Tatra Mt. Math. Publ. **26** (2003), 59–67

TATRA MOUNTAINS
*Mathematical Publications*

# A METHOD HOW TO DELETE POINTS WHICH DO NOT SUPPORT A D-OPTIMAL DESIGN

RADOSLAV HARMAN

ABSTRACT. For linear regression models with uncorrelated errors, we describe a method of deletion of design points, which do not support any D-optimal design measure. The key idea is to construct a suitable set of matrices, which contains the information matrix corresponding to an optimal design. A combination of this method with iterative algorithms speeds up computations, and leads to designs in a more concentrated form.

## 1. Introduction

Suppose that we intend to perform experiments with real-valued observations $y_i$ modeled by the linear regression formula

$$y_i = f^T(x_i)\beta + \varepsilon_i\,,$$

where $i \in \mathbb{N}$ is the index of an observation, $\beta$ is an unknown parameter from $\mathbb{R}^p$, and the values $x_i$ are chosen from a compact experimental domain $\mathfrak{X} \subseteq \mathbb{R}^m$. For any choice of $x_i$, the errors $\varepsilon_i$ are assumed to be uncorrelated, with zero mean and the same finite variance $\sigma^2$. We also assume that the function $f(\cdot)\colon \mathfrak{X} \to \mathbb{R}^p$ is continuous on $\mathfrak{X}$.

As is usual, by an (asymptotic) experimental design we understand a probability measure $\xi$ finitely supported on $\mathfrak{X}$, such that the value $\xi(x)$ indicates the relative proportion of the measurements which should be taken in $x \in \mathfrak{X}$ (see, e.g., [1], p.17). By the symbol $\Xi$ we denote the set of all designs on the experimental domain $\mathfrak{X}$. Without loss of generality we can assume that $\sigma = 1$ and define the information matrix $\mathbf{M}(\xi)$ associated with a design $\xi \in \Xi$ as

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

---

2000 Mathematics Subject Classification: 62K05.
Keywords: D-optimality, support of design measure, computaion of optimal design.
Supported by the grant VEGA No. 1/7295/20.

It can be shown that the set $\mathcal{M} = \{\mathbf{M}(\xi) \colon \xi \in \Xi\}$ of all information matrices associated with designs on $\mathfrak{X}$ is convex and compact (cf. [1], p. 29, p. 60).

To design an experiment in an optimal way, we will use the criterion of D-optimality: $\Phi \colon \mathcal{S}_+ \to \mathbb{R}$; defined

$$\Phi(\mathbf{M}) = \begin{cases} -\ln\det\mathbf{M} & \text{for } \mathbf{M} \in \mathcal{S}_{++}\,, \\ \infty & \text{for } \mathbf{M} \in \mathcal{S}_+ \backslash \mathcal{S}_{++}\,, \end{cases}$$

where $\mathcal{S}_+$ and $\mathcal{S}_{++}$ denote the set of all semipositively, resp. positively definite matrices of type $p \times p$. The function $\Phi$ is continuous and convex on $\mathcal{S}_+$ and strictly convex on $\mathcal{S}_{++}$ ([1], pp. 78–81).

A design $\xi^*$ is said to be D-optimal, if it minimizes $\Phi(\mathbf{M}(\xi))$ over all $\xi \in \Xi$, or equivalently if $\det(\mathbf{M}(\xi^*)) = \sup\limits_{\xi\in\Xi} \det(\mathbf{M}(\xi))$. The symbol $\Xi^*$ shall denote the set of all D-optimal designs. The properties of $\Phi$ and $\mathcal{M}$ imply that there always exists at least one D-optimal design. Assuming that $\mathcal{M} \cap \mathcal{S}_{++} \neq \emptyset$, it also follows that all the D-optimal designs have the same information matrix, which we denote by $\mathbf{M}^*$ and call the D-optimal information matrix.

