



Comparison of Spectral and Hadamard Bounds for D-Optimality

Chun Wa Ko
Jon Lee
Kevin Wayne

ABSTRACT: We introduce a spectral bound for D-optimal design problems, based on singular values. We compare the spectral bound to a bound based on Hadamard's inequality which was introduced by Welch. In particular, we demonstrate that (i) in general, neither bound dominates the other, (ii) the spectral bound is superior in a general situation of highly replicated designs, and (iii) the spectral bound is superior when a very accurate bound is required in situations of singularity.

KEYWORDS: D-Optimal Design

1 Introduction.

We consider the linear model

$$y_i = x_i^t \beta + e_i \quad (i \in N),$$

where each $m \times 1$ vector x_i is a potential design point with associated response y_i , the e_i are i.i.d. with $E[e_i] = 0$ and $V[e_i] = \sigma^2$, β is an $m \times 1$ vector of parameters to be estimated, and N is an n -set of indices. When desired, a constant term in the linear model can be incorporated by setting a particular coordinate of each x_i to equal one. We assume that there are m linearly independent design points amongst the n points. We consider the situation where we are given an f -subset F of N , and an integer s with $f < s < n$, and we are to choose an s -set S satisfying $F \subset S \subset N$. We incorporate upper bounds on the number of times a design point can be replicated by explicitly including multiple copies of such points in the design matrix. Let $X(S)$ be the $s \times m$ design matrix with rows x_i^t , $i \in S$, and let $D(S) := X^t(S)X(S)$. Our criterion, which we seek to maximize, is

the determinant of $D(S)$. That is, we wish to solve

$$\max\{\det(D(S)) : F \subset S \subset N, |S| = s\}.$$

This is the so-called “ $|X^t X|$ ” or “D-optimality” criterion, which has been studied extensively. Methods to search for D-optimal designs have been suggested by Dykstra (1971), Fedorov (1972), Wynn (1970, 1972), Mitchell (1974a,b). The papers by St. John and Draper (1975) and by Cook and Nachtsheim (1980) survey the state of the art up through the late 1970’s, while Atkinson and Donev (1988), Dodge et al. (1988), and Yonchev (1988) discuss more recent progress. Welch (1982) proposed the first general algorithm for finding a *provably* D-optimal design. His method is based on a general framework of combinatorial optimization called *branch-and-bound*. To implement the framework Welch used lower bounds obtained by an exchange method, and two upper bounds that he proposed. One is based on Hadamard’s inequality, and the other is based upon an iterative method for generating an optimal “continuous design”.

In Section 2, we establish a *spectral* upper bound on $\det(D(S))$, based on the singular values of a matrix. We also discuss a variant of the *Hadamard bound* of Welch. We demonstrate that (i) neither bound always dominates the other, (ii) the spectral bound dominates the Hadamard bound in situations where highly replicated designs are sought, and (iii) the spectral bound dominates the Hadamard bound when very precise bounds are required for the case in which $D(F)$ is singular.

A complete branch-and-bound algorithm using these bounds and associated computational results appears in Ko, Lee and Wayne (1994).

2 Upper Bounds.

Initially, we make the simplifying assumption that $X(F)$ has full column rank. This implies that the symmetric matrix $D(F) = X^t(F)X(F)$ is invertible, and, moreover, that it is positive definite. Let $L(F)$ be the (unique) Cholesky factor of $D(F)$. That is, the invertible matrix that has all entries equal to 0 above the main diagonal and satisfies $D(F) = L(F)L^t(F)$ (see Golub and Van Loan (1983), pg. 88, for example). Let $\phi_i(F)$ denote the Euclidean norm of $x_i^t \cdot L^{-t}(F)$, for $i \in N \setminus F$. Let τ be a bijection from $\{1, 2, \dots, n - f\}$ to $N \setminus F$, such that $\phi_{\tau(i)}(F) \geq \phi_{\tau(j)}(F)$ whenever $i \leq j$. For a matrix A , let $\sigma_i(A)$ denote the i^{th} greatest singular value of A . (The nonzero singular values of a matrix A are precisely the square roots of the nonzero eigenvalues of $A^t A$; see Golub and Van Loan, pg. 285, for example). Let $\epsilon_i(F)$ denote the i^{th} greatest number among the $\min\{n - f, m\}$ singular values of $X(N \setminus F) \cdot L^{-t}(F)$, for $1 \leq i \leq \min\{n - f, m\}$, and let

