterion and at the same time performs well under other $\text{DS}(\varepsilon)$ -criteria.

Let $\nu(p, \varepsilon)$ be the *efficiency* of the design $\xi_p = \{0, 1; p, 1 - p\}$ with respect to the $\text{DS}(\varepsilon)$ -criterion, that is

$$\nu(p, \varepsilon) = \frac{\ln P(\|\widehat{\beta}(M(\xi_{p_\varepsilon})) - \beta\| \leq \varepsilon)}{\ln P(\|\widehat{\beta}(M(\xi_p)) - \beta\| \leq \varepsilon)}$$

where ξ_{p_ε} stands for the $\text{DS}(\varepsilon)$ -optimal design. We use the *maximin* approach and wish to find p so as to maximize $\min_{\varepsilon > 0} \nu(p, \varepsilon)$. The design $\xi_{p^*} \in C$ such that

$$p^* = \text{Arg} \max_p \min_{\varepsilon > 0} \nu(p, \varepsilon)$$

is called the *efficient* design. Let us keep the same considerations and notations for the SS_ρ -criterion as well.

The numerical calculations give the following results. For the DS-criterion

$$p^* = \text{Arg} \max_{0.5 \leq p \leq 0.6} \min_{\varepsilon > 0} \nu(p, \varepsilon) = 0.539, \quad \min_{\varepsilon > 0} \nu(p^*, \varepsilon) = 0.992.$$

For the SS_ρ -criterion (squares) where $C = \{\{0, 1; p, 1 - p\}, 0.5 \leq p \leq 0.54(54)\}$ (see Table 2)

$$p^* = \text{Arg} \max_{0.5 \leq p \leq 0.54(54)} \min_{\varepsilon > 0} \nu(p, \varepsilon) = 0.521, \quad \min_{\varepsilon > 0} \nu(p^*, \varepsilon) = 0.998.$$

For the SS_ρ -criterion (rectangulars) where $C = \{\{0, 1; p, 1 - p\}, 0.5 \leq p \leq 0.75\}$ (see Table 3)

$$p^* = \text{Arg} \max_{0.5 \leq p \leq 0.75} \min_{\varepsilon > 0} \nu(p, \varepsilon) = 0.580, \quad \min_{\varepsilon > 0} \nu(p^*, \varepsilon) = 0.962.$$

At last, for the SS_ρ -criterion (ellipses) where $C = \{\{0, 1; p, 1 - p\}, 0.4 \leq p \leq 0.5\}$ (see Table 4)

$$p^* = \text{Arg} \max_{0.4 \leq p \leq 0.5} \min_{\varepsilon > 0} \nu(p, \varepsilon) = 0.461, \quad \min_{\varepsilon > 0} \nu(p^*, \varepsilon) = 0.992.$$

As one can see, in all the cases (except for rectangulars) the efficiencies of the efficient designs are very high.

Appendix

Proof of Theorem 2. We begin with two auxiliary results.

Lemma 1. Let Z be an $(k + 1) \times k$ matrix of the full rank k and $e \neq 0$ be a given vector in \mathbb{R}^{k+1} . Then under the condition $Z^T e = 0$,

$$ZZ^T = c[I_{k+1} - ee^T/\|e\|^2] \iff Z^T Z = cI_k$$

for any given $c > 0$.

Proof. (\implies) Assume that $Z^T e = 0$ and $ZZ^T = c[I_{k+1} - ee^T/\|e\|^2]$. Multiplying the last equality by Z^T on the left and by Z on the right, we get $(Z^T Z)^2 = cZ^T Z$. Since Z is of the rank k , $Z^T Z$ is positive definite. Therefore, $Z^T Z = cI_k$.

(\impliedby) Assume that $Z^T e = 0$ and $Z^T Z = cI_k$. Then

$$\begin{pmatrix} \|e\|Z^T \\ \sqrt{c}e^T \end{pmatrix} (\|e\|Z \quad \sqrt{c}e) = \begin{pmatrix} \|e\|^2 Z^T Z & 0 \\ 0 & c\|e\|^2 \end{pmatrix} = c\|e\|^2 I_{k+1}.$$

Therefore, also

$$(\|e\|Z \quad \sqrt{c}e) \begin{pmatrix} \|e\|Z^T \\ \sqrt{c}e^T \end{pmatrix} = c\|e\|^2 I_{k+1}.$$

But the last equality is equivalent to $\|e\|^2 ZZ^T + cee^T = c\|e\|^2 I_{k+1}$. The lemma is proved.

Now, we establish the D-optimal design.