The problem of finding an asymptotic D-optimal design has gained much attention in the experimental design literature. The main problem of existing iterative algorithms is their slow convergence, especially for a large dimension of the parameter. Also, the support of the design measures obtained in the iterative process does not have a tendency to shrink, that is the output of a program which implements these algorithms contains many points with small weights, or groups of points, which are very close together.

The aim of this article is to show a geometrically based method which allows us remove some unnecessary points of the experimental domain, that is some of the points of $\mathfrak{X}$, which can not support any D-optimal design measure. We remark that the iterative algorithms used for construction of a D-optimal design need to scan the experimental domain in every iteration. (This is best seen if $\mathfrak{X}$ is finite.) Therefore, if we were able to restrict our attention to a set which is smaller than $\mathfrak{X}$, we would, as a rule, gain an increase in the speed of the algorithm.

One possible approach is to delete the points $x \in \mathfrak{X}$, such that $f(x)$ does not belong to the boundary of the set $\operatorname{conv}\Big(\{f(x) \colon x \in \mathfrak{X}\} \cup \{-f(x) \colon x \in \mathfrak{X}\}\Big)$ (where conv means the convex hull; cf. [2] and [1], p. 56). However, it turns out that in majority of commonly used models this method only removes small parts of $\mathfrak{X}$, and does not help us use the knowledge gained during the iterative process of computation of a design. Moreover, the method is difficult from the computational point of view.

In this article we propose a different method which does not have these disadvantages. The main idea of the method is formulated in the following simple proposition:

**Proposition 1.** *Let $\mathcal{P} \subseteq \mathcal{S}_{++}$ be a set containing $\mathbf{M}^*$. If*

$$\sup_{\mathbf{M}\in\mathcal{P}} f^T(x)\mathbf{M}^{-1}f(x) < p$$

*for some $x \in \mathfrak{X}$, then $x$ does not support any D-optimal design, i.e., for all $\xi^* \in \Xi^*$:*

$$\xi^*\big\{x \in \mathfrak{X}\colon \sup_{\mathbf{M}\in\mathcal{P}} f^T(x)\mathbf{M}^{-1}f(x) < p\big\} = 0\,.$$

*In particular, if $\mathbf{M}^* \in \mathcal{P} = \operatorname{conv}\{\mathbf{Q}_1,\dots,\mathbf{Q}_r\}$ for some $\mathbf{Q}_1,\dots,\mathbf{Q}_r \in \mathcal{S}_{++}$, and if $\max_{i=1,\dots,r} f^T(x)\mathbf{Q}_i^{-1}f(x) < p$, then $x \in \mathfrak{X}$ does not support any D-optimal design.*

Proof. If $\xi^*$ is a D-optimal design then $\xi^*$-almost all points from $\mathfrak{X}$ satisfy the condition $f^T(x)\mathbf{M}^{*-1}f(x) = p$ (cf. [4], Theorem 1c), which entails the first part of the proposition. The second part is a consequence of the fact that the function $\psi_x\colon \mathbf{M} \to f^T(x)\mathbf{M}^{-1}f(x)$ is convex on $\mathcal{S}_{++}$ (see [1], p. 62). $\square$

Naturally, we would use the full potential of the proposition if we knew $\mathbf{M}^*$, simply setting $\mathcal{P} = \{\mathbf{M}^*\}$. However, we rarely can find the value of $\mathbf{M}^*$ exactly. Nevertheless, as we show in the next section, we are able to construct a polyhedral set $\mathcal{P} = \operatorname{conv}\{\mathbf{Q}_1,\dots,\mathbf{Q}_r\} \subseteq \mathcal{S}_{++}$ containing $\mathbf{M}^*$ *without* knowing the value of $\mathbf{M}^*$. This set shall only depend on an information matrix $\mathbf{M}_0 = \mathbf{M}(\xi_0)$, which is "close" to $\mathbf{M}^*$, i.e., on a suboptimal design $\xi_0$.