$\epsilon_i(F) := 0$, for $\min\{n - f, m\} < i \leq n - f$. We define the *Hadamard bound*

$$\mathcal{H}(F, s) := \det(D(F)) \prod_{i=1}^{s-f} (1 + \phi_{\tau(i)}^2(F)),$$

and the *spectral bound*

$$\mathcal{S}(F, s) := \det(D(F)) \prod_{i=1}^{s-f} (1 + \epsilon_i^2(F)).$$

Proposition 1

$$\max\{\det(D(S)) : F \subset S \subset N, |S| = s\} \leq \min\{\mathcal{H}(F, s), \mathcal{S}(F, s)\}.$$

Proof: For a real symmetric matrix B , let $\lambda_i(B)$ denote the i^{th} greatest eigenvalue of B . First, we note that $D(S) = D(F) + D(S \setminus F)$. Hence

$$\begin{aligned} \det(D(S)) &= \det(D(F) + D(S \setminus F)) \\ &= \det(D(F)) \cdot \det(I_{s-f} + X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F)) \\ &= \det(D(F)) \prod_{i=1}^{s-f} \lambda_i \left(I_{s-f} + X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F) \right) \\ &= \det(D(F)) \prod_{i=1}^{s-f} \left(\lambda_i \left(I_{s-f} \right) + \right. \\ &\quad \left. \lambda_i \left(X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F) \right) \right) \\ &= \det(D(F)) \prod_{i=1}^{s-f} \left(1 + \lambda_i \left(X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F) \right) \right) \\ &\leq \det(D(F)) \prod_{i=1}^{s-f} \left(1 + \lambda_i \left(X(N \setminus F) \cdot D^{-1}(F) \cdot X^t(N \setminus F) \right) \right) \\ &= \det(D(F)) \prod_{i=1}^{s-f} \left(1 + \sigma_i^2 \left(X(N \setminus F) \cdot L^{-t}(F) \right) \right) \\ &= \mathcal{S}(F, s). \end{aligned}$$

thus establishing the spectral bound. We note that the inequality above follows from the *interlacing property* of singular values (see Golub and Van Loan, pg. 286, for example).

Additionally, we note that diagonal entry indexed by i of

$$I_{s-f} + X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F)$$

is $1 + \phi_i^2$. Now, since

$$I_{s-f} + X(S \setminus F) \cdot D^{-1}(F) \cdot X^t(S \setminus F)$$

is symmetric and positive definite, its determinant is no more than the product of its diagonal elements (see Horn and Johnson (1985), pg. 477, for example), thus establishing the Hadamard bound. ■

The bound based on Hadamard's inequality in Welch has a slightly different form than ours, since in our setup we incorporate upper bounds on the number of times a design point can be replicated by explicitly including multiple copies of such points in the design matrix. Since Welch allows arbitrary replication of all design points, his bound is:

$$\det(D(F))(1 + \phi_{\tau(1)}^2(F))^{s-f} .$$

As the following example indicates, neither the Hadamard bound \mathcal{H} nor the spectral bound \mathcal{S} always dominates the other:

Example 1 Nondominance. *Let*

$$X = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{pmatrix} ,$$

and take $F = \{1, 2\}$, and $N = \{1, 2, 3, 4, 5\}$. We calculate

$$D(F) = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} , \quad \det(D(F)) = 1 ,$$

$$L(F) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} , \quad L^{-t}(F) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} ,$$

$$X(N \setminus F) \cdot L^{-t}(F) = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} .$$

We have

$$\epsilon^2(F) = (4 + \sqrt{10}, 4 - \sqrt{10}, 0) \approx (7.1623, 0.8377, 0) ,$$

and

$$\phi^2(F) = (5, 2, 1) .$$

For $s = 4$, the spectral bound is sharper: We have $\mathcal{S}(F, s) = 15$ and $\mathcal{H}(F, s) = 18$, while the D-optimal design has $S = \{1, 2, 3, 5\}$, with $\det(D(S)) = 11$. On the other hand, for $s = 3$, the Hadamard bound is sharper: In this case, we have $\mathcal{S}(F, s) = 5 + \sqrt{10} \approx 8.1623$ and $\mathcal{H}(F, s) = 6$, while the D-optimal design has $S = \{1, 2, 3\}$, with $\det(D(S)) = 6$.

