Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices

Radoslav Harman, Mária Trnovská
Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University Bratislava

Abstract

In the paper we solve the problem of $D_{\mathcal{H}}$-optimal design on a discrete experimental domain, which is formally equivalent to maximizing determinant on the convex hull of a finite number of positive semidefinite matrices. The problem of $D_{\mathcal{H}}$-optimality covers many special design settings, e.g. the D-optimal experimental design for regression models with grouped observations. For $D_{\mathcal{H}}$-optimal designs we prove several theorems generalizing known properties of standard D-optimality. Moreover, we show that $D_{\mathcal{H}}$-optimal designs can be numerically computed using a multiplicative algorithm, for which we give a proof of convergence. We illustrate the results on the problem of D-optimal augmentation of independent regression trials for the quadratic model on a rectangular grid of points in the plane.

Keywords: D-optimal design, grouped observations, D-optimal augmentation of trials, multiplicative algorithm

1 Introduction

Consider the standard homoscedastic linear regression model with uncorrelated observations Y satisfying $E(Y) = \beta^T f(t)$, where $\beta \in \mathbb{R}^m$ is an unknown vector of parameters and $f = (f_1, ..., f_m)^T$ is a vector of real-valued regression functions linearly independent on the experimental domain $\mathfrak{X} = \{x_1, ..., x_n\}$. For this model, constructing the D-optimal experimental design (see, e.g., monographs [4], or [5]) is equivalent to finding a vector of D-optimal weights, which is any solution w^* of the problem

$$\max\left\{\ln\det\left(\sum_{i=1}^{n} w_i f(x_i) f^T(x_i)\right)\middle|\, w \in \mathbb{S}_n\right\}, \tag{1}$$

where $\mathbb{S}_n$ is the unit simplex in $\mathbb{R}^n$:

$$\mathbb{S}_n = \left\{w \in \mathbb{R}^n : \sum_{i=1}^{n} w_i = 1, w_1, ..., w_n \geq 0\right\}.$$

In this paper we study a generalization of the problem (1) that can be used in a variety of less standard optimal design settings and, in the same time, exhibits similar theoretical properties as well as permits the use of efficient algorithms, such as a generalization of the multiplicative algorithm for the standard problem of D-optimality on a discrete experimental domain.

Notation: By the symbols $\mathcal{S}^m, \mathcal{S}^m_+$ and $\mathcal{S}^m_{++}$ we denote the set of all symmetric, positive semidefinite and positive definite matrices of type $m \times m$. The symbol $\succeq$ defines the Loewner partial ordering on $\mathcal{S}^m$, i.e., $\mathbf{A} \succeq \mathbf{B} \Leftrightarrow \mathbf{A} - \mathbf{B} \in \mathcal{S}^m_+$.

Let $\mathcal{H}$ be the convex hull of the set of nonzero positive semidefinite matrices $\mathbf{H}_1, ..., \mathbf{H}_n$ of type $m \times m$, such that

$$\mathcal{H} \cap \mathcal{S}^m_{++} \neq \emptyset, \tag{2}$$

i.e., $\mathcal{H}$ contains a regular matrix. Our aim is to find a vector $w^* = (w_1^*, ..., w_n^*)^T$ that solves the optimization problem

$$\max\left\{\ln\det\left(\sum_{i=1}^{n} w_i \mathbf{H}_i\right)\middle|\, w \in \mathbb{S}_n\right\}. \tag{3}$$

Note that there always exists a solution w^* of (3), and the matrix $\mathbf{M}_* = \sum_{i=1}^n w_i^* \mathbf{H}_i$ is unique and regular, which follows from compactness and convexity of $\mathcal{H}$, existence of a regular matrix in $\mathcal{H}$, and strict concavity of $\ln\det(\cdot)$ on $\mathcal{S}^m_{++}$. We will say that the vector w^* is a $D_\mathcal{H}$-optimal design, and its components $w_1^*, ..., w_n^*$ are $D_\mathcal{H}$-optimal weights corresponding to the elementary design matrices $\mathbf{H}_1, ..., \mathbf{H}_n$. The matrix $\mathbf{M}_*$ will be called the $D_\mathcal{H}$-optimal information matrix.