Lemma 2. *A design ξ with the support size $m = k + 1$ for model (3) on the experimental domain \mathcal{Z}_r is D-optimal if and only if it is a regular simplex design.*

Proof. Let $\xi = \{x^{(1)}, x^{(2)}, \dots, x^{(k+1)}; p_1, p_2, \dots, p_{k+1}\}$, $x^{(i)} = (1, z^{(i)\text{T}})^\text{T}$, $z^{(i)} \in \mathcal{Z}_r$, $i = 1, 2, \dots, k + 1$. We have

$$M(\xi) = X^\text{T} P X = \begin{pmatrix} 1 & \sum_{i=1}^{k+1} p_i z^{(i)\text{T}} \\ \sum_{i=1}^{k+1} p_i z^{(i)} & \sum_{i=1}^{k+1} p_i z^{(i)} z^{(i)\text{T}} \end{pmatrix} \quad (6)$$

where $P = \text{diag}(p_1, p_2, \dots, p_{k+1})$. Observe that

$$\det(P) \leq (1/(k+1))^{k+1}$$

and the equality holds if and only if $p_1 = p_2 = \dots = p_{k+1} = 1/(k+1)$. Therefore,

$$\det(M(\xi)) = (\det(X))^2 \det(P) \leq (1/(k+1))^{k+1} \det(X^\text{T} X).$$

This means that searching for the D-optimal design it is enough to consider only designs with equal weights. By the properties of determinant (see Rao, 1973, p. 32),

$$\det(X^\text{T} X) \leq (k+1) \det\left(\sum_{i=1}^{k+1} z^{(i)} z^{(i)\text{T}}\right) = (k+1) \det(Z^\text{T} Z)$$

where $Z = (z^{(1)}, z^{(2)}, \dots, z^{(k+1)})^\text{T}$. The upper bound is attained if and only if $\sum_{i=1}^{k+1} z^{(i)} = 0$. On the other hand, by the properties of trace we have

$$\text{tr}(Z^\text{T} Z) = \text{tr}(Z Z^\text{T}) = \sum_{i=1}^{k+1} \|z^{(i)}\|^2 \leq r^2(k+1)$$

and the upper bound is attained if and only if $\|z^{(i)}\| = r$, $i = 1, 2, \dots, k + 1$. Clearly, under the constraint $\text{tr}(Z^\text{T} Z) = \lambda_1 + \lambda_2 + \dots + \lambda_k \leq r^2(k+1)$, the value of $\det(Z^\text{T} Z) = \lambda_1 \lambda_2 \dots \lambda_k$ attains its maximum if and only if $\lambda_1 = \lambda_2 = \dots = \lambda_k = r^2(k+1)/k$. Thus,

$$\det(M(\xi)) \leq (1/(k+1))^k \det(Z^\text{T} Z) \leq (r^2/k)^k.$$

Summing up, one can note that the upper bound is attained here if and only if $M(\xi) = \text{diag}(1, r^2/k, \dots, r^2/k)$. It remains to show that only a regular simplex design has such a moment matrix.

Observe that if a design ξ has equal weights, then (6) gives

$$M(\xi) = \begin{pmatrix} 1 & (k+1)^{-1} \sum_{i=1}^{k+1} z^{(i)\text{T}} \\ (k+1)^{-1} \sum_{i=1}^{k+1} z^{(i)} & (k+1)^{-1} Z^\text{T} Z \end{pmatrix}.$$

For any regular simplex design we have $\sum_{i=1}^{k+1} z^{(i)} = 0$. This is the consequence of the following two facts: (i) $\sum_{i=1}^{k+1} z^{(i)\text{T}} z^{(j)} = 0$ for any $j = 1, 2, \dots, k+1$; (ii) the span of the vectors $\{z^{(1)}, z^{(2)}, \dots, z^{(k+1)}\}$ coincides with \mathbb{R}^k , otherwise the matrix $M(\xi)$ would be degenerate.

Furthermore, for any regular simplex design

$$ZZ^{\text{T}} = \frac{r^2}{k} \begin{pmatrix} k & -1 & \dots & -1 \\ -1 & k & \dots & -1 \\ \dots & \dots & \dots & \dots \\ -1 & -1 & \dots & k \end{pmatrix} = \frac{r^2(k+1)}{k} [I_{k+1} - ee^{\text{T}}]$$

where $e = (k+1)^{-1/2}(1, 1, \dots, 1)^{\text{T}}$. Since for any regular simplex design $Z^{\text{T}}e = 0$, applying Lemma 2 we obtain $Z^{\text{T}}Z = (r^2(k+1)/k)I_k$. Thus, we prove that any regular simplex design has the moment matrix $\text{diag}(1, r^2/k, \dots, r^2/k)$.