In the next example, we investigate how the bounds perform when we allow replication and require many design points to be selected.

Example 2 Heavily Replicated Designs. Let $F := \{0, 1, 2, \dots, f-1\}$. Fix a set of f m -vectors \tilde{x}_i , $i \in F$. We may assume that $\{\tilde{x}_i : i \in F\}$ contains m linearly independent points. Let k be a positive integer, let $N := \{0, 1, 2, \dots, (k+1)f-1\}$, and let $x_i := \tilde{x}_{i \pmod{f}}$, for $i \in N$. Let $s = s_k := f + k$. Our problem, then, is to choose $f + k$ design points from $\{\tilde{x}_i : i \in F\}$, allowing arbitrary nonzero replication, so as to produce a D-optimal design. We note that for any such problem, $\det(D(S_k))$ will behave like a polynomial in k of degree m , where S_k indicates the dependence of a D-optimal set of indices on k . In this situation we have,

$$\mathcal{H}(F, s_k) = \det(D(F)) \cdot (1 + \tilde{\phi}_{\tau(1)}^2(F))^k ,$$

where $\tilde{\phi}_{\tau(1)}(F) := \max_{i \in F} \{\|\tilde{x}_i^t \cdot L^{-t}(F)\|_2\}$. Thus, in the present situation, the Hadamard bound grows exponentially in k . On the other hand,

$$\mathcal{S}(F, s_k) = \det(D(F)) \prod_{i=1}^m (1 + k \cdot \tilde{\epsilon}_i^2(F)) ,$$

where $\tilde{\epsilon}_i(F)$ denotes the i^{th} greatest singular value of $X(F) \cdot L^{-t}(F)$. We note that in this case, the spectral bound increases as a polynomial in k of degree m .

Therefore, the spectral bound is within a constant factor of the D-optimal value, while the Hadamard bound is not within a subexponential factor.

To handle the case in which $D(F)$ is singular, we perturb the problem. Let $D_\alpha(S) = D(S) + (\alpha/n)D(N)$. Let \bar{v}^2 denote the average variance of the least squares estimators of the responses $\hat{y}_i := x_i^t \hat{\beta}$, over all $i \in N$, where $\hat{\beta} := D^{-1}(N)X^t y$. That is,

$$\bar{v}^2 = \frac{\sigma^2}{n} \sum_{i \in N} x_i^t D^{-1}(N) x_i .$$

Proposition 2 (*Mitchell*)

$$\lim_{\alpha \rightarrow 0} \frac{(\det(D_\alpha(S)) - \det(D(S))) / \det(D(S))}{\alpha} = \bar{\nu}^2 .$$

Thus, we can choose some small α (Mitchell suggests .005; Welch uses .001), and change our criteria to that of maximizing $\det(D_\alpha(S))$, where

$$D_\alpha(S) := D(S) + (\alpha/n)D(N) .$$

Then the relative error related to using $\det(D_\alpha(S))$ rather than $\det(D(S))$ will be approximately $\alpha\bar{\nu}^2$. We can view what we are doing as augmenting the design space by the n points $\sqrt{(\alpha/n)}x_i$ ($i \in N$), and forcing them into the solution.