Clearly, the standard problem (1) is a special case of (3) with elementary information matrices corresponding to individual regression trials, i.e., $\mathbf{H}_i = f(x_i) f^T(x_i)$ for all $i = 1, ..., n$. However, (3) covers *all* problems of approximate D-optimal experimental design in which there exists a mapping $\mathbf{H} : \mathfrak{X} \to \mathcal{S}^m_+$, such that the information matrix corresponding to trials in points $t_1, ..., t_k \in \mathfrak{X}$ is equal to $\sum_{i=1}^k \mathbf{H}(t_i)$, that is in any model, in which the information matrix is additive.

For instance, in the regression model with grouped observations (see [4], Section II.5.3), the information matrix of a design $w \in \mathbb{S}^n$ is given by the formula

$$\mathbf{M}(w) = \sum_{i=1}^{n} w_i \mathbf{G}_i \mathbf{K}_i^{-1} \mathbf{G}_i^T,$$

where $\mathbf{G}_i$ is an $m \times r_i$ matrix and $\mathbf{K}_i$ is a regular $r_i \times r_i$ covariance matrix of the r_i-dimensional vector of observations corresponding to the trial in the point $x_i \in \mathfrak{X}$. Hence, D-optimality for the model with grouped observations is a problem of $D_{\mathcal{H}}$-optimality with the basic information matrices $\mathbf{H}_i = \mathbf{G}_i \mathbf{K}_i^{-1} \mathbf{G}_i^T$ for $i = 1, ..., n$. In Section 6 we demonstrate that $D_{\mathcal{H}}$-optimality covers also other design problems, such as D-optimal augmentation of independent regression trials.

2 Equivalence theorem for $D_{\mathcal{H}}$-optimal designs

Consider the function $\Phi : \mathcal{S}^m_+ \to \mathbb{R} \cup \{-\infty\}$ defined $\Phi(\mathbf{M}) = \ln \det(\mathbf{M})$ for $\mathbf{M} \in \mathcal{S}^m_{++}$ and $\Phi(\mathbf{M}) = -\infty$ otherwise. We will call the function Φ the "criterion of D-optimality". It is well known that Φ is a concave function and the gradient in $\mathbf{M} \in \mathcal{S}^m_{++}$ is $\nabla\Phi(\mathbf{M}) = \mathbf{M}^{-1}$, see, e.g., [4], Section IV.2.1. Hence the directional derivative in $\mathbf{M}$ along the direction $\mathbf{H} - \mathbf{M}$, where $\mathbf{H} \in \mathcal{S}^m$, is

$$\partial\Phi(\mathbf{M}; \mathbf{H} - \mathbf{M}) = \mathrm{tr}\left(\nabla\Phi(\mathbf{M})(\mathbf{H} - \mathbf{M})\right) = \mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}) - m.$$

Clearly, a matrix $\mathbf{M}_*$ maximizes Φ on $\mathcal{H}$ iff $\mathbf{M}_*$ is regular and $\partial\Phi(\mathbf{M}_*, \mathbf{H}_i - \mathbf{M}_*) \leq 0$ for all $i = 1, ..., n$, which is equivalent to

$$\mathrm{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) \leq m \text{ for all } i = 1, ..., n.$$

That is, the optimization problem:

$$\min \left\{ \max_{i=1,...,n} \mathrm{tr}\left(\mathbf{M}^{-1}\mathbf{H}_i\right) \middle| \mathbf{M} = \sum_{i=1}^{n} w_i \mathbf{H}_i, w \in \mathbb{S}^n \right\} \tag{4}$$

has the optimal value less or equal to m, and $\mathbf{M}_* = \sum_{i=1}^n w_i^* \mathbf{H}_i$, where w_i^* are components of an optimal solution of (4), maximizes the determinant on the set $\mathcal{H}$. Moreover, notice that

$$m \geq \max_{i=1,\dots,n} \operatorname{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) \geq \sum_{i=1}^{n} w_i^* \operatorname{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i)$$

$$= \operatorname{tr}\left(\mathbf{M}_*^{-1} \sum_{i=1}^{n} w_i^* \mathbf{H}_i\right) = m, \tag{5}$$

that is $\max_{i=1,\dots,n} \operatorname{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) = m$, which means that the optimal value of the problem (4) is *exactly* m. Therefore, we know the optimal value of the problem (4), although we do not know the point at which it is attained. Moreover, from (5) we see that a coefficient w_i^* of the optimal convex combination is nonzero only if $\operatorname{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) = m$.

