# Improvements on removing non-optimal support points in $D$-optimum design algorithms

Radoslav Harman [a,1] and Luc Pronzato [b,*,2]

[a] *Department of Applied Mathematics and Statistics*
*Mlynská Dolina, 84248 Bratislava, Slovakia*

[b] *Laboratoire I3S, CNRS/UNSA, Bât. Euclide, Les Algorithmes,*
*2000 route des Lucioles, BP 121, 06903 Sophia-Antipolis Cedex, France*

---

**Abstract**

We improve the inequality used in (Pronzato, 2003) to remove points from the design space during the search for a $D$-optimum design, which can be used to accelerate design algorithms. Let $\xi$ be any design on a compact space $\mathcal{X} \subset \mathbb{R}^m$ with a nonsingular information matrix, and let $m + \epsilon$ be the maximum of the variance function $d(\xi, \mathbf{x})$ over all $\mathbf{x} \in \mathcal{X}$. We prove that any support point $\mathbf{x}_*$ of a $D$-optimum design on $\mathcal{X}$ must satisfy the inequality $d(\xi, \mathbf{x}_*) \geq m(1 + \epsilon/2 - \sqrt{\epsilon(4 + \epsilon - 4/m)}/2)$. We show that this new lower bound on $d(\xi, \mathbf{x}_*)$ is, in a sense, the best possible.

*Key words:* $D$-optimum design, design algorithm, support points
*PACS:* 62K05, 90C46

---

## 1 Introduction

We consider the same settings as in (Pronzato, 2003). Let $\mathcal{X} \subseteq \mathbb{R}^m$ be a compact design space and let $\Xi$ be the set of all designs (i.e., finitely supported

* author for correspondence
*Email addresses:* `harman@fmph.uniba.sk` (Radoslav Harman), `pronzato@i3s.unice.fr` (Luc Pronzato).

[1] The work of the first author was supported by the VEGA grant No. 1/0264/03 of the Slovak Agency.

[2] The work of the second author was partially supported by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors's view.

probability measures) on $\mathcal{X}$. For any $\xi \in \Xi$, let

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathbf{x}\mathbf{x}^\top \, \xi(\mathbf{dx})$$

denote the information matrix. Suppose that there exists a design with nonsingular information matrix and let $\Xi^+$ be the set of such designs. Let $\xi^*$ denote a $D$-optimum design, that is, a measure in $\Xi$ that maximizes $\det \mathbf{M}(\xi)$, see, e.g., (Fedorov, 1972). Note that a $D$-optimum design always exists and that the $D$-optimum information matrix $\mathbf{M}_* = \mathbf{M}(\xi^*)$ is unique. For any $\xi \in \Xi^+$ denote $d(\xi, \cdot) : \mathcal{X} \to [0, \infty)$ the variance function defined by

$$d(\xi, \mathbf{x}) = \mathbf{x}^\top \mathbf{M}^{-1}(\xi)\mathbf{x} \, .$$

The celebrated Kiefer-Wolfowitz Equivalence Theorem (1960) writes as follows.

**Theorem 1** *The following three statements are equivalent:*

***(i)*** $\xi^*$ *is $D$-optimum;*
***(ii)*** $\max_{\mathbf{x}\in\mathcal{X}} d(\xi^*, \mathbf{x}) = m$*;*
***(iii)*** $\xi^*$ *minimizes* $\max_{\mathbf{x}\in\mathcal{X}} d(\xi, \mathbf{x})$, $\xi \in \Xi$.