## 2. Construction of the polyhedral set $\mathcal{P}$

In this section, we will describe a method of construction of a polyhedral set $\mathcal{P} \subseteq \mathcal{S}_{++}$ containing the D-optimal information matrix $\mathbf{M}^*$.

Choose any matrix $\mathbf{M}_0 = \mathbf{M}(\xi_0) \in \mathcal{S}_{++}$. It is clear that $\mathbf{M}^*$ is an element of

$$\mathcal{C}_{\mathbf{M}_0} = \big\{\mathbf{A} \in \mathcal{S}_+\colon\ \det(\mathbf{A}) \ge \det(\mathbf{M}_0)\big\}\,.$$

Next, for $\xi^* \in \Xi^*$ we can write

$$\begin{aligned}\operatorname{tr}\mathbf{M}^*\mathbf{M}_0^{-1} &= \operatorname{tr}\Bigg(\sum_{x\,;\,\xi^*(x)>0} \xi^*(x)f(x)f^T(x)\mathbf{M}_0^{-1}\Bigg)\\ &= \sum_{x;\xi^*(x)>0} \xi^*(x)f^T(x)\mathbf{M}_0^{-1}f(x) \le \sup_{x\in\mathfrak{X}} f^T(x)\mathbf{M}_0^{-1}f(x)\,,\end{aligned}$$

which means that $\mathbf{M}^*$ also belongs to

$$\mathcal{H}_{\mathbf{M}_0}(c) = \{\mathbf{S} \in \mathcal{S}\colon\ \operatorname{tr}\mathbf{S}\mathbf{M}_0^{-1} \le c + p\},$$

where $c = \sup_{x \in \mathfrak{X}} f^T(x)\mathbf{M}_0^{-1} f(x) - p \ge 0$. ($\mathcal{S}$ is the set of all symmetric matrices $p \times p$.)

Notice that $\mathcal{H}_{\mathbf{M}_0}(c)$ has a simple geometric interpretation: it is the minimal half-space in $\mathcal{S}$ from all the half-spaces which contain $\mathcal{M}$ and which have the normal vector equal to $\nabla \ln \det(\mathbf{M}_0) = \mathbf{M}_0^{-1}$ (for definition of $\nabla$ see [1], p. 81). Consider the set

$$\mathcal{B}_{\mathbf{M}_0}(c) = \mathcal{C}_{\mathbf{M}_0} \cap \mathcal{H}_{\mathbf{M}_0}(c).$$

This set is closed, convex, and contains $\mathbf{M}^*$. Moreover, $\mathcal{B}_{\mathbf{M}_0}(c)$ is bounded, which is a simple consequence of a proposition proven in the sequel.

Let us introduce the following function:

$$\varphi_{\mathbf{M}_0}\colon \mathcal{S} \to \mathcal{S}; \quad \varphi_{\mathbf{M}_0}(\mathbf{S}) = \mathbf{M}_0^{-1/2}\mathbf{S}\mathbf{M}_0^{-1/2}.$$

One can see that the function $\varphi_{\mathbf{M}_0}$ is linear, regular (i.e., bijective), $\varphi_{\mathbf{M}_0}(\mathcal{S}_{++}) = \mathcal{S}_{++}$, and $\varphi_{\mathbf{M}_0}(\mathbf{M}_0) = \mathbf{I}$. This function allows us normalize the problem, because $\varphi_{\mathbf{M}_0}(\mathcal{C}_{\mathbf{M}_0}) = \mathcal{C}_{\mathbf{I}}$, $\varphi_{\mathbf{M}_0}(\mathcal{H}_{\mathbf{M}_0}(c)) = \mathcal{H}_{\mathbf{I}}(c)$, resp. $\varphi_{\mathbf{M}_0}(\mathcal{B}_{\mathbf{M}_0}(c)) = \mathcal{B}_{\mathbf{I}}(c)$, which is easy to check. Therefore, it is enough to study the set

$$\mathcal{B}_{\mathbf{I}}(c) = \bigl\{\mathbf{A} \in \mathcal{S}_+\colon\ \det(\mathbf{A}) \ge 1 \ \text{ and } \ \operatorname{tr}(\mathbf{A}) \le p + c\bigr\}.$$