We obtain a generalization of the famous Kiefer-Wolfowitz theorem of equivalence between the problems of D- and G-optimality (see [3], or [4] Section IV.2.4):

Theorem 1. *Let $w^* \in \mathbb{S}^n$. Then the following three statements are equivalent:*

(i) *w^* is a $D_{\mathcal{H}}$-optimal design;*

(ii) *w^* is a solution of the problem (4);*

(iii) $\max_{i=1,\dots,n} \operatorname{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) = m$, *where* $\mathbf{M}_* = \sum_{i=1}^{n} w_i^* \mathbf{H}_i$.

Note also that for any information matrix $\mathbf{M} \in \mathcal{H}$ we have

$$\Phi(\mathbf{M}_*) - \Phi(\mathbf{M}) \leq \operatorname{tr}((\mathbf{M}_* - \mathbf{M})\nabla\Phi(\mathbf{M}))$$

$$= \operatorname{tr}(\mathbf{M}_*\mathbf{M}^{-1}) - m = \sum_{i=1}^{n} w_i^* \operatorname{tr}\left(\mathbf{H}_i\mathbf{M}^{-1}\right) - m \leq \varepsilon_{\mathbf{M}}, \tag{6}$$

where

$$\varepsilon_{\mathbf{M}} = \max_{i=1,\dots,n} \operatorname{tr}(\mathbf{H}_i\mathbf{M}^{-1}) - m. \tag{7}$$

By Theorem 1, if $\mathbf{M}$ approaches $\mathbf{M}_*$, then $\varepsilon_{\mathbf{M}}$ converges to 0. Therefore, inequality (6) can be used to control convergence of $D_{\mathcal{H}}$-optimal design algorithms; cf. Section 5.

3 Bounds for $D_{\mathcal{H}}$-optimal weights

A well known fact in the theory of D-optimal design is that the components of the vector of D-optimal weights are bounded by $1/m$ from above (see, e.g., [5] Section 8.12). It turns out that in the case of general $D_{\mathcal{H}}$-optimality, any optimal weight w_i^* satisfies constraints determined by the rank of the corresponding matrix $\mathbf{H}_i$:

Theorem 2. *Let w^* be a $D_{\mathcal{H}}$-optimal design. Then*

$$w_i^* \leq \frac{\mathrm{rank}(\mathbf{H}_i)}{m} \textit{ for all } i = 1, ..., n.$$

Proof. Let $\mathbf{M}_* = \sum_{i=1}^n w_i^* \mathbf{H}_i$ be the $D_{\mathcal{H}}$-optimal information matrix, where w^* is a vector of $D_{\mathcal{H}}$-optimal weights. Fix a single index $i \in \{1, ..., n\}$, such that $w_i^* > 0$ and denote

$$\mathbf{N}_i = \mathbf{M}_*^{-\frac{1}{2}} \mathbf{H}_i \mathbf{M}_*^{-\frac{1}{2}}.$$

From Theorem 1 we know that $\mathrm{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) = \mathrm{tr}(\mathbf{N}_i) = m$, which means that the sum of the eigenvalues of $\mathbf{N}_i$ is m. In the same time, the number of nonzero eigenvalues of $\mathbf{N}_i$ cannot exceed $\mathrm{rank}(\mathbf{N}_i) = \mathrm{rank}(\mathbf{H}_i)$. Thus, there exists an eigenvalue λ of the matrix $\mathbf{N}_i$ such that

$$\lambda \geq m(\mathrm{rank}(\mathbf{H}_i))^{-1}.$$

Since λ is the eigenvalue of $\mathbf{N}_i$ we have $\det(\mathbf{N}_i - \lambda \mathbf{I}_m) = 0$, from which we obtain

$$0 = \det(\mathbf{H}_i - \lambda \mathbf{M}_*).$$

In other words $\mathbf{M}_* - \lambda^{-1}\mathbf{H}_i$ is singular. We will show that this implies $w_i^* \leq \lambda^{-1}$. Assume the converse, that is $w_i^* > \lambda^{-1}$. Then