Notice that

$$\int_{\mathcal{X}} d(\xi^*, \mathbf{x}) \, \xi^*(d\mathbf{x}) = \int_{\mathcal{X}} \mathbf{x}^\top \mathbf{M}_*^{-1} \mathbf{x} \, \xi^*(d\mathbf{x}) = \mathrm{trace}(\mathbf{M}_* \mathbf{M}_*^{-1}) = m \, .$$

Hence, (ii) of Theorem 1 implies that for any support point $\mathbf{x}_*$ of the design $\xi^*$ (i.e., for a point satisfying $\xi^*(\mathbf{x}_*) > 0$), we have

$$d(\xi^*, \mathbf{x}_*) = m \, . \tag{1}$$

In the next section we show that the equality (1) can be used to prove that

$$\forall \xi \in \Xi^+ \, , \;\; d(\xi, \mathbf{x}_*) \geq m \lambda_1^*(\xi)$$

where $\lambda_1^*$ depends on $\xi$ only via the maximum of $d(\xi, \cdot)$ over the design space $\mathcal{X}$. Hence, we can test candidate support points by using any finite number of design measures $\xi \in \Xi^+$, e.g., those that are generated by a design algorithm on its way towards the optimum: any point that does not pass the test defined by $\xi^k$ of iteration $k$ need not be considered for further investigations and can thus be removed from the design space.

## 2 A necessary condition for candidate support points

For $\xi$ a design in $\Xi^+$ denote $\mathbf{M} = \mathbf{M}(\xi)$,

$$\mathbf{H} = \mathbf{M}^{-1/2}\mathbf{M}_*\mathbf{M}^{-1/2}$$

and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ the eigenvalues of $\mathbf{H}$. Notice that $\lambda_1 > 0$ and that the eigenvalues depend on the design $\xi$ as well as on the $D$-optimum information matrix $\mathbf{M}_*$. Let $\mathbf{x}_*$ be a support point of a $D$-optimum design and let $\mathbf{y}_* = \mathbf{H}^{-1/2}\mathbf{M}^{-1/2}\mathbf{x}_*$. The equality (1) can be written in the form $\mathbf{y}_*^\top \mathbf{y}_* = m$ which implies:

$$d(\xi, \mathbf{x}_*) = \mathbf{x}_*^\top \mathbf{M}^{-1}\mathbf{x}_* = \mathbf{y}_*^\top \mathbf{H}\mathbf{y}_* \geq \lambda_1 \mathbf{y}_*^\top \mathbf{y}_* = m\lambda_1 \,. \tag{2}$$

To be able to use the inequality (2), we need to derive a lower bound $\lambda_1^*$ on $\lambda_1$ that does not depend on the unknown matrix $\mathbf{M}_*$.

The inequality (ii) of Theorem (1) implies

$$\begin{aligned}\sum_{i=1}^m \lambda_i^{-1} &= \mathrm{trace}(\mathbf{H}^{-1})\\ &= \mathrm{trace}(\mathbf{M}_*^{-1}\mathbf{M}) = \int_{\mathcal{X}} \mathbf{x}^\top \mathbf{M}_*^{-1}\mathbf{x}\, \xi(d\mathbf{x}) = \int_{\mathcal{X}} d(\xi^*, \mathbf{x})\, \xi(d\mathbf{x}) \leq m \,.\end{aligned}$$

Also,

$$\begin{aligned}\sum_{i=1}^m \lambda_i &= \mathrm{trace}(\mathbf{H})\\ &= \mathrm{trace}(\mathbf{M}_*\mathbf{M}^{-1}) = \int_{\mathcal{X}} \mathbf{x}^\top \mathbf{M}^{-1}\mathbf{x}\, \xi^*(d\mathbf{x}) \leq \max_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^\top \mathbf{M}^{-1}\mathbf{x} = m + \epsilon \,,\end{aligned}$$

where we used the notation

$$\epsilon = \max_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^\top \mathbf{M}^{-1}\mathbf{x} - m \geq 0 \,.$$