**Lemma 2.** *Let $p \ge 2$. Then for any $\mathbf{A} \in \mathcal{B}_{\mathbf{I}}(c)$, such that $\operatorname{tr}(\mathbf{A}) = p + c$:*

$$\left\|\mathbf{A} - \left(1 + \frac{c}{p}\right)\mathbf{I}\right\| \le r_p(c), \quad \text{where } r_p(c) = \sqrt{\frac{p-1}{p}c^2 + (p^2 - p)c}\,.$$

P r o o f. Let $\mathbf{A} \in \mathcal{B}_{\mathbf{I}}(c)$, $\operatorname{tr}(\mathbf{A}) = p + c$, $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is the orthonormal matrix of eigenvectors of $\mathbf{A}$, and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_p)$, where $\lambda_1, \dots, \lambda_p$ are corresponding eigenvalues of $\mathbf{A}$. The inequality $1 \le \det(\mathbf{A}) = \prod_k \lambda_k$ and the relation between geometric and arithmetic means implies that for any $i \ne j$, $i, j \in \{1, \dots, p\}$:

$$1 \le \left((\lambda_i\lambda_j) \prod_{k \ne i, k \ne j} \lambda_k\right)^{p-1} \le \frac{1}{p-1}\left((\lambda_i\lambda_j) + \sum_{k \ne i, k \ne j} \lambda_k\right).$$

Summing up all the $\binom{p}{2}$ inequalities for $i < j$, and noticing that $\sum_{i<j}\sum_{k \ne i, k \ne j} \lambda_k = \binom{p-1}{2}\sum_i \lambda_i$ we obtain:

$$\binom{p}{2}(p-1) - \binom{p-1}{2}\sum_i \lambda_i \le \sum_{i<j} \lambda_i\lambda_j\,.$$

From the previous inequality, the equality $\operatorname{tr}(\mathbf{A}^2) = \sum_i \lambda_i^2 = \Bigl(\sum_i \lambda_i\Bigr)^2 - 2\sum_{i<j} \lambda_i\lambda_j$, and $\operatorname{tr}(\mathbf{A}) = \sum_i \lambda_i = p + c$, we get:

$$\left\| \mathbf{A} - \Bigl(1+\frac{c}{p}\Bigr)\mathbf{I} \right\|^2 = (p+c)^2 - 2\sum_{i<j}\lambda_i\lambda_j - 2\Bigl(1+\frac{c}{p}\Bigr)(p+c) + p\Bigl(1+\frac{c}{p}\Bigr)^2$$

$$\le (p+c)^2 - 2\left(\binom{p}{2}(p-1) - \binom{p-1}{2}(p+c)\right) - 2\Bigl(1+\frac{c}{p}\Bigr)(p+c) + p\Bigl(1+\frac{c}{p}\Bigr)^2$$

$$= \frac{p-1}{p}c^2 + (p^2-p)c\,.$$

□

**Proposition 3.** *Let $p \ge 2$. Then $\mathcal{B}_{\mathbf{I}}(c)$ is a subset of $\mathcal{R} = \operatorname{conv}(\mathcal{R}_* \cup \mathcal{R}^*)$, where*

$$\mathcal{R}^* = \left\{ \mathbf{A} \in \mathcal{S}\colon \left\| \mathbf{A} - \Bigl(1+\frac{c}{p}\Bigr)\mathbf{I} \right\| \le r_p(c)\,,\ \textit{and}\ \operatorname{tr}(\mathbf{A}) = p + c \right\},$$