$$\begin{aligned} \mathbf{M}_* - \lambda^{-1}\mathbf{H}_i &= \sum_{j \neq i} w_j^* \mathbf{H}_j + (w_i^* - \lambda^{-1})\mathbf{H}_i \\ = \sum_{j \neq i} w_j^* \mathbf{H}_j + \frac{w_i^* - \lambda^{-1}}{w_i^*} w_i^* \mathbf{H}_i &\succeq \frac{w_i^* - \lambda^{-1}}{w_i^*} \sum_{j=1}^n w_j^* \mathbf{H}_j \\ &= \frac{w_i^* - \lambda^{-1}}{w_i^*} \mathbf{M}_* \in \mathcal{S}_{++}^m, \end{aligned}$$

which is a contradiction with singularity of $\mathbf{M}_* - \lambda^{-1}\mathbf{H}_i$. Therefore $w_i^* \leq \lambda^{-1} \leq \mathrm{rank}(\mathbf{H}_i) m^{-1}$. □

As a straightforward corollary of the previous theorem we obtain that if $\sum_{i=1}^n \mathrm{rank}(\mathbf{H}_i) = m$, then the $D_{\mathcal{H}}$-optimal design is simply a vector w^* with components $w_i^* = \mathrm{rank}(\mathbf{H}_i)/m$.

4 Identification of zero weights of $D_{\mathcal{H}}$-optimal designs

In this section we formulate a direct generalization the method from the paper [2], which allows us to use any regular information matrix

$\mathbf{M} = \sum_{i=1}^{n} w_i \mathbf{H}_i$ to identify indices $j \in \{1, ..., n\}$, such that $w_j^* = 0$ for any $D_{\mathcal{H}}$-optimal design w^*.

Similarly as in the paper [2], let $\mathbf{N} = \mathbf{M}^{-\frac{1}{2}} \mathbf{M}_* \mathbf{M}^{-\frac{1}{2}}$, where $\mathbf{M}_*$ is the $D_{\mathcal{H}}$-optimal information matrix, and let $0 < \lambda_1 \leq ... \leq \lambda_m$ denote the eigenvalues of $\mathbf{N}$. Let w^* be any $D_{\mathcal{H}}$-optimal design and let $j \in \{1, ..., n\}$, such $w_j^* > 0$. Set $\mathbf{Y}_j = \mathbf{N}^{-\frac{1}{2}} \mathbf{M}^{-\frac{1}{2}} \mathbf{H}_j^{\frac{1}{2}}$. We have

$$\begin{aligned} \mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_j) &= \mathrm{tr}(\mathbf{N}\mathbf{Y}_j\mathbf{Y}_j^T) \geq \lambda_1 \mathrm{tr}(\mathbf{Y}_j\mathbf{Y}_j^T) \\ &= \lambda_1 \mathrm{tr}(\mathbf{M}_*^{-1}\mathbf{H}_j) = \lambda_1 m, \end{aligned} \tag{8}$$

$$\begin{aligned} \sum_{i=1}^{m} \lambda_i^{-1} &= \mathrm{tr}(\mathbf{N}^{-1}) = \mathrm{tr}(\mathbf{M}_*^{-1}\mathbf{M}) \\ &= \sum_{i=1}^{n} w_i \mathrm{tr}(\mathbf{M}_*^{-1}\mathbf{H}_i) \leq m, \end{aligned} \tag{9}$$

$$\begin{aligned} \sum_{i=1}^{m} \lambda_i &= \mathrm{tr}(\mathbf{N}) = \mathrm{tr}(\mathbf{M}_*\mathbf{M}^{-1}) \\ &= \sum_{i=1}^{n} w_i^* \mathrm{tr}(\mathbf{H}_i\mathbf{M}^{-1}) \leq m + \varepsilon_{\mathbf{M}}, \end{aligned} \tag{10}$$

where $\varepsilon_{\mathbf{M}}$ is defined by (7). Using identical methods as in the paper [2], we can show that inequalities (9) and (10) imply

$$\lambda_1 \geq 1 + \frac{\varepsilon_{\mathbf{M}}}{2} - \frac{\sqrt{\varepsilon_{\mathbf{M}}(4 + \varepsilon_{\mathbf{M}} - 4/m)}}{2}, \tag{11}$$

which together with inequality (8) yields:

Theorem 3. *Let $w \in \mathbb{S}_n$ be any design such that $\mathbf{M} = \sum_{i=1}^{n} w_i \mathbf{H}_i$ is regular, let w^* be any $D_{\mathcal{H}}$-optimal design, and let $j \in \{1, ..., n\}$ be such that $w_j^* > 0$. Then*

$$\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_j) \geq m \left[1 + \frac{\varepsilon_{\mathbf{M}}}{2} - \frac{\sqrt{\varepsilon_{\mathbf{M}}(4 + \varepsilon_{\mathbf{M}} - 4/m)}}{2} \right]. \tag{12}$$

Therefore, using any regular information matrix $\mathbf{M}$ we can remove all the basic information matrices $\mathbf{H}_j$ for which the inequality (12) is not satisfied, since corresponding weights of the $D_{\mathcal{H}}$-optimal design must be 0. This can help us significantly speed up computation of a $D_{\mathcal{H}}$-optimal design (cf. [2]).