For $m = 1$ we directly obtain the lower bound $\lambda_1 \geq \lambda_1^* = 1$. For $m > 1$, the Lagrangian for the minimisation of $\lambda_1$ subject to $\sum_{i=1}^m \lambda_i^{-1} \leq m$ and $\sum_{i=1}^m \lambda_i \leq m + \epsilon$ is given by

$$\mathcal{L}(\lambda, \mu_1, \mu_2) = \lambda_1 + \mu_1 \left( \sum_{i=1}^m \lambda_i^{-1} - m \right) + \mu_2 \left( \sum_{i=1}^m \lambda_i - m - \epsilon \right)$$

with $\lambda = (\lambda_1, ..., \lambda_m)^\top$, $\mu_1, \mu_2 \geq 0$. The stationarity of $\mathcal{L}(\lambda, \mu_1, \mu_2)$ with respect to the $\lambda_i$'s and the Kuhn-Tucker conditions

$$\mu_1 \left( \sum_{i=1}^{m} \lambda_i^{-1} - m \right) = 0,\ \mu_2 \left( \sum_{i=1}^{m} \lambda_i - m - \epsilon \right) = 0$$

give $\lambda_i = L$ for $i = 2, ..., m$, with $\lambda_1$ and $L$ satisfying

$$\begin{aligned} &\lambda_1^{-1} + (m-1)L^{-1} = m \\ &\lambda_1 + (m-1)L = m + \epsilon\,. \end{aligned}$$

The solution is thus

$$\lambda_1^* = 1 + \frac{\epsilon}{2} - \frac{\sqrt{\epsilon(4 + \epsilon - 4/m)}}{2} \leq 1 \tag{3}$$

and $\lambda_i^* = L^* = (m-1)/(m - 1/\lambda_1^*) \geq 1$, $i = 2, ..., m$. Notice that the bound (3) gives $\lambda_1^* = 1$ when $m = 1$ and can thus be used for any dimension $m \geq 1$. By substituting $\lambda_1^*$ for $\lambda_1$ in (2) we obtain the following result.

**Theorem 2** *For any design $\xi \in \Xi^+$, any point $\mathbf{x}_* \in \mathcal{X}$ such that*

$$d(\xi, \mathbf{x}_*) < h_m(\epsilon) = m \left[ 1 + \frac{\epsilon}{2} - \frac{\sqrt{\epsilon(4 + \epsilon - 4/m)}}{2} \right] \tag{4}$$

*where $\epsilon = \max_{\mathbf{x} \in \mathcal{X}} d(\xi, \mathbf{x}) - m$, cannot be support point of a D-optimum design measure.*

The inequality in (Pronzato, 2003) uses

$$\tilde{h}_m(\epsilon) = m \left[ 1 + \frac{\epsilon}{2} - \frac{\sqrt{\epsilon(4 + \epsilon)}}{2} \right]\,. \tag{5}$$

Notice, that $m \geq h_m(\epsilon) > \tilde{h}_m(\epsilon)$ for all integer $m \geq 1$ and all $\epsilon > 0$, and that $\lim_{\epsilon \to \infty} h_m(\epsilon) = 1$ while $\lim_{\epsilon \to \infty} \tilde{h}_m(\epsilon) = 0$. The new bound is thus always stronger, especially for large values of $\epsilon$, i.e. when the design $\xi$ is far from being optimum. When $m = 1$, $h_1(\epsilon) = 1$ for any $\epsilon > 0$ and the inequality (4) cannot be improved. When $m \geq 2$, $h_m(\epsilon)$ is the tightest lower bound on the variance function $d(\xi, \mathbf{x}_*)$ at a $D$-optimal support point $\mathbf{x}_*$ that depends only on $m$ and $\epsilon$, in the sense of the following theorem.