*and*

$$\mathcal{R}_* = \mathcal{R}^* - \frac{c}{p}\mathbf{I}\,.$$

P r o o f. It is simple to verify the proposition for $c = 0$. Let $c > 0$, $\mathbf{A} \in \mathcal{B}_{\mathbf{I}}(c)$, $\mathbf{R}^* = \mathbf{A} + \bigl(1+\frac{c}{p}\bigr)\mathbf{I} - \bigl(\frac{1}{p}\operatorname{tr}(\mathbf{A})\bigr)\mathbf{I}$, $\mathbf{R}_* = \mathbf{R}^* - \frac{c}{p}\mathbf{I}$, and $\beta = \frac{1}{c}\bigl(p + c - \operatorname{tr}(\mathbf{A})\bigr)$. Notice that $\operatorname{tr}(\mathbf{R}^*) = p + c$, and $\mathbf{R}^* \ge \mathbf{A}$ in Schur ordering, therefore $\det(\mathbf{R}^*) \ge \det(\mathbf{A}) \ge 1$. Consequently, the previous lemma entails $\mathbf{R}^* \in \mathcal{R}^*$, and hence also $\mathbf{R}_* \in \mathcal{R}_*$. Obviously $\beta \in [0,1]$, because $p \le \operatorname{tr}(\mathbf{A}) \le p + c$. Moreover, $\beta\mathbf{R}_* + (\mathbf{1}-\beta)\mathbf{R}^* = \mathbf{R}^* - \beta\frac{c}{p}\mathbf{I} = \mathbf{A}$, as is easy to check. □

The set $\mathcal{R}$ forms a multi-dimensional cylinder, because the $k = \frac{1}{2}p(p+1) - 1$ dimensional circles $\mathcal{R}_*$ and $\mathcal{R}^*$ are parallel. Hence, we are able to encase $\mathcal{R}$ into a convex polyhedral set generated by $2k + 2 = p(p+1)$ symmetric matrices $\mathbf{R}_1, \dots, \mathbf{R}_{2k+2}$, where $\mathbf{R}_1, \dots, \mathbf{R}_{k+1}$ are vertices of a simplex circumscribing $\mathcal{R}_*$, and $\mathbf{R}_{k+2} = \mathbf{R}_1 + \frac{c}{p}\mathbf{I}, \dots, \mathbf{R}_{2k+2} = \mathbf{R}_{k+1} + \frac{c}{p}\mathbf{I}$ are vertices of a simplex circumscribing $\mathcal{R}^*$. (Naturally, this construction is a compromise between the volume and the number of vertices; we can construct a polyhedral superset of $\mathcal{R}$ with somewhat smaller volume but with more vertices.)

It is obvious that we can construct the matrices $\mathbf{R}_1, \dots, \mathbf{R}_{2k+2}$ as follows:

$$\mathbf{R}_i = k r_p(c)\mathbf{A}_i + \mathbf{I}\,, \quad \mathbf{R}_{i+k+1} = \mathbf{R}_i + \frac{c}{p}\mathbf{I} \qquad \text{for } i = 1, \dots, k+1\,,$$

where $\mathbf{A}_1, \dots, \mathbf{A}_{k+1}$ correspond to the vertices of a regular simplex circumscribed by the unit sphere in the $k$ dimensional linear space $\mathbf{I}^{\perp} \cap \mathcal{S}$, that is $\mathbf{A}_1, \dots, \mathbf{A}_{k+1}$ are symmetric matrices which satisfy: $\langle \mathbf{A}_i, \mathbf{A}_j \rangle = \operatorname{tr}(\mathbf{A}_i \mathbf{A}_j) = -1/k$ for all $i \neq j$, $\langle \mathbf{A}_i, \mathbf{I} \rangle = \operatorname{tr}(\mathbf{A}_i) = 0$, and $\|\mathbf{A}_i\|^2 = \operatorname{tr}(\mathbf{A}_i^2) = 1$ for all $i = 1, \dots, k+1$.