5 A multiplicative algorithm for constructing $D_{\mathcal{H}}$-optimal designs

As a simple numerical method for calculating $D_{\mathcal{H}}$-optimal designs, we will formulate a generalization of the Titterington-Torsney multiplicative algorithm (see, e.g., [6], [7]).

Let $w^{(0)} = (w_1^{(0)}, ..., w_n^{(0)})^T$ be an initial design such that $w_i^{(0)} > 0$ for all $i = 1, ..., n$. Based on the design $w^{(j)} = (w_1^{(j)}, ..., w_n^{(j)})^T$, $j \geq 0$, and its information matrix $\mathbf{M}_j = \sum_{i=1}^n w_i^{(j)} \mathbf{H}_i$, we can construct the new vector $w^{(j+1)} = (w_1^{(j+1)}, ..., w_n^{(j+1)})^T$ using the formula:

$$w_i^{(j+1)} = m^{-1} \mathrm{tr}(\mathbf{M}_j^{-1} \mathbf{H}_i) w_i^{(j)} \text{ for all } i = 1, ..., n. \tag{13}$$

(Note that (2) and positivity of $w_i^{(j)}$ implies regularity of $\mathbf{M}_j$.) Clearly

$$\sum_{i=1}^n w_i^{(j+1)} = \frac{1}{m} \mathrm{tr} \left(\mathbf{M}_j^{-1} \sum_{i=1}^n w_i^{(j)} \mathbf{H}_i \right) = 1,$$

that is $w^{(j+1)}$ is also a design. Note that the algorithm is computationally very rapid, since it calculates the inverse of an $m \times m$ matrix only once per iteration, with m being usually small (less than 10 in most optimal design problems). The speed of calculation is more influenced by the number n of support matrices, but the number of candidate support matrices can be significantly reduced during the calculation using the technique of Section 4.

In the following, we will prove that the multiplicative algorithm produces a sequence $\left(w^{(j)}\right)_{j=1}^{\infty}$ of designs that converges to the $D_{\mathcal{H}}$-optimal design in the sense that $\lim_{j \to \infty} \det(\mathbf{M}_j) = \det(\mathbf{M}_*)$, where $\mathbf{M}_*$ is the $D_{\mathcal{H}}$-optimal information matrix. The proof of convergence of the multiplicative algorithm for the standard D-optimality has been based on a technique of conditional expectations, see [4], Section V.3. For the proof of convergence of the multiplicative algorithm for general $D_{\mathcal{H}}$-optimality, we will use a different approach based on the following lemmas:

Lemma 1. *(Theorem 6.10 of [8], or Theorem IX.5.11. in [1])*
If

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \succeq 0$$

and all of the blocks are square matrices of the same size, then

$$(\det \mathbf{A}_{12})^2 \leq \det \mathbf{A}_{11} \det \mathbf{A}_{22}.$$

Lemma 2. *(Problem 28, Section 6.2 of [8])*
Let $\mathbf{A}, \mathbf{B}$ *be positive semidefinite. Then* $\det(\mathbf{A}+\mathbf{B}) \geq \det \mathbf{A}$ *and the equality occurs if and only if* $\mathbf{A}+\mathbf{B}$ *is singular or* $\mathbf{B} = 0$.

Theorem 4. *For the algorithm defined by (13) it holds* $\det(\mathbf{M}_j) \leq \det(\mathbf{M}_{j+1})$ *for all* $j \geq 0$*, with optimality if and only if* $\mathbf{M}_j = \mathbf{M}_*$*. Moreover* $\lim_{j\to\infty} \det(\mathbf{M}_j) = \det(\mathbf{M}_*)$.