Now, we show that if $M(\xi) = \text{diag}(1, r^2/k, \dots, r^2/k)$, then ξ is a regular simplex design. Indeed, in such a case from (6) it follows that

$$\sum_{i=1}^{k+1} p_i z^{(i)} = 0, \quad \sum_{i=1}^{k+1} p_i z^{(i)} z^{(i)\text{T}} = (r^2/k)I_k. \quad (7)$$

Denote $P^{1/2}Z = Y$, $P^{1/2}e = \varepsilon$, where P and e are the same as above. Then (7) implies

$$Y^{\text{T}}\varepsilon = 0, \quad Y^{\text{T}}Y = (r^2/k)I_k$$

and by Lemma 2

$$YY^{\text{T}} = (r^2/k)[I_{k+1} - \varepsilon\varepsilon^{\text{T}}/\|\varepsilon\|^2].$$

Since $\|\varepsilon\|^2 = e^{\text{T}}Pe = (k+1)^{-1}$, we obtain

$$P^{1/2}ZZ^{\text{T}}P^{1/2} = (r^2/k)[I_{k+1} - (k+1)P^{1/2}ee^{\text{T}}P^{1/2}],$$

or

$$ZZ^{\text{T}} = (r^2/k)[P^{-1} - (k+1)ee^{\text{T}}] = (r^2/k) \begin{pmatrix} 1/p_1 - 1 & -1 & \dots & -1 \\ -1 & 1/p_2 - 1 & \dots & -1 \\ \dots & \dots & \dots & \dots \\ -1 & -1 & \dots & 1/p_{k+1} - 1 \end{pmatrix}.$$

But the diagonal elements of the matrix ZZ^{T} are $\|z^{(1)}\|^2, \|z^{(2)}\|^2, \dots, \|z^{(k+1)}\|^2$. Since $\|z^{(i)}\|^2 \leq r^2$, $i = 1, 2, \dots, k+1$, we should have $p_i \geq (k+1)^{-1}$, $i = 1, 2, \dots, k+1$. Thus, $p_1 = p_2 = \dots = p_{k+1} = (k+1)^{-1}$. The lemma is proved.

Now, we come back to the proof of the theorem. By Lemma 2 and Theorem 1 of Zaigraev (2002), it follows that if the DS-optimal design exists, then it is necessary a regular simplex design.

Assume, first, that $r^2 \leq k$. By Theorem 3.2 of Liski et al. (1999), in order to prove the theorem it is enough to show that for any given design ξ the eigenvalues $\lambda_1(M(\xi))$,

$\lambda_2(M(\xi)), \dots, \lambda_{k+1}(M(\xi))$ of the moment matrix $M(\xi)$ given by (6), arranged in the descending order of magnitudes, satisfy the inequalities:

$$\begin{aligned}\lambda_{k+1}(M(\xi)) &\leq r^2/k, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) &\leq 2r^2/k, \dots, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) + \dots + \lambda_2(M(\xi)) &\leq r^2, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) + \dots + \lambda_1(M(\xi)) &\leq r^2 + 1.\end{aligned}\tag{8}$$

The last inequality is evident since

$$\text{tr}(M(\xi)) = \text{tr}(PXX^T) = \sum_{i=1}^{k+1} p_i(\|z^{(i)}\|^2 + 1) \leq r^2 + 1.$$

Denote $B = \sum_{i=1}^{k+1} p_i z^{(i)} z^{(i)T}$ in (6). By the Sturm separation theorem (see Rao, 1973, p. 64) we have the sequence of inequalities:

$$\begin{aligned}\lambda_{k+1}(M(\xi)) &\leq \lambda_k(B) \leq \lambda_k(M(\xi)) \leq \lambda_{k-1}(B) \leq \lambda_{k-1}(M(\xi)) \leq \dots \leq \lambda_2(B) \\ &\leq \lambda_2(M(\xi)) \leq \lambda_1(B) \leq \lambda_1(M(\xi)).\end{aligned}$$

Therefore, in order to prove the inequalities in (8) it is enough to show that

$$\begin{aligned}\lambda_k(B) &\leq r^2/k, \lambda_k(B) + \lambda_{k-1}(B) \leq 2r^2/k, \dots, \\ \lambda_k(B) + \lambda_{k-1}(B) + \dots + \lambda_1(B) &\leq r^2.\end{aligned}\tag{9}$$

We have $\text{tr}(B) = \sum_{i=1}^{k+1} p_i \|z^{(i)}\|^2 \leq r^2$. If $\text{tr}(B) = r^2$, then (9) holds with the equality in the last inequality (see Marshall and Olkin, 1979, Section 1.A). If $\text{tr}(B) < r^2$, then (9) holds all the more.