Transforming the vertices $\mathbf{R}_1, \dots, \mathbf{R}_{2k+2}$ by $\varphi_{\mathbf{M}_0}^{-1}$, and summarizing all the considerations above, we obtain the following theorem:

**THEOREM 4.** *Let* $\mathbf{M}_0 \in \mathcal{M} \cap \mathcal{S}_{++}$, $k = \frac{1}{2}p(p+1) - 1$,

$$c = \sup_{x \in \mathfrak{X}} f^T(x) \mathbf{M}_0^{-1} f(x) - p, \quad \text{and} \quad r = \sqrt{\frac{p-1}{p} c^2 + (p^2 - p) c}\,.$$

*Next, let* $\mathbf{A}_1, \dots, \mathbf{A}_{k+1}$ *be a set of symmetric matrices of the type* $p \times p$ *which satisfy:* $\langle \mathbf{A}_i, \mathbf{I} \rangle = 0$, $\|\mathbf{A}_i\| = 1$, $\langle \mathbf{A}_i, \mathbf{A}_j \rangle = -1/k$ *for all* $i \neq j$, $i, j = 1, \dots, k+1$. *Define*

$$\mathcal{P} = \operatorname{conv}\{\mathbf{Q}_1, \dots, \mathbf{Q}_{p(p+1)}\},$$

*where*

$$\mathbf{Q}_i = k r \mathbf{M}_0^{1/2} \mathbf{A}_i \mathbf{M}_0^{1/2} + \mathbf{M}_0, \quad \mathbf{Q}_{i+k+1} = \mathbf{Q}_i + \frac{c}{p} \mathbf{M}_0 \qquad \text{for} \;\; i = 1, \dots, k+1.$$

*Then*

$$\mathbf{M}^* \in \mathcal{P}.$$

*Moreover, if* $kr < 1$ *then*

$$\mathcal{P} \subseteq \mathcal{S}_{++}.$$

Proof. The fact that $\mathbf{M}^* \in \mathcal{P}$ follows from the Proposition 3 and the considerations above. We shall prove that $kr < 1$ implies $\mathcal{P} \subseteq \mathcal{S}_{++}$. Clearly, it is enough to guarantee that $\mathbf{A}_i + \mathbf{I}$ are positive semidefinite for all $i = 1, \dots, k+1$. To prove this, we denote the eigenvalues of $\mathbf{A}_i$ by $\gamma_1, \dots, \gamma_p$ then $1 = \operatorname{tr}(\mathbf{A}_i^2) = \sum_j \gamma_j^2$, which means $\gamma_j \geq -1$ for all $j = 1, \dots, p$. But the eigenvalues of $\mathbf{A}_i + \mathbf{I}$ are $\gamma_1 + 1, \dots, \gamma_p + 1$. Consequently, the symmetric matrix $\mathbf{A}_i + \mathbf{I}$ has all the eigenvalues nonnegative, hence it is positive semidefinite. $\square$

An important fact to notice is that $\mathbf{M}_0 \to \mathbf{M}^*$ implies $c \to 0$ which in turn guarantees that $r_p(c) \to 0$. Therefore, the condition $kr < 1$ shall be satisfied if $\mathbf{M}_0$ is close enough to $\mathbf{M}^*$, and we shall have $\mathcal{P} \subseteq \mathcal{S}_{++}$. Also, the diameter of $\mathcal{P}$ converges to 0 as $\mathbf{M}_0 \to \mathbf{M}^*$, although the convergence is relatively slow: $\operatorname{diam} \mathcal{P} \lesssim \text{constant} \cdot \sqrt{c}$ for the values of $c$ approaching 0.

The matrices $\mathbf{A}_1, \dots, \mathbf{A}_{k+1}$ in the Theorem 4 only depend on the number $p$ of parameters, so we need to compute them only once. Moreover, there exists a very fast (finite) iterative method how to find a set of such matrices, as we will outline.