**Theorem 3** *For any integer $m \geq 2$ and any $\epsilon, \delta > 0$ there exist a compact design space $\mathcal{X} \subset \mathbb{R}^m$, a design $\xi$ on $\mathcal{X}$ and a point $\mathbf{x}_* \in \mathcal{X}$ supporting a*

*$D$-optimum design on $\mathcal{X}$ such that $\epsilon = \max_{\mathbf{x}\in\mathcal{X}} d(\xi, \mathbf{x}) - m$ and*

$$d(\xi, \mathbf{x}_*) < h_m(\epsilon) + \delta \,.$$

*Proof.* Denote $h = h_m(\epsilon)$ and $k = 2^{m-1}$. Let $\mathbf{x}_1, ..., \mathbf{x}_k$ correspond to the $k$ vectors of $\mathbb{R}^m$ of the form

$$\left(\sqrt{\frac{1}{h}}, \pm\sqrt{\frac{h-1}{h(m-1)}}, \ldots, \pm\sqrt{\frac{h-1}{h(m-1)}}\right)^\top$$

and let $\mathbf{y}_1, \ldots, \mathbf{y}_k$ correspond to the $k$ vectors of the form

$$\left(\sqrt{1/m}, \pm\sqrt{1/m}, \ldots, \pm\sqrt{1/m}\right)^\top \,.$$

Take $\mathbf{x}_* = (\sqrt{b}, 0, \ldots, 0)^\top \in \mathbb{R}^m$ with $1 < b < \min\{(\epsilon + m)/h, (h+\delta)/h\}$, $\mathcal{X}$ as the finite set $\mathcal{X} = \{\mathbf{x}_1, ..., \mathbf{x}_k, \mathbf{y}_1, ..., \mathbf{y}_k, \mathbf{x}_*\}$ and let $\xi$ be the uniform probability measure on $\mathbf{x}_1, ..., \mathbf{x}_k$. Note that $\mathbf{M}(\xi)$ is a diagonal matrix with diagonal elements $(1/h, (h-1)/[h(m-1)], \ldots, (h-1)/[h(m-1)])$. We can easily verify that

$$\max_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^\top \mathbf{M}(\xi)^{-1}\mathbf{x} - m = \epsilon \text{ and } d(\xi, \mathbf{x}_*) = \mathbf{x}_*^\top \mathbf{M}(\xi)^{-1}\mathbf{x}_* = bh < h_m(\epsilon) + \delta \,.$$

The uniform probability measure $\eta$ on $\mathbf{y}_1, ..., \mathbf{y}_k$ is $D$-optimum on $\mathcal{X}/\{\mathbf{x}_*\}$, as can be directly verified by checking (ii) of the Equivalence Theorem 1. On the other hand, $\eta$ is not $D$-optimum on $\mathcal{X}$ since $\mathbf{x}_*^\top \mathbf{M}(\eta)^{-1}\mathbf{x}_* = bm > m$, which implies that $\mathbf{x}_*$ must support a $D$-optimum design on $\mathcal{X}$. ∎

In order to illustrate the improvement over the inequality derived in (Pronzato, 2003) we consider the same two examples. In both cases the design algorithm is given by the simple recursion

$$w_i^{k+1} = w_i^k \frac{d(\xi^k, \mathbf{x}_i)}{m} \,,\ i = 1, \ldots, q(k) \tag{6}$$

where $w_i^k = \xi^k(\mathbf{x}_i)$ is the weight given by the discrete design $\xi^k$ to the point $\mathbf{x}_i$, see (Titterington, 1976; Torsney, 1983). At each step $k$, any design point $\mathbf{x}_j$ satisfying $d(\xi^k, \mathbf{x}_j) < h_m(\epsilon)$, see (4), or $d(\xi^k, \mathbf{x}_j) < \tilde{h}_m(\epsilon)$, with $\tilde{h}_m(\epsilon)$ given by (5), is removed from $\mathcal{X}$, its weight is set to zero and reallocated to other $\mathbf{x}_i'$'s proportionally to $d(\xi^k, \mathbf{x}_i) - m$; the cardinality $q(k)$ of the design space is decreased in consequence. The initial design $\xi^1$ allocates weights uniformly over $\mathcal{X}$.

**Example 1:** We consider a $D$-optimum design problem for the quadratic regression problem with regressors $\mathbf{x} = (1\ x\ x^2)^\top$, $x \in [0, 1]$. The initial set $\mathcal{X}$ is given by 1000 points equally spaced in the interval.