Assume, now, that $r^2 > k$. Here, instead of (8) we should prove that

$$\begin{aligned}\lambda_{k+1}(M(\xi)) &\leq 1, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) &\leq 1 + r^2/k, \dots, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) + \dots + \lambda_2(M(\xi)) &\leq 1 + r^2(k-1)/k, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) + \dots + \lambda_1(M(\xi)) &\leq 1 + r^2.\end{aligned}\tag{10}$$

Again we have already had the last inequality. As to all the others, utilizing the extremal properties of eigenvalues (see Rao, 1973, p. 62) we obtain

$$\begin{aligned}\lambda_{k+1}(M(\xi)) &= \min_{a \in \mathbb{R}^{k+1}, \|a\|=1} a^T M(\xi) a \leq e_*^T M(\xi) e_* = 1, \\ \lambda_{k+1}(M(\xi)) + \lambda_k(M(\xi)) &= \min_{a, b \in \mathbb{R}^{k+1}, \|a\|=1, \|b\|=1, a^T b=0} (a^T M(\xi) a + b^T M(\xi) b)\end{aligned}$$

$$\leq e_*^T M(\xi) e_* + y^T M(\xi) y \leq 1 + \lambda_k(B), \dots$$

where $e_* = (1, 0, \dots, 0)^T$ and $y = (0, y_1, \dots, y_k)^T$, $\|y\| = 1$. Continuing such a process and taking into account (9), we obtain (10). The theorem is proved.

Proof of Theorem 3. First, observe that $\mathcal{X}_2 \subset \mathcal{X}_1$, $\mathcal{X}_3 \subset \mathcal{X}_1$ and both \mathcal{X}_2 and \mathcal{X}_3 contain all the orthonormal designs. Therefore, it is enough to prove the theorem only for the experimental domain \mathcal{X}_1 .

We start with proving that an orthonormal design is D-optimal. As in the proof of Theorem 2, it is easy to show that we can confine ourselves to the case of designs with equal weights. Such a design ξ has the moment matrix $M(\xi) = (1/k)X^T X$. From the Hadamard inequality (Rao, 1973, p. 56) it follows that

$$\det(M(\xi)) = (1/k^k) \det(X^T X) = (1/k^k) \det(X X^T) \leq (1/k^k) \prod_{i=1}^k \|x^{(i)}\|^2 \leq (r^2/k)^k.$$

Equality holds if and only if $X X^T = r^2 I_k$, that is if and only if the design ξ is orthonormal. Thus, an orthonormal design is D-optimal on \mathcal{X}_1 .

In order to show that an orthonormal design is DS-optimal, it remains to prove that for any given design ξ the following inequalities

$$\lambda_k(M(\xi)) \leq r^2/k,$$

$$\lambda_k(M(\xi)) + \lambda_{k-1}(M(\xi)) \leq 2r^2/k, \dots,$$

$$\lambda_k(M(\xi)) + \lambda_{k-1}(M(\xi)) + \dots + \lambda_1(M(\xi)) \leq r^2$$

hold (cf. (8)). But these inequalities are evident due to the relation

$$\text{tr}(M(\xi)) = \sum_{i=1}^k p_i \|x^{(i)}\|^2 \leq r^2.$$

The theorem is proved.

Acknowledgement

I would like to thank the referee for the constructive comments which improved the paper.

References

- A. De la Garza. Spacing of information in polynomial estimation. *Ann. Math. Stat.*, 25:123–130, 1954.
- K.-H. Li and N.N. Chan. L_p -optimality for regression designs under correlations. *Stat. Sinica*, 12:1015–1026, 2002.
- E.P. Liski, A. Luoma, and A. Zaigraev. Distance optimality design criterion in linear models. *Metrika*, 49:193–211, 1999.

- E.P. Liski, N.K. Mandal, K.R. Shah, and B.K. Sinha. *Topics in Optimal Design*. Lecture Notes in Statistics, 163, Springer-Verlag, New York, 2002.
- E.P. Liski and A. Zaigraev. A stochastic characterization of Loewner optimality design criterion in linear models. *Metrika*, 53:207–222, 2001.
- A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.
- F. Pukelsheim. *Optimal Design of Experiments*, Wiley, New York, 1993.
- C.R. Rao. *Linear Statistical Inference and Its Applications*, Wiley, New York, 2nd edition, 1973.
- J.M. Rodrigues-Diaz and J. López-Fidalgo. A bi-dimensional class of optimality criteria involving ϕ_p and characteristic criteria. *Statistics*, 37(4):325–334, 2003.
- B.K. Sinha. On the optimality of some design. *Calcutta Statistical Association Bulletin*, 20:1–20, 1970.
- A. Zaigraev. Shape optimal design criterion in linear models. *Metrika*, 56:259–273, 2002.

Author's address:

Dr. Alexander Zaigraev
Faculty of Mathematics and Computer Science
N.Copernicus University
Chopin str. 12/18
PL-87-100 Toruń
Poland

Tel. +48 56 6113372
Fax +48 56 6228979
E-mail: alzaig@mat.uni.torun.pl
<http://www.mat.uni.torun.pl/~alzaig/>