Proof. Let $\mathbf{M} = \sum_{i=1}^n w_i \mathbf{H}_i$ for some vector w of positive weights and let $\mathbf{M}^+ = \sum_{i=1}^n w_i^+ \mathbf{H}_i$, where w^+ is the vector of weights obtained from w by one step of the multiplicative algorithm, i.e. $w^+ = m^{-1}\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i) w_i$ for all $i = 1, ..., n$. Define $\alpha_i = \mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i)/m = w_i^+/w_i$ and $\tilde{\mathbf{M}} = \sum_{i=1}^n \frac{1}{\alpha_i} w_i \mathbf{H}_i$, and note that $\mathbf{M}^+ = \sum_{i=1}^n \alpha_i w_i \mathbf{H}_i$. Clearly

$$\begin{pmatrix} \mathbf{M}^+ & \mathbf{M} \\ \mathbf{M} & \tilde{\mathbf{M}} \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} \alpha_i w_i \mathbf{H}_i & w_i \mathbf{H}_i \\ w_i \mathbf{H}_i & \frac{1}{\alpha_i} w_i \mathbf{H}_i \end{pmatrix}$$

$$= \sum_{i=1}^n \begin{pmatrix} \sqrt{\alpha_i w_i}\mathbf{H}_i^{\frac{1}{2}} & \sqrt{\frac{w_i}{\alpha_i}}\mathbf{H}_i^{\frac{1}{2}} \end{pmatrix}^T \begin{pmatrix} \sqrt{\alpha_i w_i}\mathbf{H}_i^{\frac{1}{2}} & \sqrt{\frac{w_i}{\alpha_i}}\mathbf{H}_i^{\frac{1}{2}} \end{pmatrix} \succeq 0. \quad (14)$$

From (14) and Lemma 1 it immediately follows that

$$\det^2(\mathbf{M}) \leq \det(\mathbf{M}^+) \det(\tilde{\mathbf{M}}). \quad (15)$$

Let λ_i be the $i-$th eigenvalue of the matrix $\mathbf{M}^{-1}\tilde{\mathbf{M}}$. Using the inequality between geometric and arithmetic means we obtain

$$\det^{\frac{1}{m}}(\mathbf{M}^{-1}\tilde{\mathbf{M}}) = \prod_{i=1}^m \lambda_i^{\frac{1}{m}} \leq \frac{1}{m}\sum_{i=1}^m \lambda_i = \frac{1}{m}\mathrm{tr}(\mathbf{M}^{-1}\tilde{\mathbf{M}})$$

$$= \sum_{i=1}^n \frac{w_i}{\alpha_i}\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i) = \sum_{i=1}^n w_i = 1. \quad (16)$$

Consequently

$$\det(\tilde{\mathbf{M}}) \leq \det(\mathbf{M}), \quad (17)$$

which together with (15) gives the required inequality

$$\det(\mathbf{M}) \leq \det(\mathbf{M}^+). \quad (18)$$

To prove the second part of the theorem, assume that

$$\det(\mathbf{M}) = \det(\mathbf{M}^+). \quad (19)$$

From (15), (17) and (19) it follows, that

$$\det(\mathbf{M}) = \det(\mathbf{M}^+) = \det(\tilde{\mathbf{M}}).$$

Therefore we have equality in (16), which implies that the eigenvalues λ_i of the matrix $\mathbf{M}^{-1}\tilde{\mathbf{M}}$ are all equal. Moreover, since $\det(\mathbf{M}^{-1}\tilde{\mathbf{M}}) = 1$ and the matrices $\mathbf{M}^{-1}, \tilde{\mathbf{M}}$ are positive definite, we have $\mathbf{M}^{-1}\tilde{\mathbf{M}} = \mathbf{I}$, i.e. $\mathbf{M} = \tilde{\mathbf{M}}$. Using this fact, (14), and properties of the Schur complement we obtain

$$\mathbf{M}^+ - \mathbf{M} = \mathbf{M}^+ - \mathbf{M}\tilde{\mathbf{M}}^{-1}\mathbf{M} \succeq 0.$$

Hence we can apply Lemma 2 with $\mathbf{A} = \mathbf{M}$ and $\mathbf{B} = \mathbf{M}^+ - \mathbf{M}$, which, together with (19) and the positive definiteness of $\mathbf{M}^+$, implies $\mathbf{M}^+ = \mathbf{M}$, i.e.