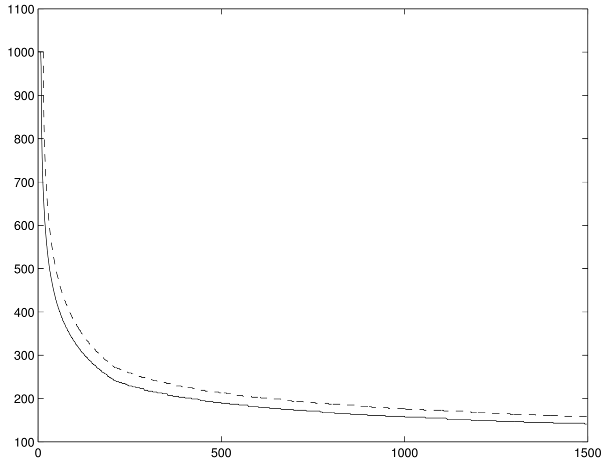


Fig. 1. $q(k)$ as a function of $k$ in Example 1.

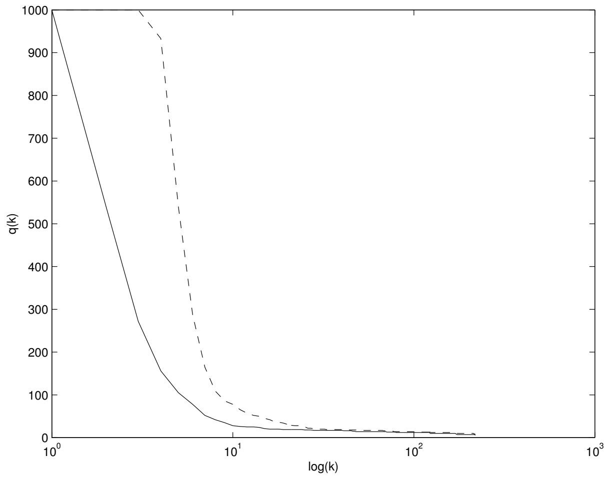


Fig. 2. $q(k)$ as a function of $\log(k)$ in Example 2.

**Example 2:** We construct the minimum covering ellipse for an initial set $\mathcal{X}$ of 1000 random points in the plane, i.i.d. $\mathcal{N}(0, \mathbf{I}_2)$ (this corresponds to a $D$-optimum design problem in $\mathbb{R}^3$, see (Titterington, 1975, 1978), so that the recursion (6) can be used).

Figure 1 (resp. 2) gives the evolution of the cardinality $q(k)$ as a function of $k$ (resp. $\log(k)$). On both figures the full line corresponds to a cardinality reduction based on (4), the dashed line is based on $\tilde{h}_m(\epsilon)$ given by (5), see (Pronzato, 2003).

## References


Fedorov, V., 1972. Theory of Optimal Experiments. Academic Press, New York.

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

Pronzato, L., 2003. Removing non-optimal support points in D-optimum design algorithms. Statistics & Probability Letters 63, 223–228.

Titterington, D., 1975. Optimal design: some geometrical espects of *D*-optimality. Biometrika 62 (2), 313–320.

Titterington, D., 1976. Algorithms for computing D-optimal designs on a finite design space. In: Proc. of the 1976 Conference on Information Science and Systems. Dept. of Electronic Engineering, John Hopkins University, Baltimore, pp. 213–216.

Titterington, D., 1978. Estimation of correlation coefficients by ellipsoidal trimming. Journal of Royal Statistical Society C27 (3), 227–234.

Torsney, B., 1983. A moment inequality and monotonicity of an algorithm. In: Kortanek, K., Fiacco, A. (Eds.), Proc. Int. Symp. on Semi-infinite Programming and Applications. Springer, Heidelberg, pp. 249–260.