Consider the following system of $2k$ equalities (the symbols $a_1, \dots, a_k$, $d_1, \dots, d_k$ represent unknowns): $a_1 = 1$, $a_1 d_1 = -\frac{1}{k}$, and $d_1^2 + \dots + d_{i-1}^2 + a_i^2 = 1$, $d_1^2 + \dots + d_{i-1}^2 + a_i d_i = -\frac{1}{k}$ for $i = 2, \dots, k$. It is not difficult to show that this system has a unique vector of solutions $(\alpha_1, \dots, \alpha_k, \delta_1, \dots, \delta_k) \in \mathbb{R}^{2k}$, such that $\alpha_i \geq 0$ for all $i = 1, \dots, k$ and that the solutions satisfy: $\delta_1^2 + \dots + \delta_{k-1}^2 - \alpha_k^2 = -\frac{1}{k}$. Notice that it is simple to derive the solutions if we keep the order of calculations given by a scheme $\alpha_1 \to \delta_1 \to \alpha_2 \to \delta_2 \to \dots \to \alpha_k \to \delta_k$. Thus, for $p = 2$ ($k = 2$) we get the vector of solutions $\big(1, \sqrt{3}/2, -1/2, -\sqrt{3}/2\big)$, for $p = 3$ ($k = 5$) we obtain $1/10 \times \big(10, \sqrt{96}, \sqrt{90}, \sqrt{80}, \sqrt{60}, -2, -\sqrt{6}, -\sqrt{10}, -\sqrt{20}, -\sqrt{60}\big)$, etc.

Let $\mathbf{B}_1, \dots, \mathbf{B}_k$ be any orthonormal basis of $\mathbf{I}^{\perp} \cap \mathcal{S}$. The method of construction of scalars $\alpha_i$ and $\delta_i$ guarantees that we can construct the vertices of the simplex as: $\mathbf{A}_1 = \alpha_1 \mathbf{B}_1$, $\mathbf{A}_i = \delta_1 \mathbf{B}_1 + \dots + \delta_{i-1} \mathbf{B}_{i-1} + \alpha_i \mathbf{B}_i$, for $i = 2, \dots, k$, and $\mathbf{A}_{k+1} = \delta_1 \mathbf{B}_1 + \dots + \delta_{k-1} \mathbf{B}_{k-1} - \alpha_k \mathbf{B}_k$.

For example, if $p = 2$ ($k = 2$) we can choose

$$\mathbf{B}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{B}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

and the method described above gives

$$\mathbf{A}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{A}_2 = \frac{1}{2\sqrt{2}} \begin{pmatrix} -1 & \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}, \quad \mathbf{A}_3 = \frac{1}{2\sqrt{2}} \begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}.$$

The computation of the matrices $\mathbf{Q}_i$ from the Theorem 4 is thus algorithmically simple and fast. Consequently, according to the Proposition 1, the matrices $\mathbf{Q}_i$ can be used to delete those points from the experimental domain, which can not support any D-optimal design.

## 3. An example how to use the deletion method

As an illustrative example we have chosen the problem of a D-optimal design for the linear regression model given by the formula

$$y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon,$$

where the values $x$ are from $\mathfrak{X} = \{0, 0.1, 0.2, \dots, 4.9, 5.0\}$.

For the computation of a D-optimal design, we will use an algorithm, which is specially suited for a discrete design space ([3]). We shall begin with the uniform initial design $\xi_0$; $\xi_0(x) = 1/51$ for all $x \in \mathfrak{X}$. Then we shall perform the iterations setting $\xi_{n+1}(x) = \frac{1}{p} \xi_n(x) \big[ f^T(x) \mathbf{M}(\xi_n)^{-1} f(x) \big]$ for all $x \in \mathfrak{X}$ at the step number $n = 0, 1, 2, \dots$

Notice that in any iteration of the algorithm we have $\xi_n(x) > 0$ for all $x \in \mathfrak{X}\backslash\{0\}$. It means that the algorithm (resp. a computer program) must take all the points (except of 0) into account at any iteration, even when the majority of weights of $\xi_n$ are negligibly small. Naturally, the computations with very small positive values take approximately as much time as those with large real numbers. The method described in the previous paragraphs allows us remove many such points, and consequently speed up the computations.