$$\sum_{i=1}^{n} \alpha_i w_i \mathbf{H}_i = \sum_{i=1}^{n} w_i \mathbf{H}_i.$$

By multiplying both sides of this equality by $\frac{1}{m}\mathbf{M}^{-1}$ and by taking the trace we have

$$\sum_{i=1}^{n} \alpha_i w_i \frac{\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i)}{m} = \sum_{i=1}^{n} w_i \frac{\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i)}{m}$$

that is $\sum_{i=1}^{n} \alpha_i^2 w_i = \sum_{i=1}^{n} \alpha_i w_i$. Since $\sum_{i=1}^{n} \alpha_i w_i = \sum_{i=1}^{n} w_i^+ = 1$, we have

$$\Big(\sum_{i=1}^{n} \alpha_i w_i\Big)^2 = \sum_{i=1}^{n} \alpha_i^2 w_i. \tag{20}$$

The equality condition of the weighted Cauchy-Schwarz inequality together with (20) implies

$$\alpha_i = \frac{\mathrm{tr}(\mathbf{M}^{-1}\mathbf{H}_i)}{m} = 1 \text{ for all } i = 1, ..., n$$

Therefore $\mathbf{M}$ must be $D_{\mathcal{H}}$-optimal by Theorem (1).

The last statement of the theorem can be proved using the same arguments as in the proof of Proposition V.6 in [4], that is, using compactness of the space $\mathbb{S}_n$ of weights and monotonicity $\det(\mathbf{M}_j) \leq \det(\mathbf{M}_{j+1})$, which is strict, unless $\mathbf{M}_j$ is optimal. $\square$

6 Example

In this section we will demonstrate a technique of how to apply our results to the problem of D-optimal approximate augmentation of a set of regression trials. Consider the quadratic regression with independent responses Y modeled by

$$\begin{aligned} E(Y) &= \beta_1 + \beta_2 u + \beta_3 v + \beta_4 u^2 + \beta_5 v^2 + \beta_6 uv \\ &= f^T(u,v)\beta = (1, u, v, u^2, v^2, uv)(\beta_1, ..., \beta_6)^T, \end{aligned}$$

where the design point $(u,v)^T$ belongs to the experimental domain $\mathfrak{X} = \{x_1, ..., x_n\}$, corresponding to the $n = 25$ point equispaced discrete grid in the square $\mathbb{I}^2 = [-1,1] \times [-1,1]$, that is

$$x_{5(i-1)+j} = \left(\frac{i-3}{2}, \frac{j-3}{2}\right)^T, \text{ for } i,j \in \{1, ..., 5\}.$$

Assume that we have already performed k trials uniformly on $\mathfrak{X}$, i.e., we have performed $k/n \in \mathbb{N}$ trials in each design point (for instance in order to verify validity of the model). Our aim is to perform $\gamma k \in \mathbb{N}$ additional trials in a way that maximizes determinant of the final information matrix.

Let $\mathbf{M}(x) = f(x)f^T(x)$ for all $x \in \mathfrak{X}$. In accord with the methodology of approximate design of experiments, we shall solve the problem

$$\mathrm{argmax}_{w \in S_n} \ln\det\left(\sum_{j=1}^n \frac{k}{n}\mathbf{M}(x_j) + \sum_{i=1}^n \gamma k w_i \mathbf{M}(x_i)\right), \tag{21}$$

where w_i is the proportion of the γk additional trials to be performed in x_i. The problem (21) is clearly equivalent to the problem of $D_{\mathcal{H}}$-optimality with

$$\mathbf{H}_i = \sum_{j=1}^n \frac{1}{n}\mathbf{M}(x_j) + \gamma \mathbf{M}(x_i), \text{ for } i = 1, ..., n.$$

Due to symmetries of the quadratic model and the experimental domain $\mathfrak{X}$, the D-optimal augmentation design is supported on at most 9 points, with optimal weights $w_1^* = w_5^* = w_{21}^* = w_{25}^*$ corresponding to the vertices of $\mathbb{I}^2$, optimal weights $w_3^* = w_{11}^* = w_{15}^* = w_{23}^*$ corresponding to the midpoints of edges of $\mathbb{I}^2$, and an optimal weight w_{13}^* corresponding to the central point $(0,0)$. Figure 1 exhibits dependence of the optimal weights on the augmentation factor γ. Notice that if we add less than about 50% of trials, then the optimal method is to perform them only in the vertices of the square, and if we add between around 50% and 150% of trials, then we should only perform them in the vertices and edge midpoints of the square. Naturally, for $\gamma \to \infty$ the initial phase of experimentation becomes negligible and the D-optimal augmentation design converges to the standard D-optimal design.