For the sake of precision, let $L_\alpha(F)L_\alpha^t(F)$ be the Cholesky factorization of $D_\alpha(F)$. Let $\epsilon_i(F, \alpha)$ denote the i^{th} greatest number among the $\min\{n - f, m\}$ singular values of $X(N \setminus F) \cdot L_\alpha^{-t}(F)$, for $1 \leq i \leq \min\{n - f, m\}$, and let $\epsilon_i(F, \alpha) := 0$, for $\min\{n - f, m\} < i \leq n - f$. We define the *spectral bound* for the perturbed problem as

$$S_\alpha(F, s) := \det(D_\alpha(F)) \prod_{i=1}^{s-f} (1 + \epsilon_i^2(F, \alpha)) .$$

Similarly, let $\phi_i(F, \alpha)$ denote the Euclidean norm of $x_i^t \cdot L_\alpha^{-t}(F)$, for $i \in N \setminus F$. Let τ be a bijection from $\{1, 2, \dots, n - f\}$ to $N \setminus F$, such that $\phi_{\tau(i)}(F, \alpha) \geq \phi_{\tau(j)}(F, \alpha)$ whenever $i \leq j$. We define the *Hadamard bound* for the perturbed problem as

$$\mathcal{H}_\alpha(F, s) := \det(D_\alpha(F)) \prod_{i=1}^{s-f} (1 + \phi_{\tau(i)}^2(F, \alpha)) ,$$

In the following example, we study how the bounds may perform when we require a high degree of precision (α near 0) in the singular case.

Example 3 *Precise Bounds in the Singular Case. Let*

$$X = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

and take $F = \{1\}$, $N = \{1, 2, 3, 4\}$ and $s = 3$. We have

$$D_\alpha(F) = \begin{pmatrix} 1 + 3\alpha/4 & 1 \\ 1 & 1 + 3\alpha/4 \end{pmatrix} ,$$

which has determinant $3\alpha(8 + 3\alpha)/16$. We have

$$\epsilon^2(F, \alpha) = \left(\frac{4}{\alpha}, \frac{4}{8 + 3\alpha}, 0 \right),$$

and

$$\phi^2(F, \alpha) = \left(\frac{8}{3\alpha}, \frac{16 + 12\alpha}{24\alpha + 9\alpha^2}, \frac{16 + 12\alpha}{24\alpha + 9\alpha^2} \right).$$

It follows that

$$S_\alpha(F, s) = \frac{9(4 + \alpha)^2}{16} \rightarrow 9 \quad (\text{as } \alpha \rightarrow 0).$$

On the other hand,

$$\mathcal{H}_\alpha(F, s) = 7 + \frac{8}{3\alpha} + \frac{15\alpha}{4} + \frac{9\alpha^2}{16} \rightarrow \infty \quad (\text{as } \alpha \rightarrow 0).$$

We note that $S = \{1, 2, 3\}$ is D -optimal, with $\det(D(S)) = 6$.

Next, we will demonstrate that the behavior of the spectral bound in Example 3 is not an anomaly. That is, we will establish that the spectral bound always converges as α vanishes. For a matrix A , let $\rho(A)$ denote the rank of A .

Lemma 1 *There exist constants a_i , for $m - \rho(D(F)) \leq i \leq m$, such that*

$$\det(D_\alpha(N)) = \sum_{i=m-\rho(D(F))}^m a_i \alpha^i.$$

Proof.

$$\begin{aligned} \det(D_\alpha(N)) &= \det\left(\frac{\alpha}{n}D(N)\right) \cdot \det\left(I + \frac{n}{\alpha}X(F) \cdot D^{-1}(N) \cdot X^t(F)\right) \\ &= \alpha^m \det\left(\frac{1}{n}D(N)\right) \cdot \prod_{i=1}^{\rho(D(F))} \left(1 + \frac{1}{\alpha} \lambda_i(nX(F) \cdot D^{-1}(N) \cdot X^t(F))\right), \end{aligned}$$

with the upper limit on the index of the product being justified by noting that

$$\rho(nX(F) \cdot D^{-1}(N) \cdot X^t(F)) \leq \rho(D(F)).$$

Now

$$\prod_{i=1}^{\rho(D(F))} \left(1 + \frac{1}{\alpha} \lambda_i(nX(F) \cdot D^{-1}(N) \cdot X^t(F))\right) = \sum_{i=0}^{\rho(D(F))} c_i \left(\frac{1}{\alpha^i}\right),$$

for some c_i , $0 \leq i \leq \rho(D(F))$. The result follows. ■

Lemma 2 *There exist constants b_j , for $0 \leq j \leq m - \rho(D(F))$, such that*

$$\prod_{i=1}^{s-f} (1 + \epsilon_i^2(F, \alpha)) \leq \sum_{j=0}^{m-\rho(D(F))} b_j \left(\frac{1}{\alpha^j}\right).$$