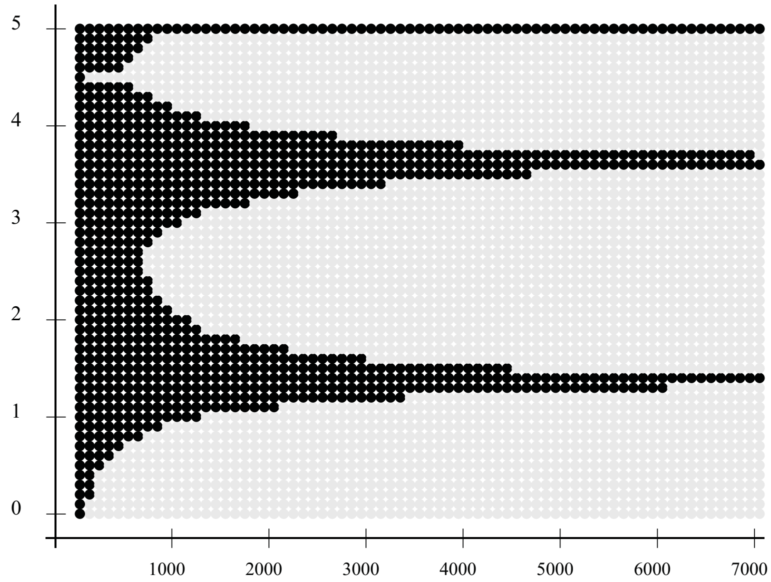


FIGURE 1

More precisely, we incorporated a modification to the algorithm, such that at every 100th step, we remove in accord with the Proposition 1 all the points $x \in \mathfrak{X}$ for which $\max_{i=1,\dots,12} f^T(x)\mathbf{Q}_i^{-1} f(x) < p$, where the matrices $\mathbf{Q}_i$ are defined in the Theorem 4. After each removal of points from the support of $\xi_n$, we standardized the remaining measure to 1.

At the Figure 1 we see which points were removed (grey dots) and kept (black dots) at each iteration. (The vertical axis corresponds to $\mathfrak{X}$ and the horizontal axis represents the number of iterations.) Notice that we finally arrived at a 3-point support, which is the smallest size possible. After the last deletion, the convergence of weights was very rapid, shortly arriving at the optimal design—the uniform probability on $\{1.4, 3.6, 5.0\}$—within the limits of the numerical precision of the software.

While the overall speed of convergence was observed to be higher using the deletion method, the number of iterations needed to obtain a given precision was almost the same as for the unmodified algorithm (except of the very final stage with the support identified by the deletion method exactly). The deletions usually remove points with very small weights, hence they do not change the quality of a design significantly.

## REFERENCES


[1] PÁZMAN, A.: *Foundations of Optimum Experimental Design*, Reidel, Dordrecht, 1986.

[2] SILVEY, S. D.: *Discussion on the papers by H. P. Wynn and P. J. Laycock*, J. Roy. Statist. Soc. **B 34** (1972), 174–175.

[3] TITTERINGTON, D. M.: *Algorithms for computing D-optimal designs on a finite design space*, in: Proceedings of the 1976 Conference on Information Science and Systems, John Hopkins University, Baltimore, 1976, pp. 213–216.

[4] WHITTLE, P.: *Some general points in the theory of optimal design*, J. Roy. Statist. Soc. **B 35** (1973), 123–130.


Received April 10, 2002

*Department of Probability and Statistics*
*Faculty of Mathematics, Physics and Informatics*
*Comenius University*
*Mlynská dolina*
*SK–842 48 Bratislava*
*SLOVAKIA*

*E-mail*: harman@fmph.uniba.sk