References

[1] Bhatia, R., 1996. Matrix Analysis (Graduate Texts in Mathematics). Springer.

[2] Harman, R., Pronzato, L., 2007. Improvements on removing nonoptimal support points in D-optimum design algorithms. Statistics & Probability Letters 77, 90-94

[3] Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. Canadian Journal of Mathematics 12, 363–366.

[4] Pázman, A., 1986. Foundations of optimum experimental design. Reidel.

[5] Pukelsheim, F., 1993. Optimal design of experiments, John Wiley & Sons.

[6] Titterington, D., 1976. Algorithms for computing D-optimal designs on a finite design space. In: Proceedings of the 1976 Conference on Information Science and Systems. Department of Electronic Engineering, John Hopkins University, Baltimore, 213-216.

[7] Torsney, B., 1983. A moment inequality and monotonicity of an algorithm. In: Kortanek, K., Fiacco, A. (Eds.), Proceedings of the International Symposium on Semi-Infinite Programming and Applications. Springer, Heidelberg, pp. 249-260.

[8] Zhang, F., 1999. Matrix Theory, Basic Results and Techniques, Springer Verlag.

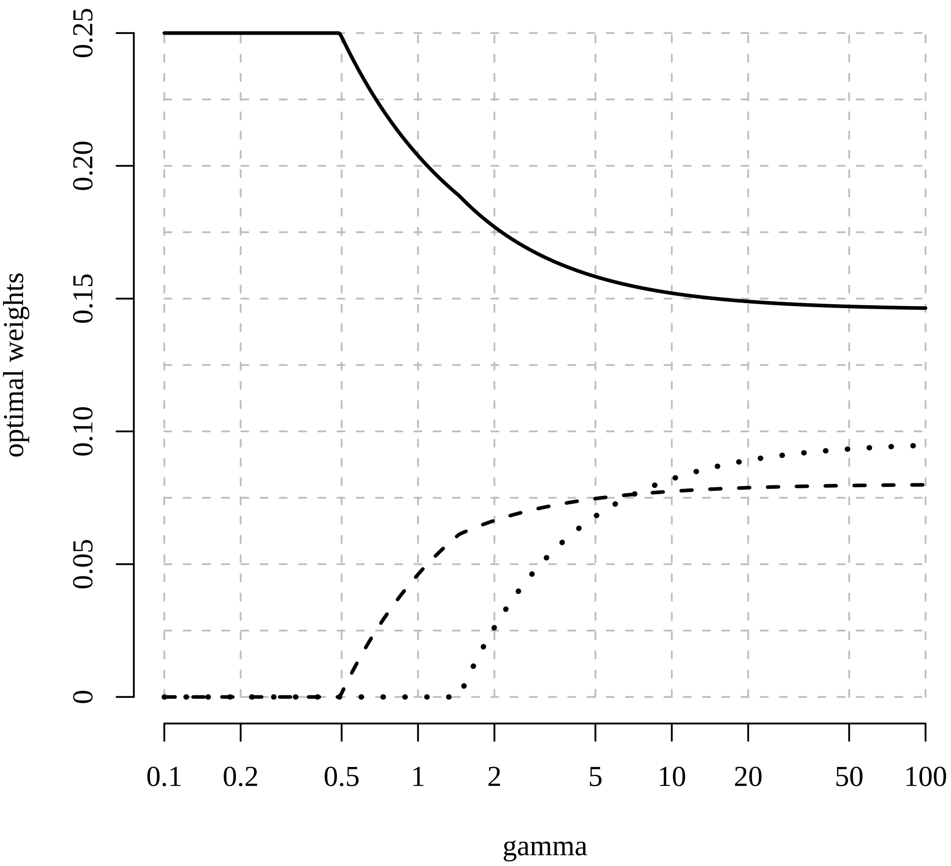

Figure 1: Optimal weights depending on the augmentation factor γ corresponding to the support points at the vertices of the square $\mathbb{I}^2$ (solid line), midpoints of the edges of the square $\mathbb{I}^2$ (dashed line) and the central point $(0, 0)$ (dotted line).