Proof. We have

$$\epsilon_i^2(F, \alpha) \leq \sigma_1(X(N \setminus F)) \cdot \sigma_1(X^t(N \setminus F)) \cdot \lambda_i \left(\left(D(F) + \frac{\alpha}{n} D(N) \right)^{-1} \right).$$

Now,

$$\begin{aligned} \lambda_i \left(\left(D(F) + \frac{\alpha}{n} D(N) \right)^{-1} \right) &= \frac{\lambda_{m-i+1}^{-1} \left(D(F) + \frac{\alpha}{n} D(N) \right)}{1} \\ &\leq \frac{1}{\lambda_{m-i+1}(D(F)) + \alpha \cdot \lambda_m \left(\frac{1}{n} D(N) \right)}. \end{aligned}$$

Hence,

$$\epsilon_i^2(F, \alpha) \leq \frac{c}{\lambda_{m-i+1}(D(F)) + b\alpha},$$

for some constants c and b . Now,

$$\lambda_{m-i+1}(D(F)) \begin{cases} = 0, & \text{for } 1 \leq i \leq m - \rho(D(F)); \\ > 0, & \text{for } m - \rho(D(F)) < i \leq m. \end{cases}$$

Therefore, there exists a constant C , such that

$$\prod_{i=1}^{s-f} (1 + \epsilon_i^2(F, \alpha)) \leq C \cdot \left(1 + \frac{c}{b\alpha} \right)^{m-\rho(D(F))}.$$

The result follows. ■

Combining Lemmata 1 and 2, we immediately have the following result.

Proposition 3

$$\lim_{\alpha \rightarrow 0} \mathcal{S}_\alpha(F, s) < \infty.$$

We can glean from the proofs of Lemmata 1 and 2 a recipe for constructing an upper bound on the limit of Proposition 3. It would be interesting to find an efficient method for calculating the limit exactly.

Acknowledgments: The research of Jon Lee was supported in part by NSF Grant DMI-9401424.

References

- A.C. Atkinson and A.N. Donev (1988). Algorithms, Exact Designs and Blocking in Response Surface and Mixture Designs, in *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov and H.P. Wynn (editors), Elsevier Science Publishers B.V., Amsterdam, pp. 61-69.
- R.D. Cook and C.J. Nachtsheim (1980). A Comparison of Algorithms for Constructing Exact D-Optimal Designs, *Technometrics* **22**, 315-324.
- Y. Dodge, V.V. Fedorov and H.P. Wynn (1988). Optimal Design of Experiments: An Overview, in *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov and H.P. Wynn (editors), Elsevier Science Publishers B.V., Amsterdam, pp. 1-11.
- O. Dykstra, Jr. (1971). The Augmentation of Experimental Data to Maximize $|X'X|$, *Technometrics* **13**, 682-688.
- V.V. Fedorov (1972). *Theory of Optimal Experiments*, Academic Press, New York.
- G.H. Golub and C.F. Van Loan (1983). *Matrix Computations*, The Johns Hopkins University Press, Baltimore.
- R.A. Horn and C.R. Johnson (1985). *Matrix Analysis*, Cambridge University Press, Cambridge.
- C.W. Ko, J. Lee and K. Wayne (1994). A Spectral Bound for D-Optimality, manuscript.
- T.J. Mitchell (1974a). An Algorithm for the Construction of "D-Optimal" Experimental Designs, *Technometrics* **16**, 203-210.
- T.J. Mitchell (1974b). Computer Construction of "D-Optimal" First-Order Designs, *Technometrics* **16**, 211-220.
- R.C. St. John and N.R. Draper (1975). D-Optimality for Regression Designs: A Review, *Technometrics* **17**, 15-23.
- W.J. Welch (1982). Branch-and-Bound Search for Experimental Designs Based on D Optimality and Other Criteria, *Technometrics* **24**, 41-48.
- H. Yonchev (1988). New Computer Procedures for Constructing D-Optimal Designs, in *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov and H.P. Wynn (editors), Elsevier Science Publishers B.V., Amsterdam, pp. 71-80.